

Introduction to Sound Recording

Geoff Martin, B.Mus., M.Mus., Ph.D.

March 14, 2004

Contents

Opening materials	xvii
0.1 Preface	xvii
0.2 Thanks	xix
0.3 Autobiography	xx
0.4 Recommended Reading	xxi
0.4.1 General Information	xxi
0.4.2 Sound Recording	xxi
0.4.3 Analog Electronics	xxi
0.4.4 Psychoacoustics	xxi
0.4.5 Acoustics	xxii
0.4.6 Digital Audio and DSP	xxii
0.4.7 Electroacoustic Measurements	xxii
0.5 Why is this book free?	xxiii
1 Introductory Materials	1
1.1 Geometry	1
1.1.1 Right Triangles	1
1.1.2 Slope	3
1.2 Exponents	5
1.3 Logarithms	6
1.3.1 Warning	6
1.4 Trigonometric Functions	8
1.4.1 Radians	15
1.4.2 Phase vs. Addition	16
1.5 Complex Numbers	19
1.5.1 Whole Numbers and Integers	19
1.5.2 Rational Numbers	19
1.5.3 Irrational Numbers	19
1.5.4 Real Numbers	20

1.5.5	Imaginary Numbers	20
1.5.6	Complex numbers	21
1.5.7	Complex Math Part 1 – Addition	21
1.5.8	Complex Math Part 2 – Multiplication	22
1.5.9	Complex Math Part 3 – Some Identity Rules	22
1.5.10	Complex Math Part 4 – Inverse Rules	24
1.5.11	Complex Conjugates	25
1.5.12	Absolute Value (also known as the Modulus)	26
1.5.13	Complex notation or... Who cares?	27
1.6	Euler’s Identity	31
1.6.1	Who cares?	32
1.7	Binary, Decimal and Hexadecimal	34
1.7.1	Decimal (Base 10)	34
1.7.2	Binary (Base 2)	35
1.7.3	Hexadecimal (Base 16)	37
1.7.4	Other Bases	40
1.8	Intuitive Calculus	41
1.8.1	Function	41
1.8.2	Limit	41
1.8.3	Derivation and Slope	43
1.8.4	Sigma - Σ	48
1.8.5	Delta - Δ	48
1.8.6	Integration and Area	48
1.8.7	Suggested Reading List	53
2	Analog Electronics	55
2.1	Basic Electrical Concepts	55
2.1.1	Introduction	55
2.1.2	Current and EMF (Voltage)	56
2.1.3	Resistance and Ohm’s Law	58
2.1.4	Power and Watt’s Law	60
2.1.5	Alternating vs. Direct Current	61
2.1.6	RMS	62
2.1.7	Suggested Reading List	68
2.2	The Decibel	69
2.2.1	Gain	69
2.2.2	Power and Bels	69
2.2.3	dBspl	72
2.2.4	dBm	73
2.2.5	dBV	73

2.2.6	dBu	74
2.2.7	dB FS	75
2.2.8	Addendum: “Professional” vs. “Consumer” Levels	75
2.2.9	The Summary	75
2.3	Basic Circuits / Series vs. Parallel	77
2.3.1	Introduction	77
2.3.2	Series circuits – from the point of view of the current	77
2.3.3	Series circuits – from the point of view of the voltage	78
2.3.4	Parallel circuits – from the point of view of the voltage	79
2.3.5	Parallel circuits – from the point of view of the current	80
2.3.6	Suggested Reading List	81
2.4	Capacitors	82
2.4.1	Suggested Reading List	88
2.5	Passive RC Filters	89
2.5.1	Another way to consider this...	94
2.5.2	Suggested Reading List	95
2.6	Electromagnetism	96
2.6.1	Suggested Reading List	99
2.7	Inductors	100
2.7.1	Impedance	105
2.7.2	RL Filters	106
2.7.3	Inductors in Series and Parallel	106
2.7.4	Inductors vs. Capacitors	107
2.7.5	Suggested Reading List	107
2.8	Transformers	108
2.8.1	Suggested Reading List	109
2.9	Diodes and Semiconductors	110
2.9.1	The geeky stuff:	118
2.9.2	Zener Diodes	120
2.9.3	Suggested Reading List	122
2.10	Rectifiers and Power Supplies	123
2.11	Transistors	130
2.11.1	Suggested Reading List	130
2.12	Basic Transistor Circuits	131
2.12.1	Suggested Reading List	131
2.13	Operational Amplifiers	132
2.13.1	Introduction	132
2.13.2	Comparators	132
2.13.3	Inverting Amplifier	134
2.13.4	Non-Inverting Amplifier	136

2.13.5	Voltage Follower	137
2.13.6	Leftovers	138
2.13.7	Mixing Amplifier	138
2.13.8	Differential Amplifier	139
2.13.9	Suggested Reading List	139
2.14	Op Amp Characteristics and Specifications	141
2.14.1	Introduction	141
2.14.2	Maximum Supply Voltage	141
2.14.3	Maximum Differential Input Voltage	141
2.14.4	Output Short-Circuit Duration	141
2.14.5	Input Resistance	141
2.14.6	Common Mode Rejection Ratio (CMRR)	142
2.14.7	Input Voltage Range (Operating Common-Mode Range)	142
2.14.8	Output Voltage Swing	142
2.14.9	Output Resistance	142
2.14.10	Open Loop Voltage Gain	143
2.14.11	Gain Bandwidth Product (GBP)	143
2.14.12	Slew Rate	143
2.14.13	Suggested Reading List	144
2.15	Practical Op Amp Applications	145
2.15.1	Introduction	145
2.15.2	Active Filters	145
2.15.3	Higher-order filters	148
2.15.4	Bandpass filters	149
2.15.5	Suggested Reading List	149
3	Acoustics	151
3.1	Introduction	151
3.1.1	Pressure	151
3.1.2	Simple Harmonic Motion	153
3.1.3	Damping	154
3.1.4	Harmonics	155
3.1.5	Overtones	156
3.1.6	Longitudinal vs. Transverse Waves	157
3.1.7	Displacement vs. Velocity	158
3.1.8	Amplitude	159
3.1.9	Frequency and Period	160
3.1.10	Angular frequency	161
3.1.11	Negative Frequency	162
3.1.12	Speed of Sound	163

3.1.13	Wavelength	164
3.1.14	Acoustic Wavenumber	164
3.1.15	Wave Addition and Subtraction	165
3.1.16	Time vs. Frequency	169
3.1.17	Noise Spectra	169
3.1.18	Amplitude vs. Distance	171
3.1.19	Free Field	172
3.1.20	Diffuse Field	173
3.1.21	Acoustic Impedance	174
3.1.22	Power	177
3.1.23	Intensity	178
3.2	Acoustic Reflection and Absorption	180
3.2.1	Impedance Boundaries	181
3.2.2	Reflection and Transmission Coefficients	183
3.2.3	Absorption Coefficient	183
3.2.4	Comb Filtering Caused by a Specular Reflection	184
3.2.5	Huygen's Wave Theory	186
3.3	Specular and Diffused Reflections	187
3.3.1	Specular Reflections	187
3.3.2	Diffused Reflections	188
3.3.3	Reading List	196
3.4	Diffraction of Sound Waves	197
3.4.1	Reading List	197
3.5	Struck and Plucked Strings	198
3.5.1	Travelling waves and reflections	198
3.5.2	Standing waves (aka Resonant Frequencies)	199
3.5.3	Impulse Response vs. Resonance	203
3.5.4	Tuning the string	206
3.5.5	Harmonics in the real-world	207
3.5.6	Reading List	209
3.6	Waves in a Tube (Woodwind Instruments)	210
3.6.1	Waveguides	210
3.6.2	Resonance in a closed pipe	210
3.6.3	Open Pipes	216
3.6.4	Reading List	218
3.7	Helmholtz Resonators	219
3.7.1	Reading List	222
3.8	Characteristics of a Vibrating Surface	223
3.8.1	Reading List	223
3.9	Woodwind Instruments	225

3.9.1	Reading List	225
3.10	Brass Instruments	226
3.10.1	Reading List	226
3.11	Bowed Strings	227
3.11.1	Reading List	229
3.12	Percussion Instruments	230
3.12.1	Reading List	230
3.13	Vocal Sound Production	231
3.13.1	Reading List	231
3.14	Directional Characteristics of Instruments	232
3.14.1	Reading List	232
3.15	Pitch and Tuning Systems	233
3.15.1	Introduction	233
3.15.2	Pythagorean Temperament	234
3.15.3	Just Temperament	236
3.15.4	Meantone Temperaments	238
3.15.5	Equal Temperament	238
3.15.6	Cents	239
3.15.7	Suggested Reading List	240
3.16	Room Acoustics – Reflections, Resonance and Reverberation	241
3.16.1	Early Reflections	241
3.16.2	Reverberation	244
3.16.3	Resonance and Room Modes	247
3.16.4	Schroeder frequency	250
3.16.5	Room Radius (aka Critical Distance)	250
3.16.6	Reading List	251
3.17	Sound Transmission	252
3.17.1	Airborne sound	252
3.17.2	Impact noise	252
3.17.3	Reading List	252
3.18	Desirable Characteristics of Rooms and Halls	253
3.18.1	Reading List	253
4	Psychoacoustics and perception	255
4.1	What’s the difference?	255
4.2	How your ears work	256
4.3	Human Response Characteristics	257
4.3.1	Introduction	257
4.3.2	Frequency Range	257
4.3.3	Dynamic Range	257

4.4	Loudness	259
4.4.1	Equal loudness contours	259
4.4.2	Phons (also known as “Phones”)	260
4.4.3	Sones	261
4.5	Weighting Curves	262
4.6	Masking	264
4.7	Localization	265
4.7.1	Cone of Confusion	266
4.8	Precedence Effect	267
4.8.1	Suggested Reading List	267
4.9	Distance Perception	268
4.9.1	Reflection patterns	268
4.9.2	Direct-to-reverberant ratio	268
4.9.3	Sound pressure level	269
4.9.4	High-frequency content	269
4.9.5	Suggested Reading List	269
4.10	Hearing Loss and Tinnitus	270
4.10.1	Suggested Reading List	270
4.11	Perception of sound qualities	271
4.11.1	Suggested Reading List	271
5	Electroacoustics	273
5.1	Filters and Equalizers	273
5.1.1	Introduction	273
5.1.2	Filters	274
5.1.3	Equalizers	280
5.1.4	Summary	290
5.1.5	Phase response	290
5.1.6	Applications	294
5.1.7	Spectral sculpting	294
5.1.8	Loudness	296
5.1.9	Noise Reduction	297
5.1.10	Dynamic Equalization	298
5.1.11	Further reading	299
5.2	Compressors, Limiters, Expanders and Gates	300
5.2.1	What a compressor does.	300
5.2.2	How compressors compress	318
5.2.3	Suggested Reading List	327
5.3	Analog Tape	328
5.3.1	Suggested Reading List	328

5.4	Noise Sources	329
5.4.1	Introduction	329
5.4.2	EMI Transmission	329
5.4.3	Suggested Reading List	331
5.5	Reducing Noise – Shielding, Balancing and Grounding	332
5.5.1	Shielding	332
5.5.2	Balanced transmission lines	332
5.5.3	Grounding	334
5.5.4	Suggested Reading List	338
5.6	Microphones – Transducer type	340
5.6.1	Introduction	340
5.6.2	Dynamic Microphones	340
5.6.3	Condenser Microphones	346
5.6.4	RF Condenser Microphones	346
5.6.5	Suggested Reading List	346
5.7	Microphones – Directional Characteristics	347
5.7.1	Introduction	347
5.7.2	Pressure Transducers	347
5.7.3	Pressure Gradient Transducers	353
5.7.4	Combinations of Pressure and Pressure Gradient	358
5.7.5	General Sensitivity Equation	360
5.7.6	Do-It-Yourself Polar Patterns	366
5.7.7	The Influence of Polar Pattern on Frequency Response	369
5.7.8	Proximity Effect	373
5.7.9	Acceptance Angle	376
5.7.10	Random-Energy Response (RER)	376
5.7.11	Random-Energy Efficiency (REE)	377
5.7.12	Directivity Factor (DRF)	379
5.7.13	Distance Factor (DSF)	381
5.7.14	Variable Pattern Microphones	383
5.7.15	Suggested Reading List	383
5.8	Loudspeakers – Transducer type	385
5.8.1	Introduction	385
5.8.2	Ribbon Loudspeakers	385
5.8.3	Moving Coil Loudspeakers	387
5.8.4	Electrostatic Loudspeakers	391
5.9	Loudspeaker acoustics	396
5.9.1	Enclosures	396
5.9.2	Other Issues	404
5.10	Power Amplifiers	409

5.10.1	Introduction	409
6	Electroacoustic Measurements	411
6.1	Tools of the Trade	411
6.1.1	Digital Multimeter	411
6.1.2	True RMS Meter	414
6.1.3	Oscilloscope	415
6.1.4	Function Generators	420
6.1.5	Automated Testing Equipment	422
6.1.6	Swept Sine Tone	422
6.1.7	MLS-Based Measurements	422
6.1.8	Suggested Reading List	425
6.2	Electrical Measurements	427
6.2.1	Output impedance	427
6.2.2	Input Impedance	427
6.2.3	Gain	428
6.2.4	Gain Linearity	428
6.2.5	Frequency Response	429
6.2.6	Bandwidth	429
6.2.7	Phase Response	429
6.2.8	Group Delay	430
6.2.9	Slew Rate	430
6.2.10	Maximum Output	430
6.2.11	Noise measurements	431
6.2.12	Dynamic Range	432
6.2.13	Signal to Noise Ratio (SNR or S/N Ratio)	432
6.2.14	Equivalent Input Noise (EIN)	432
6.2.15	Common Mode Rejection Ratio (CMRR)	433
6.2.16	Total Harmonic Distortion (THD)	434
6.2.17	Total Harmonic Distortion + Noise (THD+N)	435
6.2.18	Crosstalk	436
6.2.19	Intermodulation Distortion (IMD)	436
6.2.20	DC Offset	436
6.2.21	Filter Q	436
6.2.22	Impulse Response	436
6.2.23	Cross-correlation	436
6.2.24	Interaural Cross-correlation	436
6.2.25	Suggested Reading List	436
6.3	Loudspeaker Measurements	437
6.3.1	Frequency Response	438

6.3.2	Phase Response	438
6.3.3	Distortion	438
6.3.4	Off-Axis Response	438
6.3.5	Power Response	438
6.3.6	Impedance Curve	438
6.3.7	Sensitivity	438
6.3.8	Suggested Reading List	438
6.4	Microphone Measurements	439
6.4.1	Polar Pattern	439
6.4.2	Frequency Response	439
6.4.3	Phase Response	439
6.4.4	Off-Axis Response	439
6.4.5	Sensitivity	439
6.4.6	Equivalent Noise Level	439
6.4.7	Impedance	439
6.4.8	Suggested Reading List	439
6.5	Acoustical Measurements	440
6.5.1	Room Modes	440
6.5.2	Reflections	440
6.5.3	Reflection and Absorbtion Coefficients	440
6.5.4	Reverberation Time (RT60)	440
6.5.5	Noise Level	440
6.5.6	Sound Transmission	440
6.5.7	Intelligibility	440
6.5.8	Suggested Reading List	440
7	Digital Audio	441
7.1	How to make a (PCM) digital signal	441
7.1.1	The basics of analog to digital conversion	441
7.1.2	The basics of digital to analog conversion	445
7.1.3	Aliasing	445
7.1.4	Binary numbers and bits	449
7.1.5	Two's complement	449
7.1.6	Suggested Reading List	450
7.2	Quantization Error and Dither	452
7.2.1	Dither	457
7.2.2	When to use dither	464
7.2.3	Suggested Reading List	465
7.3	Aliasing	466
7.3.1	Antialiasing filters	466

7.3.2	Suggested Reading List	466
7.4	Advanced A/D and D/A conversion	471
7.4.1	Suggested Reading List	471
7.5	Digital Signal Transmission Protocols	472
7.5.1	Introduction	472
7.5.2	What's being sent?	473
7.5.3	Some more info about AES/EBU	476
7.5.4	S/PDIF	478
7.5.5	Some Terms That You Should Know...	478
7.5.6	Jitter	479
7.5.7	Suggested Reading List	479
7.6	Jitter	481
7.6.1	When to worry about jitter	482
7.6.2	What causes jitter, and how to reduce it	482
7.6.3	Suggested Reading List	482
7.7	Fixed- vs. Floating Point	483
7.7.1	Suggested Reading List	483
7.8	Noise Shaping	484
7.8.1	Suggested Reading List	484
7.9	"High-Resolution" Audio	485
7.9.1	"What is "High-Resolution Audio?"	486
7.9.2	Wither High-Resolution audio?	487
7.9.3	Is it worth it?	491
7.9.4	Suggested Reading List	491
7.10	Data Compression and Reduction	492
7.10.1	History	492
7.10.2	ATRAC	492
7.10.3	PASC	492
7.10.4	Dolby	492
7.10.5	MPEG	492
7.10.6	MLP	492
7.10.7	Suggested Reading List	493
7.11	Transmission Techniques for Internet Distribution	494
7.11.1	Suggested Reading List	494
7.12	CD – Compact Disc	495
7.12.1	Sampling rate	495
7.12.2	Word length	496
7.12.3	Storage capacity (recording time)	497
7.12.4	Physical construction	497
7.12.5	Eight-to-Fourteen Modulation (EFM)	498

7.12.6 Suggested Reading List	502
7.13 DVD-Audio	503
7.13.1 Suggested Reading List	503
7.14 SACD and DSD	504
7.14.1 Suggested Reading List	504
7.15 Hard Disk Recording	505
7.15.1 Bandwidth and disk space	505
7.15.2 Suggested Reading List	506
7.16 Digital Audio File Formats	507
7.16.1 AIFF	507
7.16.2 WAV	507
7.16.3 SDII	507
7.16.4 mu-law	507
7.16.5 Suggested Reading List	507
8 Digital Signal Processing	509
8.1 Introduction to DSP	509
8.1.1 Block Diagrams and Notation	510
8.1.2 Normalized Frequency	511
8.1.3 Suggested Reading List	512
8.2 FFT's, DFT's and the Relationship Between the Time and Frequency Domains	513
8.2.1 Fourier in a Nutshell	513
8.2.2 Discrete Fourier Transforms	516
8.2.3 A couple of more details on detail	521
8.2.4 Redundancy in the DFT Result	522
8.2.5 What's the use?	523
8.2.6 Suggested Reading List	523
8.3 Windowing Functions	524
8.3.1 Rectangular	530
8.3.2 Hanning	532
8.3.3 Hamming	534
8.3.4 Comparisons	536
8.3.5 Suggested Reading List	537
8.4 FIR Filters	538
8.4.1 Comb Filters	538
8.4.2 Frequency Response Deviation	544
8.4.3 Impulse Response vs. Frequency Response	546
8.4.4 More complicated filters	547
8.4.5 Suggested Reading List	548

8.5	IIR Filters	549
8.5.1	Comb Filters	549
8.5.2	Danger!	553
8.5.3	Biquadratic	553
8.5.4	Allpass filters	559
8.5.5	Suggested Reading List	560
8.6	Convolution	561
8.6.1	Correlation	567
8.6.2	Autocorrelation	567
8.6.3	Suggested Reading List	567
8.7	The z-domain	568
8.7.1	Introduction	568
8.7.2	Filter frequency response	569
8.7.3	The z-plane	572
8.7.4	Transfer function	573
8.7.5	Calculating filters as polynomials	576
8.7.6	Zeros	577
8.7.7	Poles	584
8.7.8	Suggested Reading List	589
9	Audio Recording	591
9.1	Levels and Metering	591
9.1.1	Introduction	591
9.1.2	Digital Gear in the PCM World	594
9.1.3	Analog electronics	596
9.1.4	Analog tape	597
9.1.5	Meters	599
9.1.6	Gain Management	616
9.1.7	Phase and Correlation Meters	616
9.1.8	Suggested Reading List	616
9.2	Monitoring Configuration and Calibration	619
9.2.1	Standard operating levels	619
9.2.2	Channels are not Loudspeakers	620
9.2.3	Bass management	621
9.2.4	Configuration	623
9.2.5	Calibration	632
9.2.6	Monitoring system equalization	638
9.2.7	Suggested Reading List	638
9.3	Introduction to Stereo Microphone Technique	639
9.3.1	Panning	639

9.3.2	Coincident techniques (X-Y)	640
9.3.3	Spaced techniques (A-B)	644
9.3.4	Near-coincident techniques	645
9.3.5	More complicated techniques	647
9.3.6	Suggested Reading List	649
9.4	General Response Characteristics of Microphone Pairs	650
9.4.1	Phantom Images Revisited	650
9.4.2	Interchannel Differences	654
9.4.3	Summed power response	690
9.4.4	Correlation	711
9.4.5	Conclusions	723
9.5	Matrixed Microphone Techniques	725
9.5.1	MS	725
9.5.2	Ambisonics	732
9.5.3	Suggested Reading List	745
9.6	Introduction to Surround Microphone Technique	746
9.6.1	Advantages to Surround	746
9.6.2	Common pitfalls	748
9.6.3	Fukada Tree	752
9.6.4	OCT Surround	753
9.6.5	OCT Front System + IRT Cross	753
9.6.6	OCT Front System + Hamasaki Square	755
9.6.7	Klepko Technique	755
9.6.8	Corey / Martin Tree	756
9.6.9	Michael Williams	756
9.6.10	Suggested Reading List	757
9.7	Introduction to Time Code	758
9.7.1	Introduction	758
9.7.2	Frame Rates	758
9.7.3	How to Count in Time Code	760
9.7.4	SMPTE/EBU Time Code Formats	760
9.7.5	Time Code Encoding	765
9.7.6	Annex – Time Code Bit Assignments	772
9.7.7	Suggested Reading List	777
10	Conclusions and Opinions	779
10.0.8	Suggested Reading List	780

11 Reference Information	781
11.1 ISO Frequency Centres	781
11.2 Scientific Notation and Prefixes	783
11.2.1 Scientific notation	783
11.2.2 Prefixes	783
11.3 Resistor Colour Codes	785
12 Hints and tips	787

Opening materials

0.1 Preface

Once upon a time I went to McGill University to try to get into the Master’s program in sound recording at the Faculty of Music. In order to accomplish this task, students are required to do a “qualifying year” of introductory courses (although it felt more like obstacle courses...) to see who *really* wants to get in the program. Looking back, there is no question that I learned more in that year than in any other single year of my life. In particular, two memories stand out.

One was my professor – a guy named Peter Cook who now works at the CBC in Toronto as a digital editor. Peter is one of those teachers who doesn’t know everything, and doesn’t pretend to know everything – but if you ask him a question about something he doesn’t understand, he’ll show up at the next week’s class with a reading list where the answer to your question can be found. That kind of enthusiasm in a teacher cannot be replaced by any other quality. I definitely wouldn’t have gotten as much out of that year without him. A good piece of advice that I recently read for university students is that you don’t choose courses, you choose professors.

The second thing was a book by John Woram called the Sound Recording Handbook (the 1989 edition). This book not only proved to be the most amazing introduction to sound recording for a novice idiot like myself, but it continued to be the most used book on my shelf for the following 10 years. In fact, I was still using it as a primary reference when I was studying for my doctoral comprehensives 10 years later. Regrettably, that book is no longer in print – so if you can find a copy (remember – the 1989 edition... no other...) buy (or steal) it and guard it with your life.

Since then, I have seen a lot of students go through various stages of becoming a recording engineer at McGill and in other places and I’ve lamented the lack of a decently-priced but academically valuable textbook for these people. There have been a couple of books that have hit the market, but

they're either too thin, too full of errors, too simplistic or too expensive. (I won't name any names here, but if you ask me in person, I'll tell you...)

This is why I'm writing this book. From the beginning, I intended it to be freely accessible to anyone that was interested enough to read it. I can't guarantee that it's completely free of errors – so if you find any, please let me know and I'll make the appropriate corrections as soon as I can. The tone of this book is pretty colloquial – that's intentional – I'm trying to make the concepts presented here as accessible as possible without reducing the level of the content, so it can make a good introduction that covers a lot of ground. I'll admit that it doesn't make a great reference because there are too many analogies and stories in here – essentially too low a signal to noise ratio to make a decent book for someone that already understands the concepts.

Note that the book isn't done yet – in fact, in keeping with everything else you'll find on the web, it will probably never be finished. You'll find many places where I've made notes to myself on what will be added where. Also, there's a couple of explanations in here that don't make much sense – even to me... so they'll get fixed later. Finally, there are a lot of references missing. These will be added in the next update – I promise...

If you think that I've left any important subjects out of the Table of Contents, please let me know by email at geoff.martin@tonmeister.ca.

0.2 Thanks

There are a lot of people to thank for helping me out with this project. I have to start with a big thanks to Peter Cook for getting me started on the right track in the first place. To Wieslaw Woszczyk for allowing me to teach courses at McGill back when I was still just a novice idiot – the best way to learn something is to have to teach it to someone else. To Alain Terriault, Gilbert Soulodre and Michel Lavoie for answering my dumb questions about electronics when I was just starting to get the hang of exactly what things like a capacitor or an op amp do in a circuit. To Brian Sarvis and Kypros Christodoulides who were patient enough to put up with my subsequent dumb questions regarding what things like diodes and transistors do in a circuit. To Jason Corey for putting up with me running down many a wrong track looking for answers to some of the questions found in here. To Mark Ballora for patiently answering questions about DSP and trying (unsuccessfully) to get me to understand Shakespeare. Finally to Philippe Depalle – once upon a time I was taking a course in DSP for dummies at McGill and, about two-thirds of the way through the semester, Philippe guest-taught for one class. In that class, I wound up taking more notes than I had for all classes in the previous part of the semester combined. Since then, Philippe came to be a full-time faculty at McGill and I was lucky enough to have him as a thesis advisor. Basically, when I have any questions about anything, Philippe is the person I ask.

I also have to thank a number of people who have proofread some of the stuff you'll find here and have offered assistance and corrections – either with or without being asked. In alphabetical order, these folks are Bruce Bartlett, Peter Cook, Goran Finnberg, John La Grou, George Massenburg, Bert Noeth, Ray Rayburn, Eberhard Sengpiel and Greg Simmons.

Also on the list of thanks are people who have given permission to use their materials. Thanks to Claudia Haase and Thomas Lischker at RTW Radio-Technische (www.rtw.de) for their kind permission to use graphics from their product line for the section on levels and meters. Also to George Massenburg (www.massenburg.com) for permission to duplicate a chapter from the GML manual on equalizers that I wrote for him.

There are also a large number of people who have emailed me, either to ask questions about things that I didn't explain well enough the first time, or to make suggestions regarding additions to the book. I'll list those people in a later update of the text – but thanks to you if you're in that group.

Finally, thanks to Jonathan Sheaffer and The Jordan Valley Academic College (www.yarden.ac.il) for hosting the space to put this file for now.

0.3 Autobiography

Just in case you're wondering "Who does this guy think he is?" I'll tell you... This is the bio I usually send out when someone asks me for one. It's reasonably up-to-date.

Originally from St. John's, Newfoundland, Geoff Martin completed his B.Mus. in pipe organ at Memorial University of Newfoundland in 1990. He is a graduate of McGill's Masters program in Sound Recording and, in 2001, he completed his doctoral studies in which he developed a method of simulating reflections from quadratic residue diffusers for multichannel virtual acoustic environments.

Following completion of his doctorate, Geoff was a Faculty Lecturer for McGill's Music Technology area, where he taught courses in new media, electronics, and electroacoustics. In addition, he was a member of the development team for McGill's new Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT). He taught electroacoustic music composition and conducted the contemporary music ensemble at the University of Ottawa. He has also been a regular member of the visiting faculty in the Music and Sound Department at the Banff Centre for the Arts. He is presently a researcher in acoustics and perception at Bang and Olufsen a/s in Denmark where he has worked since the Fall of 2002. He maintains an active musical career as an organist, choral conductor and composer.

Geoff has been a member of the [Audio Engineering Society](#) since 1990 and has served on the executive for the Montreal Student Chapter for two years. He was the Papers Chair for the 24th International Conference of the Audio Engineering Society titled "Multichannel Audio: The New Reality" held at The Banff Centre in Alberta, Canada. He is presently the chair of the AES Technical Committee on Microphones and Applications.



Figure 1: What I used to look like once upon a time.

0.4 Recommended Reading

There are a couple of books that I would highly recommend, either because they're full of great explanations of the concepts that you'll need to know, or because they're great reference books. Some of these aren't in print any more

0.4.1 General Information

Ballou, G., ed. (1987) *Handbook for Sound Engineers: The New Audio Cyclopedia*, Howard W. Sams & Company, Indianapolis.

Rane's online dictionary of audio terms at www.rane.com

0.4.2 Sound Recording

Woram, John M. (1989) *Sound Recording Handbook*, Howard W. Sams & Company, Indianapolis. (This is the only edition of this book that I can recommend. I don't know the previous edition, and the subsequent one wasn't remotely as good.)

Eargle, John (1986) *Handbook of Recording Engineering*, Van Nostrand Reinhold, New York.

0.4.3 Analog Electronics

Jung, Walter G., *IC Op-Amp Cookbook*, Prentice Hall Inc.

Gayakwad, Ramakant A., (1983) *Op-amps and Linear Integrated Circuit Technology*, Prentice-Hall Inc.

0.4.4 Psychoacoustics

Moore, B. C. J. (1997) *An Introduction to the Psychology of Hearing*, Academic Press, San Diego, 4th Edition.

Blauert, J. (1997) *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, Revised Edition.

Bregman, A. S. (1990) *Auditory Scene Analysis : The Perceptual Organization of Sound*, MIT Press. Cambridge.

Zwicker, E., & Fastl, H. (1999) *Psychoacoustics: Facts and Models*, Springer, Berlin.

0.4.5 Acoustics

Morfeý, C. L. (2001). *Dictionary of Acoustics*, Academic Press, San Diego.

Kinsler L. E., Frey, A. R., Coppens, A. B., & Sanders, J. V. (1982) *Fundamentals of Acoustics*, John Wiley & Sons, New York, 3rd edition.

Hall, D. E. (1980) *Musical Acoustics: An Introduction*, Wadsworth Publishing, Belmont.

Kutruuff, K. H. (1991) *Room Acoustics*, Elsevier Science Publishers, Essex.

0.4.6 Digital Audio and DSP

Roads, C., ed. (1996) *The Computer Music Tutorial*, MIT Press, Cambridge.

Strawn, J., editor (1985). *Digital Audio Signal Processing: An Anthology*, William Kaufmann, Inc., Los Altos.

Smith, Steven W. *The Scientist and Engineers Guide to Digital Signal Processing* (www.dspguide.com)

Steiglitz, K. (1996) *A DSP Primer : With Applications to Digital Audio and Computer Music*, Addison-Wesley, Menlo Park.

Zlzer, U. (1997) *Digital Audio Signal Processing*, John Wiley & Sons, Chichester.

Anything written by Julius Smith

0.4.7 Electroacoustic Measurements

Mezler, Bob (1993) *Audio Measurement Handbook*, Audio Precision, Beaverton (available at a very reasonable price from www.ap.com)

Anything written by Julian Dunn

0.5 Why is this book free?

My friends ask me why in the hell I spend so much time writing a book that I don't make any money on. There are lots of reasons starting with the three biggies:

1. If I did try to sell this through a publisher, I sure as hell wouldn't make enough money to make it worth my while. This thing has taken many hours over many years to write.
2. By the time the thing actually got out on the market, I'd be long gone and everything here would be obsolete. It takes a long long time to get something published... and
3. It's really difficult to do regular updates on a hard copy of a book. You probably wouldn't want me to be dropping by your house penciling in corrections and additions in a book that you spent a lot of cash on. So, it's easier to do it this way.

I've recently been introduced to the concept of "charityware" which I'm using as the model for this book. So, if you use it, and you think that it's worth some cash, please donate whatever you would have spent on it in a bookstore to your local cancer research foundation. I've put together a small list of suggestions below. Any amount will be appreciated. Alternatively, you could just send me a postcard from wherever you live. My address is Geoff Martin, Grønnegade 17, DK-7830 Vinderup, Denmark

I might try and sell the next one. For this one, you have my permission to use, copy and distribute this book as long as you obey the following paragraph:

Copyright Geoff Martin 1999-2003. Users are authorized to copy the information in this document for their personal research or educational use for non-profit activities provided this paragraph is included. There is no limit on the number of copies that can be made. The material cannot be duplicated or used in printed or electronic form for a commercial purpose unless appropriate bibliographic references are used. - www.tonmeister.ca

Chapter 1

Introductory Materials

1.1 Geometry

It may initially seem that a section explaining geometry is a very strange place to start a book on sound recording, but as we'll see later, it's actually the best place to start. In order to understand many of the concepts in the chapters on acoustics, electronics, digital signal processing and electroacoustics, you'll need to have a *very* firm and intuitive grasp of a couple of simple geometrical concepts. In particular, these two concepts are the *right triangle* and the concept of *slope*.

1.1.1 Right Triangles

I'll assume at this point that you know what a triangle is. If you do not understand what a triangle is, then I would recommend backing up a bit from this textbook and reading other tomes such as *Trevor Draws a Triangle* and the immortal classic, *Baby's First Book of Euclidian Geometry*. (Okay, I made those titles up... let's move on, shall we?)

Once upon a time, a Greek by the name of Pythagoreas had a minor obsession with triangles. (He also believed in reincarnation and thought that if you were *really* bad, you might come back as a bean, so the Pythagoreans (his followers) didn't eat meat or beans... You can look it up if you don't believe me.) Interestingly, Pythagoreas, like many other Greeks at the time, recognized the direct link between mathematics and musical acoustics, so you'll see his name popping up all over the place as we go through this book.

Anyways, back to triangles. The first thing that we have to define is something called a *right triangle*. This is just a regular old everyday triangle with one specific characteristic. One of its angles is a *right angle* meaning

that it's 90° as is shown in Figure 1.1. One other new word to learn. The side opposite the right angle (in Figure 1.1, that would be side a) is called the *hypotenuse* of the triangle.

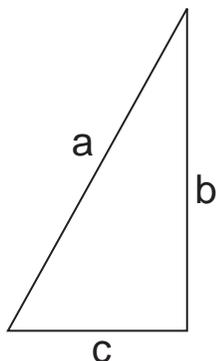


Figure 1.1: A right triangle with sides of lengths a , b and c . Note that side a is called the hypotenuse of the triangle.

One of the things Pythagoras discovered was that if you take a right triangle and make a square from each of its sides as is shown in Figure 1.2, then the sum of the areas of the two smaller squares is equal to the area of the big square.

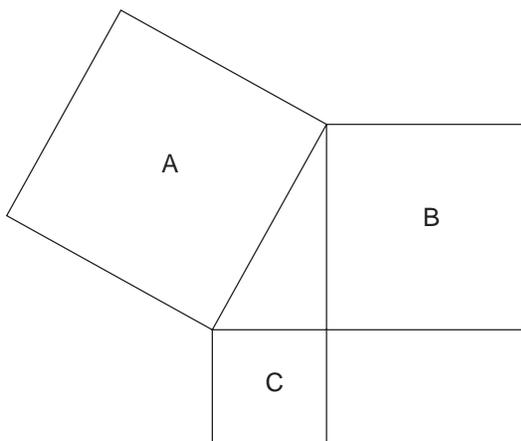


Figure 1.2: Three squares of areas A , B and C created by making squares out of the sides of a right triangle of arbitrary dimensions. $A = B + C$

So, looking at Figure 1.2, then we can say that $A = B + C$. We also

should know that the area of a square is equal to the square of the length of one of its sides. Looking at Figures 1.1 and 1.2 this means that $A = a^2$, $B = b^2$, and $C = c^2$.

Therefore, we can put this information together to arrive at a standard equation for right triangles known as the *Pythagorean Theorem*, shown in Equation 1.2.

$$a^2 = b^2 + c^2 \quad (1.1)$$

and therefore

$$a = \sqrt{b^2 + c^2} \quad (1.2)$$

This equation is probably the single most important key to understanding the concepts presented in this book, so you'd better remember it.

1.1.2 Slope

Let's go downhill skiing. One of the big questions when you're a beginner downhill skier is "how steep is the hill?" Well, there is a mathematical way to calculate the answer to this question. Essentially, another way to ask the same question is "how many metres do I drop for every metre that I ski forward?" The more you drop over a given distance, the steeper the slope of the hill.

So, what we're talking about when we discuss the slope of the hill is how much it rises (or drops) for a given run. Mathematically, the slope is written as a ratio of these two values as is shown in Equation 1.3.

$$\text{slope} = \frac{\text{rise}}{\text{run}} \quad (1.3)$$

but if we wanted to be a little more technical about this, then we would talk about the ratio of the difference in the y-value (the rise) for a given difference in the x-value (the run), so we'd write it like this:

$$\text{slope} = \frac{\Delta y}{\Delta x} \quad (1.4)$$

Where Δ is a symbol (it's the Greek capital letter *delta*) commonly used to indicate a difference or a change.

Let's just think about this a little more for a couple of minutes and consider some different slopes.

When there is no rise or drop for a change in horizontal distance, (like sailing on the ocean with no waves) then the value of Δy is 0, so the slope is 0.

When you're climbing a sheer rock face that drops straight down, then the value of Δx is 0 for a large change in y therefore the slope is ∞ .

If the change in x and y are both positive (so, you are going forwards and up at the same time) then the slope is positive. In other words, the line goes up from left to right on a graph.

If the change in y is negative while the change in x is positive, then the slope is negative. In other words, you're going downhill forwards, or you're looking at a graph of a line that goes downwards from left to right.

If you look at a real textbook on geometry then you'll see a slightly different equation for slope that looks like Equation 1.5, but we won't bother with this one. If you compare it to Equation 1.4, then you'll see that, apart from the k they're identical, and that the k is just a sort of altitude reading.

$$y = mx + k \tag{1.5}$$

where m is the slope.

1.2 Exponents

An *exponent* is just a lazy way to write down that you want to multiply a number by itself.

If I say 10^2 , then this is just a short notation for “10 multiplied by itself 2 times” – therefore, it’s $10 * 10 = 100$. For example, $3^4 = 3 * 3 * 3 * 3 = 81$.

Sometime’s you’ll see a negative number as the exponent. This simply means that you have to include a little division in your calculation. Whenever you see a negative exponent, you just have to divide 1 by the same thing without the negative sign. For example, $10^{-2} = \frac{1}{10^2}$

1.3 Logarithms

Once upon a time you learned to do multiplication, after which someone explained that you can use division to do the reverse. For example:

if

$$A = B * C \tag{1.6}$$

then

$$\frac{A}{B} = C \tag{1.7}$$

and

$$\frac{A}{C} = B \tag{1.8}$$

Logarithms sort of work in the same way, except that they are the backwards version of an exponent. (Just as division is the backwards version of multiplication.) Logarithms (or *logs*) work like this:

If $10^2 = 100$ then $\log_{10} 100 = 2$

Actually, it's:

If $A^B = C$ then $\log_A C = B$

Now we have to go through some properties of logarithms.

$\log_{10} 10 = 1$ or $\log_{10} 10^1 = 1$

$\log_{10} 100 = 2$ or $\log_{10} 10^2 = 2$

$\log_{10} 1000 = 3$ or $\log_{10} 10^3 = 3$

This should come as no great surprise – you can check them on your calculator if you don't believe me. Now, let's play with these three equations.

$\log_{10} 1000 = 3$

$\log_{10} 10^3 = 3$

$3 * \log_{10} 10 = 3$

Therefore:

$\log_C A^B = B * \log_C A$

1.3.1 Warning

I once learned that you should never assume, because when you assume you make an ass out of you and me... (get it? ass—u—me... okay... dumb joke). One small problem with logarithms is the way they're written. People usually don't write the base of the log so you'll see things like $\log(3)$ written which usually means $\log_{10} 3$ – if the base isn't written, it's assumed to be 10.

This also holds true on most calculators. Punch in 100 and hit LOG and see if you get 2 as an answer – you probably will. Unfortunately, this assumption is not true if you’re using a computer to calculate your logarithms. For example, if you’re using MATLAB and you type `log(100)` and hit the RETURN button, you’ll get the answer 4.6052. This is because MATLAB assumes that you mean base e (a number close to 2.7182) instead of base 10. So, if you’re using MATLAB, you’ll have to type in `log10(100)` to indicate that the logarithm is in base 10. If you’re in Mathematica, you’ll have to use `Log[10, 100]` to mean the same thing.

Note that many textbooks write `log` and mean \log_{10} just like your calculator. When the books want you to use \log_e like your computer they’ll write “`ln`” (pronounced “lawn”) meaning the *natural logarithm*.

The moral of the story is: BEWARE! Verify that you know the base of the logarithm before you get too many wrong answers and have to do it all again.

1.4 Trigonometric Functions

I've got an idea for a great invention. I'm going to get a flat piece of wood and cut it out in the shape of a circle. In the centre of the circle, I'm going to drill a hole and stick a dowel of wood in there. I think I'm going to call my new invention a *wheel*.

Okay, okay, so I ran down to the patent office and found out that the wheel has already been invented... (Apparently, Bill Gates just bought the patent rights from Michael Jackson last year...) But, since it's such a great idea let's look at one anyway. Let's drill another hole out on the edge of the wheel and stick in a handle so that it looks like the contraption in Figure 1.3. If I turn the wheel, the handle goes around in circles.

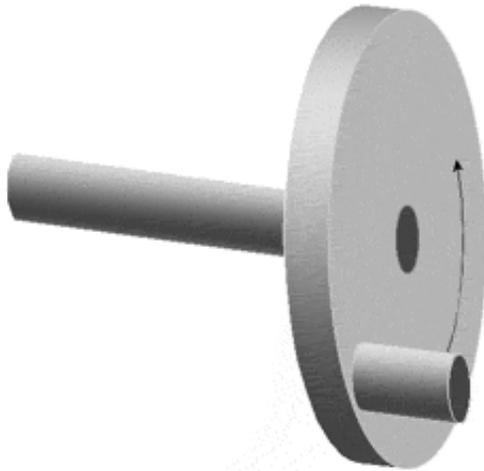


Figure 1.3: Wheel rotating counterclockwise when viewed from the side of the handle that's sticking out on the right.

Now let's think of an animation of the rotating wheel. In addition, we'll look at the height of the handle relative to the centre of the wheel. As the wheel rotates, the handle will obviously go up and down, but it will follow a specific pattern over time. If that height is graphed in time as the wheel rotates, we get a nice wave as is shown in Figure 1.4.

That nice wave tells us a couple of things about the wheel:

Firstly, if we assume that the handle is right on the edge of the wheel, it tells us the diameter of the wheel itself. The total height of the wave from the positive peak to negative trough is a measurement of the total vertical

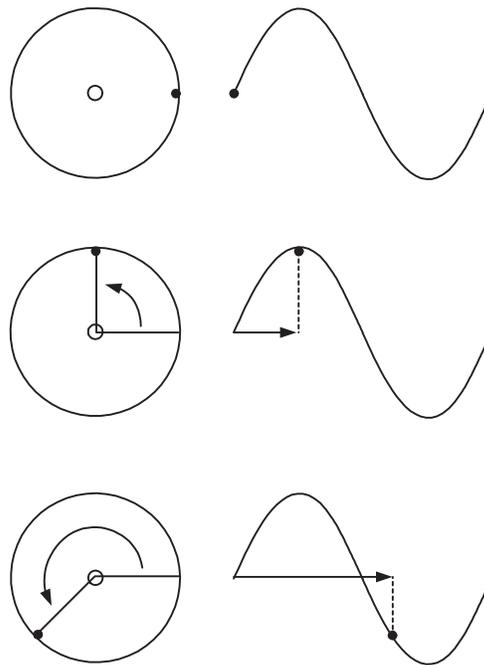


Figure 1.5: Graphs showing the relationship between the angle of rotation of the wheel and the waveform's X-axis.

motion.

There's one important thing that the wave *isn't* telling us – the direction of rotation of the wheel. If the wheel were turning clockwise instead of counterclockwise, then the wave would look exactly the same as is shown in Figure 1.6.

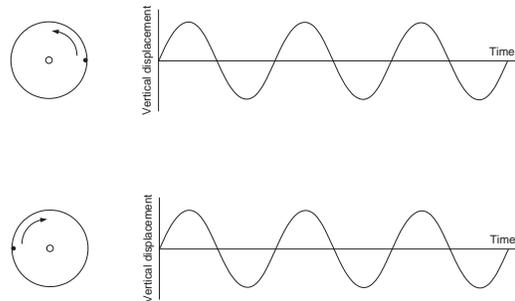


Figure 1.6: Two wheels rotating at the same speed in opposite directions resulting in the same waveform.

So, how do we get this piece of information? Well, as it stands now, we're just getting the height information of the handle. That's the same as if we were sitting to the side of the wheel, looking at its edge, watching the handle bob up and down, but not knowing anything about it going from side to side. In order to get this information, we'll have to look at the wheel along its edge from below. This will result in two waves – the sine wave that we saw above, and a second wave that shows the horizontal displacement of the handle over time as is shown in Figure 1.7.

As can be seen in this diagram, if the wheel were turning in the opposite direction as in the example in Figure 1.6, then although the vertical displacement would be the same, the horizontal displacement would be opposite, and we'd know right away that the wheel was turning in the opposite direction.

This second waveform is called a *cosine* wave (because it's the *compliment* of the *sine* wave). Notice how, whenever the sine wave is at a maximum or a minimum, the cosine wave is at 0 – in the middle of its movement. The opposite is also true – whenever the cosine is at a maximum or a minimum, the sine wave is at 0. The four points that we talked about earlier (regarding what the sine wave tells us) are also true for the cosine – we know the diameter of the wheel, the speed of its rotation, and the horizontal (not vertical) displacement of the handle at a given time or angle of rotation.

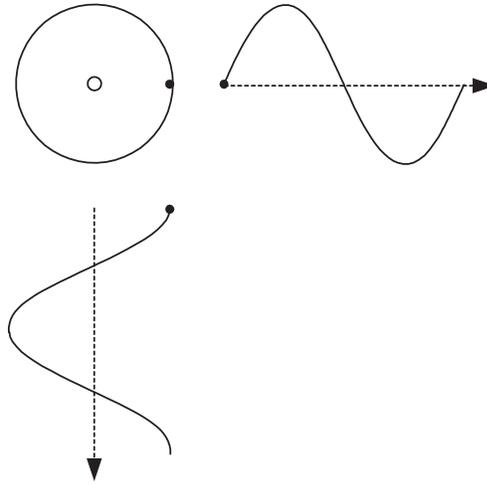


Figure 1.7: Graphs showing the relationship between the angle of rotation of the wheel and the vertical and horizontal displacements of the handle.

Keep in mind as well that if we only knew the cosine, we still wouldn't know the direction of rotation of the wheel – we need to know the simultaneous values of the sine and the cosine to know whether the wheel is going clockwise or counterclockwise.

Now then, let's assume for a moment that the circle has a radius of 1. (1 centimeter, 1 foot... it doesn't matter so long as we keep thinking in the same units for the rest of this little chat.) If that's the case then the maximum value of the sine wave will be 1 and the minimum will be -1. The same holds true for the cosine wave. Also, looking back at Figure 1.5, we can see that the value of the sine is 1 when the angle of rotation (also known as the *phase* angle) is 90° . At the same time, the value of the cosine is 0 (because there's 0 horizontal displacement at 90°). Using this, we can complete Table 1.1:

In fact, if you get out your calculator and start looking for the Sine (“sin” on a calculator) and the Cosine (“cos”) for every angle between 0 and 359° (no point in checking 360 because it'll be the same as 0 – you've made a full rotation at that point...) and plot each value, you'll get a graph that looks like Figure 1.8.

As can be seen in Figure 1.8, the sine and cosine intersect at 45° (with a value of 0.707 or $\frac{1}{\sqrt{2}}$) and at 215° (with a value of -0.707 or $-\frac{1}{\sqrt{2}}$). Also, you can see from this graph that a cosine is essentially a sine wave, but 90°

Phase (degrees)	Vertical displacement (Sine)	Horizontal displacement (Cosine)
0°	0	1
45°	0.707	0.707
90°	1	0
135°	0.707	-0.707
180°	0	-1
225°	-0.707	-0.707
270°	-1	0
315°	-0.707	0.707

Table 1.1: Values of sine and cosine for various angles

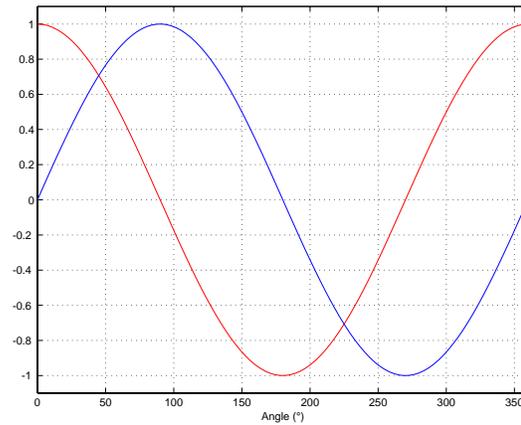


Figure 1.8: The relationship between Sine (blue) and Cosine (red) for angles from 0° to 359°.

earlier. That is to say that the value of a cosine at any angle is the same as the value of the sine 90° later. These two things provide a small clue as to another way of looking at this relationship.

Look at the first 90° of rotation of the handle. If we draw a line from the centre of the wheel to the location of the handle at a given angle, and then add lines showing the vertical and horizontal displacements as in Figure 1.7, then we get a triangle like the one shown in Figure 1.9.

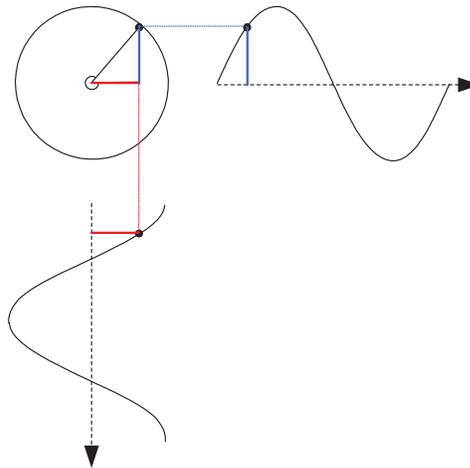


Figure 1.9: A right triangle within the rotating wheel. Notice that the value of the sine wave is the green vertical leg of the triangle, the value of the cosine is the red horizontal leg of the triangle and the diameter of the wheel (and therefore the peak values of both the sine and cosine) is the hypotenuse.

Now, if the radius of the wheel (the hypotenuse of the triangle) is 1, then the vertical line is the sine of the inside angle indicated with a red arrow. Likewise, the horizontal leg of the triangle is the cosine of the angle.

Also, we know from Pythagoras that the square of the hypotenuse of a right triangle is equal to the sum of the squares of the other two sides (remember $a^2 + b^2 = c^2$ where c is the length of the hypotenuse). In other words, in the case of our triangle above where the hypotenuse is equal to 1, then the sin of the angle squared + the cosine of the angle squared = 1 squared... This is a rule (shown below) that is true for any angle.

$$\sin^2 \alpha + \cos^2 \alpha = 1 \quad (1.9)$$

where α is any angle.

Since this is true, then when the angle is 45° , then we know that the right triangle is isosceles – meaning that the two legs other than the hypotenuse are of equal length (take a look at the graph in Figure 1.8). Not only are they the same length, but, their squares add up to 1. Remember that $a^2 + b^2 = c^2$ and that $c^2 = 1$. Therefore, with a little bit of math, we can see that the value of the sine and the cosine when the angle is 45° is $\frac{1}{\sqrt{2}}$ because it's the square root of $\frac{1}{2}$ and $\sqrt{\frac{1}{2}} = \frac{\sqrt{1}}{\sqrt{2}} = \frac{1}{\sqrt{2}}$.

1.4.1 Radians

Once upon a time, someone discovered that there is a relationship between the radius of a circle and its circumference. It turned out that, no matter how big or small the circle, the circumference was equal to the radius multiplied by 2 and multiplied again by the number 3.141592645... That number was given the name “pi” (written π) and people have been fascinated by it ever since. In fact, the number doesn't stop where I said it did – it keeps going for at least a billion places without repeating itself... but 9 places after the decimal is plenty for our purposes.

So, now we have a new little equation:

$$\text{Circumference} = 2 * \pi * r \tag{1.10}$$

where r is the radius of the circle and π is 3.141592645...

Normally we measure angles in degrees where there are 360° in a full circle, however, in order to measure this way, we need a protractor to tell us where the degrees are. There's another way to measure angles using only a ruler and a piece of string...

Let's go back to the circle above with a radius of 1. Since we have the new equation, we know that the circumference is equal to $2 * \pi * r$ – but $r = 1$, so the circumference is $2 * \pi$ (say “two pi”). Now, we can measure angles using the circumference – instead of saying that there are 360° in a circle, we can say that there are 2π *radians*. We call them radians because they're based on the radius. Since the circumference of the circle is $2\pi r$ and there are 2π radians in the circle, then 1 radian is the angle where the corresponding arc on the circle is equal to the length of the radius of the same circle.

Using radians is just like using degrees – you just have to put your calculator into a different mode. Look for a way of getting it into “RAD” instead of “DEG” (RADians instead of DEGrees). Now, remember that there are 2π radians in a circle which is the same as saying 360 degrees.

Therefore, 180° which is half of the circle is equal to π radians. 90° is $\frac{\pi}{2}$ radians and so on. You should be able to use these interchangeably in day to day conversation.

1.4.2 Phase vs. Addition

If I take any two sinusoidal waves that have the same frequency, but they have different amplitudes, and they're offset in phase, and I add them, the result will be a sinusoidal wave with the same frequency with another amplitude and phase. For example, take a look at Figure 1.10. The top plot shows one period of a 1 kHz sinusoidal wave starting with a phase of 0 radians and a peak amplitude of 0.5. The second plot shows one period of a 1 kHz sinusoidal wave starting with a phase of $\frac{\pi}{4}$ radians (45°) and a peak amplitude of 0.8. If these two waves are added together, the result, shown on the bottom, is one period of a 1 kHz sinusoidal wave with a different peak amplitude and starting phase. The important thing that I'm trying to get across here is that the frequency and wave shape stay the same – only the amplitude and phase change.

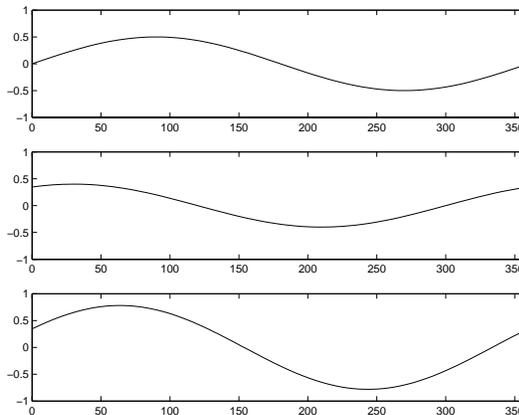


Figure 1.10: Adding two sinusoids with the same frequency and different phases. The sum of the top two waveforms is the bottom waveform.

So what? Well, most recording engineers talk about phase. They'll say things like "a sine wave, 135° late" which looks like the curve shown in Figure 1.11.

If we wanted to be a little geeky about this, we could use the equation below to say the same thing:

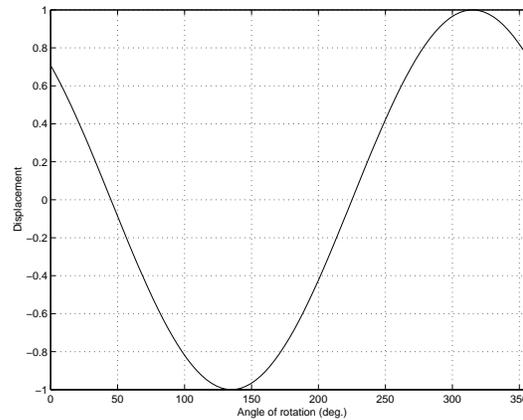


Figure 1.11: A sine wave, starting at a phase of 135° .

$$y(n) = A \sin(n + \phi) \quad (1.11)$$

which means the value of y at a given value of n is equal to A multiplied by the sine of the sum of the values n and ϕ . In other words, the amplitude y at angle n equals the sine of the angle n added to a constant value ϕ and the peak value will be A . In the above example, $y(n)$ would be equal to $1 * \sin(n + 135^\circ)$ where n can be any value.

Now, we have to be a little more geeky than that, even... We have to talk about cosine waves instead of sine waves. We've already seen that these are really the same thing, just 90° apart, so we can already figure out that a sine wave that's starting 135° late is the same as a cosine wave that's starting 45° late.

Now that we've made that transition, there is another way to describe a wave. If we scale the sine and cosine components correctly and add them together, the result will be a sinusoidal wave at any phase and amplitude we want. Take a look at the equation below:

$$A \cos(n + \phi) = a \cos(n) - b \sin(n) \quad (1.12)$$

where A is the amplitude

ϕ is the phase angle

$$a = A \cos(\phi)$$

$$b = A \sin(\phi)$$

What does this mean? Well, all it means is that we can now specify values for a and b and, using this equation, wind up with a sinusoidal waveform of any amplitude and phase that we want. Essentially, we just have an alternate way of describing the waveform.

For example, where you used to say “A cosine wave with a peak amplitude of 0.93 and $\frac{\pi}{3}$ radians (60°) late” you can now say:

$$A = 0.93$$

$$\phi = \frac{\pi}{3}$$

$$a = 0.93 * \cos\left(\frac{\pi}{3}\right) = 0.93 * 0.5 = 0.4650$$

$$b = 0.93 * \sin\left(\frac{\pi}{3}\right) = 0.93 * 0.8660 = 0.8054$$

Therefore

$$A \cos(n + 27^\circ) = 0.4650 \cos(n) - 0.8054 \sin(n) \quad (1.13)$$

So we could say that it’s the combination of an upside-down sine wave with a peak amplitude of 0.8054 and a cosine wave with a peak amplitude of 0.4650. We’ll see in Chapter 1.5 how to write this a little more easily.

Remember that, if you’re used to thinking in terms of a peak amplitude and a fixed phase offset, then this might seem less intuitive. However, if your job is to build a synthesizer that makes a sinusoidal wave with a given phase offset, you’d much rather just add an appropriately scaled cosine and sine rather than having to build a delay.

1.5 Complex Numbers

1.5.1 Whole Numbers and Integers

Once upon a time you learned how to count. You were probably taught to count your fingers... 1, 2, 3, 4 and so on. Although no one told you so at the time, you were being taught a set of numbers called *whole numbers*.

Sometime after that, you were probably taught that there's one number that gets tacked on before the ones you already knew – the number 0.

A little later, sometime after you learned about money and the fact that we don't have enough, you were taught negative numbers... -1, -2, -3 and so on. These are the numbers that are less than 0.

That collection of numbers is called *integers* – all “countable” numbers that are negative, zero and positive. So the collection is typically written

... -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5 ...

1.5.2 Rational Numbers

Eventually, after you learned about counting and numbers, you were taught how to divide. When someone said “20 divided by 5 equals 4” then they meant “if you have 20 sticks, then you could put those sticks in 4 piles with 5 sticks in each pile.” Eventually, you learned that the division of one number by another can be written as a *fraction* like $\frac{3}{1}$ or $\frac{20}{5}$ or $\frac{5}{4}$ or $\frac{1}{3}$.

If you do that division the old-fashioned way, you get numbers like this:

$$3/1 = 3.000000000 \text{ etc...}$$

$$20/5 = 4.000000000 \text{ etc...}$$

$$5/4 = 1.200000000 \text{ etc...}$$

$$1/3 = 0.333333333 \text{ etc...}$$

The thing that I'm trying to point out here is that eventually, these numbers start repeating sometime after the decimal point. These numbers are called *rational numbers*.

1.5.3 Irrational Numbers

What happens if you have a number that doesn't start repeating, no matter how many numbers you have? Take a number like the square root of 2 for example. This is a number that, when you multiply it by itself, results in the number 2. This number is approximately 1.4142. But, if we multiply 1.4142 by 1.4142, we get 1.99996164 – so 1.4142 isn't exactly the square root of 2. In fact, if we started calculating the exact square root of 2, we'd result in a number that keeps going forever after the decimal place and never repeats.

Numbers like this (π is another one...) that never repeat after the decimal are called *irrational numbers*

1.5.4 Real Numbers

All of these number types – rational numbers (which includes integers) and irrational numbers fall under the general heading of *real numbers*. The fact that these are called “real” implies immediately that there is a classification of numbers that are “unreal” – in fact this is the case, but we call them “imaginary” instead.

1.5.5 Imaginary Numbers

Let’s think about the idea of a square root. The square root of a number is another number which, when multiplied by itself is the first number. For example, 3 is the square root of 9 because $3 * 3 = 9$. Let’s consider this a little further: a positive number multiplied by itself is a positive number (for example, $4 * 4 = 16$... 4 is positive and 16 is also positive). A negative number multiplied by itself is also positive (i.e. $-4 * -4 = 16$).

Now, in the first case, the square root of 16 is 4 because $4 * 4 = 16$. (Some people would be really picky and they’ll tell you that 16 has two roots: 4 and -4. Those people are slightly geeky, but technically correct.) There’s just one small snag – what if you were asked for the square root of a negative number? There is no such thing as a number which, when multiplied by itself results in a negative number. So asking for the square root of -16 doesn’t make sense. In fact, if you try to do this on your calculator, it’ll probably tell you that it gets an error instead of producing an answer.

Mathematicians as a general rule don’t like loose ends – they aren’t the type of people who leave things lying around... and having something as simple as the square root of a negative number lying around unanswered got on their nerves so they had a bunch of committee meetings and decided to do something about it. Their answer was to invent a new number called i (although some people – namely physicists and engineers – call it j just to screw everyone up... we’ll stick with j for now, but we might change later...)

“What is j ?” I hear you cry. Well, j is the square root of -1. Of course, there is no number that is the square root of -1, but since that answer is inadequate, j will do the trick.

“Why is it called j ?” I hear you cry. That one is simple – j stands for “imaginary” because the square root of -1 isn’t real, it’s an imaginary number that someone just made up.

Now, remember that $j * j = -1$. This is useful for any square root of any negative number, you just calculate the square root of the number pretending that it was positive, and then stick an j after it. So, since the square root of 16, abbreviated $\sqrt{16} = 4$ and $\sqrt{-1} = j$, then $\sqrt{-16} = j4$. Let's do a couple:

$$\sqrt{-9} = j3 \quad (1.14)$$

$$\sqrt{-4} = j2 \quad (1.15)$$

Another way to think of this is $\sqrt{-a} = \sqrt{-1 * a} = \sqrt{-1} * \sqrt{a} = j\sqrt{a}$ so:

$$\sqrt{-9} = \sqrt{-1} * \sqrt{9} = j * \sqrt{9} = j3 \quad (1.16)$$

Of course, this also means that

$$j3 * j3 = (3 * 3) * (j * j) = -1 * 9 = -9 \quad (1.17)$$

1.5.6 Complex numbers

Now that we have real and imaginary numbers, we can combine them to create a *complex number*. Remember that you can't just mix real numbers with imaginary ones – you keep them separate most of the time, so you see numbers like

$$3 + j2$$

This is an example of a complex number that contains a real component (the 3) and an imaginary component (the $j2$). In many cases, these numbers are further abbreviated with a single Greek character, like α or β , so you'll see things like

$$\alpha = 3 + j2$$

but for the purposes of what we'll do in this book, I'm going to stick with either writing the complex number the long way or I'll use a bold character so, instead, I'll use

$$\mathbf{A} = 3 + j2$$

1.5.7 Complex Math Part 1 – Addition

Let's say that you have to add complex numbers. In this case, you have to keep the real and imaginary components separate, but you just add the separate components separately. For example:

$$(5 + j3) + (2 + j4) \quad (1.18)$$

$$(5 + 2) + (j3 + j4) \quad (1.19)$$

$$(5 + 2) + j(3 + 4) \quad (1.20)$$

$$7 + j7 \quad (1.21)$$

If you'd like the short-cut rule, it's

$$(a + jb) + (c + jd) = (a + c) + j(b + d) \quad (1.22)$$

1.5.8 Complex Math Part 2 – Multiplication

The multiplication of two complex numbers is similar to multiplying regular old real numbers. For example:

$$(5 + j3) * (2 + j4) \quad (1.23)$$

$$((5 * 2) + (j3 * j4)) + ((5 * j4) + (j3 * 2)) \quad (1.24)$$

$$((5 * 2) + (j * j * 3 * 4)) + (j(5 * 4) + j(3 * 2)) \quad (1.25)$$

$$(10 + (12 * -1)) + (j20 + j6) \quad (1.26)$$

$$(10 - 12) + j26 \quad (1.27)$$

$$- 2 + j26 \quad (1.28)$$

The shortened rule is:

$$(a + jb)(c + jd) = (ac - bd) + j(ad + bc) \quad (1.29)$$

1.5.9 Complex Math Part 3 – Some Identity Rules

There are a couple of basic rules that we can get out of the way at this point when it comes to complex numbers. These are similar to their corresponding rules in normal mathematics.

Commutative Laws

This law says that the order of the numbers doesn't matter when you add or multiply. For example, $3 + 5$ is the same as $5 + 3$, and $3 * 5 = 5 * 3$. In the case of complex math:

$$(a + jb) + (c + jd) = (c + jd) + (a + jb) \quad (1.30)$$

and

$$(a + jb) * (c + jd) = (c + jd) * (a + jb) \quad (1.31)$$

Associative Laws

This law says that, when you're adding more than two numbers, it doesn't matter which two you do first. For example $(2 + 3) + 5 = 2 + (3 + 5)$. The same holds true for multiplication.

$$((a + jb) + (c + jd)) + (e + jf) = (a + jb) + ((c + jd) + (e + jf)) \quad (1.32)$$

and

$$((a + jb) * (c + jd)) * (e + jf) = (a + jb) * ((c + jd) * (e + jf)) \quad (1.33)$$

Distributive Laws

This law says that, when you're multiplying a number by the sum of two other numbers, it's the same as adding the results of multiplying the numbers one at a time. For example, $2 * (3 + 4) = (2 * 3) + (2 * 4)$. In the case of complex math:

$$(a + jb) * ((c + jd) + (e + jf)) = ((a + jb) * (c + jd)) + ((a + jb) * (e + jf)) \quad (1.34)$$

Identity Laws

These are laws that are pretty obvious, but sometimes they help out. The corresponding laws in normal math are $x + 0 = x$ and $x * 1 = x$.

$$(a + jb) + 0 = (a + jb) \quad (1.35)$$

and

$$(a + jb) * 1 = (a + jb) \quad (1.36)$$

1.5.10 Complex Math Part 4 – Inverse Rules

Additive Inverse

Every number has what is known as an *additive inverse* – a matching number which when added to its partner equals 0. For example, the additive inverse of 2 is -2 because $2 + -2 = 0$. This additive inverse can be found by multiplying the number by -1 so, in the case of complex numbers:

$$(a + jb) + (-a - jb) = 0 \quad (1.37)$$

Therefore, the additive inverse of $(a + jb)$ is $(-a - jb)$

Multiplicative Inverse

Similarly to additive inverses, every number has a matching number, which, when the two are multiplied, equals 1. The only exception to this is the number 0, so, if x does not equal 0, then the *multiplicative inverse* of x is $\frac{1}{x}$ because $x * \frac{1}{x} = * \frac{x}{x} = 1$. Some books write this as $x * x^{-1} = 1$ because $x^{-1} = \frac{1}{x}$. In the case of complex math, things are unfortunately a little different because 1 divided by a complex number is... well... complex.

if $(a + jb)$ is not equal to 0 then:

$$\frac{1}{a + jb} = \frac{a - jb}{a^2 + b^2} \quad (1.38)$$

We won't worry too much about how that's calculated, but we can establish that it's true by doing the following:

$$(a + jb) * \frac{a - jb}{a^2 + b^2} \quad (1.39)$$

$$\frac{(a + jb)(a - jb)}{a^2 + b^2} \quad (1.40)$$

$$\frac{a^2 + b^2}{a^2 + b^2} \quad (1.41)$$

$$1 \quad (1.42)$$

There's one interesting thing that results from this rule. What if we looked for the multiplicative inverse of j ? In other words, what is $\frac{1}{j}$? Well, let's use the rule above and plug in $(0 + 1j)$.

$$\frac{1}{a + jb} = \frac{a - jb}{a^2 + b^2} \quad (1.43)$$

$$\frac{1}{0 + j1} = \frac{0 - j1}{0^2 + 1^2} \quad (1.44)$$

$$\frac{1}{j} = \frac{-1j}{1} \quad (1.45)$$

$$\frac{1}{j} = -j \quad (1.46)$$

Weird, but true.

Dividing complex numbers

The multiplicative inverse rule can be used if you're trying to divide. For example, $\frac{a}{b}$ is the same as $a * \frac{1}{b}$ therefore:

$$\frac{a + jb}{c + jd} \quad (1.47)$$

$$(a + jb) * \frac{1}{c + jd} \quad (1.48)$$

$$\frac{(a + jb)(c - jd)}{c^2 + d^2} \quad (1.49)$$

1.5.11 Complex Conjugates

Complex math has an extra relationship that doesn't really have a corresponding equivalent in normal math. This relationship is called a *complex conjugate* but don't worry – it's not terribly complicated. The complex conjugate of a complex number is defined as the same number, but with the polarity of the imaginary component inverted. So:

the complex conjugate of $(a + bj)$ is $(a - bj)$

A complex conjugate is usually abbreviated by drawing a line over the number. So:

$$\overline{(a + jb)} = (a - jb) \quad (1.50)$$

1.5.12 Absolute Value (also known as the Modulus)

The absolute value of a complex number is a little weirder than what we usually think of as an absolute value. In order to understand this, we have to look at complex numbers a little differently:

Remember that $j * j = -1$. Also, remember that, if we have a cosine wave and we delay it by 90° and then delay it by another 90° , it's the same as inverting the polarity of the cosine, in other words, multiplying the cosine by -1. So, we can think of the imaginary component of a complex number as being a real number that's been rotated by 90° , we can picture it as is shown in Figure 1.12.

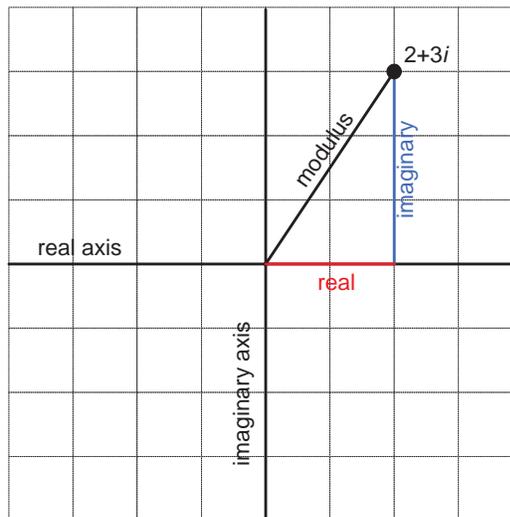


Figure 1.12: The relationship between the real and imaginary components for the number $(2 + 3j)$. Notice that the X and Y axes have been labeled the “real” and “imaginary” axes.

Notice that Figure 1.12 actually winds up showing three things. It shows the real component along the x-axis, the imaginary component along the y-axis, and the *absolute value* or *modulus* of the complex number as the hypotenuse of the triangle.

This should make the calculation for determining the modulus of the complex number almost obvious. Since it's the length of the hypotenuse of the right triangle formed by the real and imaginary components, and since we already know the Pythagorean theorem then the modulus of the complex number $(a + jb)$ is

$$\text{modulus} = \sqrt{a^2 + b^2} \quad (1.51)$$

Given the values of the real and imaginary components, we can also calculate the angle of the hypotenuse from horizontal using the equation

$$\phi = \arctan \frac{\text{imaginary}}{\text{real}} \quad (1.52)$$

$$\phi = \arctan \frac{b}{a} \quad (1.53)$$

This will come in handy later.

1.5.13 Complex notation or... Who cares?

This is probably the most important question for us. Imaginary numbers are great for mathematicians who like wrapping up loose ends that are incurred when a student asks “what’s the square root of -1?” but what use are complex numbers for people in audio? Well, it turns out that they’re used all the time, by the people doing analog electronics as well as the people working on digital signal processing. We’ll get into how they apply to each specific field in a little more detail once we know what we’re talking about, but let’s do a little right now to get a taste.

In the chapter that introduces the trigonometric functions sine and cosine, we looked at how both functions are just one-dimensional representations of a two-dimensional rotation of a wheel. Essentially, the cosine is the horizontal displacement of a point on the wheel as it rotates. The sine is the vertical displacement of the same point at the same time. Also, if we know one of these two components, we know

1. the diameter of the wheel and
2. how fast it’s rotating, but we need to know both components to know
3. the direction of rotation.

At any given moment in time, if we froze the wheel, we’d have some contribution of these two components – a cosine component and a sine component for a given angle of rotation. Since these two components are effectively identical functions that are 90° apart (for example, a sine wave is the same as a cosine that’s been delayed by 90°) and since we’re thinking of the real and imaginary components in a complex number as being 90° apart,

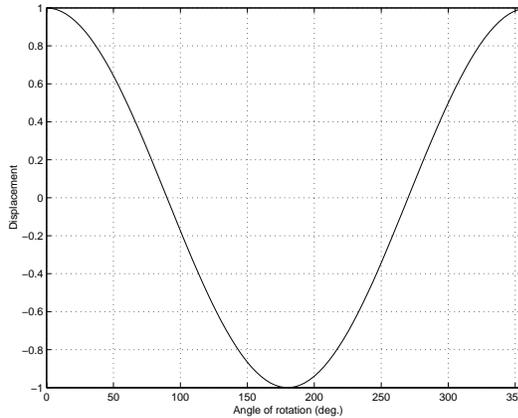


Figure 1.13: A signal consisting only of a cosine wave

then we can use complex math to describe the contributions of the sine and cosine components to a signal.

Huh? Let's look at an example. If the signal we wanted to look at a signal that consisted only of a cosine wave as is shown in Figure 1.13, then we'd know that the signal had 100% cosine and 0% sine. So, if we express the cosine component as the real component and the sine as the imaginary, then what we have is:

$$1 + 0j \quad (1.54)$$

If the signal was an upside-down cosine, then the complex notation for it would be $(-1 + 0j)$ because it would essentially be a cosine * -1 and no sine component. Similarly, if the signal was a sine wave, it would be notated as $(0 - 1j)$.

This last statement should raise at least one eyebrow... Why is the complex notation for a positive sine wave $(0 - 1j)$? In other words, why is there a negative sign there to represent a positive sine component? Well... Actually there is no good explanation for this at this point in the book, but it should become clear when we discuss a concept known as the Fourier Transform in Section 8.2.

This is fine, but what if the signal looks like a sinusoidal wave that's been delayed a little like the one in Figure 1.14?

This signal was created by a specific combination of a sine and cosine wave. In fact, it's 70.7% sine and 70.7% cosine. (If you don't know how I arrived that those numbers, check out Equation 1.11.) How would you

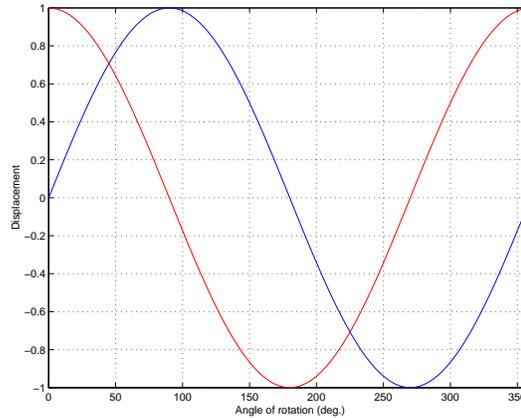


Figure 1.14: A signal consisting of a combination of attenuated cosine and sine waves with the same frequency.

express this using complex notation? Well, you just look at the relative contributions of the two components as before:

$$0.707 - 0.707j \quad (1.55)$$

It's interesting to notice that, although Figure 1.14 is actually a combination of a cosine and a sine with a specific ratio of amplitudes (in this case, both at 0.707 of “normal”), the result looks like a sine wave that's been shifted in phase by -45° or $-\frac{\pi}{4}$ radians (or a cosine that's been phase-shifted by 45° or $\frac{\pi}{4}$ radians). In fact, this is the case – any phase-shifted sine wave can be expressed as the combination of its sine and cosine components with a specific amplitude relationship.

Therefore, any sinusoidal waveform with any phase can be simplified into its two elemental components, the cosine (or real) and the sine (or imaginary). Once the signal is broken into these two constituent components, it cannot be further simplified.

If we look at the example at the end of Section 1.4, we calculated using the equation

$$A \cos(n + \phi) = a \cos(n) - b \sin(n) \quad (1.56)$$

that a cosine wave with a peak amplitude of 0.93 and a delay of $\frac{\pi}{3}$ radians was equivalent to the combination of a cosine wave with a peak amplitude of 0.4650 and an upside-down sine wave with a peak amplitude of 0.8054. Since

the cosine is the real component and the sine is the imaginary component, this can be expressed using the complex number as follows:

$$0.93 \cos\left(n + \frac{\pi}{3}\right) = 0.4650 \cos(n) - 0.8054 \sin(n) \quad (1.57)$$

which is represented as

$$0.4650 + j 0.8054$$

which is a much simpler way of doing things. (Notice that I flipped the “-” sign to a “+.”) For more information on this, check out *The Scientist and Engineer’s Guide to Digital Signal Processing* available at www.dspguide.com

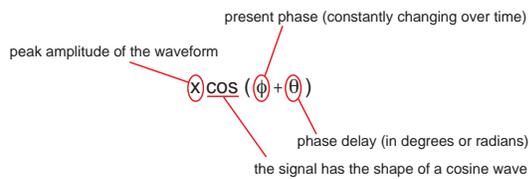


Figure 1.15: A quick guide to what things mean when you see a cosine wave expressed as this type of equation.

1.6 Euler's Identity

So far, we've looked at logarithms with a base of 10. As we've seen, a logarithm is just an exponent backwards, so if $A^B = C$ then $\log_A C = B$. Therefore $\log_{10} 100 = 2$.

There is a beast in mathematics known as a *natural logarithm* which is just like a regular old everyday logarithm except that the base is a very specific number – it's e . “What's e ?” I hear you cry... Well, just like π is an irrational number close to 3.14159, e is an irrational number that is pretty close to 2.718281828459045235360287 but it keeps on going after the decimal place forever and ever. (If it didn't, it wouldn't be irrational, would it?) How someone arrived at that number is pretty much inconsequential – particular if we want to avoid using calculus, but if you'd like to calculate it, and if you've got a lot of time on your hands, the math is:

$$e = \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \dots \quad (1.58)$$

(If you're not familiar with the mathematical expression “!” you don't have to panic! It's short for *factorial* and it means that you multiply all the whole numbers up to and including the number. For example, $5! = 1 * 2 * 3 * 4 * 5$.)

How is this e useful to us? Well, there are a number of reasons, but one in particular. It turns out that if we raise e to an exponent x , we get the following.

$$e^x = \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots \quad (1.59)$$

Unfortunately, this isn't really useful to us. However, if we raise e to an exponent that is an imaginary number, something different happens.

$$e^{j\theta} = \cos(\theta) + j \sin(\theta) \quad (1.60)$$

This is known as *Euler's identity* or *Euler's formula*.

Notice now that, by putting an i up there in the exponent, we have an equation that links trigonometric functions to an algebraic function. This *identity*, first proved by a man named *Leonhard Euler*, is really useful to us.

There's just a couple of extra things to take note of:

Since $\cos(\pi) = -1$ and $\sin(\pi) = 0$ then:

$$e^{j\pi} = \cos(\pi) + j \sin(\pi) \quad (1.61)$$

$$e^{j\pi} = -1 + 0 \quad (1.62)$$

therefore

$$e^{j\pi} + 1 = 0 \quad (1.63)$$

1.6.1 Who cares?

Here's where the beauty of all this math actually becomes apparent. What we've essentially done is to make things look complicated in order to simplify working with the equations.

We saw in the last two chapters how an arbitrary wave like a cosine with a peak amplitude of 0.93 and $\frac{\pi}{3}$ radians late could be expressed in a number of different ways. We can say $0.93 \cos(n + \frac{\pi}{3})$ or $0.4650 \cos(n) - 0.8054 \sin(n)$ or we can represent it with the complex number $0.4650 + j0.8054$. I argued at the time that using these weird complex notations would make life simpler. For example, it's easier to add two complex numbers to get a third complex number than it is to try and add two waves with different amplitudes and delays. However, if you were the arguing type, you'd have pointed out that multiplying two complex numbers really isn't all that attractive a proposition. This is where Euler becomes our friend.

Using Euler's identity, we can convert the complex representation of our waveform into a complex exponential notation as shown below

$$0.93 \cos(n + \frac{\pi}{3}) = 0.4650 \cos(n) - 0.8054 \sin(n) \quad (1.64)$$

Which is represented as

$$0.4650 + j0.8054 = 0.93e^{j\frac{\pi}{3}} \quad (1.65)$$

There's a *really* important thing to remember here. The two values shown in Equation 1.65 are only *representations* of the values shown in Equation 1.64. They are not the same thing mathematically.

In other words, if you calculated $0.93 \cos(n + \frac{\pi}{3})$ and $0.4650 \cos(n) - 0.8054 \sin(n)$, you'd get the same answer. If you calculated $0.4650 + j0.8054$ and $0.93e^{j\frac{\pi}{3}}$, you'd get the same answer. BUT the two answers that you just got would not be the same. We just use the notation to make life a little simpler.

The nice thing about this, and the thing to remember is the way that the 0.93 and the $\frac{\pi}{3}$ on the left-hand side of Equation 1.64 correspond to the same numbers on the right-hand side of Equation 1.65.

Of course, now the question is “Why the hell do we go through all of this hassle?” Well, the answer lies in the simplicity of dealing with complex numbers represented as exponents, but I will leave it to other books to explain this. A good place to start with this question is *The Scientist and Engineer’s Guide to Digital Signal Processing* by Steven W. Smith and found at www.dspguide.com.

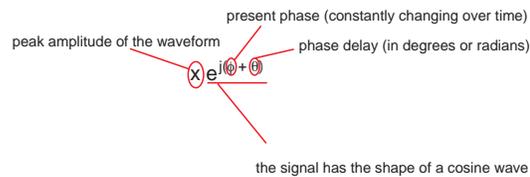


Figure 1.16: A quick guide to what things mean when you see a cosine wave expressed in exponential notation.

1.7 Binary, Decimal and Hexadecimal

1.7.1 Decimal (Base 10)

Let's count. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ... Fun isn't it? When we count we're using the words "one," "two" and so on (or the symbols "1," "2" and so on) to represent a quantity of things. Two eyes, two ears, two teeth... The word "two" or the symbol "2" is just used to denote how many of something we have.

One of the great things about our numbering system is that we don't need a symbol for every quantity – we recycle. What I mean by this is, we have individual symbols that we can write down which indicate the quantities zero through to nine. Once we get past nine, however, we have to use two symbols to denote the number ten – we write "one zero" but we mean "ten." This is very efficient for lots of reasons, not the least of which is the fact that, if we had to have a different symbol for every number, our laptop computers would have to be much bigger to accommodate the keyboard.

This raises a couple of issues:

1. why do we only have ten of those symbols?
2. how does the recycling system work?

We have ten symbols because most of us have ten fingers. When we learned to count, we started by counting our fingers. In fact, another word for "finger" is "digit" which is why we use the word *digit* for the symbols that we use for counting – the digit "0" represents the quantity (or the number) "zero."

How does the system work? This won't come as a surprise, but we'll go through it anyway... Let's look at the number 7354. What does this represent? Well, one way to write is to say "seven thousand, three hundred and fifty-four." In fact, this tells us right away how the system works. Each digit represents a different "power" of ten... I'll explain:

So, we can see that if we write a string of digits together, each of the digits is multiplied by a power of ten where the placement of the digit in question determines the exponent. The right-most digit multiplied by the 0th power of ten, the next digit to the left is multiplied by the 1st power of ten, the next is multiplied by the 2nd power of ten and so on until you run out of digits. Also, we can see why we're taught phrases like "the thousand's place" – the digit 7 in the number above is multiplied by 1000 (10^3 because of its location in the written number – it's in the "thousand's place")

7	3	5	4
Thousand's place	Hundred's place	Ten's place	One's place
$7 * 1000+$	$3 * 100+$	$5 * 10+$	$4 * 1$
$7 * 10^3+$	$3 * 10^2+$	$5 * 10^1+$	$4 * 10^0$
	$= 7354$		

Table 1.2: An illustration of how the location of a digit within a number determines the power of ten by which it's multiplied.

This is a very efficient method of writing down a number because each time you add an extra digit, you increase the number of possible numbers you can represent by a factor of ten. For example, if I have three digits, I can represent a total of one thousand different numbers (000 – 999). If I add one more digit and make it a four-digit number, I can represent ten thousand different numbers (0000 – 9999) – an increase by a factor of ten.

This particular property of the system makes some specific mathematical functions very easy. If you want to multiply a number by ten, you just stick a “0” on the right end of it. For example, 346 multiplied by ten is 3460. By adding a zero, you shift all the digits to the left and the multiplication is automatic. In fact, what you're doing here is using the way you write the number to do the multiplication for you – by shifting digits, you wind up multiplying the digits by new powers of ten in your head when you read the number aloud.

Similarly, if you don't mind a little inaccuracy, you can divide by ten by removing the right-most digit. This is a little less effective because it's not perfect – you are throwing some details away – but it's pretty close. For example, 346 divided by ten is pretty close to 34.

We typically call this system the *decimal* numbering system (because the root “dec” means “ten” – therefore words like “decimate” – to reduce in number by a power of ten). There are those among us, however, who like our lives to be a little more ordered – they use a different name for this system. They call it *base 10* – indicating that there are a total of ten different digits at our disposal and that the location of the digits in a number correspond to some power of 10.

1.7.2 Binary (Base 2)

Imagine if we all had only two fingers instead of ten. We would probably have wound up with a different way of writing down numbers. We have ten digits (0 – 9) because we have ten fingers. If we only had two fingers, then

it's reasonable to assume that we'd only have two digits – 0 and 1. This means that we have to re-think the way we construct the recycling of digits to make bigger numbers.

Let's count using this new two-digit system...

000 (zero)

001 (one)

Now what? Well, it's the same as in decimal, we increase our number to two digits and keep going.

010 (two)

011 (three)

100 (four)

This time, because we only have two digits, we multiply the digits in different locations by powers of two instead of powers of ten. So, for a big number like 10011, we can follow the same procedure as we did for base 10.

<i>1</i>	<i>0</i>	<i>0</i>	<i>1</i>	<i>1</i>
Sixteen's place	Eight's place	Four's place	Two's place	One's place
<i>1</i> * 2 ⁴ +	<i>0</i> * 2 ³ +	<i>0</i> * 2 ² +	<i>1</i> * 2 ¹ +	<i>1</i> * 2 ⁰
1 * 16 +	0 * 8 +	0 * 4 +	1 * 2 +	1 * 1
16 +	0 +	0 +	2 +	1
		= 19		

Table 1.3: A breakdown of an arbitrary binary number, converting it to a decimal number. Note that binary and decimal are used simultaneously in this table: to distinguish the two, all binary numbers are in italics.

So, the binary number 10011 represents the same quantity as the decimal number 19. Remember, all we've done is to change the method by which we're writing down the same thing. "19" in base 10 and "10011" in base 2 both mean "nineteen."

This would be a good time to point out that if we add an extra digit to our binary number, we increase the number of quantities we can represent by a factor of two. For example: if we have a three-digit binary number, we can represent a total of eight different numbers (000 – 111 or zero to seven). If we add an extra digit and make it a four-digit number we can represent sixteen different quantities (0000 – 1111 or zero to fifteen).

There are a lot of reasons why this system is good. For example, let's say that you had to send a number to a friend using only a flashlight to communicate. One smart way to do this would be to flash the light on and off with a short "on" corresponding to a "0" and a long "on" corresponding

to a “1” – so the number nineteen would be “long – short – short – long – long.” Of course another way would be to bang your flashlight on a table nineteen times – or you could write the number “19” on the flashlight and throw it to your friend...

In the world of digital audio we typically use slightly different names for things in the binary world. For starters, we call the binary digits *bits* (get it? *binary* digits). Also, we don’t call a binary number a “number” – we call it a *binary word*. Therefore “1011010100101011” is a sixteen-bit binary word.

Just like in the decimal system, there are a couple of quick math tasks that can be accomplished by shifting the bits (programmers call this *bit shifting* but be careful when you’re saying this out loud to use it in context, thereby avoiding people getting confused and thinking that you’re talking about dogs...)

If we take a binary word and bit shift it one place to the left (i.e. 111 becoming 1110) we are multiplying by two (in the previous example, 111 is seven, 1110 is fourteen).

Similarly, if we bit shift to the right, removing the right-most bit, we are dividing by two quickly, but frequently inaccurately. (i.e. 111 is seven, 11 is 3 – nearly half of seven) Note that, if it’s an even number (therefore ending in a “0”) a bit shift to the right will be a perfect division by zero (i.e. 1110 is fourteen, 111 is seven – half of fourteen). So if we bit shift for division, half of the time we’ll get the correct answer and the other half of the time we’ll wind up ignoring a remainder of a half. (sorry... I couldn’t resist)

1.7.3 Hexadecimal (Base 16)

The advantage of binary is that we only need two digits (0 and 1) or two “states” (on and off, short and long etc...) to represent the binary word. The disadvantage of binary is that it takes so many bits to represent a big number. For example, the fifteen-bit word 111111111111111 is the same as the decimal number 32768. I only need five digits to express in decimal the same number that takes me fifteen digits in binary. So, let’s invent a system that has too much of a good thing: we’ll go the other way and pretend we all have sixteen fingers.

Let’s count again: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 ... now what? We can’t go to “10” yet because we have six more fingers left... so we need some new symbols to take care of the gap. We’ll use letters! A, B, C, D, E, F, 10, 11, 12, huh?

Remember, we’re assuming that we have sixteen digits to use, therefore

there have to be sixteen individual symbols to represent the numbers zero through to fifteen. Therefore, we can make the Table 1.4:

Number	Decimal	Hexadecimal	Binary
zero	0	0	0000
one	1	1	0001
two	2	2	0010
three	3	3	0011
four	4	4	0100
five	5	5	0101
six	6	6	0110
seven	7	7	0111
eight	8	8	1000
nine	9	9	1001
ten	10	A	1010
eleven	11	B	1011
twelve	12	C	1100
thirteen	13	D	1101
fourteen	14	E	1110
fifteen	15	F	1111
sixteen	16	10	10000

Table 1.4: The numbers zero through to sixteen and the corresponding representations in decimal, hexadecimal, and binary.

So, now we wind up with these strange numbers that include the letters *A* through *F*. So we'll see something like *3D4A*. What number is this exactly?

If this seems a little confusing at this point, don't panic. It does for everyone. I think that the confusion with hexadecimal arises from the fact that it's so close to decimal – you can have the number 246 in decimal and the number 246 in hexadecimal – but these are not the same number, so you have to translate. (for example, the German word for “poison” is “Gift” – so if you're reading in German, this is not a word that you should think in English. An English “gift” and a German “Gift” are different things... hopefully...)

Of course, this raises the question “Why would we use such a confusing system in the first place!?” The answer actually lies back in the binary system. All of our computers and DSP and digital audio everything use the binary system to fire numbers around. This is inescapable. The problem is that those binary words are just so long to write down that, if you had

<i>3</i>	<i>D</i>	<i>4</i>	<i>A</i>
4096's place	256's place	16's place	1's place
$3 * 16^3 +$	$D * 16^2 +$	$4 * 16^1 +$	$A * 16^0$
$3 * 4096 +$	$D * 256 +$	$4 * 16 +$	$A * 1$
$3 * 4096 +$	$13 * 256 +$	$4 * 16 +$	$10 * 1$
12288 +	3328 +	64 +	10
	= 15690		

Table 1.5: A breakdown of an arbitrary hexadecimal number, converting it to a decimal number. Note that hexadecimal and decimal are used simultaneously in this table: to distinguish the two, all hexadecimal numbers are in italics.

to write them in a book, you'd waste a lot of paper. You could translate the numbers into decimal, but there's no correlation between binary and decimal – it's difficult to translate. However, check back to Table 3. Notice that going from the number fifteen to the number sixteen results in the hexadecimal number going from a 1-digit number to a 2-digit number. Also notice that, at the same time, the binary word goes from 4 bits to 5. This is where the magic lies. A single hexadecimal digit (0 – F) corresponds directly to a four-bit binary word (0000 – 1111). Not only this, but if you have a longer binary word, you can slice it up into four-bit sections and represent each section with its corresponding hexadecimal digit. For example, take the number 38069:

1001010010110101

Slice this number into 4-bit sections (start slicing from the right)

1001 0100 1011 0101

Now, look up the corresponding hexadecimal equivalents for each 4-bit section using Table 1.4:

9 4 B 5

94B5

So, as can be seen from the example above, there is a direct relationship between each 4-bit “slice” of the binary word and a corresponding hexadecimal number. If we were to try to convert the binary word into decimal, it would be much more of a headache. Since this translation is so simple, and because we use one quarter the number of digits, you'll often see hexadecimal used to denote numbers that are actually sent through the computer as a binary word.

1.7.4 Other Bases

This chapter only describes 3 different numbering systems, base 10, base 2 and base 16, but there are lots of others – even some that used to be used, that aren't anymore, but still have vestigages hidden in our language. “Four score and seven years ago” means “ $4 * 20 + 7$ years ago” or “87 years ago.” This is a leftover from the days when time was measured in base 20. The concept of a week is a form of base 7: 3 weeks is $3 * 7 = 21$ days. 12 inches in a foot (base 12), 14 pounds in a stone (base 14), 12 eggs in a dozen (base 12), 3 feet in a yard (base 3). We use these strange systems all the time without even really knowing it.

1.8 Intuitive Calculus

Warning! This chapter is not intended to teach you how to “do” calculus. It’s just here to give you an intuitive feel for what’s being calculated when you see a nasty-looking equation. If you want to learn calculus, this is probably not going to help you at all...

Calculus can be thought of as math that can cope with the idea of infinity. In normal, everyday math, we generally stick with finite numbers because they’re easier to deal with. Whenever a problem starts to use infinity, we just bail out and say something like “that’s undefined” or “you can’t do that.” For example, let’s think about division for a moment. If you take the equation $y = \frac{1}{x}$, then you know that, the smaller x gets, the bigger y becomes. The problem is that, as x approaches 0, y approaches infinity which makes people nervous. Consequently, if $x = 0$, then we just back away and say “you can’t divide by zero” and your calculator gives you a little complaint like “ERROR.” Calculus lets us cope with this minor problem.

1.8.1 Function

When you have an *equation*, you are saying that something is equal to something else. For example, look at Equation 1.66.

$$y = 2x \tag{1.66}$$

This says that the value of y is calculated by getting a value for x and multiplying that by 2 (I *really* hope that this is not coming as a surprise...).

A *function* is basically the busy side of an equation. Frequently, you will see books talking about a function $f(x)$ which just means that you do something to x to come up with something else. For example, in Equation 1.66, the function $f(x)$ (when you read this out loud, you say “ f of x ”) is $2x$. Of course, this can be much more complicated, but it can still just be thought of as a function - do some math using x and you’ll get your answer.

1.8.2 Limit

Let’s pretend that we’re standing in a concert hall with a very long reverb time. If I clap my hands and create a sound, it starts dying away as soon as I’ve stopped making the noise. Each second that goes by, the sound in the concert hall gets closer and closer to no sound.

One interesting thing about this model is that, if it were true, the sound pressure level would be reduced each second and, since half of something

can never equal nothing, there will be sound in the room forever, always getting quieter and quieter. The level of sound is always getting closer and closer to 0, but it never actually gets there.

This is the idea behind a *limit*. In this particular example, the limit of the sound pressure level in the decaying reverb is 0 – we never get to 0, but we always get closer to it.

There are lots of things in nature that follow this idea of a limit. The radioactive half-life of a material is one example. (The remaining principal on my house loan feels like another example...)

We won't do anything with limits directly, but we'll use them below. Just remember that a limit is a boundary that is never reached, but you can always get closer to it.

Think about Equation 1.67.

$$y = \frac{1}{x} \tag{1.67}$$

This is a pretty simple equation that says that y is inversely proportional to x . Therefore, if x gets bigger, y gets smaller. For example, if we calculate the value of y in this equation for a number of values of x , we'll get a graph that looks like Figure 1.17.

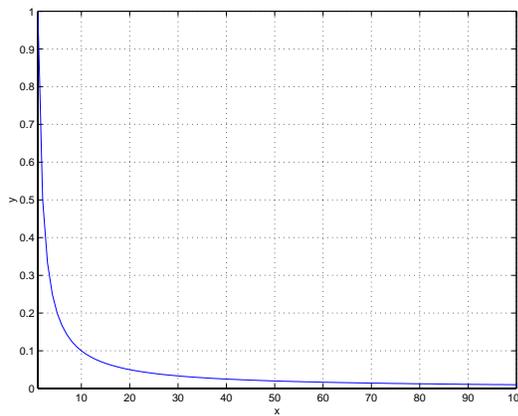


Figure 1.17: A graph of Equation 1.67

As x gets bigger and bigger, y will get closer and closer to 0, but it will never reach it. If $x = \infty$ then $y = 0$, but you don't have an ∞ button on your calculator. If x is less than ∞ but greater than 0, then y has to be a number that is greater than 0 because 1 divided by something can never be

nothing.

This is exactly the idea of a *limit*, the first concept to learn when delving into the world of calculus. It's the number that a function (like $\frac{1}{x}$ for example) gets closer and closer to, but never reaches. In the case of the function $\frac{1}{x}$, its limit is 0 as x approaches ∞ .) For example, take a look at Equation 1.68.

$$z = \lim_{x \rightarrow \infty} \frac{1}{x} \quad (1.68)$$

Equation 1.68 says “ z is equal to the number that the function $\frac{1}{x}$ approaches as x gets closer to ∞ .” This does not mean that x will ever get to ∞ , but that it will forever get closer to it.

In the example above, x is getting closer and closer to ∞ but this isn't always the case in a limit. For example, Equation 1.69 shows that you can have a limit where a number is getting closer and closer to something other than ∞ .

$$\lim_{x \rightarrow 0} \frac{\sin(x)}{x} = 1 \quad (1.69)$$

If $x = 0$, then we get a nasty number from $f(x)$, but as x approaches 0, then $f(x)$ approaches 1 because $\sin(x)$ gets closer and closer to x as x gets closer and closer to 0.

As I said in the introduction, calculus is just math than can cope with infinity. In the case of limits, we're talking about numbers that we get infinitely close to, but never reach.

1.8.3 Derivation and Slope

Back in Chapter 1.1 we looked at how to find the slope of a straight line, but what about if the line is a curve? Well, this gets us into a small problem because, if the line is curved, the its slope is different in different places. For example, take a look at the sinusoidal curve in Figure 1.18.

Our goal is to find the slope at the point marked on the plot where $x = 2$. In other words, we're looking for the slope of the *tangent* of the curve at $x = 2$ as is shown in Figure 1.19.

We could estimate the slope at this point by drawing a line through two points where $x_1 = 2$ and $x_2 = 3$ and then measuring the rise and run for that line as is shown in Figure 1.20.

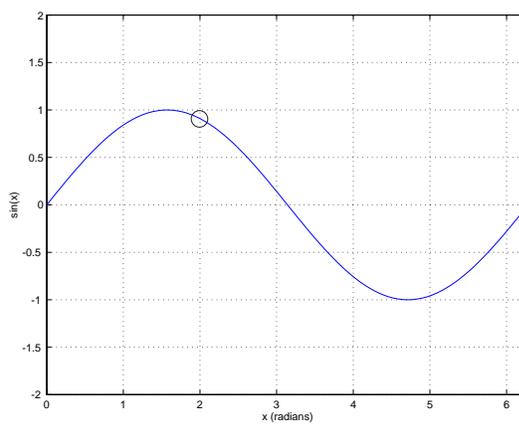


Figure 1.18: A graph of $y = \sin(x)$ showing the point $x = 2$ where we want to find the slope of the graph.

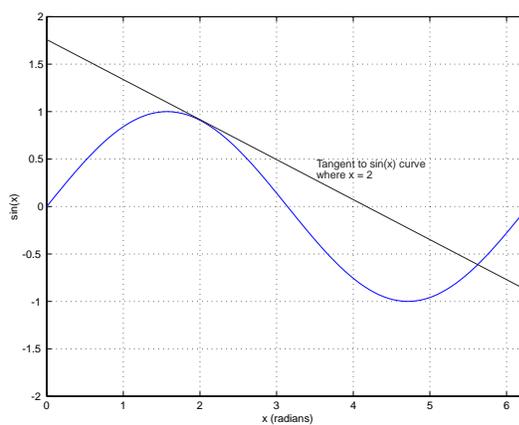


Figure 1.19: A graph of $y = \sin(x)$ showing the tangent to the curve at the point $x = 2$. The slope of the tangent is the slope of the curve at that point.

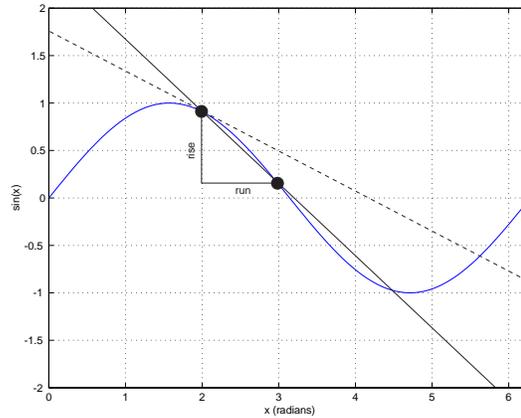


Figure 1.20: A graph of $y = \sin(x)$ showing an estimate to the tangent to the curve at the point $x = 2$ by drawing a line through the curve at $x_1 = 2$ and $x_2 = 3$.

As we can see in this example, the run for the line we've created is 1 (because it's $3 - 2$) and the rise is -0.768 (because it's $\sin(3) - \sin(2)$). Therefore the slope is -0.768 .

This method of approximating will give us a slope that is pretty close to the slope at the point we're interested in, but how do we make a better approximation? One way to do it is to reduce the distance between the point we're looking for and the points where we're drawing the line. For example, looking at Figure 1.21, we've changed the nearby points to $x_1 = 2$ and $x_2 = 2.5$, which gives us a line with a slope of -0.622 . This is a little closer to the real answer, but still not perfect.

As we get the points closer and closer together, the slope of the line gets closer and closer to the right answer. When the two points are infinitely close to $x = 2$ (therefore, they are at $x_1 = 2$ and $x_2 = 2$ because two points that are infinitely close are in the same place) then the slope of the line is the slope of the curve at that point.

What we're doing here is using the idea of a limit – as the run of the line (the horizontal distance between our two points) approaches the limit of 0, then the slope of the line approaches the slope of the curve at the point we're looking at.

This is the idea behind something called the *derivative* of a function. The curve that we're looking at can be described by an equation where the value of y is determined by some math that we do using the value of x . For example, if the equation is

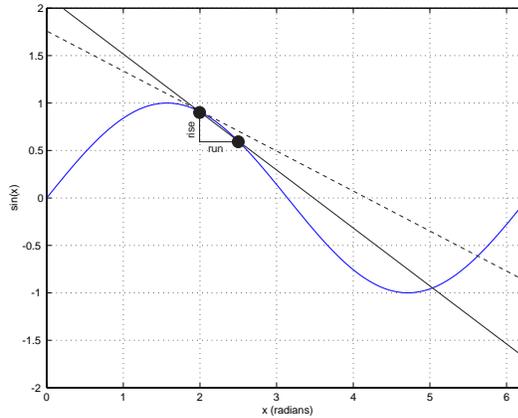


Figure 1.21: A graph of $y = \sin(x)$ showing a better estimate to the tangent to the curve at the point $x = 2$ by drawing a line through the curve at $x_1 = 2$ and $x_2 = 2.5$.

$$y = \sin(x) \quad (1.70)$$

then we get the curve seen above in Figure 1.18. In this particular case, given a value of x , we can figure out the value of y . As a result we say that y is a function of x or

$$y = f(x) \quad (1.71)$$

So, the derivative of $f(x)$ is just another equation that gives you the slope of the curve at any value of x . In mathematical language, the derivative of $f(x)$ is written in one of two ways. This simplest is if we just we write $f'(x)$ which means “the derivative (or the slope) of the function $f(x)$ ” (remember: *derivative* is just a fancy word for *slope*).

If you’re dealing with an equation where $y = f(x)$ as we’ve seen above in this chapter, then you’re looking for the “derivative of y with respect to x .” This is just a fancier way of saying “what’s the slope of $f(x)$?” We don’t need to say “ f of x ” because it’s easier to say “ y ” but we need to say “with respect to x ” because the slope changes as x changes. Therefore there is a relationship between the derivative of y and the value of x . If you want to use mathematical language to write “derivative of y with respect to x ,” you write

$$\frac{dy}{dx} \quad (1.72)$$

But, always remember, if $y = f(x)$ then

$$\frac{dy}{dx} = f'(x) \quad (1.73)$$

There is one important thing to remember when you see this symbol. $\frac{dy}{dx}$ is one thing that means “the derivative of y with respect to x .” It is not something divided by something else. This is a stand-alone symbol that doesn’t have anything to do with division.

Let’s look at a practical example. Any introductory calculus book will tell you that the derivative of a sine function is a cosine. (We don’t need to ask why at this point, but if you think back to the spinning wheel and the horizontal and vertical components of its movement, then it might make sense intuitively.) What does this mean? Well, that means that the slope of a sine wave at some value of x is equal to the cosine of the same value of x . This is written as is shown in Equation 1.74.

$$f'(\sin(x)) = \cos(x) \quad (1.74)$$

Just remember, if somebody walks up to you on the street and asks “what’s the derivative of the curve at $x = 2$?” what they’re saying is “what’s the slope of the curve at $x = 2$?”

If you want to calculate a derivative, then you can use the idea of a limit to do it. Think back a bit to when we were trying to find the tangent of a sine wave by plotting lines that crossed the sine wave at points closer and closer to the point we were interested in. Mathematically, what we were doing was finding the limit of the slope of the line, as the run between the two points approached 0. This is described in Equation 1.75

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \quad (1.75)$$

Huh? What Equation 1.75 says is that we’re looking for the value that the slope of the line drawn between two points separated in the x -axis by the value h approaches as h approaches 0. For example, in Figure 1.20, $h = 1$. In Figure 1.21, $h = 0.5$. Remember that h is just the “run” and $f(x+h) - f(x)$ is just the rise of the triangle shown in those plots. As h approaches 0, the slope of the hypotenuse gets closer and closer to the answer that we’re looking for.

Just to make things really miserable, I’ll let you know that you can have beasts like the vicious *double derivative* written $f''(x)$. This just means that you’re looking for the slope of the slope of a function. So, we’ve already seen

that the derivative of a sine function is a cosine function (or $f'(sin(x)) = cos(x)$), therefore the double derivative of a sine function is the derivative of a cosine function (or $f''(sin(x)) = f'(cos(x))$). It just so happens that $f'(cos(x)) = -sin(x)$, therefore $f''(sin(x)) = -sin(x)$.

1.8.4 Sigma - Σ

A Σ (the Greek capital letter Sigma) in an equation is just a lazy way of writing “the sum of” whatever follows it. The stuff under and over it give you an indication of when to start and when to stop adding. Let’s look at a simple example shown in Equation 1.76...

$$y = \sum_{x=1}^{10} x \quad (1.76)$$

Equation 1.76 says “ y equals the sum of all of the values of x from $x = 1$ to $x = 10$. The Σ sign says “add up everything from...” the “ $x = 1$ ” at the bottom says where to start adding, the “10” on top says when to stop, and the “ x ” after the Σ tells you what you’re adding. So:

$$\sum_{x=1}^{10} x = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 = 55 \quad (1.77)$$

This, of course, can get a little more complicated, as is shown in Equation 1.78 but if you don’t panic, it should be reasonably easy to figure out what you’re adding.

$$\sum_{x=1}^5 \sin(x) = \sin(1) + \sin(2) + \sin(3) + \sin(4) + \sin(5) \quad (1.78)$$

1.8.5 Delta - Δ

A Δ (a Greek capital Delta) means “the change in...” or “the difference in.” So, Δx means “the change in x .” This is useful if you’re looking at a value that’s changing like the speed of a car when you’re accelerating.

1.8.6 Integration and Area

If I asked you to calculate the area of a rectangle with the dimensions 2 cm x 3 cm, you’d probably already know that you calculate this by multiplying

the length by the width. Therefore $2 \text{ cm} \times 3 \text{ cm} = 6 \text{ cm}^2$ or 6 square centimeters.

If you were asked to calculate the area of a triangle, the math would be a little bit more complicated, but not much more.

What happens when you have to find the area of a more complicated shape that includes a curved side? That's the problem we're going to look at in this section.

Let's look at the graph of the function $y = \sin(x)$ from $x = 0$ to $x = \pi$ as is shown in Figure 1.22.

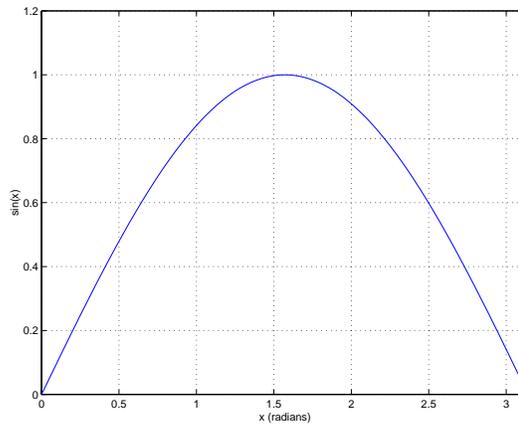


Figure 1.22: $y = \sin(x)$ for $0 \leq x \leq \pi$

We've seen in the previous section how to express the equation for finding the slope of this function, but what if we wanted to find the area under the graph? How do we find this? Well, the way we found the slope was initially to break the curve into shorter and shorter straight line components. We'll do basically the same thing here, but we'll make rectangles by creating simplified slices of the area under the curve.

For example, compare Figure 1.22 to Figure 1.23. What we've done in the second one is to consider the sine wave as 10 discrete values, and draw a rectangle for each. The width of each rectangle is equal to $\frac{1}{n}$ of the total length of the curve where n is the number of rectangles. For example, in Figure 1.23, there are 10 rectangles (so $n = 10$) in a total width of π (the length of the curve on the x-axis) so each rectangle is $\frac{\pi}{10}$ wide. The height of each rectangle is the value of the function (in our case, $\sin(x)$) for the series of values of x , $\frac{0\pi}{n}$, $\frac{1\pi}{n}$, $\frac{2\pi}{n}$, and so on up to $\frac{n\pi}{n}$.

If we add the areas of these 10 rectangles together, we'll get an approxi-

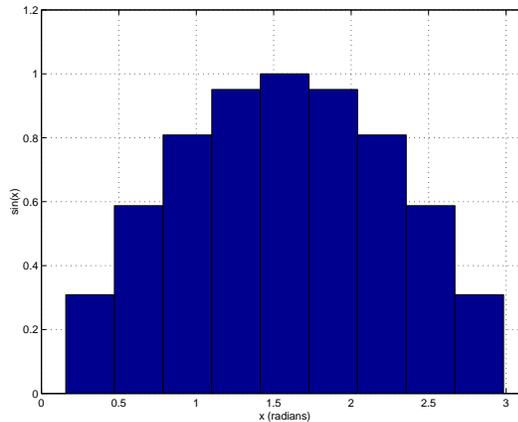


Figure 1.23: $y = \sin(x)$ for $0 \leq x \leq \pi$ divided into 10 rectangles

mation of the area under the curve. How do we express this mathematically? This is shown in Equation ??.

$$A = \sum_{i=0}^9 \sin\left(\frac{i}{10}\right) \frac{\pi}{10} \quad (1.79)$$

What does this mess mean? Well, we already know that n is the number of rectangles. The A_n on the left side of the equation means “the area contained in n rectangles.” The right side of the equation says that we’re going to add up the product of the sin of some changing number and $\frac{\pi}{10}$ ten times (because the number under the \sum is 1 and the number above is 10).

Therefore, Equation ?? written out the long way is shown in Equation 1.80

$$A = \sin\left(\frac{0\pi}{10}\right) \frac{\pi}{10} + \sin\left(\frac{1\pi}{10}\right) \frac{\pi}{10} + \sin\left(\frac{2\pi}{10}\right) \frac{\pi}{10} + \dots + \sin\left(\frac{9\pi}{10}\right) \frac{\pi}{10} \quad (1.80)$$

Each component in this sum of ten components that are added together is the product of the height ($\sin\left(\frac{i\pi}{10}\right)$) and width ($\frac{\pi}{10}$) of each rectangle. (All of those 10’s in there are there because we have divided the shape into 10 rectangles.) Therefore, each component is the area of a rectangle, which, when all added together give an approximation of the area under the curve.

There is a general equation that describes the way we divided up the area called the *Riemann Sum*. This is the general method of cutting a shape into

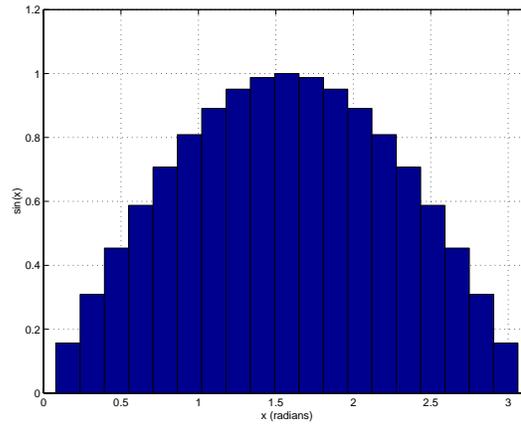


Figure 1.24: $y = \sin(x)$ for $0 \leq x \leq \pi$ divided into 20 rectangles

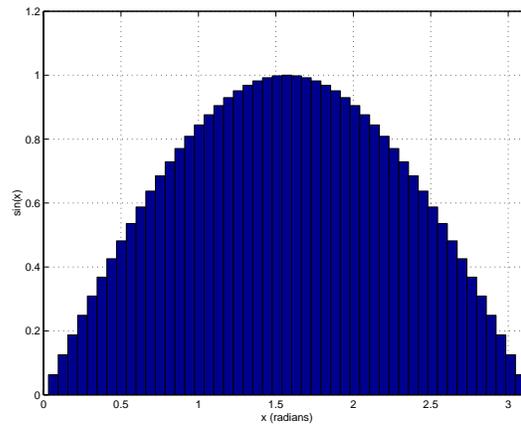


Figure 1.25: $y = \sin(x)$ for $0 \leq x \leq \pi$ divided into 50 rectangles

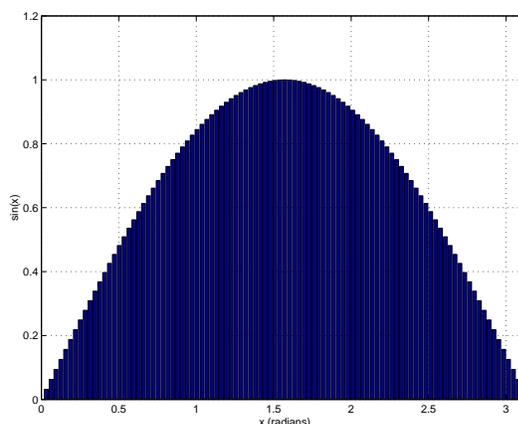


Figure 1.26: $y = \sin(x)$ for $0 \leq x \leq \pi$ divided into 100 rectangles

a number of rectangles to find an approximation of its area. If you want to learn about this, you'll just have to look in a calculus textbook.

Using rectangular slices of the area under the curve gives you an approximation of that area. The more rectangles you use, the better the approximation. As the number of rectangles approaches infinity, the approximation approaches the actual area under the curve.

There is a notation that says the same thing as Equation ?? but with an infinite number of rectangles. This is shown in Equation 1.81.

$$\int_0^{\pi} \sin(x) dx \quad (1.81)$$

What does this weird notation mean? The simplified version is that you're adding the areas of infinitely thin rectangles under the curve $\sin(x)$ from $x = 0$ to $x = \pi$. The 0 and π are indicated below and above the \int sign. On the right side of the equation you'll see a dx which means that's it's the x that's changing from 0 to π .

This equation is called an *integral* which is just a fancy word for “the area under a function.” Essentially, just like “derivative” is another word for “slope,” “integral” is another word for “area.”

The general form of an integral is shown in Equation 1.82.

$$A = \int_a^b f(x) dx \quad (1.82)$$

Compare Equation 1.82 to Figure 1.27. What it's saying is that the area

A is equal to all of the vertical “slices” of the space under the curve $f(x)$ from a to b . Each of these vertical slices has a width of dx . Essentially, dx has an infinitely small width (in other words, the width is 0) but there are an infinite number of the slices, so the whole thing adds up to a number.

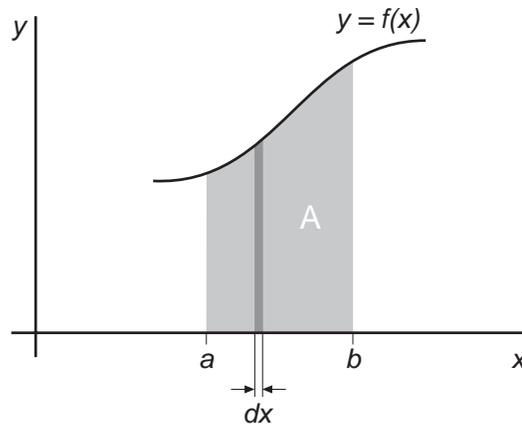


Figure 1.27: A graphic representation of what is meant by each of the components in Equation 1.82

1.8.7 Suggested Reading List

Chapter 2

Analog Electronics

2.1 Basic Electrical Concepts

2.1.1 Introduction

Everything, everywhere is made of molecules, which in turn are made of atoms (Except in the case of the elements in which the molecules *are* atoms). Atoms are made of two things, electrons and a nucleus. The nucleus is made of protons and neutrons. Each of these three particles has a specific charge or electrical capacity. Electrons have a negative charge, protons have a positive charge, and neutrons have no charge. As a result, the electrons, which orbit around the nucleus like planets around the sun, don't go flying off into the next atom because their negative charge attracts to the positive charge of the protons. Just as gravity ensures that the planets keep orbiting around the sun and don't go flying off into space, charge keeps the electrons orbiting the nucleus.

There is a slight difference between the orbits of the planets and the orbits of the electrons. In the case of the solar system, every planet maintains a unique orbit – each being a different distance from the sun, forming roughly concentric ellipses from Mercury out to Pluto (or sometimes Neptune). In an atom, the electrons group together into what are called valence shells. A valence shell is much like an orbit that is shared by a number of electrons. Different valence shells have different numbers of electrons which all add up to the total number in the atom. Different atoms have a different number of electrons, depending on the substance. This number can be found up in a periodic table of the elements. For example, the element hydrogen, which is number 1 on the periodic table, has 1 electron in each of its atoms; copper, on the other hand, is number 29 and therefore has 29 electrons.

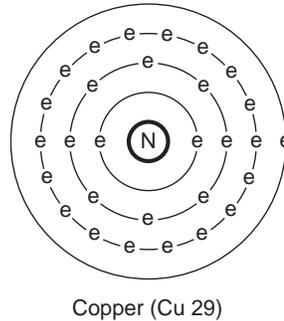


Figure 2.1: The structure of a copper atom showing the arrangement of the electrons in valence shells orbiting the nucleus.

Each valence shell likes to have a specific number of electrons in it to be stable. The inside shell is “full” when it has 2 electrons. The number of electrons required in each shell outside that one is a little complicated but is well explained in any high-school chemistry textbook.

Let’s look at a diagram of two atoms. As can be seen in the helium atom on the left, all of the valence shells are full, the copper atom, on the other hand, has an empty space in one of its shells. This difference between the two atom structures give the two substances very different characteristics.

In the case of the helium atom, since all the valence shells are full, the atom is very stable. The nucleus holds on to its electrons very tightly and will not let go without a great deal of persuasion, nor will it accept any new stray electrons. The copper atom, in comparison, has an empty space waiting to be filled by an electron nudged in its general direction. The questions are, how does one “nudge” an electron, and where does it go when released? The answers are rather simple: we push the electron out of the atom with another electron from an adjacent atom. The new electron takes its place and the now free particle moves to the next atom to fill its empty space.

So essentially, if we have a wire of copper, and we add some electrons to one end of it, and give the electrons on the other end somewhere to go, then we can have a flow of particles through the metal.

2.1.2 Current and EMF (Voltage)

I have a slight plumbing problem in my kitchen. I have two sinks, side by side, one for washing dishes and one for putting the washed dishes in to dry.

The drain of the two sinks feed into a single drain which has built up some rust inside over the years (I live in an old building).

When I fill up one sink with water and pull the plug, the water goes down the drain, but can't get down through the bottom drain as quickly as it should, so the result is that my second sink fills up with water coming up through its drain from the other sink.

Why does this happen? Well – the first answer is “gravity” – but there's more to it than that. Think of the two sinks as two tanks joined by a pipe at their bottoms. We'll put different amounts of water in each sink.

The water in the sink on the left weighs a lot – you can prove this by trying to lift the tank. So, the water is pushing down on the tank – but we also have to consider that the water at the top of the tank is pushing down on the water at the bottom. Thus there is more water *pressure* at the bottom than the top. Think of it as the molecules being squeezed together at the bottom of the tank – a result of the weight of all the water above it. Since there's more water in the tank on the left, there is more pressure at the bottom of the left tank than there is in the right tank.

Now consider the pipe. On the left end, we have the water pressure trying to push the water through the pipe, on the right end, we also have pressure pushing against the water, but less so than on the left. The result is that the water flows through the pipe from left to right. This continues until the pressure at both ends of the pipe is the same – or, we have the same water level in each tank.

We also have to think about how much water flows through the pipe in a given amount of time. If the difference in water pressure between the two ends is quite high, then the water will flow quite quickly through the pipe. If the difference in pressure is small, then only a small amount of water will flow. Thus the flow of the water (the volume which passes a point in the pipe per amount of time) is proportional on the pressure *difference*. If the pressure difference goes up, then the flow goes up.

The same can be said of electricity, or the flow of electrons through a wire. If we connect two “tanks” of electrons, one at either end of the wire, and one “tank” has more electrons (or more pressure) than the other, then the electrons will flow through the wire, bumping from atom to atom, until the two tanks reach the same level. Normally we call the two tanks a battery. Batteries have two terminals – one is the opening to a tank full of too many electrons (the negative terminal – because electrons are negative) and the other the opening to a tank with too few electrons (the positive terminal). If we connect a wire between the two terminals (don't try this at home!) then the surplus electrons at the negative terminal will flow through

to the positive terminal until the two terminals have the same number of electrons in them. The number of surplus electrons in the tank determines the “pressure” or *voltage* (abbreviated V and measured in volts) being put on the terminal. (Note: once upon a time, people used to call this *electromotive force* or *EMF* but as knowledge increases from generation to generation, so does laziness, apparently... So, most people today call it voltage instead of EMF.) The more electrons, the more voltage, or electrical pressure. The flow of electrons in the wire is called *current* (abbreviated I and measured in amperes or amps) and is actually a specific number of electrons passing a point in the wire every second (6,250,000,000,000,000 to be precise). (Note: some people call this “amperage” – but it’s not common enough to be the standard... yet...) If we increase the voltage (pressure) difference between the two ends of the wire, then the current (flow) will increase, just as the water in our pipe between the two tanks.

There’s one important point to remember when you’re talking about current. Due to a bad guess on the part of Benjamin Franklin, current flows in the opposite direction to the electrons in the wire, so while the electrons are flowing from the negative to the positive terminal, the current is flowing from positive to negative. This system is called *conventional current* theory. There are some books out there that follow the flow of electrons – and therefore say that current flows from negative to positive. It really doesn’t matter which system you’re using, so long as you know which is which.

Let’s now replace the two tanks by a pump with pipe connecting its output to its input – that way, we won’t run out of water. When the pump is turned on, it acts just like the fan in chapter 1 – it decreases the water pressure at its input in order to increase the pressure at its output. The water in the pipe doesn’t enjoy having different pressures at different points in the same pipe so it tries to equalize by moving some water molecules out of the high pressure area and into the low pressure area. This creates water flow through the pipe, and the process continues until the pump is switched off.

2.1.3 Resistance and Ohm’s Law

Let’s complicate matters a little by putting a constriction in the pipe – a small section where the diameter of the tube is narrower than anywhere else in the system. If we keep the same pressure difference created by the pump, then less water will go through because of the restriction – therefore, in order to get the same amount of water through the pipe as before, we’ll have to increase the pressure difference. So the higher the pressure difference, the

higher the flow; the greater the restriction, the smaller the flow.

We'll also have a situation where the pressure at the input to the restriction is different than that at the output. This is because the water molecules are bunching up at the point where they are trying to get through the smaller pipe. In fact the pressure at the output of the pump will be the same as the input of the restriction while the pressure at the input of the pump will match the output of the restriction. We could also say that there is a drop in pressure across the smaller diameter pipe.

We can have almost exactly the same scenario with electricity instead of water. The electrical equivalent to the restriction is called a *resistor*. It's a small component which resists the current, or flow of electrons. If we place a resistor in the wire, like the restriction in the pipe, we'll reduce the current as is shown in Figure 2.2.

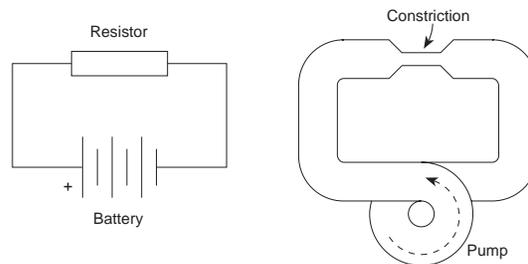


Figure 2.2: Equivalent situations showing the analogy between an electrical circuit with a battery and resistor to a plumbing network consisting of a pump and a constriction in the pipe. In both cases the flow (of electrical current or water, depending) runs clockwise around the loop.

In order to get the same current as with a single wire, we'll have to increase the voltage difference. Therefore, the higher the voltage difference, the higher the current; bigger the resistor, the smaller the current. Just as in the case of the water, there is a drop in voltage across the resistor. The voltage at the output of the resistor is lower than that at its input. Normally this is expressed as an equation called *Ohm's Law* which goes like this:

$$\text{Voltage} = \text{Current} * \text{Resistance}$$

or

$$V = IR \tag{2.1}$$

where V is in volts, I is in amps and R is in ohms (abbreviated Ω).

We use this equation to define how much resistance we have. The rule is that 1 volt of potential difference across a resistor will make 1 amp of

current flow through it if the resistor has a value of 1 ohm (abbreviated Ω). An ohm is simply a measurement of how much the flow of electrons is resisted.

The equation is also used to calculate one component of an electrical circuit given the other two. For example, if you know the current through and the value of a resistor, the voltage drop across it can be calculated.

Everything resists the flow of electrons to different degrees. Copper doesn't resist very much at all – in other words, it *conducts* electricity, so it is used as wiring; rubber, on the other hand, has a very high resistance, in fact it has an effectively infinite resistance so we call it an *insulator*

2.1.4 Power and Watt's Law

If we return to the example of a pump creating flow through a pipe, it's pretty obvious that this little system is useless for anything other than wasting the energy required to run the pump. If we were smart we'd put some kind of turbine or waterwheel in the pipe which would be connected to something which does work – any kind of work. Once upon a time a similar system was used to cut logs – connect the waterwheel to a circular saw; nowadays we do things like connecting generators to the turbine to generate electricity to power circular saws. In either case, we're using the energy or power in the moving water to do work of some sort.

How can we measure or calculate the amount of work our waterwheel is capable of doing? Well, there are two variables involved with which we are concerned – the pressure behind the water and the quantity of water flowing through the pipe and turbine. If there is more pressure, there is more energy in the water to do work; if there is more flow, then there is more water to do work.

Electricity can be used in the same fashion – we can put a small device in the wire between the two battery terminals which will convert the power in the electrical current into some useful work like brewing coffee or powering an electric stapler. We have the same equivalent components to concern us, except now they are named current and voltage. The higher the voltage or the higher the current, the more work which can be performed by our system – therefore the *power* it has.

This relationship can be expressed by an equation called *Watt's Law* which is as follows:

$$\text{Power} = \text{Voltage} * \text{Current}$$

or

$$P = VI \quad (2.2)$$

where P is in *watts*, V is in volts and I is in amps.

Just as Ohm's law defines the ohm, Watt's law defines the watt to be the amount of power consumed by a device which, when supplied with 1 volt of difference across its terminals will use 1 amp of current.

We can create a variation on Watt's law by combining it with Ohm's law as follows:

$$P = VI \text{ and } V = IR$$

therefore

$$P = (IR)I \quad (2.3)$$

$$P = I^2R \quad (2.4)$$

and

$$P = V\frac{V}{R} \quad (2.5)$$

$$P = \frac{V^2}{R} \quad (2.6)$$

Note that, as is shown in the equation above on the right, the power is proportional to the square of the voltage. This gem of wisdom will come in handy later.

2.1.5 Alternating vs. Direct Current

So far we have been talking about a constant supply of voltage – one that doesn't change over time, such as a battery before it starts to run down. This is what is commonly know of as direct current or DC which is to say that there is no change in voltage over a period of time. This is not the kind of electricity found coming out of the sockets in your wall at home. The electricity supplied by the hydro company changes over short periods of time (it changes over long periods of time as well, but that's an entirely different story...) Every second, the voltage difference between the two terminals in your wall socket fluctuates between about -170 V and 170 V sixty times a second. This brings up two important points to discuss.

Firstly, the negative voltage... All a negative voltage means is that the electrons are flowing in a direction opposite to that being measured. There

are more electrons in the tested point in the circuit than there are in the reference point, therefore more negative charge. If you think of this in terms of the two tanks of water – if we’re sitting at the bottom of the empty tank, and we measure the relative pressure of the full one, its pressure will be more, and therefore positive relative to your reference. If you’re at the bottom of the full tank and you measure the pressure at the bottom of the empty one, you’ll find that it’s less than your reference and therefore negative. (Two other analogies to completely confuse you... it’s like describing someone by their height. It doesn’t matter how tall or short someone is – if you say they’re tall, it probably means that they’re taller than you.

Secondly, the idea that the voltage is fluctuating. When you plug your coffee maker into the wall, you’ll notice that the plug has two terminals. One is a reference voltage which stays constant (normally called a “cold” wire in this case...) and one is the “hot” wire which changes in voltage relative to the cold wire. The device in the coffee maker which is doing the work is connected with each of these two wires. When the voltage in the hot wire is positive in comparison to the cold wire, the current flows from hot through the coffee maker to cold. One one-hundred and twentieth of a second later the hot wire is negative compared to the cold, the current flows from cold to hot. This is commonly known as alternating current or AC.

So remember, alternating current means that both the voltage and the current are changing in time.

2.1.6 RMS

Look at a light bulb. Not directly – you’ll hurt your eyes – actually let’s just think of a lightbulb. I turn on the switch on my wall and that closes a connection which sends electricity to the bulb. That electricity flows through the bulb which is slightly resistive. The result of the resistance in the bulb is that it has to burn off power which it does by heating up – so much that it starts to glow. But remember, the electricity which I’m sending to the bulb is not constant – it’s fluctuating up and down between -170 and 170 volts. Since it takes a little while for the bulb to heat up and cool down, it’s always lagging behind the voltage change – actually, it’s so slow that it stays virtually constant in temperature and therefore brightness.

This tendency is true for any resistor in an AC circuit. The resistor does not respond to instantaneous voltage values – instead, it burns off an average amount of power over time. That average is essentially an equivalent DC voltage that would result in the same power dissipation. The question is, how do we calculate it?

First we'll begin by looking at the average voltage delivered to your lightbulb by the hydro company. If we average the voltage for the full 360° of the sine wave that they provide to the outlets in your house, you'd wind up with 0 V – because the voltage is negative as much as it's positive in the full wave – it's symmetrical around 0 V. This is not a good way for the hydro company to decide on how much to bill you, because your monthly cost would be 0 dollars. (Sounds good to me – but bad to the hydro company...)

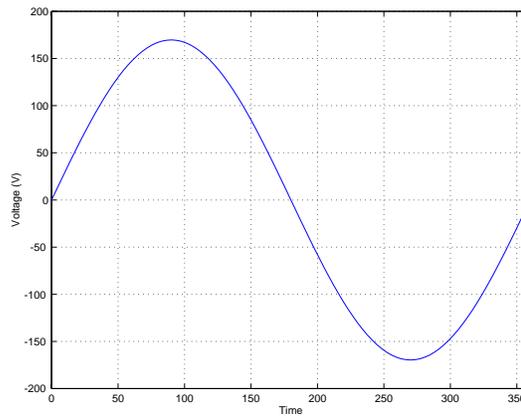


Figure 2.3: A full cycle of a sinusoidal AC waveform with a level of $170V_p$. Note that the total average for this waveform would be 0 V because the negative voltage is identically opposite to the positive voltage.

What if we only consider one half of a cycle of the 60 Hz waveform? Therefore, the voltage curve looks like the first half of a sine wave. There are 180° in this section of the wave. If we were to measure the voltage at each degree of the wave, add the results together and divide by 180 (in other words, find the average voltage) we would come up with a number which is 63.6% of the peak value of the wave. For example, the hydro company gives me a 170 volt peak sine wave. Therefore, the average voltage which I receive for the positive half of each wave is $170 \text{ V} * 0.636$ or 108.1 V as is shown in Figure 2.4.

This does not, however give me the equivalent DC voltage level which would match my AC power usage, because our calculation did not bring power into account. In order to find this level, we have to complicate matters a little. We know from Watt's law and Ohm's law that $P = V^2/R$. Therefore, if we have an AC wave of $170V_{peak}$ in a circuit containing a 1Ω resistor, the peak power consumption is

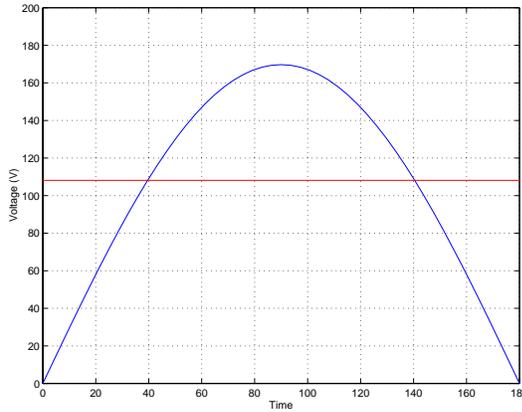


Figure 2.4: The positive half of one cycle of a sinusoidal AC waveform with a level of $170V_p$. Note that the total average for this waveform would be 63.6% of the peak voltage as is shown by the red horizontal line at 108.1 V (63.6% of $170V_p$).

$$\frac{170V^2}{1\Omega} = 28900Watts \quad (2.7)$$

But this is the power consumption for one point in time, when the voltage level is actually at 170 V. The rest of the time, the voltage is either swinging on its way up to 170 V or on its way down from 170 V. The power consumption curve would no longer be a sine wave, but a \sin^2 wave. Think of it as taking all of those 180 voltage measurements and squaring each one. From this list of 180 numbers (the instantaneous power consumption for each of the 180°) we can find the average power consumed for a half of a waveform. This number turns out to be 0.5 of the peak power, or, in the above case, $0.5 \cdot 28900$ Watts, or 14450 W as is shown in Figure 2.5.

This gives us the average power consumption of the resistor, but what is the equivalent DC voltage which would result in this consumption? We find this by using Watt's law in reverse as follows:

$$P = \frac{V^2}{R} \quad (2.8)$$

$$14450W = \frac{V^2}{1\Omega} \quad (2.9)$$

$$\sqrt{14450} = V \quad (2.10)$$

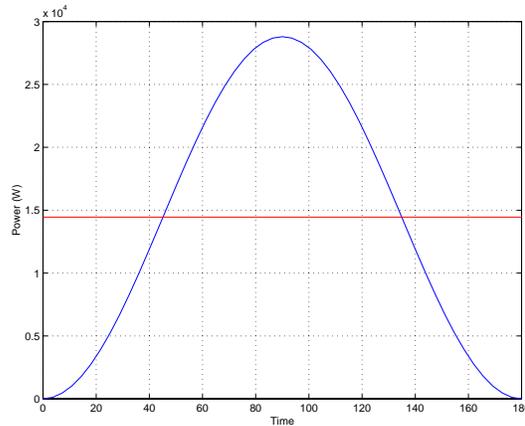


Figure 2.5: The power consumed during the positive half of one cycle of a sinusoidal AC waveform with a level of $170V_p$ and a resistance of $1\ \Omega$. Note that the total average for this waveform would be 50% of the peak power as is shown by the red horizontal line at 14450 W (50% of 28900 W).

$$V = 120V \quad (2.11)$$

Therefore, 120 V DC would result in the same power consumption over a period of time as a 170 V AC wave. This equivalent is called the Root Mean Square or RMS of the AC voltage. We call it this because it's the square root of the mean (or average) of the square of the original voltage.

In other words, a lightbulb in a lamp plugged into the wall (remember, it's being fed $170V_{peak}$ AC sine wave) will be exactly as bright if it's fed 120 V DC.

Just for a point of reference, the RMS value of a sine wave is always 0.707 of the peak value and the RMS value of a square wave (with a 50% duty cycle) is always the peak value. If you use other waveforms, the relationship between the peak value and the RMS value changes.

This relationship between the RMS and the peak value of a waveform is called the *crest factor*. This is a number that describes the ratio of the peak to the RMS of the signal, therefore

$$Crest\ factor = \frac{V_{peak}}{V_{RMS}} \quad (2.12)$$

So, the crest factor of a sine wave is 1.41. The crest factor of a square wave is 1.

This causes a small problem when you're using a digital volt meter. The reading on these devices ostensibly show you the RMS value of the AC waveform you're measuring, but they don't really measure the RMS value. They measure the peak value of the wave, and then multiply that value by 0.707 – therefore they're assuming that you're measuring a sine wave. If the waveform is anything other than a sine, then the measurement will be incorrect (unless you've thrown out a ton of money on a True RMS multimeter...)

RMS Time Constants

There's just one small problem with this explanation. We're talking about an RMS value of an alternating voltage being determined in part by an average of the instantaneous voltages over a period of time called the *time constant*. In Figure 2.5, we're assuming that the signal is averaged for at least one half of one cycle for the sine wave. If the average is taken for anything more than one half of a cycle, then our math will work out fine. What if this wasn't the case, however? What if the time constant was shorter than one half of a cycle?

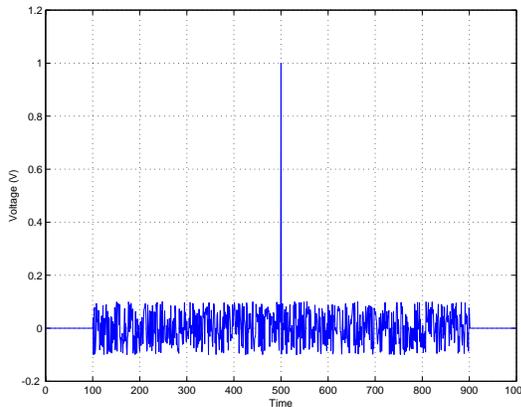


Figure 2.6: An arbitrary voltage signal with a short spike.

Take a look at the signal in Figure 2.6. This signal usually has a pretty low level, but there's a spike in the middle of it. This signal is comprised of a string of 1000 values, numbered from 1 to 1000. If we assume that this a voltage level, then it can be converted to a power value by squaring it (we'll keep assuming that the resistance is 1 Ω). That power curve is shown in Figure 2.7.

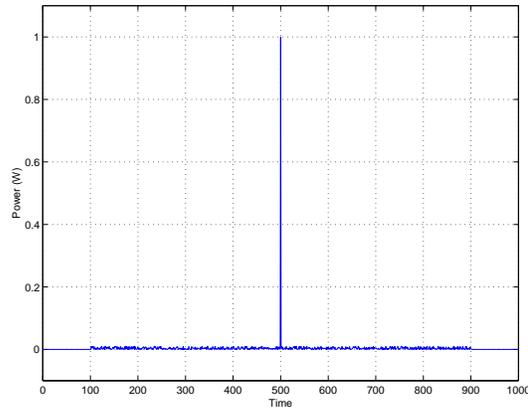


Figure 2.7: The power dissipation resulting from the signal in Figure 2.6 being sent through a $1\ \Omega$ resistor.

Now, let's make a running average of the values in this signal. One way to do this would be to take all 1000 values that are plotted in Figure 2.7 and find the average. Instead, let's use an average of 100 values (the length of this *window* in time is our *time constant*). So, the first average will be the values 1 to 100. The second average will be 2 to 101 and so on until we get to the average of values 901 to 1000. If these averages are plotted, they'll look like the graph in Figure 2.8.

There are a couple of things to note about this signal. Firstly, notice how the signal gradually ramps in at the beginning. This is because, as the time window that we're using for the average gradually "slides" over the transition from no signal to a low-level signal, the total average gradually increases. Also notice that what was a very short, very high level spike in the signal in the instantaneous power curve becomes a very wide (in fact, the width of the time constant), much lower-level signal (notice the scale of the y-axis). This is because the short spike is just getting thrown into an average with a lot of low-level signals, so the RMS value is much lower. Finally, the end ramps out just as the beginning ramped in for the same reasons.

So, we can now see that the RMS value is potentially much smaller than the peak value, but that this relationship is highly dependent on the time constant of the RMS detection. The shorter the time constant, the closer the RMS value is to the instantaneous peak value (in fact, if the time constant was infinitely short, then the RMS would equal the peak...).

The moral of the story is that it's not enough to just know that you're

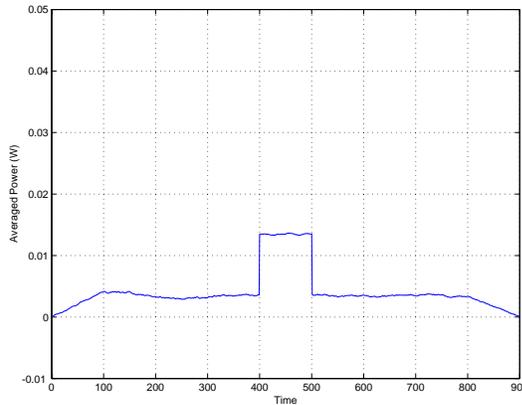


Figure 2.8: The average power dissipation of 100 consecutive values from the curve in Figure 2.7. For example, value 1 in this plot is the average of values 1 to 100 in Figure 2.7, value 2 is the average of values 2 to 101 in Figure 2.7 and so on.

being given the RMS value, you'll also need to know what the time constant of that RMS value is.

2.1.7 Suggested Reading List

Fundamentals of Service: Electrical Systems, John Deere Service Publication

Basic Electricity, R. Miller

The Incredible Illustrated Electricity Book, D.R. Riso

Elements of Electricity, W.H. Timbie

Introduction to Electronic Technology, R.J. Romanek

Electricity and Basic Electronics, S.R. Matt

Electricity: Principles and Applications, R. J. Fowler

Basic Electricity: Theory and Practice, M. Kaufman and J.A. Wilson

2.2 The Decibel

Thanks to Mr. Ray Rayburn for his proofreading of and suggestions for this section.

2.2.1 Gain

Lesson 1 for almost all recording engineers comes from the classic movie “Spinal Tap” [] where we all learned that the only reason for buying any piece of audio gear is to make things louder (“It goes all the way up to 11...”)
The amount by which a device makes a signal louder or quieter is called the *gain* of the device. If the output of the device is two times the amplitude of the input, then we say that the device has a gain of 2. This can be easily calculated using Equation 2.13.

$$gain = \frac{amplitude_{out}}{amplitude_{in}} \quad (2.13)$$

Note that you can use gain for evil as well as good - you can have a gain of less than 1 (but more than 0) which means that the output is quieter than the input.

If the gain equals 1, then the output is identical to the input.

If the gain is 0, then this means that the output of the device is 0, regardless of the input.

Finally, if the device has a negative gain, then the output will have an opposite polarity compared to the input. (As you go through this section, you should always keep in mind that a negative gain is different from a gain with a negative value in dB... but we’ll straighten this out as we go along.

(Incidentally, Lesson 2, entitled “How to wrap a microphone cable with one hand while holding a chili dog and a styrofoam cup of coffee in the other hand and not spill anything on your *Twisted Sister World Tour* T-shirt” will be addressed in a later chapter.)

2.2.2 Power and Bels

Sound in the air is a change in pressure. The greater the change, the louder the sound. The softest sound you can hear according to the books is 20×10^{-6} Pascals (abbreviated “Pa”) (it doesn’t really matter how big a Pa is – you just need to know the number for now...). The loudest sound you can tolerate without screaming in pain is about 200000000×10^{-6} Pa (or 200 Pa). This ratio of the loudest sound to the softest sound is therefore a 10,000,000:1

ratio (the loudest sound is 10,000,000 times louder than the softest). This range is simply too big. So a group of people at Bell Labs decided to represent the same scale with smaller numbers. They arrived at a unit of measurement called the Bel (named after Alexander Graham Bell – hence the capital B.) The Bel is a measurement of power difference. It's really just the *logarithm* of the ratio of two powers (Power1:Power2 or $\frac{Power1}{Power2}$). So to find out the difference in two power measurements measured in Bels (B). We use the following equation.

$$\Delta Power(Bels) = \log\left(\frac{Power1}{Power2}\right) \quad (2.14)$$

Let's leave the subject for a minute and talk about measurements. Our basic unit of length is the metre (m). If I were to talk about the distance between the wall and me, I would measure that distance in metres. If I were to talk about the distance between Vancouver and me, I would not use metres, I would use kilometres. Why? Because if I were to measure the distance between Vancouver and me in metres the number would be something like 5,000,000 m. This number is too big, so I say I'll measure it in kilometres. I know that 1 km = 1000 m therefore the distance between Vancouver and me is 5 000 000 m / 1 000 m/km = 5 000 km. The same would apply if I were measuring the length of a pencil. I would not use metres because the number would be something like 0.15 m. It's easier to think in centimetres or millimetres for small distances – all we're really doing is making the number look nicer.

The same applies to Bels. It turns out that if we use the above equation, we'll start getting small numbers. Too small for comfort; so instead of using Bels, we use decibels or dB. Now all we have to do is convert.

There are 10 dB in a Bel, so if we know the number of Bels, the number of decibels is just 10 times that. So:

$$1dB = \frac{1Bel}{10} \quad (2.15)$$

$$\Delta Power(dB) = 10 \log\left(\frac{Power1}{Power2}\right) \quad (2.16)$$

So that's how you calculate dB when you have two different amounts of power and you want to find the difference between them. The point that I'm trying to overemphasize thus far is that we are dealing with power measurements. We know that power is measured in watts (Joules per second) so we

use the above equation only when the ratio is comparing two measurements in watts.

What if we wanted to calculate the difference between two voltages (or electrical pressures)? Well, Watt's Law says that:

$$Power = \frac{Voltage^2}{Resistance} \quad (2.17)$$

or

$$P = \frac{V^2}{R} \quad (2.18)$$

Therefore, if we know our two voltages (V1 and V2) and we know the resistance stays the same:

$$\Delta Power(dB) = 10 \log \left(\frac{Power1}{Power2} \right) \quad (2.19)$$

$$= 10 \log \left(\frac{\frac{V1^2}{R}}{\frac{V2^2}{R}} \right) \quad (2.20)$$

$$= 10 \log \left(\frac{V1^2}{R} * \frac{R}{V2^2} \right) \quad (2.21)$$

$$= 10 \log \left(\frac{V1^2}{V2^2} \right) \quad (2.22)$$

$$= 10 \log \left(\frac{V1}{V2} \right)^2 \quad (2.23)$$

$$= 2 * 10 \log \left(\frac{V1}{V2} \right) \quad (2.24)$$

$$(because \log A^B = B * \log A) \quad (2.25)$$

$$= 20 \log \left(\frac{V1}{V2} \right) \quad (2.26)$$

That's it! (Finally!) So, the moral of the story is, if you want to compare two voltages and express the difference in dB, you have to go through that last equation.

Remember, voltage is analogous to pressure. So if you want to compare two pressures (like $20 * 10^{-6}$ Pa and $200000000 * 10^{-6}$ Pa) you have to use the same equation, just substitute V1 and V2 with P1 and P2 like this:

$$\Delta Power(dB) = 2 * 10 \log \left(\frac{Pressure1}{Pressure2} \right) \quad (2.27)$$

This is all well and good if you have two measurements (of power, voltage or pressure) to compare with each other, but what about all those books that say something like “a jet at takeoff is 140 dB loud.” What does that mean? Well, what it really means is “the sound a jet makes when it’s taking off is 140 dB louder than...” Doesn’t make a great deal of sense... Louder than what? The first measurement was the sound pressure of the jet taking off, but what was the second measurement with which it’s compared?

This is where we get into variations on the dB. There are a number of different types of dB which have references (second measurements) already supplied for you. We’ll do them one by one.

2.2.3 dBspl

The dBspl is a measurement of sound pressure (spl stand for Sound Pressure Level). What you do is take a measurement of, say, the sound pressure of a jet at takeoff (measured in Pa). This provides *Pressure1*. Our reference *Pressure2* is given as the sound pressure of the softest sound you can hear, which we have already said is $20 * 10^{-6}$ Pa.

Let’s say we go to the end of an airport runway with a sound pressure meter and measure a jet as it flies overhead. Let’s also say that, hypothetically, the sound pressure turns out to be 200 Pa. Let’s also say we want to calculate this into dBspl. So, the sound of a jet at takeoff is :

$$Pressure(dBspl) = 20 \log \left(\frac{Pressure1}{Reference} \right) \quad (2.28)$$

$$= 20 \log \left(\frac{Pressure1}{20 * 10^{-6} Pa} \right) \quad (2.29)$$

$$= 20 \log \left(\frac{200 Pa}{20 * 10^{-6} Pa} \right) \quad (2.30)$$

$$= 20 \log 10000000 \quad (2.31)$$

$$= 20 \log 10^7 \quad (2.32)$$

$$= 20 * 7 \quad (2.33)$$

$$= 140 dBspl \quad (2.34)$$

So what we’re saying is that a jet taking off is 140 dBspl which means “the sound pressure of a jet taking off is 140 dB louder than the softest sound I can hear.”

2.2.4 dBm

When you're measuring sound pressure levels, you use a reference based on the threshold of hearing (20×10^{-6} Pa) which is fine, but what if you want to measure the electrical power output of a piece of audio equipment? What is the reference that you use to compare your measurement? Well, in 1939, a bunch of people sat down at a table and decided that when the needles on their equipment read 0 VU, then the power output of the device in question should be 0.001 W or 1 milliwatt (mW). Now, remember that the power in watts is dependent on two things – the voltage and the resistance (Watt's law again). Back in 1939, the impedance of the input of every piece of audio gear was 600Ω . If you were Sony in 1939 and you wanted to build a tape deck or an amplifier or anything else with an input, the impedance across the input wire and the ground in the connector would have to be 600Ω .

As a result, people today (including me until my error was spotted by Ray Rayburn) believe that the dBm measurement uses two standard references – 1 mW across a 600Ω impedance. This is only partially the case. We use the 1 mW, but not the 600Ω . To quote John Woram, "...the dBm may be correctly used with any convenient resistance or impedance." [Woram, 1989]

By the way, the m stands for milliwatt.

Now this is important: since your reference is in mW we're dealing with power. Decibels are a measurement of a power difference, therefore you use the following equation:

$$Power(dBm) = 10 \log \left(\frac{Power1}{1mW_{RMS}} \right) \quad (2.35)$$

Where Power1 is measured in mW_{RMS} .

What's so important? There's a 10 in there and not a 20. It would be 20 if we were measuring pressure, either sound or electrical, but we're not. We're measuring power.

2.2.5 dBV

Nowadays, the 600Ω specification doesn't apply anymore. The input impedance of a tape deck you pick up off the shelf tomorrow could be anything – but it's likely to be pretty high, somewhere around 10 k Ω . When the impedance is high, the dissipated power is low, because power is inversely proportional to the resistance. Therefore, there may be times when your power measurement is quite low, even though your voltage is pretty high. In this case, it makes more sense to measure the voltage rather than the power. Now we

need a new reference, one in volts rather than watts. Well, there's actually two references... The first one is $1V_{RMS}$. When you use this reference, your measurement is in dBV.

So, you measure the voltage output of your piece of gear – let's say a mixer, for example, and compare that measurement with the $1V_{RMS}$ reference, using the following equation.

$$Voltage(dBV) = 20 \log \left(\frac{Voltage1}{1V_{RMS}} \right) \quad (2.36)$$

Where Voltage1 is measured in V_{RMS} .

Now this time it's a 20 instead of a 10 because we're measuring pressure and not power. Also note that the dBV does not imply a measurement across a specific impedance.

2.2.6 dBu

Let's think back to the 1mW into 600Ω situation. What will be the voltage required to generate 1mW in a 600Ω resistor?

$$P = \frac{V^2}{R} \quad (2.37)$$

$$\text{therefore } V^2 = P * R \quad (2.38)$$

$$V = \sqrt{P * R} \quad (2.39)$$

$$= \sqrt{1mW_{RMS} * 600\Omega} \quad (2.40)$$

$$= \sqrt{0.001W_{RMS} * 600\Omega} \quad (2.41)$$

$$= \sqrt{0.6} \quad (2.42)$$

$$= 0.774596669V_{RMS} \quad (2.43)$$

Therefore, the voltage required to generate the reference power was about $0.775V_{RMS}$. Nowadays, we don't use the 600Ω impedance anymore, but the rounded-off value of $0.775V_{RMS}$ was kept as a standard reference. So, if you use $0.775V_{RMS}$ as your reference voltage in the equation like this:

$$Voltage(dBu) = 20 \log \left(\frac{Voltage1}{0.775V_{RMS}} \right) \quad (2.44)$$

your unit of measure is called dBu. Where Voltage1 is measured in V_{RMS} .

(It used to be called dBv, but people kept mixing up dBv with dBV and that couldn't continue, so they changed the dBv to dBu instead. You'll still

see dBv occasionally – it is exactly the same as dBu... just different names for the same thing.)

Remember – we’re still measuring pressure so it’s a 20 instead of a 10, and, like the dBV measurement, there is no specified impedance.

2.2.7 dB FS

The *dB FS* designation is used for digital signals, so we won’t talk about them here. They’re discussed later in Chapter 9.1.

2.2.8 Addendum: “Professional” vs. “Consumer” Levels

Once upon a time you may have learned that “professional” gear ran at a nominal operating level of +4 dB compared to “consumer” gear at only -10 dB. (Nowadays, this seems to be the only distinction between the two...) What few people ever notice is that this is not a 14 dB difference in level. If you take a piece of consumer gear outputting what it thinks is 0 dB VU (0 dB on the VU meter), and you plug it into a piece of pro gear, you’ll find that the level is not -14 dB but -11.79 dB VU... The reason for this is that the professional level is +4 dBu and the consumer level -10 dBV. Therefore we have two separate reference voltages for each measurement.

0 dB VU on a piece of pro gear is +4 dBu which in turn translates to an actual voltage level of $1.228V_{RMS}$. In comparison, 0 dB VU on a piece of consumer gear is -10 dBV, or $0.316V_{RMS}$. If we compare these two voltages in terms of decibels, the result is a difference of 11.79 dB.

2.2.9 The Summary

dBspl

$$Pressure(dBspl) = 20 \log \left(\frac{Pressure1}{20 * 10^{-6}} \right) \quad (2.45)$$

where Pressure1 is measured in Pa.

dBm

$$Power(dBm) = 10 \log \left(\frac{Power1}{1mW_{RMS}} \right) \quad (2.46)$$

where Power1 is measured in mW_{RMS} .

dBV

$$Voltage(dBV) = 20 \log \left(\frac{Voltage1}{1V_{RMS}} \right) \quad (2.47)$$

where Voltage1 is measured in V_{RMS} .

dBu

$$Voltage(dBu) = 20 \log \left(\frac{Voltage1}{0.775V_{RMS}} \right) \quad (2.48)$$

where Voltage1 is measured in V_{RMS} .

2.3 Basic Circuits / Series vs. Parallel

2.3.1 Introduction

Picture it – you’re getting a shower one morning and someone in the bathroom downstairs flushes the toilet... what happens? You scream in pain because you’re suddenly deprived of cold water in your comfortable hot / cold mix... Not good. Why does this happen? It’s simple... It’s because you were forced to share cold water without being asked for permission. Essentially, we can pretend (for the purposes of this example...) that there is a steady pressure pushing the flow of water into your house. When the shower and the toilet are asking for water from the same source (the intake to your house) then the water that was normally flowing through only one source suddenly goes in two directions. The flow is split between two paths.

How much water is going to flow down each path? That depends on the amount of resistance the water “sees” going down each path. The toilet is probably going to have a lower “resistance” to impede the flow of water than your shower head, and so more water flows through the toilet than the shower. If the resistance of the shower was smaller than the toilet, then you would only be mildly uncomfortable instead of jumping through the shower curtain to get away from the boiling water...

2.3.2 Series circuits – from the point of view of the current

Think back to the tanks connected by a pipe with a restriction in it described in Chapter 2.1. All of the water flowing from one tank to the other must flow through the pipe, and therefore, through the restriction. The flow of the water is, in fact, determined by the resistance of the restriction (we’re assuming that the pipe does not impede the flow of water... just the restriction...)

What would happen if we put a second restriction on the same pipe? The water flowing through the pipe from tank to tank now “sees” a single, bigger resistance, and therefore flows more slowly through the entire pipe.

The same is true of an electrical circuit. If we connect 2 resistors, end to end, and connect them to a battery as is shown in the diagram below, the current must flow from the positive battery terminal through the first resistor, through the second resistor, and into the negative terminal of the battery. Therefore the current leaving the positive terminal “sees” one big resistance equivalent to the added resistances of the two resistors.

What we are looking at here is called an example of resistors connected in *series*. What this essentially means is that there are a series of resistors

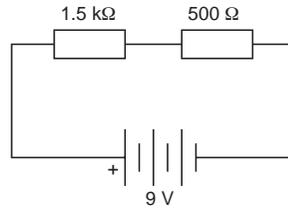


Figure 2.9: Two resistors and a battery, all connected in series.

that the current must flow through in order to make its way through the entire circuit.

So, how much current will flow through this system? That depends on the two resistances. If we have a 9 V battery, a 1.5 kohm resistor and a 500 Ω resistor, then the total resistance is 2 kohms. From there we can just use Ohm's law to figure out the total current running through the system.

$$V = IR \quad (2.49)$$

$$\textit{therefore} \quad (2.50)$$

$$I = \frac{V}{R} \quad (2.51)$$

$$= \frac{9V}{1500\Omega + 500\Omega} \quad (2.52)$$

$$= \frac{9}{2000\Omega} \quad (2.53)$$

$$= 0.0045A \quad (2.54)$$

$$= 4.5mA \quad (2.55)$$

Remember that this is not only the current flowing through the entire system, it's also therefore the current running through each of the two resistors. This piece of information allows us to go on to calculate the amount of voltage drop across each resistor.

2.3.3 Series circuits – from the point of view of the voltage

Since we know the amount of current flowing through each of the two resistors, we can use Ohm's law to calculate the voltage drop across each of them. Going back to our original example used above, and if we label our resistors. ($R_1 = 1.5k\Omega$ and $R_2 = 500\Omega$)

$V_1 = I_1 R_1$ (This means the voltage drop across $R_1 =$ the current through R_1 times the resistance of R_1)

$$V_1 = 4.5mA * 1.5k\Omega \quad (2.56)$$

$$= 0.0045 * 1500 \quad (2.57)$$

$$= 6.75V \quad (2.58)$$

Therefore the voltage drop across R_1 is 6.75 V.

Now we can do the same thing for R_2 (remember, it's the same current...)

$$V_2 = I_2 R_2 \quad (2.59)$$

$$= 4.5mA * 500\Omega \quad (2.60)$$

$$= 0.0045A * 500\Omega \quad (2.61)$$

$$= 2.25V \quad (2.62)$$

So the voltage drop across R_2 is 2.25 V. An interesting thing to note here is that the voltage drop across R_1 and the voltage drop across R_2 , when added together, equal 9 V. In fact, in this particular case, we could have simply calculated one of the two voltage drops and then simply subtracted it from the voltage produced by the battery to find the second voltage drop.

2.3.4 Parallel circuits – from the point of view of the voltage

Now let's connect the same two resistors to the battery slightly differently. We'll put them side by side, parallel to each other, as shown in the diagram below. This configuration is called *parallel* resistors and their effect on the current and voltage in the circuit is somewhat different than when they were in series...

Look at the connections between the resistors and the battery. They are directly connected, therefore we know that the battery is ensuring that there is a 9 V voltage drop across each of the resistors. This is a state imposed by the battery, and you simply expected to accept it as a given... (just kidding...)

The voltage difference across the battery terminals is 9 V – this is a given fact which doesn't change whether they are connected with a resistor or not. If we connect two parallel resistors across the terminals, they still stay 9 V apart.

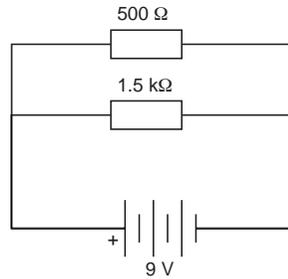


Figure 2.10: Two resistors and a battery, all in parallel.

If this is causing some difficulty, think back to the example at the top of this page where we had a shower running while a toilet was flushing in the same house. The water pressure supplied to the house didn't change... It's the same thing with a battery and two parallel resistors.

2.3.5 Parallel circuits – from the point of view of the current

Since we know the amount of voltage applied across each resistor (in this case, they're both 9 V) then we can again use Ohm's law to determine the amount of current flowing through each of the two resistors.

$$I_1 = \frac{V_1}{R_1} \quad (2.63)$$

$$= \frac{9V}{1.5k\Omega} \quad (2.64)$$

$$= \frac{9V}{1500\Omega} \quad (2.65)$$

$$= 0.006A \quad (2.66)$$

$$= 6mA \quad (2.67)$$

$$I_2 = \frac{V_2}{R_2} \quad (2.68)$$

$$= \frac{9V}{500\Omega} \quad (2.69)$$

$$= 0.018A \quad (2.70)$$

$$= 18mA \quad (2.71)$$

One way to calculate the total current coming out of the battery here is to calculate the two individual currents going through the resistors, and

adding them together. This will work, and then from there, we can calculate backwards to figure out what the equivalent resistance of the pair of resistors would be. If we did that whole procedure, we would find that the reciprocal of the total resistance is equal to the sum of the reciprocals of the individual resistors. (huh?) It's like this...

$$\frac{1}{R_{Total}} = \frac{1}{R_1} + \frac{1}{R_2} \quad (2.72)$$

$$\frac{1}{R_{Total}} = \frac{1}{R_1} * \frac{R_2}{R_2} + \frac{1}{R_2} * \frac{R_1}{R_1} \quad (2.73)$$

$$\frac{1}{R_{Total}} = \frac{R_2}{R_1 R_2} + \frac{R_1}{R_2 R_1} \quad (2.74)$$

$$\frac{1}{R_{Total}} = \frac{R_1 + R_2}{R_1 R_2} \quad (2.75)$$

$$R_{Total} = \frac{R_1 R_2}{R_1 + R_2} \quad (2.76)$$

2.3.6 Suggested Reading List

2.4 Capacitors

Let's go back a couple of chapters to the concept of a water pump sending water out its output through a pipe which has a constriction in it back to the input of the pump. We equated this system with a battery pushing current through a wire and resistor. Now, we're replacing the restriction in the water pipe with a couple of waterbeds... Stay with me here... This will make sense, I promise...

If the input of the water pump is connected to one of the waterbeds and the output of the pump is connected to the other waterbed, and the output waterbed is placed on top of the input waterbed, what will happen? Well, if we assume that the two waterbeds have the same amount of water in them before we turn on the pump (therefore the water pressure in the two are the same... sort of...) , then, after the pump is turned on, the water is drained from the bottom waterbed and placed in the top waterbed. This means that we have a change in the pressure difference between the two beds (The upper waterbed having the higher pressure). This difference will increase until we run out of water for the pump to move. The work the pump is doing is assisted by the fact that, as the top waterbed gets heavier, the water is pushed out of the bottom waterbed... Now, what does this have to do with electricity?

We're going to take the original circuit with the resistor and the battery and we're going to add a device called a *capacitor* in series with the resistor. A capacitor is a device with two metal plates that are placed very close together, but without touching (the waterbeds...). There's a wire coming off of each of the two plates (the pipes). Each of these plates, then can act as a reservoir for electrons – we can push extra ones into a plate, making it negative (by connecting the negative terminal of a battery to it...), or we can take electrons out, making the plate positive (by connecting the positive terminal of the battery to it...). Remember though that electrons, and the lack-of-electrons (holes) are mutually attracted to each other. As a result, the extra electrons in the negative plate are attracted to the holes in the positive plate. This means that the electrons and holes line up on the sides of the plates closest to the opposite plate – trying desperately to get across the gap... The narrower the gap, the more attraction, therefore the more electrons and holes we can pack in the plates... Also, the bigger the plates, the more electrons and holes we can get in there...

This device has the capacity to store quantities of electrons and holes – that's why we call them capacitors. The value of the capacitor, measured in Farads (abbreviated F) is a measure of its capacity... (we'll leave it at

that for now...) That capacitance is determined by two things essentially – both physical attributes of the device. The first is the size of the plates – the bigger the plates, the bigger the capacitance. For big caps, we take a couple of sheets of metal foil with a piece of non-conductive material sandwiched between them (called the dielectric) and roll the whole thing up like a sleeping bag being stored for a hike – put the whole thing in a little can and let the wires stick out of the ends (or one end...). The second attribute controlling the capacitance is the gap between the plates (the smaller the gap, the bigger the capacitance).

The reason we use these capacitors is because of a little property that they have which could almost be considered a problem... you can't dump all the electrons you want to through the wire into the plate instantaneously... It takes a little bit of time, especially if we restrict the current flow a bit with a resistor. Let's take a circuit as an example. We'll connect a switch, a resistor, and a capacitor all in series with a battery, as is shown in Figure 2.11.

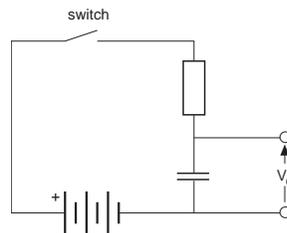


Figure 2.11: A battery, switch, resistor, and capacitor, all in series.

Just before we close the switch, let's assume that the two plates of the capacitor have the same number of electrons and holes in them – therefore they are at the same potential – so the voltage across the capacitor is 0 V. When we close the switch, the electrons in the negative terminal want to flow to the top plate of the cap to meet the holes flowing into the bottom plate. Therefore, when we first close the switch, we get a surge of current through the circuit which gradually decreases as the voltage across the capacitor is increased. The more the capacitor fills with holes and electrons, the higher the voltage across it, and therefore the smaller the voltage across the resistor – this in turn means a smaller current.

If we were to graph this change in the flow of current over time, it would look like Figure 2.12:

As you can see, the longer in time after the switch has been closed, the

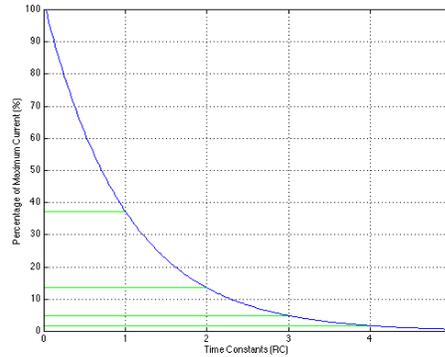


Figure 2.12: The change in current flowing through the resistor and into the top plate of the capacitor after the switch is closed.

smaller the current. The graph of the change in voltage over time would be exactly opposite to this as is shown in Figure 2.13.

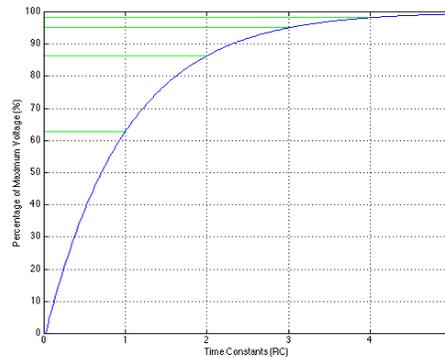


Figure 2.13: The change in voltage across the capacitor after the switch is closed.

You may notice that in most books, the time axis of the graph is not marked in seconds but in something that looks like a T – it’s called Tau (that’s a Greek letter and not a Chinese word, in case you’re thinking that I’m going to make a joke about Winnie the Pooh... It’s also pronounced differently – say “tao” not “dao”). This Tao is the symbol for something called a time constant, which is determined by the value of the capacitor and the resistor, as in Equation 2.77 :

$$\tau = RC \quad (2.77)$$

As you can see, if either the resistance or the capacitance is increased, the *RC time constant* goes up. “But what’s a time constant?” I hear you cry... Well, a time constant is the time it takes for the voltage to reach 63.2 percent of the voltage applied to the capacitor. After 2 time constants, we’ve gone up 63.2 percent and then 63.2 percent of the remaining 36.8 percent, which means we’re at 86.5 percent... Once we get to 5 time constants, we’re at 99.3 percent of the voltage and we can consider ourselves to have reached our destination. (In fact, we never really get there... we just keep approaching the voltage forever...)

So, this is all very well if our voltage source is providing us with a suddenly applied DC, but what would happen if we replaced our battery with a square wave and monitored the voltage across and the current flowing into the capacitor? Well, the output would look something like Figure 2.14 (assuming that the period of the square wave = 10 time constants...)

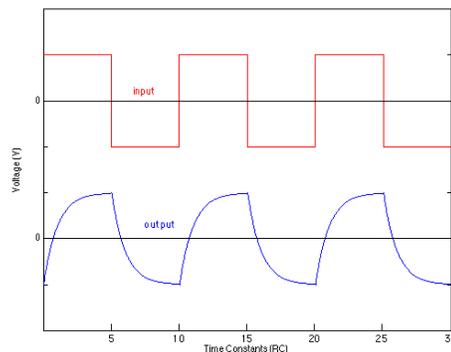


Figure 2.14: The change in voltage across a capacitor over time if the DC source in Figure 2.11 were changed to a square wave generator.

What’s going on? Well, the voltage is applied to the capacitor, and it starts charging, initially demanding lots of current through the resistor, but asking for less and less all the time... When the voltage drops to the lower half of the square wave, the capacitor starts charging (or discharging) to the new value, initially demanding lots of current in the opposite direction and slowly reaching the voltage. Since I said that the period of the square wave is 10 time constants, the voltage of the capacitor just reaches the voltage of

the function generator (5 time constants...) when the square wave goes to the other value.

Consider that, since the circuit is rounding off the square edges of the initially applied square wave, it must be doing something to the frequency response – but we'll worry about that later.

Let's now apply an AC sine wave to the input of the same circuit and look at what's going on at the output. The voltage of the function generator is always changing, and therefore the capacitor is always being asked to change the voltage across it. However, it is not changing nearly as quickly as it was with the square wave. If the change in voltage over time is quite slow (therefore, a low frequency sine wave) the current required to bring the capacitor to its new (but always changing) voltage will be small. The higher the frequency of the sine wave at the input, the more quickly the capacitor must change to the new voltage, therefore the more current it demands. Therefore, the current flowing through the circuit is dependent on the frequency – the higher the frequency, the higher the current. If we think of this another way, we could pretend that the capacitor is a resistor which changes in value as the frequency changes – the lower the frequency, the bigger the resistor, because the smaller the current. This isn't really what's going on, but we'll work that out in a minute.

The lower the frequency, the lower the current – the smaller the capacitor the lower the current (because it needs less current to change to the new voltage than a bigger capacitor). Therefore, we have a new equation which describes this relationship:

$$X_C = \frac{1}{2\pi fC} = \frac{1}{\omega C} \quad (2.78)$$

Where f is the frequency in Hz, C is the capacitance in Farads, and π is 3.14159264...

What's X_C ? It's something called the *capacitive reactance* of the capacitor, and it's expressed in Ω . It's not the same as resistance for two reasons – firstly, resistance burns power if it's resisting the flow of current... when current is impeded by capacitive reactance, there is no power lost. It's also different from a resistor because there is a different relationship between the voltage and the current flowing through (or into) the device. For resistors, Ohm's Law tells us that $V=IR$, therefore if the resistor stays the same and the voltage goes up, the current goes up at the same time. Therefore, we can say that, when an AC voltage is applied to a resistor, the flow of current through the resistor is in phase with the voltage. (when V is 0, I is 0, when V is maximum, I is maximum and so on...) In a capacitive circuit (one

where the reactance of the capacitor is much greater than the resistance of the resistor and the two are in series...) the current precedes the voltage (remember the time constant curves... voltage changes slowly, current changes quickly...) by 90° . This also means that the voltage across the resistor is 90° ahead of the voltage across the capacitor (because the voltage across the resistor is in phase with the current through it and into the capacitor)

As far as the function generator is concerned, it doesn't know whether the current it's being asked to supply is determined by resistance or reactance... all it sees is some THING out there, impeding the current flow differently at different frequencies (the lower the frequency, the higher the *impedance*...) This impedance is not simply the addition of the resistance and the reactance, because the two are not in phase with each other... in fact they're 90° out of phase. The way we calculate the total impedance of the circuit is by finding the square root of the sum of the squares of the resistance and the reactance or :

$$Z = \sqrt{R^2 + X_C^2} \quad (2.79)$$

Where Z is the impedance of the RC combination, R is the resistance of the resistor, and X_C is the capacitive reactance, all expressed in Ω .

Remember back to Pythagoreas – that same equation above is the one we use to find the length of the hypotenuse of a right triangle (a triangle whose legs are 90° apart...) when we know the lengths of the legs. Get it? Voltages are 90° apart, legs are 90° apart... If you don't get it, not to worry, it's explained in Chapter 6.

Also, remember that. as frequency goes up, the X_C goes down, and therefore the Z goes down. If the frequency is 0 Hz (or DC...) then the X_C is infinity Ω , and the circuit is no longer closed – no current will flow. This will come in handy in the next chapter.

As for the combination of capacitors in series and parallel, it's exactly the same equations as for resistors except that they're opposite. If you put two capacitors in parallel – the total capacitance is bigger... in fact it's the addition of the two capacitances (because you're effectively making the plates bigger...) Therefore, in order to calculate the total capacitance for a number of capacitors connected in parallel, you use Equation 2.80.

$$C_{total} = C_1 + C_2 + \dots + C_n \quad (2.80)$$

If the capacitors are in series, then you use the equation

$$\frac{1}{C_{total}} = \frac{1}{C_1} + \frac{1}{C_2} + \dots + \frac{1}{C_n} \quad (2.81)$$

Note that both of these equations are very similar to the ones for resistors, except that we use them “backwards.” That is to say that the equations for series resistors is the same as for parallel capacitors, and the one for parallel resistors is the same as for series capacitors.

2.4.1 Suggested Reading List

2.5 Passive RC Filters

In the last chapter, we looked at the circuit in Figure 2.15 and we talked about the impedance of the RC combination as it related to frequency. Now we'll talk about how to harness that incredible power to screw up your audio signal.

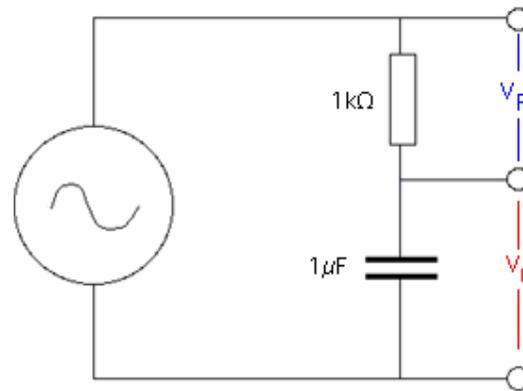


Figure 2.15: A first-order RC circuit. This is called “first-order” because there is only 1 reactive component in it (the capacitor).

In this circuit, the lower the frequency, the higher the impedance, and therefore the lower the current flowing through the circuit. If we're at 0 Hz, there is no current flowing through the circuit. If we're at infinity Hz (this is very high...) then the capacitor has a capacitive reactance of 0Ω , and the impedance of the circuit is the resistance of the resistor. This can be seen in Figure 2.16.

We also talked about how, at low frequencies, the circuit is considered to be capacitive (because the capacitive reactance is MUCH greater than the resistor value and therefore the resistor is negligible in comparison).

When the circuit is capacitive, the current flowing through the resistor into the capacitor is changing faster than the voltage across the capacitor. We said last week, that, in this case, the current is 90 degree ahead of the voltage. This also means that the voltage across the resistor (which is in phase with the current) is 90° ahead of the voltage across the capacitor. This is shown in Figure 2.17.

Let's look at the voltage across the capacitor as we change the voltage. At very low frequencies, the capacitor has a very high capacitive reactance, therefore the resistance of the resistor is negligible in comparison. If we

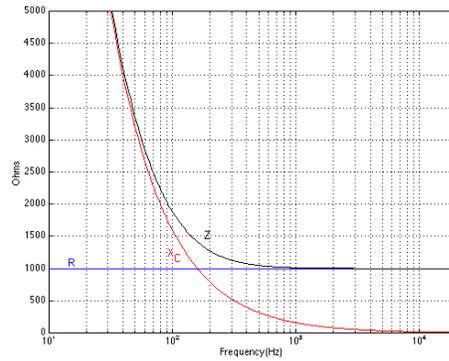


Figure 2.16: The resistance, capacitive reactance and total impedance of the above circuit. Notice that there is one frequency where X_C is equal to R .

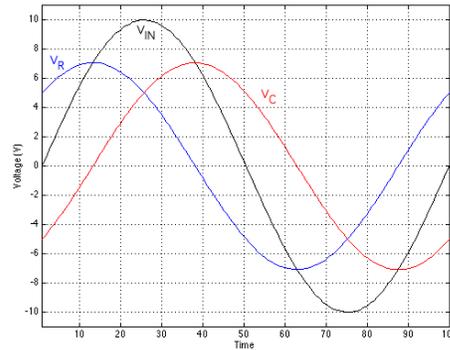


Figure 2.17: The relationship in the time domain between V_{IN} (the voltage difference across the function generator), V_R (the voltage across the resistor), and V_C (the voltage across the capacitor). Time is passing from left to right, so V_C is later than V_{IN} which is preceded by V_R . Note as well that V_R is 90° ahead of V_C .

consider the circuit to be a voltage divider (where the voltage is divided between the capacitor and the resistor) then there will be a much larger voltage drop across the capacitor than the resistor. At DC (0 Hz) the X_C is infinite, and the voltage across the capacitor is equal to the voltage output from the function generator. Another way to consider this is, if X_C is infinite, then there is no current flowing through the resistor, therefore there is no voltage drop across it (because $0 \text{ V} = 0 \text{ A} * R$). If the frequency is higher, then the reactance is lower, and we have a smaller voltage drop across the capacitor. The higher we go, the lower the voltage drop until, at infinity Hz, we have 0 V.

If we were to plot the voltage drop across the capacitor relative to the frequency, it would, therefore produce a graph like Figure 2.18.

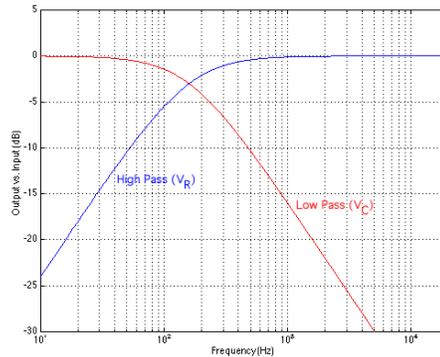


Figure 2.18: The output level of the voltages across the resistor and capacitor relative to the voltage of the sine wave generator in Figure 2.15 for various frequencies.

Note that we're specifying the voltage as a level relative to the input of the circuit, expressed in dB. The frequency at which the output (the voltage drop across the capacitor) is 3 dB below the input (that is to say -3 dB) is called the *cutoff frequency* (f_c) of the circuit. (We may as well start calling it a filter, since it's filtering different frequencies differently... since it allows low frequencies to pass through unchanged, we'll call it a low-pass filter.)

The f_c of the low-pass filter can be calculated if you know the values of the resistor and the capacitor. The equation is shown in Equation 2.82:

$$f_c = \frac{1}{2\pi RC} \quad (2.82)$$

(Note that if we put the values of the resistor and the capacitor from

Figure 1 – $R = 1000\Omega$ and $C = 1$ microfarad) into this equation, we get 159 Hz – the same frequency where $R = X_C$)

Where f_c is expressed in Hz, R is in Ω and C is in Farads.

At frequencies below $f_c / 10$ (1 *decade* below f_c – musicians like to think in octaves – 2 times the frequency – engineers like to think in decades, 10 times the frequency) we consider the output to be equal to the input – therefore at 0 dB. At frequencies 1 decade above f_c and higher, we drop 6 dB in amplitude every time we go up 1 octave, so we say that we have a slope of -6 dB per octave (this is also expressed as -20 dB per decade – means the same thing)

We also have to consider, however that the change in voltage across the capacitor isn't always keeping up with the change in voltage across the function generator. In fact, at higher frequencies, it lags behind the input voltage by 90° . Up to 1 decade below f_c , we are in phase with the input, at f_c , we are 45° behind the input voltage, and at 1 decade above f_c and higher, we are lagging by 90° . The resulting graph looks like Figure 2.19 :

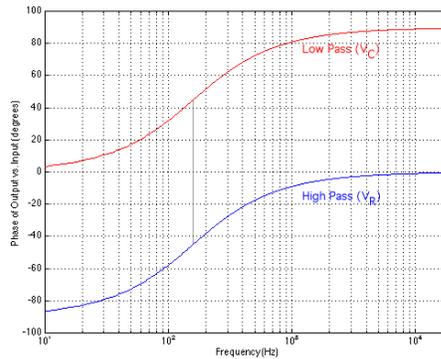


Figure 2.19: The phase relationship between the voltages across the resistor and the capacitor relative to the voltage across the function generator with different frequencies.

As is evident in the graph, a lag in the sine wave is expressed as a positive phase, therefore the voltage across the capacitor goes from 0° to 90° relative to the input voltage.

While all that is going on, what's happening across the resistor? Well, since we're considering that this circuit is a fancy type of voltage divider, we can say that if the voltage across the capacitor is high, the voltage across the resistor is low – if the voltage across the capacitor is low, then the voltage across the resistor is high. Another way to consider this is to say that if the

frequency is low, then the current through the circuit is low (because X_C is high) and therefore V_r is low. If the frequency is high, the current is high and V_r is high.

The result is Figure 4, showing the voltage across the resistor relative to frequency. Again, we're plotting the amplitude of the voltage as it relates to the input voltage, in dB.

Now, of course, we're looking at a high-pass filter. The f_c is again the frequency where we're at -3 dB relative to the input, and the equation to calculate it is the same as for the low-pass filter.

$$f_c = \frac{1}{2\pi RC} \quad (2.83)$$

The slope of the filter is now 6 dB per octave (20 dB per decade) because we increase by 6 dB as we go up one octave... That slope holds true for frequencies up to 1 decade below the f_c . At frequencies above f_c , we are at 0 dB relative to the input.

The phase response is also similar but different. Now the sine wave that we see across the resistor is *ahead* of the input. This is because, as we said before, the current feeding the capacitor precedes its voltage by 90° . At extremely low frequencies, we've established that the voltage across the capacitor is in phase with the input – but the current precedes that by 90° ... therefore the voltage across the resistor must precede the voltage across the capacitor (and therefore the voltage across the input) by 90° (up to $f_c / 10$)...

Again, at f_c , the voltage across the resistor is 45° away from the input, but this time it is ahead, not behind.

Finally, at $f_c * 10$ and above, the voltage across the resistor is in phase with the input. This all results in the phase response graph shown in Figure 5.

As you can see, the voltage across the resistor and the voltage across the capacitor are always 90° out of phase with each other, but their relationships with the input voltage change.

There's only one thing left that we have to discuss... this is an apparent conflict in what we have learned (though it isn't really a conflict...) We know that the f_c is the point where the voltage across the capacitor and the voltage across the resistor are both -3 dB relative to the input. Therefore the two voltages are equal – yet, when we add them together, we go up by 3 dB and not 6 dB as we would expect. This is because the two waves are 90° apart – if they were in phase, they would add to produce a gain of 6 dB. Since they are out of phase by 90° , their sum is 3 dB.

2.5.1 Another way to consider this...

We know that the voltage across the capacitor and the voltage across the resistor are always 90° apart at all frequencies, regardless of their phase relationships to the input voltage.

Consider the Resistance and the Capacitive reactance as both providing components of the impedance, but 90° apart. Therefore, we can plot the relationship between these three using a right triangle as is shown in Figure 2.21.

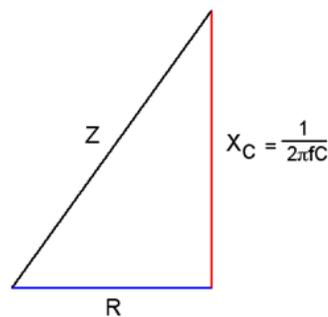


Figure 2.20: The triangle representing the relationship between the resistance, capacitive reactance and the impedance of the circuit. Note that, as frequency changes, only R remains constant.

At this point, it should be easy to see why the impedance is the square root of the sum of the squares of R and X_C . In addition, it becomes intuitive that, as the frequency goes to infinity Hz, X_C goes to zero and the hypotenuse of the triangle, Z, becomes the same as R. If the frequency goes to 0 Hz (DC), X_C goes to infinity Ω as does Z.

Go back to the concept of a voltage divider using two resistors. Remember that the ratio of the two resistances is the same as the ratio of the voltages across the two resistors.

$$\frac{R_1}{R_2} = \frac{V_1}{V_2} \quad (2.84)$$

If we consider the RC circuit in Figure 2.15, we can treat the two components in a similar manner, however the phase change must be taken into consideration. Figure 2.21 shows a triangle exactly the same as that in Figure 2.20 – now showing the relationship between the input voltage, and the voltages across the resistor and the capacitor.

So, once again, we can see that, as the frequency goes up, the voltage across the capacitor goes down until, at infinity Hz, the voltage across the

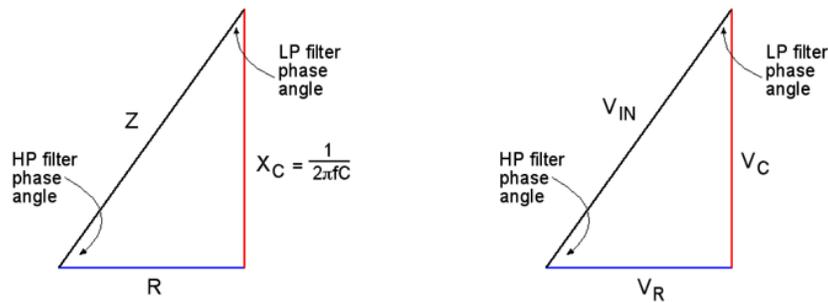


Figure 2.21: The triangle representing the relationship between the input voltage and the outputs of the high-pass and low-pass filters. Note that, as frequency changes, only V_{IN} remains constant.

cap is 0 V and $V_{IN} = V_R$.

Notice as well, that this triangle gives us the phase relationships of the voltages. The voltage across the Resistor and the Capacitor are always 90° apart, but the phase of these two voltages in relation to the input voltage changes according to the value of the capacitive inductance which is, in turn, determined by the capacitance and the frequency.

So, now we can see that, as the frequency goes down, the voltage across the resistor goes down, the voltage across the resistor approaches the input voltage, the phase of the low-pass filter approaches 90° and the phase of the high-pass filter approaches 0° . As the frequency goes up, the voltage across the capacitor goes down, the voltage across the resistor approaches the input voltage and the phase of the low-pass filter approaches 0° and the phase of the high-pass filter approaches 90° .

TO BE ADDED LATER: Discussion here about how to use complex numbers to describe the contributions of the components in this system

2.5.2 Suggested Reading List

2.6 Electromagnetism

Once upon a time, you did an experiment, probably around grade 3 or so, where you put a piece of paper on top of a bar magnet and sprinkled iron filings on the paper. The result was a pretty pattern that spread from pole to pole of the magnet. The iron filings were aligning themselves along what are called *magnetic lines of force*. These lines of force spread out around a magnet and have some effect on the things around them (like iron filings and compasses for example...) These lines of force have a direction – they go from the north pole of the magnet to the south pole as shown in Figures 2.22 and 2.23.

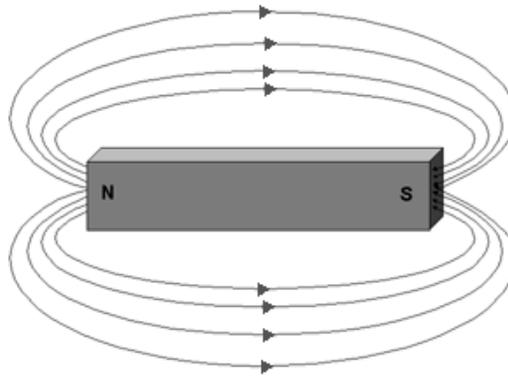


Figure 2.22: Magnetic lines of force around a bar magnet.

It turns out that there is a relationship between current in a wire and magnetic lines of force. If we send current through a wire, we generate magnetic lines of force that rotate around the wire. The more current, the more the lines of force expand out from the wire. The direction of the magnetic lines of force can be calculated using what is probably the first calculator you ever used... your right hand... Look at Figure 2.24. As you can see, if your thumb points in the direction of the current and you wrap your fingers around the wire, the direction your fingers wrap is the direction of the magnetic field. (You may be asking yourself “so what!?” – but we’ll get there...)

Let’s then, take this wire and make a spring with it so that the wire at one point in the section of spring that we’ve made is adjacent to another point on the same wire. The direction of the magnetic field in each section of the wire is then reinforced by the direction of the adjacent bits of wire



Figure 2.23: Magnetic lines of force around a horseshoe magnet.

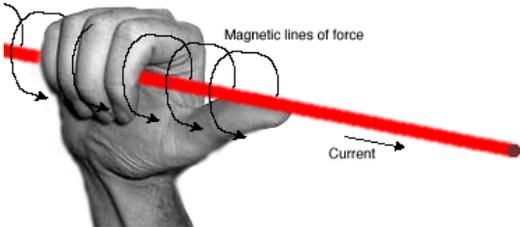


Figure 2.24: Right hand being used to show the direction of rotation of the magnetic lines of force when you know the direction of the current.

and the whole thing acts as one big magnetic field generator. When this happens, as you can see below, the coil has a total magnetic field similar to the bar magnet in the diagram above.

We can use our right hand again to figure out which end of the coil is north and which is south. If you wrap your fingers around the coil in the direction of the current, you will find that your thumb is pointing north, as is shown in Figure 2.25. Remember again, that, if we increase the current through the wire, then the magnetic lines of force move farther away from the coil.

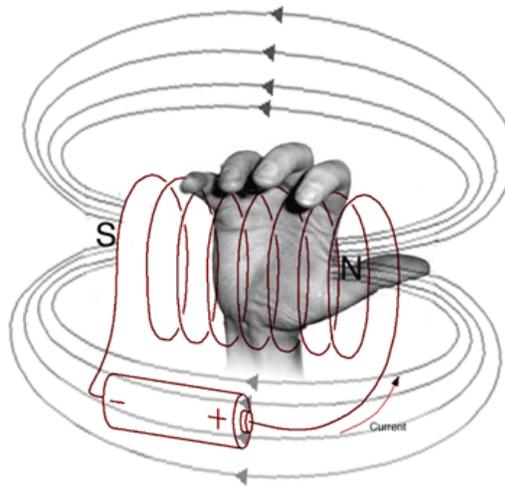


Figure 2.25: Right hand being used to find the polarity of the magnetic field around a coil of wire (the thumb is pointing towards the North pole) when you know the direction of the current around the coil (the fingers are wrapping around the coil in the same direction as the current).

One more interesting relationship between magnetism and current is that if we move a wire in a magnetic field, the movement will create a current in the wire. Essentially, as we cut through the magnetic lines of force, we cause the electrons to move in the wire. The faster we move the wire, the more current we generate. Again, our right hand helps us determine which way the current is going to flow. If you hold your hand as is shown in Figure 2.26, point your index finger in the direction of the magnetic lines of force (N to S...) and your thumb in the direction of the movement of the wire relative to the lines of force, your middle finger will point in the direction of the current.

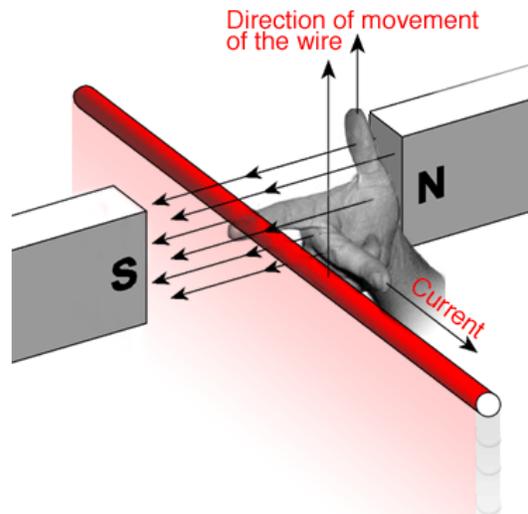


Figure 2.26: Right hand being used to find the current through a wire when it is moving in a magnetic field. The index finger points in the direction of the lines of magnetic force, the thumb is pointing in the direction of movement of the wire and the middle finger indicates the direction of current in the wire.

2.6.1 Suggested Reading List

Thanks to Mr. Francois Goupil for his most excellent hand-modelling.

2.7 Inductors

We saw in Section 2.6 that if you have a piece of wire moving through a magnetic field, you will induce current in the wire. The direction of the current is dependent on the direction of the magnetic lines of force and the direction of movement of the wire. Figure 2.30 shows an example of this effect.

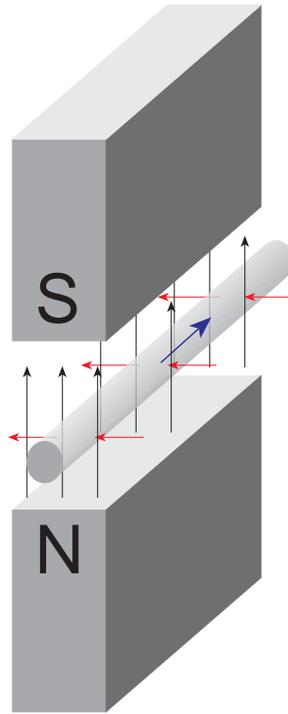


Figure 2.27: A wire moving through a constant magnetic field causing a current to be induced in the wire. The black arrows show the direction of the magnetic field, the red arrows show the direction of the movement of the wire and the blue arrow shows the direction of the electrical current.

We also saw that the reverse is true. If you have a piece of wire with current running through it, then you create a magnetic field around the wire with the magnetic lines of force going in circles around it. The direction of the magnetic lines of force is dependent on the direction of the current. The strength of the magnetic field and, therefore, the distance it extends from the wire is dependent on the amount of current. An example of this is shown in Figure 2.28 where we see two different wires with two different magnetic fields due to two different currents.

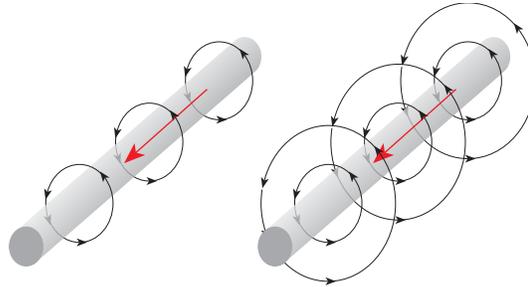


Figure 2.28: Two independent wires with current running through them. The wire on the right has a higher current going through it than the one on the left. Consequently, the magnetic field around it is stronger and therefore extends further from the wire.

What happens if we combine these two effects? Let's take a piece of wire and connect it to a circuit that let's us have current running through it. Then, we'll put a second piece of wire next to the first piece as is shown in Figure 2.29. Finally, we'll increase the current over time, so that the magnetic field expands outwards from the first wire. What will happen? The magnetic lines of force will expand outwards from around the first wire and cut through the second wire. This is essentially the same as if we had a constant magnetic field between two magnets and we moved the wire through it – we're just moving the magnetic field instead of the wire. Consequently, we'll induce a current in the second wire.

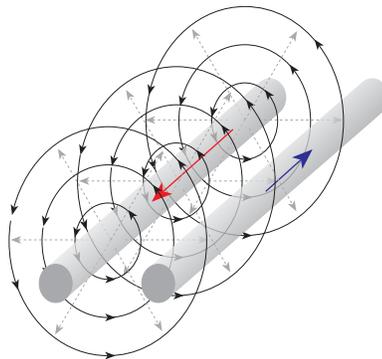


Figure 2.29: The wire on the right has a current induced in it because the magnetic field around the wire on the right is expanding. This is happening because the current through the wire on the left is increasing. Notice that the induced current is in the opposite direction to the current in the left wire. This diagram is essentially the same as the one shown in Figure 2.30.

Now let's go a step further and put a current through the wire on the

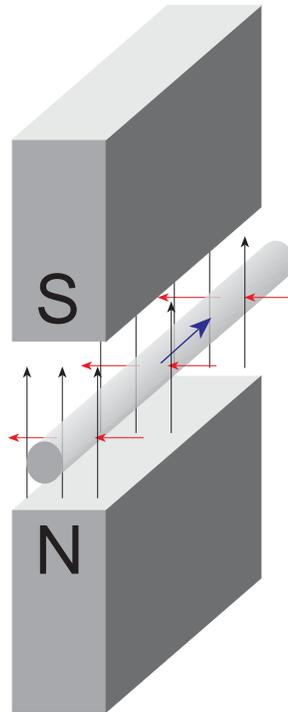


Figure 2.30: The wire on the right has a current induced in it because the magnetic field around the wire on the right is expanding. This is happening because the current through the wire on the left is increasing. Notice that the induced current is in the opposite direction to the current in the left wire. This diagram is essentially the same as the one shown in [Figure 2.29](#)

right that is always changing – the most common form of this signal in the electrical world is a sinusoidal waveform that alternates back and forth between positive and negative current (meaning that it changes direction). Figure 2.31 shows the result when we put an everyday AC signal into the wire on the left in Figure 2.30.

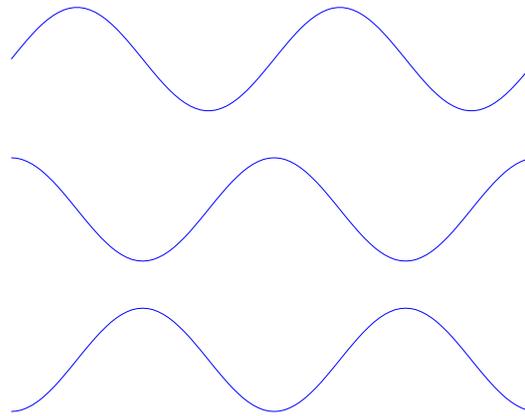


Figure 2.31: The relationship between the current in the two wires. The top plot shows the current in the wire on the left in Figure 2.30. The middle plot also shows the rate of change (the slope) of the top plot, therefore it is a plot of the velocity (the speed and direction of travel) of the magnetic field. The bottom plot shows the induced current in the wire on the right in the same figure. Note that the plots are not drawn on any particular scale in the vertical axes – so you shouldn't assume that you'll get the same current in both wires, but the phase relationship is correct.

Let's take a piece of wire and wind it into a coil consisting of two turns as is shown in Figure 2.32. One thing to beware of is that we aren't just wrapping naked wire in a coil - we have to make sure that adjacent sections of the wire don't touch each other, so we insulate the wire using a thin insulation.

Now think about Figure 2.30 as being just the top two adjacent sections of wire in the coil in Figure 2.32. This should raise a question or two. As we saw in Figure 2.30, increasing the current in one of the wires results in a current in the other wire *in the opposite direction*. If these two wires are actually just two sections of the same coil of wire, then the current we're putting through the coil goes through the whole length of wire. However, if we increase that current, then we induce a current in the opposite direction on the adjacent wires in the coil, which, as we know, is the same wire. Therefore, by increasing the current in the wire, we increase the induced current pushing in the opposite direction, opposing the current that we're

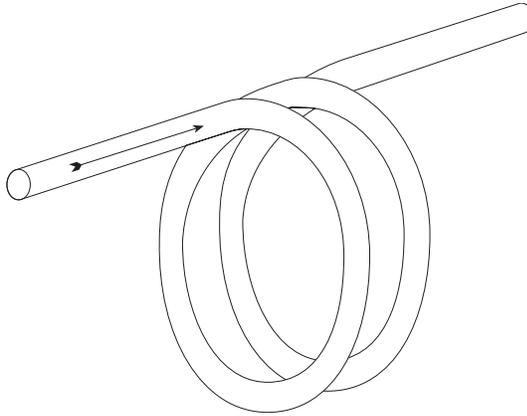


Figure 2.32: A piece of wire wound into a coil with only two turns.

putting in the wire. This opposing current results in a measurable voltage difference across the coil that is called *back electromotive force* or *back EMF*. This back EMF is proportional to the *change* (and therefore the slope, if we're looking at a graph) in the current, not the current itself, since it's proportional to the speed at which the wire is cutting through the magnetic field. Therefore, the amount that the coil (which we'll now start calling an *inductor* because we are inducing a current in the opposite direction), opposes the AC voltage applied to it is proportional to the frequency, since the higher the frequency, the faster the change in voltage.

This effect generates something called *inductive reactance*, abbreviated X_L which is measured in Ω . It's similar to capacitive reactance in that it opposes a change in current (whereas a capacitor opposes a change in voltage) without consuming power. This time, however, the roles of current and voltage are reversed in that, when we apply a change in current to the inductor, the current through it changes slowly, but the voltage across it changes quickly.

The inductance of an inductor is given in *Farads*, abbreviated F . Generally speaking, the bigger the inductance in Farads, the bigger the inductor has to be physically. There is one trick that is used to make the inductor a little more efficient, and that is to wrap the coil around an iron core. Remember that the wire in the coil is insulated (usually with a kind of varnish to keep the insulation really thin, and therefore getting the sections of wire as close together as possible to increase efficiency. Since the wire is insulated, the only thing that the iron core is doing is to act as a conductor for

the magnetic field. This may sound a little strange at first – so far we have only talked about materials as being conductors or insulators of electrical current. However, materials can also be classified as how well they conduct magnetic fields – iron is a very good magnetic conductor.

Similar to a capacitor, the inductive reactance of an inductor is dependent on the inductance of the device, in Farads and the frequency of the sinusoidal signal being sent through it. This is shown in Equation 2.85.

$$X_L = 2\pi fL = \omega L \quad (2.85)$$

Where L is the inductance of the inductor, in Farads. As can be seen, the inductive reactance, X_L , is proportional to both frequency and inductance (unlike a capacitor, in which X_C is inversely proportional to both frequency and capacitance).

2.7.1 Impedance

Let's put an inductor in series with a resistor as is shown in Figure 2.33.

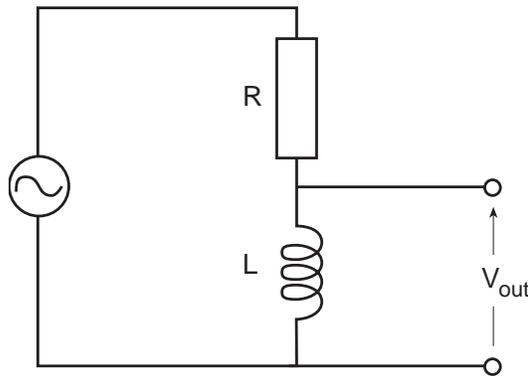


Figure 2.33: A resistor and an inductor in series.

Just like the case of a capacitor and a resistor in series (see section 2.4), the resulting load on the signal generator is an impedance, the result of a combination of a resistance and an inductance. Similar to what we saw with capacitors, there will be a phase difference of 90° between the voltages across the inductor and the resistor. However, unlike the capacitor, the voltage across the inductor is 90° *ahead* of the voltage across the resistor.

Since the resistance and the inductive reactance are 90° apart, we can calculate the total impedance – the load on the signal generator using the

Pythagorean Theorem shown in Equation 2.86 and explained in Section 2.5.

$$Z = \sqrt{R^2 + X_L^2} \quad (2.86)$$

2.7.2 RL Filters

We saw in Section 2.5 that we can build a filter using the relationship between the resistance of a resistor and the capacitive inductance of a capacitor. The same can be done using a resistor and an inductor, making an RL filter instead of an RC filter.

Connect an inductor and a resistor in series as is shown in Figure 2.33 and look at the voltage difference across the inductor as you change the frequency of the signal generator. If the frequency is very low, then the reactance of the inductor is practically 0Ω , so you get almost no voltage difference across it – therefore no output from the circuit. The higher the frequency, the higher the reactance. At some frequency, the reactance of the inductor will be the same as the resistance of the resistor, and the voltages across the two components are the same. However, since they are 90° apart, the voltage across either one will be 0.707 of the input voltage (or -3 dB). As we go higher in frequency, the reactance goes higher and higher and we get a higher and higher voltage difference across the inductor.

This should all sound very familiar. What we have done is to create a first-order high-pass filter using a resistor and an inductor, therefore it's called an RL filter. If we wanted a low-pass filter, then we use the voltage across the resistor as the output.

The cutoff frequency of an RL filter is calculated using Equation 2.87.

$$f_c = \frac{1}{2\pi RL} \quad (2.87)$$

2.7.3 Inductors in Series and Parallel

Finding the total inductance of a number of inductors connected in series or parallel behave the same way as resistors.

If the inductors are connected in series, then you add the individual inductances as in Equation 2.88.

$$L_{total} = L_1 + L_2 + L_3 + \cdots + L_n \quad (2.88)$$

If the inductors are connected in parallel, then you use Equation 2.89.

$$L_{total} = \frac{1}{\frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} + \dots + \frac{1}{L_n}} \quad (2.89)$$

2.7.4 Inductors vs. Capacitors

So, if we can build a filter using either an RC circuit or an RL circuit, which should we use, and why? They give the same frequency and phase responses, so what's the difference?

The knee-jerk answer is that we should use an RC circuit instead of an RL circuit. This is simply because inductors are bigger and heavier than capacitors. You'll notice on older equipment that RL circuits were frequently used. This is because capacitor manufacturing wasn't great - capacitors would leak electrolytic over time, thus changing their capacitance and the characteristics of the filter. An inductor is just a coil, so it doesn't change over time. However, modern capacitors are much more stable over long periods of time, so we can trust them in circuits.

2.7.5 Suggested Reading List

2.8 Transformers

If we take our coil and wrap it around a bar of iron, the iron acts as a conductor for the magnetic lines of force (not a conductor for the electricity... our wire is insulated...) therefore the lines are concentrated within the bar (the second right hand rule still applies for figuring out which way is north – but remember that if we’re using an AC waveform, the magnetic is changing in strength and polarity according to the change in the current.) Better yet, we can bend our bar around so it looks like a donut (mmmmmm donuts...) – that way the lines of force are most concentrated all the time. If we then wrap another coil around the bar (now donut-shaped, also known by topologists as *toroidal*) then the magnetic lines of force expanding and contracting around the bar will cut through the second coil. This will generate an alternating current in the second coil, just because it’s sitting there in a moving magnetic field. The relationship between these two coils is interesting...

It turns out (we’ll find out in a minute that that was just a pun...) that the power that we send into the input coil (called the primary coil) of this thing (called a *transformer*) is equal to the power that we get out of the second coil (called the secondary coil). (this is not entirely true – if it were, that would mean that the transformer is 100 percent efficient, which is not the case... but we’ll pretend that it is...)

Also, the ratio of the primary voltage to the secondary voltage is equal to the ratio of the number of turns of wire in the primary coil to the number of turns of wire in the secondary coil. This can also be expressed as an equation :

$$\frac{V_{primary}}{V_{secondary}} = \frac{Turns_{primary}}{Turns_{secondary}} \quad (2.90)$$

Given these two things... we can therefore figure out how much current is flowing into the transformer based on how much current is demanded of the secondary coil. Looking at the diagram below :

We know that we have $120V_{rms}$ applied to the primary coil. We therefore know that the voltage across the secondary coil and therefore across the resistor, is $12V_{rms}$ because $\frac{120V_{rms}}{12V_{rms}} = \frac{10Turns}{1Turn}$.

If we have $12V_{rms}$ across a 15 kohm resistor, then there is $0.8mA_{rms}$ flowing through it ($V=IR$). Therefore the power consumed by the resistor (therefore the power output of the secondary coil) is $9.6mW_{rms}$ ($P=VI$). Therefore the power input of the transformer is also $9.6mW_{rms}$. Therefore the current flowing into the primary coil is $0.08mA_{rms}$.

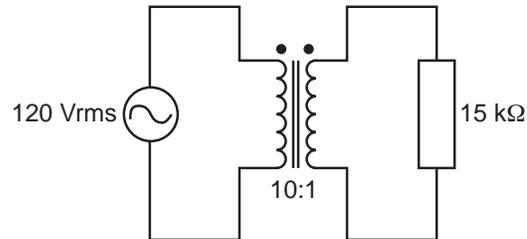


Figure 2.34: A 120 Vrms AC voltage source connected to the primary coil of a transformer with a 10:1 turns ratio. The secondary coil is connected to (or “loaded with”) a 15 kΩ resistor.

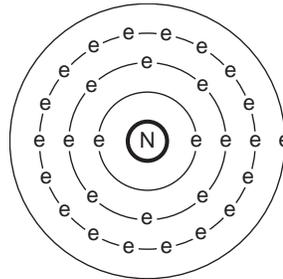
Note that, since the voltage went *down* by a factor of 10 (the turns ratio of the transformer) as we went from input to output, the current went *up* by the same factor of 10. This is the result of the $power_{in}$ being equal to the $power_{out}$.

You can have more than 1 secondary coil on a transformer. In fact, you can have as many as you want – the power into the primary coil will still be equal to the power of all of the secondary coils. We can also take a tap off the secondary coil at its half-way point. This is exactly the same as if we had two secondary coils with exactly the same number of turns, connected to each other. In this case, the centre tap (wire connected to the the half-way point on the coil) is always half-way in voltage between the two outside legs of the coil. If, therefore, we use the centre tap as our reference, arbitrarily called 0 V (or ground...) then the two ends of the coil are always an equal voltage “away” from the ground, but in opposite directions – therefore the two AC waves will be opposite in polarity.

2.8.1 Suggested Reading List

2.9 Diodes and Semiconductors

Back in the second chapter, we said that some substances have too few electrons in their outer valence shell. These electrons are therefore free to move on to the adjacent atom if we give them a nudge with an extra electron from a negative battery terminal – we then call this substance a conductor, because it conducts the flow of electricity.



Copper (Cu 29)

Figure 2.35: Diagram of a copper atom. Notice that there is one lone electron out there in the outer shell.

If the outer valence shell has too many electrons, they don't like moving (too many things to pack... moving vans are too expensive, and all their friends go to this school...) so they don't – we call those substances insulators.

There is a class of substances that have an in-between number of electrons in their outer valence shell. These substances (like silicon, germanium and carbon...) are neither conductors nor insulators – they lie somewhere in between so we call them *semiconductors* (not insulators nor conductors...)

Now take a look at Figure 2.37. As you can see, when you put a bunch of silicon atoms together, they start sharing outer electrons – that way each atom “thinks” that it has 8 electrons in its outer shell, but each one needs the 4 adjacent atoms to accomplish this

Compare Figure 2.36, which shows the structure of Silicon, to Figures 2.38 and 2.39 which show Arsenic and Gallium. The interesting thing about Arsenic is that it has 5 electrons in its outer shell (1 more than 4). Gallium has 3 electrons in its outer shell. We'll see why this is interesting in a second...

Recipe :

1 cup arsenic

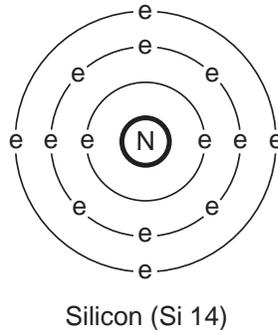


Figure 2.36: Diagram of a silicon atom. Notice that there are 4 electrons in the outer shell.

999 999 cups silicon

Stir well and bake until done.

Serves lots.

If you follow those steps carefully (do NOT sue me because I told you to play with arsenic...) you get a substance called arsenic doped silicon. Since arsenic has one too many electrons floating around its outer shell, then this new substance will have 1 extra electron per 1 000 000 atoms. We call this substance N-silicon or an *N-type* material (because electrons are negative) in spite of the fact that the substance really doesn't have a negative charge (because the extra electrons are equalled in opposite charge by the extra proton in each Arsenic atom's nucleus).

If you repeat the same recipe, replacing the arsenic with gallium, you wind up with 1 atom per 1 000 000 with 1 too few electrons. This makes P-silicon or *P-type* material (P because it's positive... Well, it's not really positive because the Gallium atom has only 3 protons in it, so it's balanced.)

Now, let's take a chunk of N-type material and glue it to a similarly sized chunk of P-type material as in the diagram below. (We'll also run a wire out of each chunk...) The extra electrons in the N-type material near the joint between the two materials see some new homes to move into across the street (the lack of electrons in the P-type material) and pack up and move out – this creates a barrier of sorts around the joint where there are no extra electrons floating around – therefore it is an insulating barrier. This doesn't happen for the entire collection of electrons and holes in the two materials because the physical attraction just isn't great enough.

If we connect the materials to a battery as shown below, with the positive terminal connected to the N-type material and the negative terminal to the

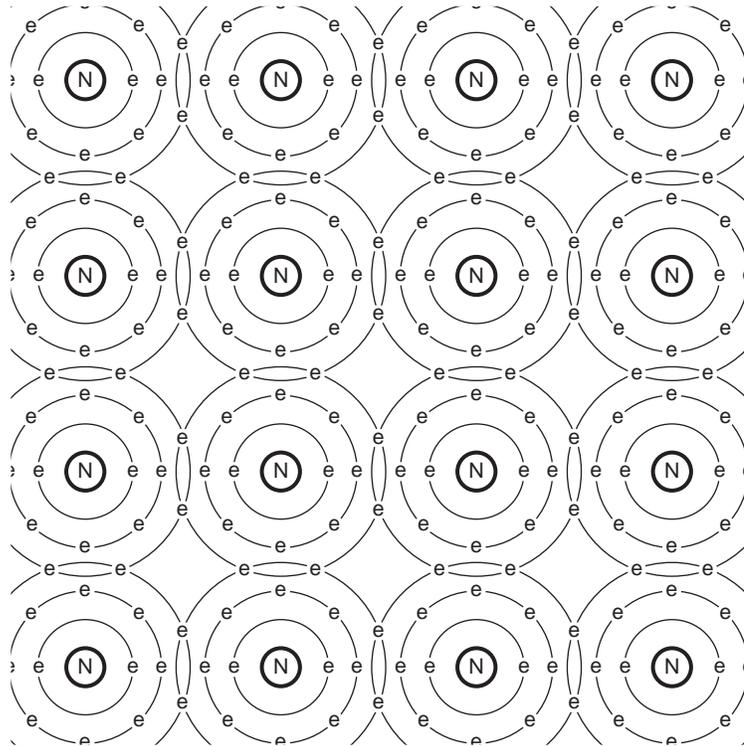
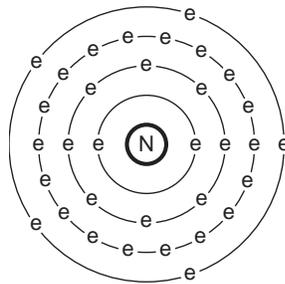
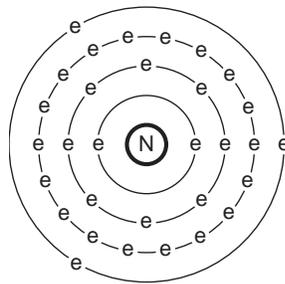


Figure 2.37: Diagram of a collection of silicon atoms. Note the sharing of outer electrons.



Arsenic (As 33)

Figure 2.38: Diagram of a Arsenic atom. Notice that there are 5 electrons in the outer shell.



Gallium (Ga 31)

Figure 2.39: Diagram of a Gallium atom. Notice that there are 3 electrons in the outer shell.

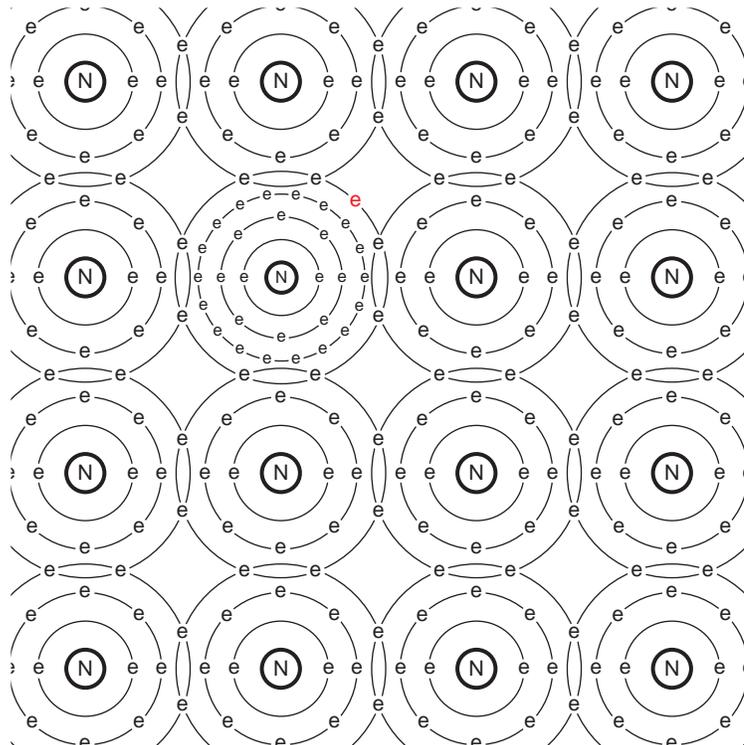


Figure 2.40: Diagram of N-type material. Notice that the extra “unattached” electron orbiting the Arsenic atom.

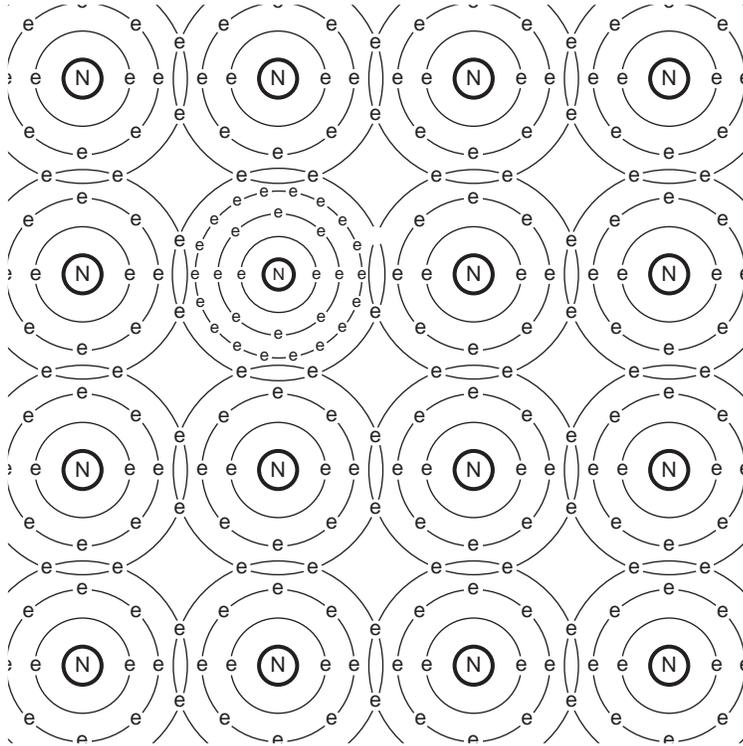


Figure 2.41: Diagram of P-type material. Notice that the “missing” electron orbiting the Gallium atom.

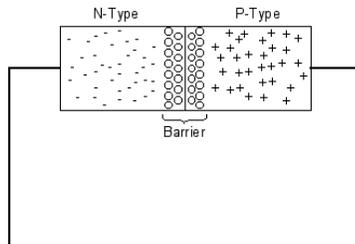


Figure 2.42: A chunk of N-type material glued to a chunk of P-type material.

P-type material through a resistor, the electrons in the N-type material will get attracted to the holes in the positive terminal and the electrons in the negative terminal will move into the holes in the P-type material. Once this happens, there are no spare electrons floating around, and no current can pass through the system. This situation is called *reverse-biasing* the device (called a *diode*). When the circuit is connect in this way, no current flows through the diode.

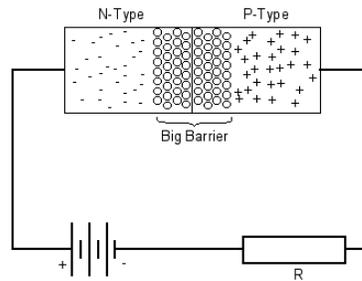


Figure 2.43: A reverse-biased diode. Note that no current will flow through the circuit.

If we connect the battery the other way, with the negative terminal to the N-type material and the positive terminal to the P-type material, then a completely different situation occurs. The extra electrons in the battery terminal push the electrons in the N-type material across the barrier, into the P-type material where they are drawn into the positive terminal. At the same time, of course, the holes (and therefore the current) is flowing in the opposite direction. This situation is called *forward biasing* the diode, which allows current to pass through it. There's only one catch with this electrical one-way gate. The diode needs some amount of voltage across it to open up and stay open. If it's a silicon diode (as most are...) then, you'll see a drop of about 0.6 V or 0.7 V across it – irrespective of current or voltage applied to the circuit (the remainder of the voltage drop will be across the resistor, which also determines the current). If the diode is made of doped germanium, then you'll only need about 0.3 V to get things running.

Of course, we don't draw all of those -'s and +'s on a schematic, we just indicate a diode using a little triangle with a cap on it. The arrow points in the direction of the current flow, so Figure 2.45 below is a schematic representation of the forward-biased diode in the circuit shown in Figure 2.44.

Now, remember that the diode “costs” 0.6 V to stay open, therefore if

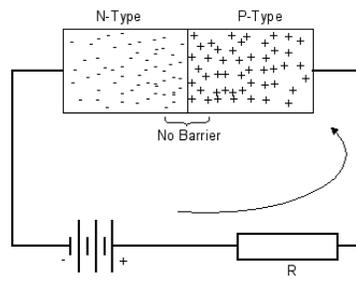


Figure 2.44: A forward-biased diode with current flowing through it.

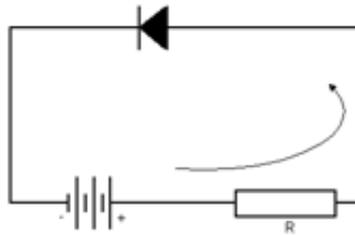


Figure 2.45: A forward-biased diode with current flowing through it just like in Figure 2.44. Note that current flows in the direction of the arrow, and will not flow in the opposite direction.

the voltage supply in Figure 2.45 is a 10 V DC source, we will see a 9.4 V DC difference across the resistor (because 0.6 V is lost across the diode). If the voltage source is 100 V DC, then there will be a 99.4 V DC difference across the resistor.

2.9.1 The geeky stuff:

Let's think back to the top of the page. The way we first considered a diode was as a 1-way gate for current. If we were to graph the characteristics of this "ideal diode" the result would look like Figure 2.46.

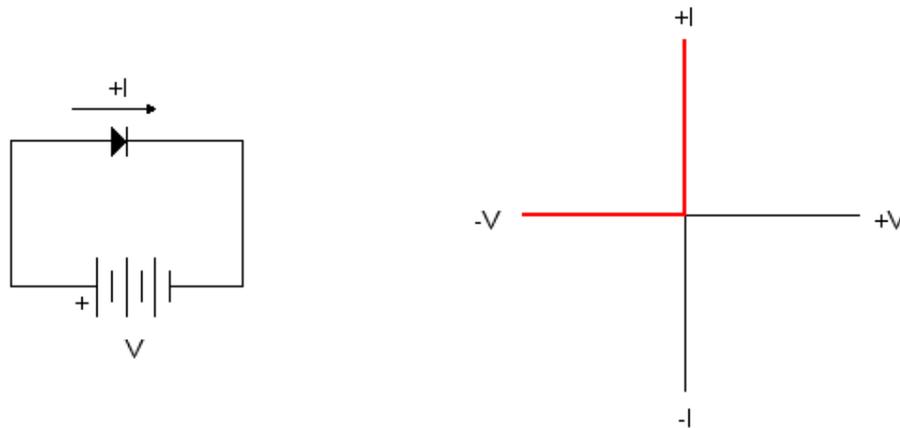


Figure 2.46: A graph showing the characteristics of an ideal diode. Notice how, if the voltage is positive (and therefore the diode is forward biased) you can have an infinite current going through the diode. Also, it does not require any voltage to open the diode – if the positive voltage is anything over 0V, the current will flow. If the voltage is negative (and therefore the diode is reverse biased), no current flows through the system. Also note that the current goes infinite because there are no resistors in the circuit. If a resistor were placed in series with the diode, it would act as a “current limiter” resulting in a current that went to a finite amount determined by the supply voltage and the resistance. ($V=IR...$)

We then learned that this didn't really represent the real world – in fact, the diode needs a little voltage in order to turn on – about 0.6 V or so. Therefore, a better representation of this characteristic is shown in Figure 2.47.

In fact, this isn't really a good representation of the real-world case either. The problem with this graph is that it leaves out a couple of important little details about the diode's behaviour. Take a look at Figure 2.48 which shows a good representation of the real-world diode.

There are a couple of things to notice here.

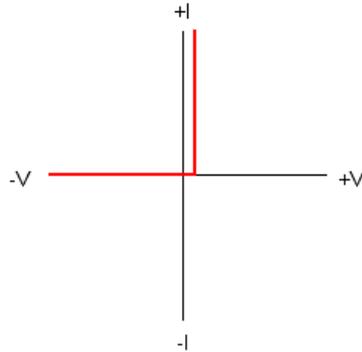


Figure 2.47: A graph showing the characteristics of a diode with slightly more realistic characteristics. Notice how you have to have about a small voltage applied to the diode before you can get current through it. The specific amount of voltage that you'll need depends on the materials used to make the diode as well as the particular characteristics of the one that you buy (in a package of 5 diodes, every one will be slightly different). If the voltage is negative (and therefore the diode is reverse biased), no current flows through the system.

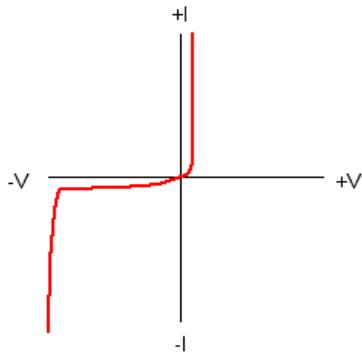


Figure 2.48: A graph showing the real-world characteristics of a diode.

Notice that the small turn-on voltage is still there, but notice that the current doesn't suddenly go from 0 amps to infinity amps when we pass that voltage. In fact, as soon as we go past 0 V, a very small amount of current will start to trickle through the diode. This amount is negligible in day-to-day operation, but it does exist. As we get closer and closer to the turn-on voltage, the amount of current trickling through gets bigger and bigger until it goes very big very quickly. The result on the graph is that the plot near the turn-on voltage is a curve, not a right angle as is shown in the simplification in Figure 2.47.

Also notice that when the diode is reverse-biased (and therefore the voltage is negative) there is also a very small amount of trickle current back through the diode. We tend to think of a reverse-biased diode as being a perfect switch that doesn't permit any current back through it, but in the real world, a little will get through.

The last thing to notice is the big swing to negative infinity amps when the voltage gets very negative. This is called the *reverse breakdown voltage* and you do not want to reach it. This is the point where the diode goes up in smoke and current runs through it whether you want it to or not.

2.9.2 Zener Diodes

There's a special type of diode that is supposed to be reversed biased. This device, called a *Zener Diode* has the characteristics shown in Figure 2.49.

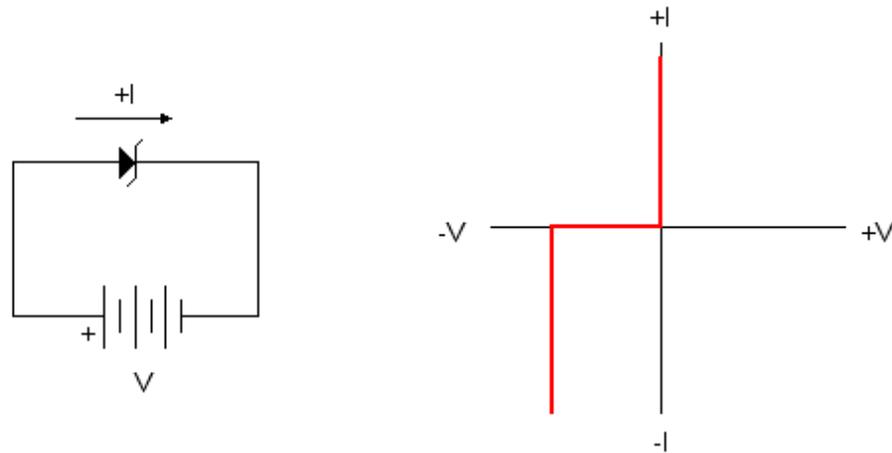


Figure 2.49: A graph showing the idealized characteristics of a zener diode.

Now, the *breakdown voltage* of the zener is a predicted value – in addition, when the zener breaks down and lets current through it in the opposite direction, it doesn't go up in smoke. When you buy a zener, it's rated for its breakdown voltage. So now, if you put the zener in series with a resistor and the zener is reverse biased, if the voltage applied to the circuit is bigger than the rated breakdown voltage, then the zener will ensure that the voltage across it is its rated voltage and the remaining voltage is across the resistor. This is useful in a bunch of applications that we'll see later on.

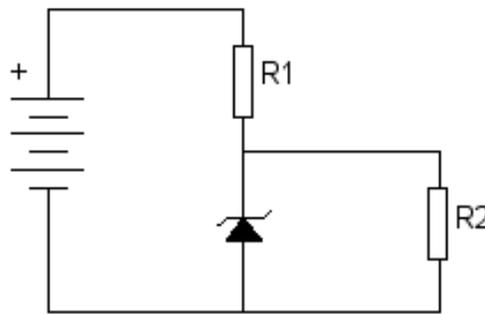


Figure 2.50: A circuit showing how a zener diode can be used to “regulate” one voltage to another. In order for this circuit to work, the voltage supply on the left must be a higher voltage than the rated breakdown voltage of the zener. The voltage across the resistor on the right will be the rated breakdown voltage of the zener. The voltage across the resistor on the left will be equal to the level of the voltage supply minus the rated breakdown voltage of the zener diode. Also, we're assuming that R1 is small compared to R2 so that the voltage across R2 without the zener in place would be normally bigger than the breakdown voltage of the zener.

So, for example, if the rated breakdown voltage of the zener diode in Figure 2.50 is 5.6 V, and the voltage supply is a 9 V battery, then the voltage across R2 will be 5.6V (because you can't have a higher voltage than that across the zener) and the voltage across R1 will be $9V - 5.6V = 3.4V$. (Remember, we're assuming that the voltage across R2 would be bigger than 5.6V if the zener wasn't there...)

Note as well that the graph in Figure 2.49 shows the characteristics of an ideal zener diode. The real-world characteristics suffer from the same trickle problems as normal diodes. For more info on this, a good book to look at is the book by Madhu in the Suggested Reading List below.

2.9.3 Suggested Reading List

Electronics: Circuits and Systems Swaminathan Madhu 1985 by Howard W. Sams and Co. Inc. ISBN 0-672-21984-0

2.10 Rectifiers and Power Supplies

What use are diodes to us? Well, what happens if we replace the battery from last chapter's circuit with an AC source as shown in the schematic below?

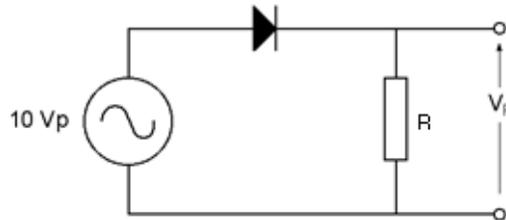


Figure 2.51: Circuit with AC voltage source, 1 diode and 1 resistor.

Now, when the voltage output of the function generator is positive relative to ground, it is pushing current through the forward-biased diode and we see current flowing through the resistor to ground. There's just the small issue of the 0.6 V drop across the diode, so until the voltage of the function generator reaches 0.6 V, there is no current, after that, the voltage drop across the resistor is 0.6 V less than the function generator's voltage level until we get back to 0.6 V on the way down...

When the voltage of the function generator is on the negative half of the wave, the diode is reverse-biased and no current flows, therefore there is no voltage drop across the resistor.

This circuit is called a *half-wave rectifier* because it takes a wave that is alternating between positive and negative voltages and turns it into a wave that has only positive voltages – but it throws away half of the wave...

If we connect 4 diodes as shown in the diagram below, we can use our AC signal more efficiently.

Now, when the output at the top of the function generator is positive, the current is pushed through to the diodes and sees two ways to go – one diode (the green one) will allow current through, while the other (red) one, which is reverse biased, will not. The current flows through the green diode to a junction where it chooses between a resistor and another reverse-biased diode (the blue one) ... so it goes through the resistor (note the direction of the current) and on to another junction between two diodes. Again, one of these diodes is reverse-biased (red) so it goes through the other one (yellow) back to the ground of the function generator.

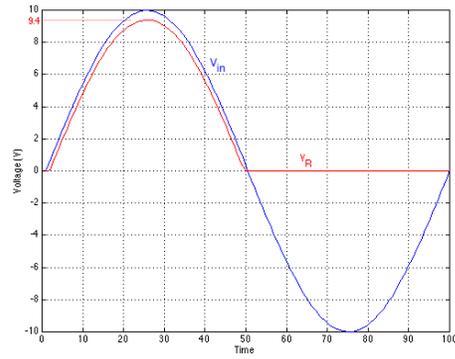


Figure 2.52: Comparison of voltage of function generator in blue and voltage across the resistor in red.

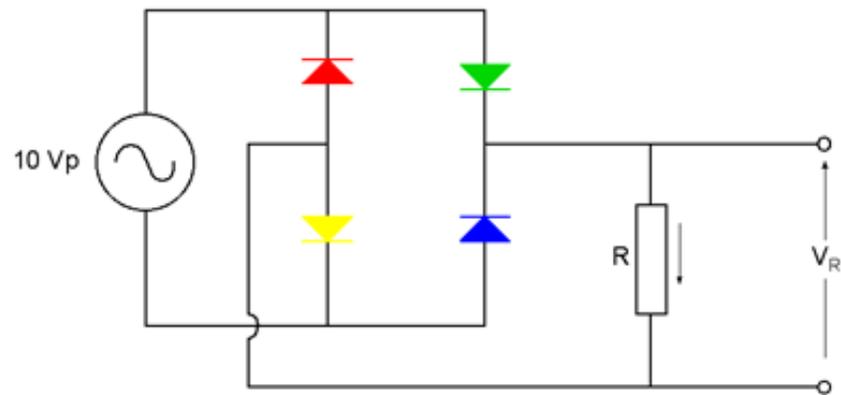


Figure 2.53: A smarter circuit for rectifying an AC waveform.

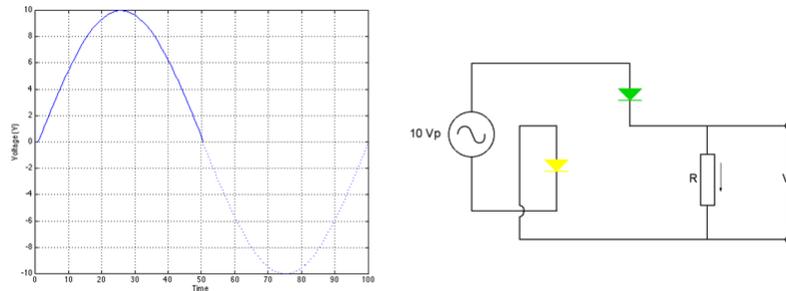


Figure 2.54: The portion of the circuit that has current flow for the positive portion of the waveform. Note that the current is flowing downwards through the resistor, therefore V_R will be positive.

When the function generator is outputting a negative voltage, the current follows a different path. The current flows from ground through the blue diode, through the resistor (note that the direction of the current flow is the same – therefore the voltage drop is of the same polarity) through the red diode back to the function generator.

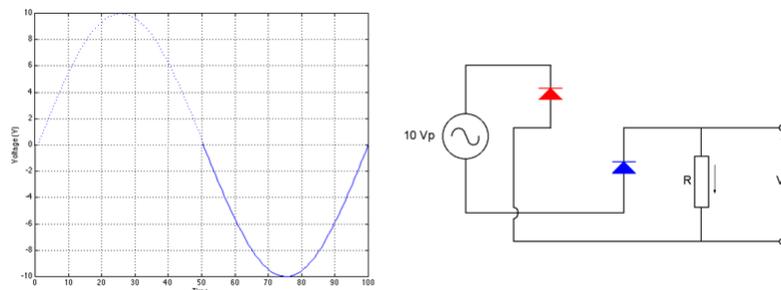


Figure 2.55: The portion of the circuit that has current flow for the positive portion of the waveform. Note that the current is flowing downwards through the resistor, therefore V_R will still be positive.

The important thing to notice after all that tracing of signal is that the voltage drop across the resistor was positive whether the output of the function generator was positive or negative. Therefore, we are using this circuit to fold the negative half of the original AC waveform up into the positive side of the fence. This circuit is therefore called a *full-wave rectifier* (actually, this particular arrangement of diodes has a specific name – a *bridge rectifier*) Remember that at any given time, the current is flowing through

two diodes and the resistor, therefore the voltage drop across the resistor will be 1.2 V less than the input voltage (0.6 V per diode – we’re assuming silicon...)

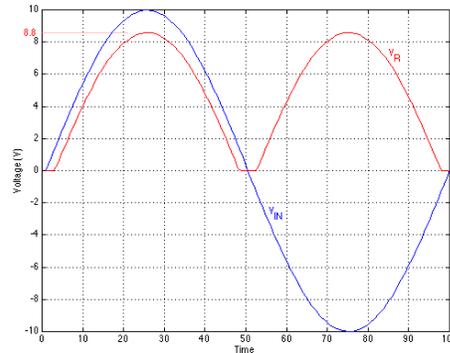


Figure 2.56: The input voltage and the output voltage of the bridge rectifier. Note that the output voltage is 0 V when the absolute value of V_{IN} is less than 1.2 V, or 1.2 V below the input voltage when the absolute value of V_{IN} is greater than 1.2 V.

Now, we have this weird bumpy wave – what do we do with it? Easy... if we run it through a type of low-pass filter to get rid of the spiky bits at the bottom of the waveform, we can turn this thing into something smoother. We won’t use a “normal” low-pass filter from two weeks ago, however... we’ll just put a capacitor in parallel with the resistor. What will this do? Well, when the voltage potential of the capacitor is less than the output of the bridge rectifier, the current will flow into the capacitor to charge it up to the same voltage as the output of the rectifier. This charging current will be quite high, but that’s okay for now... trust me... When the voltage of the bridge rectifier drops down, the capacitor can’t discharge back into it, because the diodes are now reverse-biased, so the capacitor discharges through the resistor according to their time constant (remember?). Hopefully, before it gets time to discharge, the voltage of the bridge rectifier comes back up and charges up the capacitor again and the whole cycle repeats itself.

The end result is that the voltage across the resistor is now a slightly weird AC with a DC offset, as is shown in Figure 2.57.

The width of the AC of this wave is given as a peak-peak measurement which is a percentage of the DC content of the wave. The smaller the percentage, the smoother and therefore better, the waveform.

If we know the value of the capacitor and the resistor, we can calculate

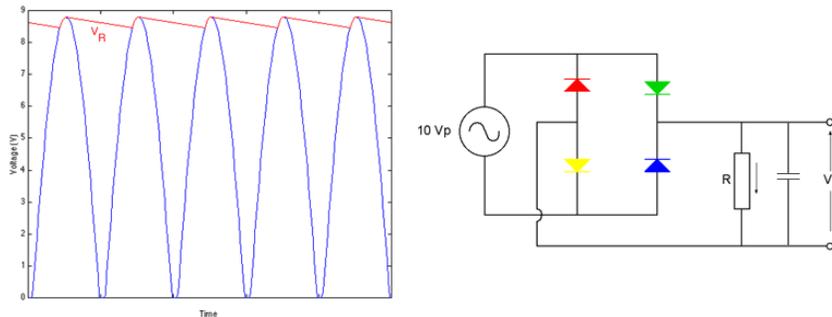


Figure 2.57: “DC” output of filtered bridge rectifier output showing ripple caused by the capacitor slowly discharging between cycles.

the ripple using the Equation 2.91 :

$$Ripple_{peak-peak} = \frac{100\%}{4fRC\sqrt{3}} \quad (2.91)$$

where f is the frequency of the original waveform in Hz, R is the value of the resistor in Ω and C is the value of the capacitor.

All we need to do, therefore, to make the ripple smaller, is to make the capacitor bigger (the resistor is really not a resistor in a real power supply, its actually something like a lightbulb or a portable CD player...)

Generally, in a real power supply, we'd add one more thing called a *voltage regulator* as is shown in Figures 8 and 9. This is a magic little device which, when fed a voltage above what you want, will give you what you want, burning off the excess as heat. They come in two flavours, negative and positive, the positive ones are designated 78XX where XX is the voltage (for example, a 7812 is a + 12 V regulator) the negative ones are designated 79XX (ditto... 7918 is a -18 V regulator.) These chips have 3 pins, one input, one ground and one output. You feed too much voltage into the input (i.e. 8.5 V into a 7805) and the chip looks at its ground, gives you exactly the right voltage at the output and gets toasty... If you use these things (you will) you'll have to bolt it to a little radiator or to the chassis of whatever you're building so that the heat will dissipate.

A couple of things about regulators : if you reverse-bias them (i.e. try and send voltage in its output) you'll break it... probably gonna see a bit of smoke too... Also, they get cranky if you demand too much current from their output. Be nice. (This is why you won't see regulators in a power supply for a power amp which needs lots-o-current...)

So, now you know how to build a real-live AC to DC power supply just like the pros. Just use an appropriate transformer instead of a function generator, plug the thing into the wall (fuses are your friend...) and throw away your batteries. The schematic below is a typical power supply of any device built before switching power supplies were invented... (we're not going to even try to figure out how THEY work...)

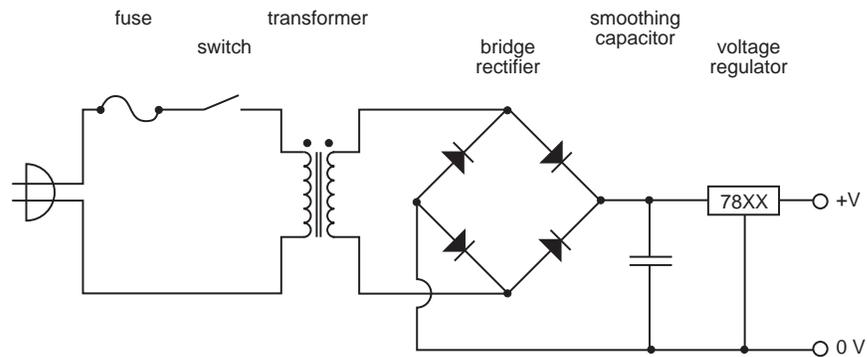


Figure 2.58: Unipolar power supply.

Below is another variation of the same power supply, but this one uses the centre-tap as the ground, so we get symmetrical negative and positive DC voltages output from the regulators.

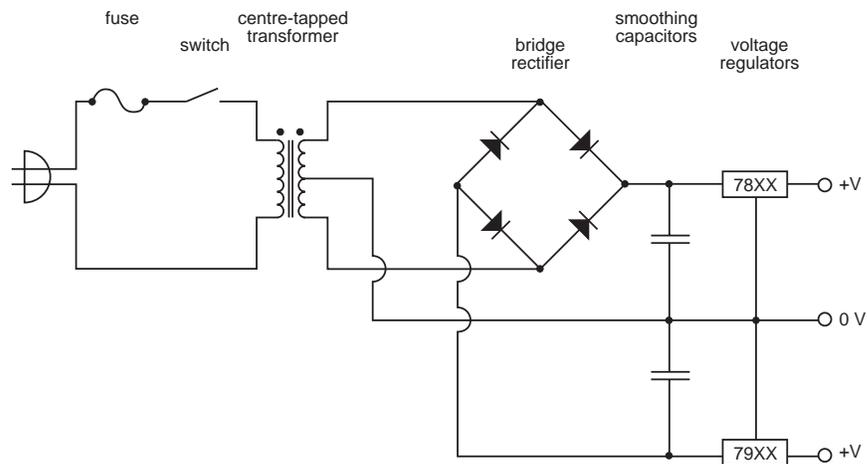


Figure 2.59: Bipolar power supply.

Suggested Reading List

2.11 Transistors

NOT YET WRITTEN

2.11.1 Suggested Reading List

2.12 Basic Transistor Circuits

NOT YET WRITTEN

2.12.1 Suggested Reading List

2.13 Operational Amplifiers

2.13.1 Introduction

If we take enough transistors and build a circuit out of them we can construct a device with three very useful characteristics. In no particular order, these are

1. infinite gain
2. infinite input impedance
3. zero output impedance

At the outset, these don't appear to be very useful, however, the device, called an *operational amplifier* or *op amp*, is used in almost every audio component built today. It has two inputs (one labeled “positive” or “non-inverting” and the other “negative” or “inverting”) and one output. The op amp measures the difference between the voltages applied to the two input “legs” (the positive minus the negative), multiplies this difference by a gain of infinity, and generates this voltage level at its output. Of course, this would mean that, if there was any difference at all between the two input legs, then the output would swing to either infinity volts (if the level of the non-inverting input was greater than that of the inverting input) or negative infinity volts (if the reverse were true). Since we obviously can't produce a level of either infinity or negative infinity volts, the op amp tries to do it, but hits a maximum value determined by the power supply rails that feed it. This could be either a battery or an AC to DC power supply such as the ones we looked at in Chapter 9.

We're not going to delve into how an op amp works or why – for the purposes of this course, our time is far better spent simply diving in and looking at how it's used. The simplest way to start looking at audio circuits which employ op amps is to consider a couple of possible configurations, each of which are, in fact, small circuits in and of themselves which can be combined like Legos to create a larger circuit.

2.13.2 Comparators

The first configuration we'll look at is a circuit called a *comparator*. You won't find this configuration in many audio circuits (blinking light circuits excepted) but it's a good way to start thinking about these devices.

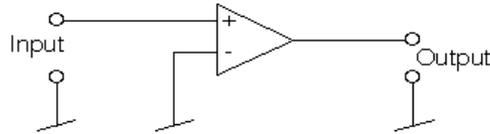


Figure 2.60: An op amp being used to create a comparator circuit.

Looking at the above schematic, you'll see that the inverting input of the op amp is connected directly to ground, therefore, it remains at a constant 0 V reference level. The audio signal is fed to the non-inverting input.

The result of this circuit can have three possible states.

1. If the audio signal at the non-inverting input is *exactly* 0 V then it will be the same as the level of the voltage at the inverting input. The op amp then subtracts 0 V (at the inverting input, because it's connected to ground) from 0 V (at the non-inverting input, because we said that the audio signal was at 0 V) and multiplies the difference of 0 V by infinity and arrives at 0 V at the output. (okay, okay – I know. 0 multiplied by infinity is really equal to any real number, but, in this case, that real number will always be 0)
2. If the audio signal is greater than 0 V, then the op amp will subtract 0 V from a positive number, arriving at a positive value, multiply that result by infinity and have an output of positive infinity (actually, as high as the op amp can go, which will really be the voltage of the positive power supply rail)
3. If the audio signal is less than 0 V, then the op amp will subtract 0 V from a negative number, arriving at a negative value, multiply that result by infinity and have an output of negative infinity (actually, as low as the op amp can go, which will really be the voltage of the negative power supply rail)

So, if we feed a sine wave with a level of 1 V_p and a frequency of 1 kHz into this comparator and power it with a 15 V power supply, what we'll see at the output is a 1 kHz square wave with a level of 15 V_p.

Unless you're a big fan of square waves or very ugly distortion pedals, this circuit will not be terribly useful to your audio circuits with one notable exception with we'll discuss later. So, how do we use op amps to our

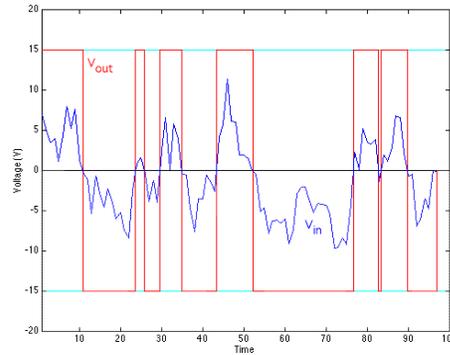


Figure 2.61: The output voltage vs. input voltage of the comparator circuit in Figure 1.

advantage? Well, the problem is that the infinite gain has to be tamed – and luckily this can be done with the helps of just a few resistors.

2.13.3 Inverting Amplifier

Take a look at the following circuit

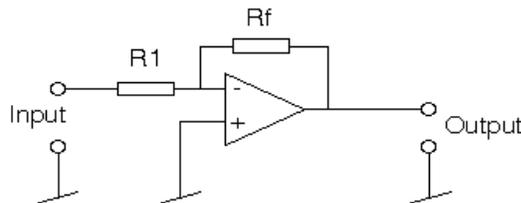


Figure 2.62: An op amp in an inverting amplifier configuration.

What do we have here? The non-inverting input of the op amp is permanently connected to ground, therefore it remains at a constant level of 0 V. The non-inverting input, on the other hand, is connected to a couple of resistors in a configuration that sort of resembles a voltage divider. The audio signal feeds into the input resistor labeled R1. Since the input impedance of the op amp is infinite, all of the current travelling through R1 caused by the input must be directed through the second resistor labeled Rf and therefore to the output. The big “problem” here is that the output of the op amp is connected to its own input and anyone who works in audio knows

that this means one thing... the dreaded monster known as feedback (which explains the “f” in “Rf” – it’s sometimes known as the feedback resistor). Well, it turns out that, in this particular case, feedback is your friend – this is because it is a special brand of feedback known as negative feedback.

There are a number of ways to conceptualize what’s happening in this circuit. Let’s apply a +1 V DC signal to the input R1 which we’ll assume to be 1 k – what will happen?

Let’s assume for a moment that the voltage at the inverting input of the op amp is 0 V. Using Ohm’s Law, we know that there is 1 mA of current flowing through R1. Since the input impedance of the op amp is infinity, there will be no current flowing into the amplifier – therefore all of the current must flow through Rf (which we’ll also make 1 kΩ) as well. Again using Ohm’s Law, we know that there’s 1 mA flowing through a 1 kΩ resistor, therefore there is a 1 V drop across it. This then means that the voltage at the output of Rf (and therefore the op amp) is -1 V. Of course, this magic wouldn’t happen without the op amp doing something...

Another way of looking at this is to say that the op amp “sees” 1 V coming in its inverting input – therefore it swings its output to negative infinity. That negative infinity volt output, however, comes back into the inverting input through Rf which causes the output to swing to positive infinity which comes back into the inverting input through Rf which causes the output to swing to negative infinity and so on and so on... All of this swinging back and forth between positive and negative infinity looks after itself, causing the 1 V input to make it through, inverted to -1 V.

One little thing that’s useful to know – remember that assumption that the level of the inverting input stays at 0 V? It’s actually a good assumption. In fact, if the voltage level of the inverting input was anything other than 0 V, the output would swing to one of the voltage rails. We can consider the inverting input to be at a virtual ground – “ground” because it stays at 0 V but “virtual” because it isn’t actually connected to ground.

What happens when we change the value of the two resistors? We change the gain of the circuit. In the above example, with both R1 and Rf were 1 kΩ and this resulted in a gain of -1. In order to achieve different gains we follow the below equation:

$$Gain = -\frac{R_f}{R_1} \quad (2.92)$$

2.13.4 Non-Inverting Amplifier

There are a couple of advantages of using the inverting amplifier – you can have any gain you want (as long as it's negative or zero) and it only requires two resistors. The disadvantage is that it inverts the polarity of the signal – so in order to maintain the correct polarity of a device, you need an even number of inverting amplifier circuits, thus increasing the noise generated by extra components.

It is possible to use a single op amp in a non-inverting configuration as shown in the schematic below.

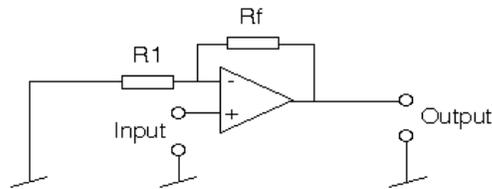


Figure 2.63: An op amp in a non-inverting amplifier configuration.

Notice that this circuit is similar to the inverting op amp configuration in that there is a feedback resistor, however, in this case, the input to R1 is connected to ground and the signal is fed to the non-inverting input of the op amp.

We know that the voltage at the non-inverting input of the op amp is the same as the voltage of the signal, for now, let's say +1 V again. Following an assumption that we made in the case of the inverting amplifier configuration, we can say that the level of the two input legs of the op amp are always matched. If this is the case, then let's follow the current through R1 and Rf. The voltage across R1 is equal to the voltage of the signal, therefore there is 1 mA of current flowing through R1, but this time from right to left. Since the impedance of the input of the op amp is infinite, all of the current flowing through R1 must be the same as is flowing through Rf. If there is 1 mA of current flowing through Rf, then there is a 1 V difference across it. Since the voltage at the input leg side of the op amp is +1 V, and there is another 1 V across Rf, then the voltage at the output of the op amp is +2 V, therefore the circuit has a gain of 2.

The result of this circuit is a device which can amplify signals without inverting the polarity of the original input voltage. The only drawback of this circuit lies in the fact that the voltages at the input legs of the op amp

must be the same. If the value of the feedback resistor is 0Ω , in other words, a piece of wire, then the output will equal the input voltage, therefore the gain of the circuit will be 1. If the value of the feedback resistor is greater than 0Ω , then the gain of the circuit will be greater than 1. Therefore the minimum gain of this circuit is 1 – so we cannot attenuate the signal as we can with the inverting amplifier configuration.

Following the above schematic, the equation for determining the gain of the circuit is

$$Gain = 1 + \frac{R_f}{R_1} \quad (2.93)$$

2.13.5 Voltage Follower

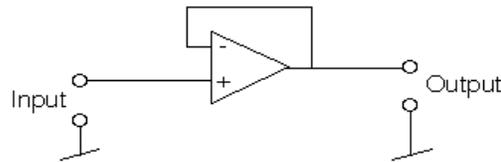


Figure 2.64: An op amp in a voltage follower configuration.

There is a special case of the non-inverting amplifier configuration which is used frequently in audio circuitry to isolate signals within devices. In this case, the feedback “resistor” has a value of 0Ω – a wire. We also omit the connection between the inverting input leg and the ground plane. This resistor is effectively unnecessary since setting the value of R_f to 0 makes the gain equation go immediately to 1. Changing the value of R_1 will have no effect on the gain, whether it’s infinite or finite, as long as it’s not 0Ω . Omitting the R_1 resistor makes the value infinity.

This circuit will have an output which is identical to the input voltage, therefore it is called a *voltage follower* (also known as a *buffer*) since it follows the signal level. The question then is, “what use is it?” Well, it’s very useful if you want to isolate a signal so that whatever you connect the output of the circuit to has no effect on the circuit itself. We’ll go into this a little farther in a later chapter.

2.13.6 Leftovers

There is one of the three characteristics of op amps that we mentioned up front that we haven't talked about since. This is the output impedance, which was stated to be 0. Why is this important? The answer to this lies in two places. The first is the simple voltage divider, the second is a block diagram of an op amp. If you look at the diagram below, you'll see that the op amp contains what can be considered as a function generator which outputs through an internal resistor (the output impedance) to the world. If we add an external load to the output of the op amp then we create a voltage divider. If the internal impedance of the op amp is anything other than 0Ω , then the output voltage of the amplifier will drop whenever a load is applied to it. For example, if the output impedance of the op amp was 100Ω , and we attached a 100Ω resistor to its output, then the voltage level of the output would be cut in half.

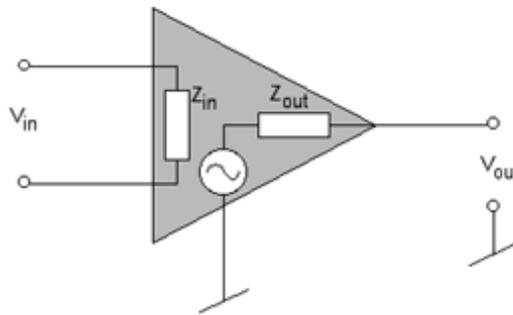


Figure 2.65: A block diagram of the internal workings of an op amp.

2.13.7 Mixing Amplifier

How to build a mixer: take an inverting amplifier circuit and add a second input (through a second resistor, of course...). In fact, you could add as many extra inputs (each with its own resistor) as you wanted to. The current flowing through each individual resistor to the virtual ground winds up at the bottleneck and the total current flows through the feedback resistor.

The total voltage at the output of the circuit will be:

$$V_{out} = -\left(V_1 \frac{R_f}{R_1} + V_2 \frac{R_f}{R_2} + \dots + V_n \frac{R_f}{R_n}\right) \quad (2.94)$$

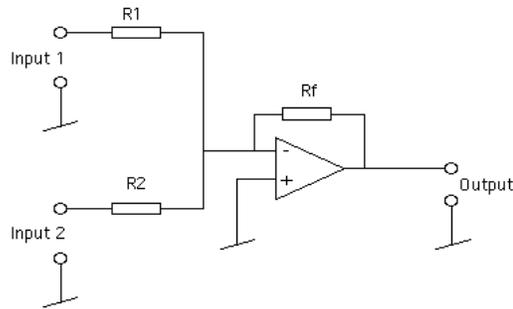


Figure 2.66: An op amp in an inverting mixing amplifier configuration

As you can see, each input voltage is inverted in polarity (the negative sign at the beginning of the right side of the equation looks after that) and individually multiplied by its gain determined by the relationship between its input resistor and the feedback resistor. This, of course, is a very simple mixing circuit (a Euphonix console it isn't...) but it will work quite effectively with a minimum of parts.

2.13.8 Differential Amplifier

The mixing circuit above adds a number of different signals and outputs the result (albeit inverted in polarity and possibly with a little gain added for flavour...). It may also be useful (for a number of reasons that we'll talk about later) to subtract signals instead. This, of course, could be done by inverting a signal before adding them (adding a negative is the same as subtracting...) but you could be a little more efficient by using the fact that an op amp is a differential amplifier all by itself.

You will notice in the above diagram that there are two inputs to a single op amp. The result of the circuit is the difference between the two inputs as can be seen in the following equation. Usually the gain of the two inputs of the amplifier are set to be equal because the usual use of this circuit is for balanced inputs which we'll discuss in a later chapter.

2.13.9 Suggested Reading List

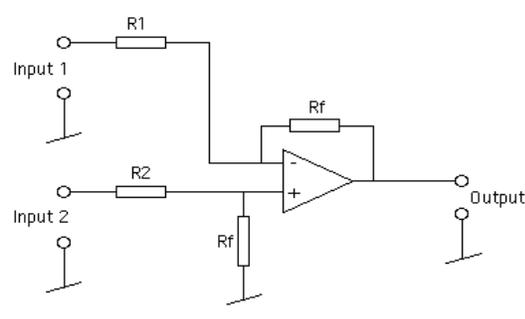


Figure 2.67: An op amp in a simple differential amplifier configuration

2.14 Op Amp Characteristics and Specifications

2.14.1 Introduction

If you get the technical information about an op amp (from the manufacturer's datasheets or their website) you'll see a bunch of specifications that tell you exactly what to expect from that particular device. Following is a list of specs and explanations for some of the more important characteristics that you should worry about.

I've taken most of these from a book called the *IC Op-Amp Cookbook* written by Walter G. Jung. It's used to be published by SAMS, but now it's published by Prentice Hall (ISBN 0-13-889601-1) You should own a copy of this book if you're planning on doing anything with op amps.

2.14.2 Maximum Supply Voltage

The supply voltage is the maximum voltage which you can apply to the power supply inputs of the op amp before it fails. Depending on who made the amplifier and which model it is, this can vary from ± 1 V to ± 40 V.

2.14.3 Maximum Differential Input Voltage

In theory, there is no connection at all between the two inputs of an operational amplifier. If the difference in voltage between the two inputs of the op amp is excessive (on the order of up to ± 30 V) then things inside the op amp break down and current starts to flow between the two inputs. This is bad – and it happens when the difference in the voltages applied to the two inputs is equal to the maximum differential input voltage.

2.14.4 Output Short-Circuit Duration

If you make a direct connection between the output of the amplifier and either the ground or one of the power supply rails, eventually the op amp will fail. The amount of time this will take is called the output short-circuit duration, and is different depending on where the connection is made. (whether it's to ground or one of the power supply rails)

2.14.5 Input Resistance

If you connect one of the two input legs of the op amp to ground and then measure the resistance between the other input leg and ground, you'll find that it's not infinite as we said it would be in the previous chapter. In fact,

it'll be somewhere up around $1M\Omega$ give or take. This value is known as the "input resistance."

2.14.6 Common Mode Rejection Ratio (CMRR)

In theory, if we send exactly the same signal to both input legs of an op amp and look at the output, we should see a constant level of 0 V. This is because the op amp is subtracting the signal from itself and giving you the result (0 V) multiplied by some gain. In practice, however, you won't see a constant 0 V at the output – you'll see an attenuated version of the signal that you're sending into the amplifier's inputs. The ratio between the level of the input signal and the level of the resulting output is a measurement of how much the op amp is able to reject a common mode signal (in other words, a signal which is common to both input legs). This is particularly useful (as we'll see later) in rejecting noise and other unwanted signals on a transmission line.

The higher this number the better the rejection, and you should see values in the area of about 100 dB. Be careful though – op amps have very different CMRR's for different frequencies (lower frequencies reject better) so take a look at what frequency you're talking about when you're looking at CMRR values.

2.14.7 Input Voltage Range (Operating Common-Mode Range)

This is simply the range of voltage that you can send to the input terminals while ensuring that the op amp behaves as you expect it to. If you exceed the input voltage range, the amplifier could do some unexpected things on you.

2.14.8 Output Voltage Swing

The output voltage swing is the maximum peak voltage that the output can produce before it starts clipping. This voltage is dependent on the voltage supplied to the op amp – the higher the supply voltage the higher the output voltage swing. A manufacturer's datasheet will specify an output voltage swing for a specific supply voltage – usually ± 15 V for audio op amps.

2.14.9 Output Resistance

We said in the last chapter that one of the characteristics of op amps is that they have an output impedance of 0Ω . This wasn't really true... In fact,

the output impedance of an op amp will vary from less than 100Ω to about $10k\Omega$. Usually, an op amp intended for audio purposes will have an output impedance in the lower end of this scale – usually about 50Ω to 100Ω or so. This measurement is taken without using a feedback loop on the op amp, and with small signal levels above a few hundred Hz.

2.14.10 Open Loop Voltage Gain

Usually, we use op amps with a feedback loop. This is in order to control what we said in the last chapter was the op amp’s infinite gain. In fact, an op amp does not have an infinite gain – but it’s pretty high. The gain of the op amp when no feedback loop is connected is somewhere up around 100 dB. Of course, this is also dependent on the load that the output is driving, the smaller the load, the smaller the output and therefore the smaller the gain of the entire device.

The open loop voltage gain is a measurement of the gain of the op amp driving a specified load when there’s no feedback loop in place (hence the “open loop”). 100 dB is a typical value for this specification. Its value doesn’t change much with changes in temperature (unlike some other specifications) but it does vary with frequency. As you go up in frequency, the gain of the op amp decreases.

2.14.11 Gain Bandwidth Product (GBP)

Eventually, if you keep going up in frequency as you’re measuring the open-loop voltage gain of the op amp, you’ll reach a frequency where the gain of the system is 1. That is to say that the input is the same level as the output. If we take the gain at that frequency (1) and multiply it by the frequency (or bandwidth) we’ll get the Gain Bandwidth Product. This will then tell you an important characteristic of the op amp, since the gain of the op amp rolls off at a rate of 6 dB/octave. If you take a given frequency and divide it by the GBP, you’ll find the open-loop gain at that frequency.

2.14.12 Slew Rate

In theory, an op amp is able to accurately and instantaneously output a signal which is a copy of the input, with some amount of gain applied. If, however, the resulting output signal is quite large, and there are very fast, large changes in that value, the op amp simply won’t be able to deliver on time. For example, if we try to make the amplifier’s output swing from -10 V to $+10\text{ V}$ instantaneously, it can’t do it – not quite as fast as

we'd like, anyway... The maximum rate at which the op amp is able to change to a different voltage is called the "slew rate" because it's the rate at which the amplifier can slew to a different value. It's usually expressed in V/microsecond – the bigger the number, the faster the op amp. The faster the op amp, the better it is able to accurately reflect transient changes in the audio signal.

The slew rate of different op amps varies widely. Typically, you'll want to see about 5 V/microsec or more.

2.14.13 Suggested Reading List

IC Op-Amp Cookbook Walter G. Jung Prentice Hall ISBN 0-13-889601-1

2.15 Practical Op Amp Applications

2.15.1 Introduction

2.15.2 Active Filters

Discussion of the advantages of an active filter instead of a passive one.

Butterworth filters

emphWhy use a Butterworth instead of anything else? What's the difference between different filters? Talk about phase responses and magnitude responses.

Low Pass (First Order)

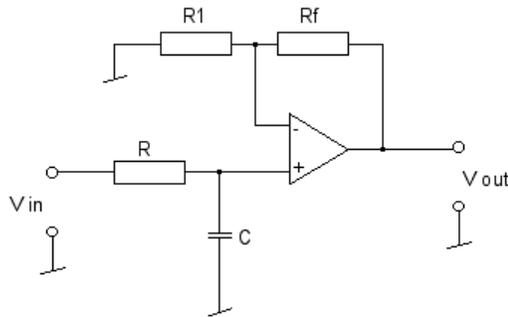


Figure 2.68: First order low-pass Butterworth filter

Equations:

$$G_F = 1 + \frac{R_F}{R_1} \quad (2.95)$$

$$f_c = \frac{1}{2\pi RC} \quad (2.96)$$

Where G_F is the passband gain of the filter

f_c is the cutoff frequency of the filter

$$\left| \frac{V_{out}}{V_{in}} \right| = \frac{G_F}{\sqrt{1 + (f/f_c)^2}} \quad (2.97)$$

$$\phi = -\tan^{-1} \frac{f}{f_c} \quad (2.98)$$

Where ϕ is the phase angle

High Pass (First Order)

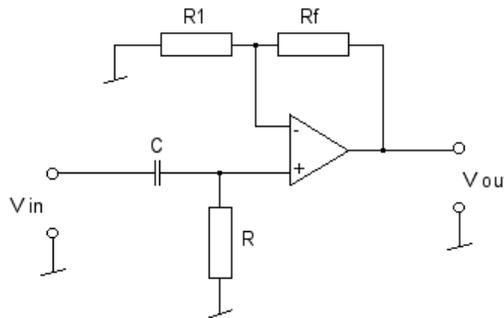


Figure 2.69: First order high-pass Butterworth filter

Equations:

$$G_F = 1 + \frac{R_F}{R_1} \quad (2.99)$$

$$f_c = \frac{1}{2\pi RC} \quad (2.100)$$

Where G_F is the passband gain of the filter

f_c is the cutoff frequency of the filter

$$\left| \frac{V_{out}}{V_{in}} \right| = \frac{G_F(f/f_c)}{\sqrt{1 + (f/f_c)^2}} \quad (2.101)$$

How to design a first-order high pass Butterworth filter:

Determine your cutoff frequency

Select a mylar or tantalum capacitor with a value of less than about 1 μ F.

Calculate the value of R using $R = \frac{1}{2\pi f_c C}$

Select values for R1 and R2 (on the order of 10k Ω or so) to deliver your desired passband gain.

Low Pass (Second Order)

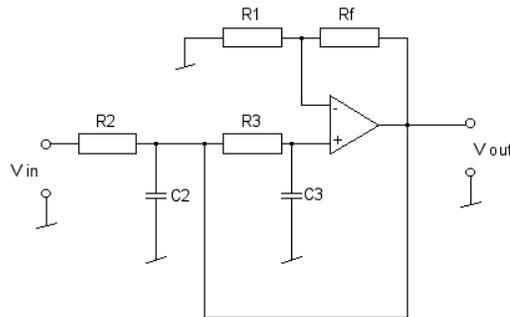


Figure 2.70: Second order low-pass Butterworth filter

Equations:

$$G_F = 1 + \frac{R_F}{R_1} \quad (2.102)$$

$$f_c = \frac{1}{2\pi\sqrt{R_2 R_3 C_2 C_3}} \quad (2.103)$$

Where G_F is the passband gain of the filter
 f_c is the cutoff frequency of the filter

$$\left| \frac{V_{out}}{V_{in}} \right| = \frac{G_F}{\sqrt{1 + (f/f_c)^4}} \quad (2.104)$$

Equations:

$$G_F = 1 + \frac{R_F}{R_1} \quad (2.105)$$

Note: G_F must equal 1.586 in order to have a true Butterworth response.

$$f_c = \frac{1}{2\pi\sqrt{R_2 R_3 C_2 C_3}} \quad (2.106)$$

Where G_F is the passband gain of the filter
 f_c is the cutoff frequency of the filter

$$\left| \frac{V_{out}}{V_{in}} \right| = \frac{G_F}{\sqrt{1 + (f_c/f)^4}} \quad (2.107)$$

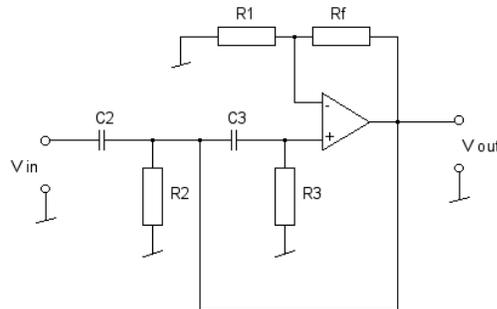


Figure 2.71: Second order high-pass Butterworth filter

How to design a second-order high pass Butterworth filter:

Determine your cutoff frequency

Make $R_2 = R_3 = R$ and $C_2 = C_3 = C$

Select a mylar or tantalum capacitor with a value of less than about 1 μ F.

Calculate the value of R using $R = \frac{1}{2\pi f_c C}$

Choose a value of R1 that's less than 100k Ω

Make $R_f = 0.586 R_1$. Note that this makes your passband gain approximately equal to 1.586. This is necessary to guarantee a Butterworth response.

2.15.3 Higher-order filters

In order to make higher-order filters, you simply have to connect the filters that you want in series. For example, if you want a third-order filter, you just take the output of a first-order filter and feed it into the input of a second-order filter. For a fourth-order filter, you just cascade two second-order filters and so on. This isn't exactly as simple as it seems, however, because you have to pay attention to the relationship of the cutoff frequencies and passband gains of the various stages in order to make the total filter behave properly. For more information on this, I'm afraid that you'll have to look elsewhere for now... However, if you're not overly concerned with pinpoint accuracy of the constructed filter to the theoretical calculated response, you can just use what you already know and come pretty close to making a workable high-order filter. Just don't call it a "Butterworth" and don't expect it to be perfect.

2.15.4 Bandpass filters

Bandpass filters are those that permit a band of frequencies to pass through the filter, while attenuating signals in frequency bands above and below the passband. These types of filters should be sub-divided into two categories: those with a wide passband, and those with a narrow passband.

If the passband of the bandpass filter is wide, then we can create it using a high-pass filter in series with a low-pass filter. The result will be a signal whose low-frequency content will be attenuated by the high-pass filter and whose high-frequency content will be attenuated by the low-pass filter. The result is that the band of frequencies between the two cutoff frequencies will be permitted to pass relatively unaffected. One of the nice aspects of this design is that you can have different orders of filters for your high and low pass (although if they are different, then you have a bit of a weird bandpass filter...)

If the passband of the bandpass filter is narrow (if the Q is greater than 10), then you'll have to build a different circuit that relies on the resonance of the filter in order to work. Take a look at the book by Gayakwad in the Reading List (or any decent book on op amp applications) to see how this is done for now... I'll come back to it some other day.

2.15.5 Suggested Reading List

Op-amps and Linear Integrated Circuit Technology Ramakant A. Gayakwad
1983 by Prentice-Hall Inc. ISBN 0-13-637355-0

Chapter 3

Acoustics

3.1 Introduction

3.1.1 Pressure

If you listen to the radio in the mornings, they'll give you the news, the sports, the traffic and the weather. Part of the weather report is to tell you that the barometric pressure is something around 100 kPa (kilopascals). What does this mean? Well, the air particles around you are all under pressure due to things like gravity and the weight of the air particles above them and other meteorological things that are outside the scope of this book. That pressure determines the amount of physical space between molecules in the air. When there's a higher barometric pressure, there's less space between the molecules than there is on a day with a lower barometric pressure.

We call this the *stasis pressure* and abbreviate it \wp_o .

When all of the particles in a gaseous medium (like air) in a given volume (like a room) are at normal pressure, then the gas is said to be at its *volume density* (also known as the *constant equilibrium density*), abbreviated ρ_o , and measured in kg/m^3 . Remember that this is actually kilograms of air per cubic metre – if you were able to trap a cubic metre and weigh it, you'd find out that it is about XXX kg.

These molecules like to stay at the same pressure all over, so if you bunch them up in one place in a room somehow, they'll move around to try and equalize the difference. This is kind of like when you pour a glass of water into a bucket, the water level of the entire bucket equalizes and therefore rises, rather than the water from the glass all bunching up in a little mound of water where you poured it in...

Let's think of this as a practical example. We'll hang the piece of paper in front of a fan. If we turn on the fan, we're essentially increasing the pressure of the air particles in front of the blades. The fan does this by removing air particles from the space behind it, thus reducing the pressure of the particles behind the blades, and putting them in front. Since the pressure in front of the fan is greater than any other place in the room, we have a situation where there is a greater air pressure on one side of the piece of paper than the other. The obvious result is that the paper moves away from the fan.

This is a large-scale example of how you hear sound. Let's say hypothetically for a moment, that you are sitting alone in a sealed room on a day when the barometric pressure is 100 kPa. Let's also say that you have a clarinet with you and that you play a concert A. What physically happens to convert air coming out of your mouth into a concert A coming in your ears?

To begin with, let's pretend that a clarinet is just a tube with a hole in each end. One of the holes has a springy piece of wood next to it which, if you press on it, will close up the hole.

1. When you blow into the hole, you bunch up the air particles and create a little area of high pressure inside the mouthpiece.
2. Blowing into the hole with the reed on it also has the effect of pushing the reed against the hole and sealing it so that no more air can enter the clarinet.
3. At that point the little high pressure area moves down the clarinet and leaves a low pressure behind it.
4. Remember that the reed is springy, and it doesn't like being pushed up against the hole in the mouthpiece, so it bounces back and lets more air in.
5. Now the cycle repeat and goes back to step 1 all over again.
6. In the meantime, all of those high and low pressure areas move down the clarinet and radiate out the bell into the room like ripples on a lake when you throw in a rock.
7. From there, they get to your ear and push your eardrum in and out (high pressure pushes in, low pressure pulls out)

Those little fluctuations in the air pressure are small variations in the stasis pressure. They're usually very small, never more than about ± 1 Pa (though we'll elaborate on that later...). At any given moment at a specific location, we can measure the *instantaneous pressure*, \wp , which will be close to the stasis pressure, but slightly different because there's a sound source causing it to change.

Once we know the stasis pressure and the instantaneous pressure, we can use these to figure out the *instantaneous amplitude* of the sound level, (also called the *acoustic pressure* or the *excess pressure*) abbreviated p , using Equation 3.1.

$$p = \wp - \wp_o \quad (3.1)$$

To see an animation of what this looks like, check out www.gmi.edu/~drussell/Demos/waves/wavemotion.html.

A sinusoidal oscillation of this pressure reaches a maximum peak pressure P which determines the *sound pressure level* or *SPL*. In air, this level is typically expressed in decibels as a logarithmic ratio of the *effective pressure* P_e referenced to the *threshold of hearing*, the commonly-accepted lowest sound pressure level audible by humans at 1 kHz, 20 microPascals, using Equation 3.2 [Woram, 1989]. The intricacies of this equation have already been discussed in Section 2.2 on decibels.

$$SPL = 20 \log_{10} \left[\frac{P_e}{20 * 10^{-6} Pa} \right] \quad (3.2)$$

Note that, for sinusoidal waveforms, the effective pressure can be calculated from the peak pressure using Equation 3.3. (If this doesn't sound familiar, it should – re-read Section 2.1.6 on RMS.)

$$P_e = \frac{P}{\sqrt{2}} \quad (3.3)$$

3.1.2 Simple Harmonic Motion

Take a weight (a little one...) and hang it on the end of a Slinky which is attached to the ceiling and wait for it to stop bouncing.

Measure the length of the Slinky. This length is determined by the weight and the strength of the Slinky. If you use a bigger weight, the Slinky will be longer – if the slinky is stronger, it will be better able to support the weight and therefore be shorter.

This is the point where the “system” is at rest or stasis.

Pull down on the weight a little bit and let go. The Slinky will pull the weight up to the stasis point and pass it.

By the time the whole thing slows down, the weight will be too high and will want to come back down to the stasis point, which it will do, stopping at the point where we let it go in the first place (or almost anyway...)

If we attached a pen to the weight and ran piece of paper along by it as it sat there bobbing up and down, the line it would draw a sinusoidal waveform. The picture the weight would draw is a graph of the vertical position of the weight (the y-axis) as it relates to time (the x-axis).

If the graph is a perfect sinusoidal shape, then we call the system (the Slinky and the weight on the end) a *simple harmonic oscillator*.

3.1.3 Damping

Let's look at that system I just described. We'll put a weight hung on a spring as is shown in Figure 3.1

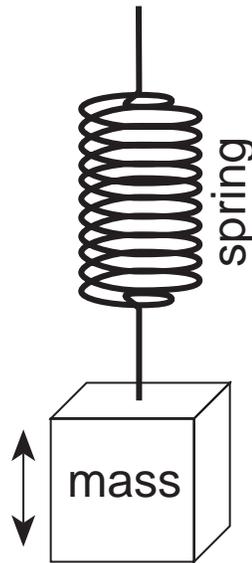


Figure 3.1: A mass supported by a spring.

If there was no such thing as air friction, and if the spring was perfect, then, if you started the mass bobbing up and down, then it would continue doing that forever. And, since, as we saw in the previous section, that this is a simple harmonic oscillator, then if we graph its vertical displacement

over time, then we get a perfect sinusoidal waveform as shown in Figure 3.2

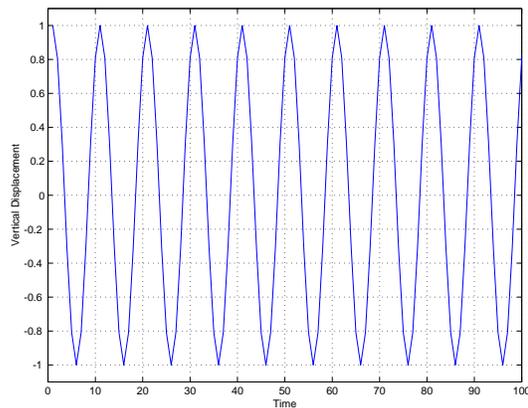


Figure 3.2: The vertical displacement of the mass versus time if there is no loss of energy due to friction. Notice that the frequency and amplitude of the oscillation never change. The mass will bob up and down exactly the same, forever.

In real life, however, there is friction. The mass pushes through the air and loses energy on each bob up and down. Eventually, it loses so much energy that it stops moving. An example of this behaviour is shown in Figure 3.3

There is a technical term that describes the difference between these two situations. The system with friction, shown in Figure 3.3 is called a *damped oscillator*. Since the oscillator is damped, then it loses energy over time. The higher the damping, the faster it loses energy. For example, if the same mass and spring were put in water, the system would be more highly damped than if it were in air. If they're put in oil, the system is more highly damped than it is in water.

Since a system with friction is said to be damped, then the system without friction is therefore called an *undamped oscillator*.

3.1.4 Harmonics

If we go back to the clarinet example, it's pretty obvious that the pressure wave that comes out the bell won't be a sine wave. This is because the clarinet reed is doing more than simply opening and closing – it's also wiggling and flapping a bit – on top of all that, the body of the clarinet is resonating various frequencies as well (more on this topic later), so what comes out is a bunch of different frequencies simultaneously.

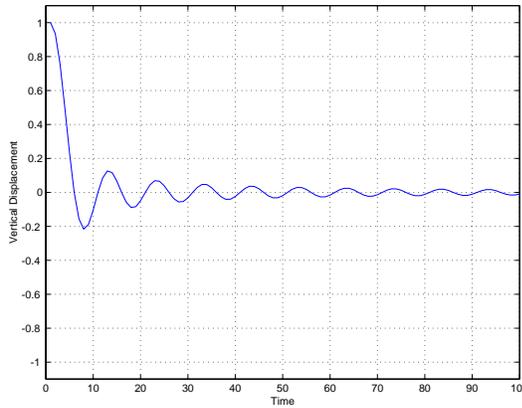


Figure 3.3: The vertical displacement of the mass versus time if there is loss of energy due to friction. Notice that the fundamental frequency of the oscillation never changes, but that the amplitude decays over time. Eventually, the mass will bob up and down so little that it can be considered to be stopped.

We call these other frequencies *harmonics* which are mathematically related to the bottom frequency (called the *fundamental*) by simple multiplication... The first harmonic is the fundamental. The second harmonic is twice the frequency of the fundamental, the third harmonic is three times the frequency of the fundamental and so on. (this is an oversimplification which we'll straighten out later...)

3.1.5 Overtones

Some people call the fundamental and its overtones *overtones* but you have to be careful here. There is a common misconception that overtones are harmonics and vice versa. In fact, in some books, you'll see people saying that the first overtone is the second harmonic, the second overtone is the third harmonic and so on. This is not necessarily the case. A sound's overtones are the harmonics that it contains, which is not necessarily all harmonics. As we'll see later, not all instruments' sounds contain all harmonics of the fundamental. There are particular cases, for example, where an instrument's sound will only contain the odd harmonics of the fundamental. In this particular case, the first overtone is the third harmonic, the second overtone is the fifth harmonic and so on.

In other words, harmonics are a mathematical idea – frequencies that are related to a fundamental frequency whereas overtones are the frequencies

that are produced by the sound source.

Another example showing that overtones are not harmonics occurs in many percussion instruments such as bells where the overtones have no harmonic relationship with the fundamental frequency – which is why these overtones are said to be *enharmonically related*.

3.1.6 Longitudinal vs. Transverse Waves

There are basically three types of waves used to transmit energy through a medium or substance.

1. Transverse
2. Longitudinal
3. Torsional

We're only really concerned with the first two.

Transverse waves are the kind we see every day in ropes and puddles. They're the kind where the motion of the particles is perpendicular to the direction of the wave propagation as can be seen in Figure 3.4. What does this mean? It's easy to see if we go fishing... A boat on the surface of the ocean will sit there bobbing up and down as the waves roll past it. The waves are travelling towards the shore along the surface of the water, but the water itself only moves up and down, not sideways (we know this because the boat would move sideways as well if the water was doing so...) So, as the water molecules move vertically, the wave propagates horizontally.

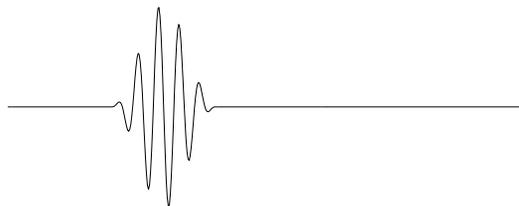


Figure 3.4: A snapshot of a transverse wave on a string. Think of the wave as moving from left to right – but remember that the string is really only moving up and down.

Longitudinal waves are a little tougher to see. They involve the compression (bunching together) and refraction (pulling apart) of the particles in the medium such that the motion of the particles is parallel with the direction of propagation of the wave. The easiest way to see a longitudinal wave is to stretch out a Slinky between two people, squeeze together a small section of it and let go. The compressed part will appear to move back and forth bouncing between the two ends of the spring. This is essentially the way sound travels through air particles.

Torsional waves don't apply to anything we're doing in this book, but they're wave in which the particles rotate around the axis along which the wave propagates (like a twisting rod). This type of wave can be seen on a Shive wave machine at physics demonstrations and science and technology museums.

3.1.7 Displacement vs. Velocity

Think back to our original discussions concerning sound. We said that there are really two things moving in a sound wave – the air molecules (which are compressing and expanding) and the pressure wave which propagates outwardly from the sound source. We compared this to a wave moving along a rope. The rope moves up and down, but the wave moves in another direction entirely.

Let's now think of this difference in terms of displacement and velocity – not of the sound wave itself (which is about 344 m/s at room temperature) but of the air molecules.

When a sound wave goes by a bunch of molecules, they compress and expand. In other words, they move closer together, then stop moving, then move further apart, then stop moving, then move closer together and so on. When the displacement is at its absolute maximum, the molecules are at the point where they're stopped and about to head back towards a low pressure. When the displacement is 0 (and therefore at whatever barometric pressure the radio said it was this morning) the molecules are moving as fast as they can. If the displacement is at a maximum in the opposite direction, the molecules are stopped again.

When pressure is 0, the particle velocity is at a maximum (or a minimum) whereas when pressure is at a maximum (or a minimum) the particle velocity is 0.

This is identical to swinging on a playground swing. When you're at the highest point off the ground, you're stopped and about to head in the direction from which you just came. Therefore at the point of maximum

displacement, you have a velocity of 0. When you're at the point closest to the ground (where you started before you were moving) your velocity is highest.

So, in addition to measurements like instantaneous pressure, we can also talk about an *instantaneous particle velocity*, u . In addition, a sinusoidal oscillation results in a *peak particle velocity*, U .

Always remember that the particle velocity is dependent on the *change* in displacement, therefore it is equivalent to the instantaneous slope (or the partial derivative) of the displacement function. As a result, the velocity wave precedes the displacement wave by $\frac{\pi}{2}$ radians (90°) as is shown in Figure 3.5.

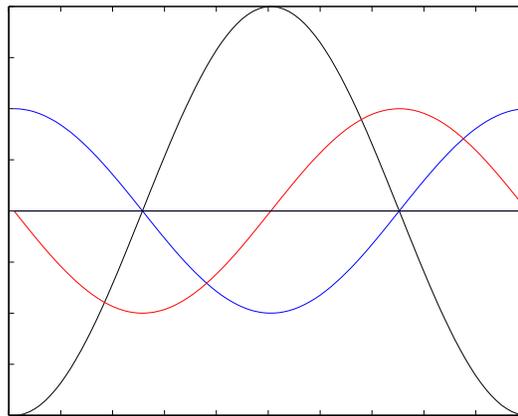


Figure 3.5: The relationship between the displacement (in blue), the velocity (in red) and the acceleration (in black) of a particle or a pendulum. Note that none of these is on any particular scale – the important things to notice are the relationships between the zero, maximum and minimum points on the two graphs as well as their relative instantaneous slopes.

One other important thing to note here is that the velocity is also related to frequency (which is discussed below). If we maintain the same peak pressure, the higher the frequency, the faster the particles have to move back and forth, therefore the higher the peak velocity. So, remember that particle velocity is proportional both to pressure (and therefore displacement) and frequency.

3.1.8 Amplitude

The amplitude of a wave is simply an measurement of the height of the wave if it's transverse, or the amount of compression and refraction if it's

longitudinal. In terms of sound, it's measured in Pascals, since sound waves are variation in atmospheric pressure. If we were measuring waves on the ocean, the unit of measurement would be metres.

There are a number of methods of defining the amplitude measurement – we'll be using three, and you have to be careful not to confuse them.

1. *Peak Pressure* – This is a measurement of the difference between the maximum value of the wave and the point of equilibrium.
2. *Peak to Peak Pressure* – This is a measurement of the difference between the minimum and maximum values of the wave.
3. *Effective Pressure* – This is a measurement based on the amount of power in the wave. It's equivalent to 0.707 of the Peak value if the signal is a sinusoidal wave. In other cases, the relationship between the effective pressure and the Peak value is different (we've already talked about this in Section 2.1.6 – except there, it's called the *RMS* value instead of the effective value).

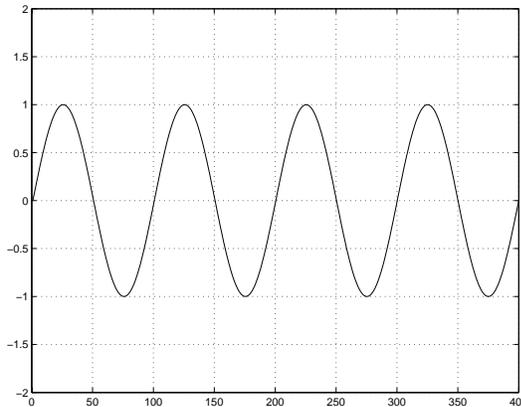


Figure 3.6: A sinusoidal pressure wave with a peak amplitude of 1, a peak-peak amplitude of 2 and an effective pressure of 0.707.

3.1.9 Frequency and Period

Go back to the clarinet example. If we play a concert A, then it just so happens that the reed is opening and closing at a rate of 440 time per

second. This therefore means that there are 440 cycles between a high and a low pressure coming out of the bell of the clarinet each second.

We normally use the term *Hertz* (indicated *Hz*) to indicate the number of cycles per second in sound waves. Therefore 440 cycles per second is more commonly known as a *frequency* of 440 Hz.

In order to find the frequency of a note one *octave* above this pitch, multiply by 2 (1 octave = twice the frequency). One octave below is one-half of the frequency.

In order to find the frequency of a note one *decade* above this pitch, multiply by 10 (1 decade = ten times the frequency). One decade below is one-tenth of the frequency.

Always remember that a complete cycle consists of a high and a low pressure. One cycle is measured from a point on the wave to the next identical point on the wave (i.e. the positive-going zero crossing to the next positive-going zero crossing or maximum to maximum...)

If we know the frequency of a sound wave (i.e. 440 Hz), then we can calculate how long it takes a single cycle to exit the bell of the clarinet. If there are 440 cycles each second, then it takes 1/440th of a second to produce 1 cycle.

The usual equation for calculating this amount of time (known as the *period*) is:

$$T = \frac{1}{f} \quad (3.4)$$

where T is the period and f is the frequency

3.1.10 Angular frequency

For this section, it's important to remember two things.

1. As we saw in Section 1.4, a sound wave is essentially just a “side view” of a rotating wheel. Therefore the higher the frequency, the more revolutions per second the wheel turns.
2. Angles can be measured in radians instead of degrees. Also, that this means that we're measuring the angle in terms of the radius of the circle.

We now know that the frequency of a sinusoidal sound wave is a measure of how many times a second the wave repeats itself. However, if we think of

the wave as a rotating wheel, then this means that the wheel makes a full revolution the same number of times per second.

We also know that one full revolution of the wheel is 360° or 2π radians.

Consequently, if we multiply the frequency of the sound wave by 2π , we get the number of radians the wheel turns each second. This value is called the *angular frequency* or the *radian frequency* and is abbreviated ω .

$$\omega = 2\pi f \quad (3.5)$$

The angular frequency can also be used to determine the phase of the signal at any given moment in time. Let's say for a moment that we have a sine wave with a frequency of 1 Hz, therefore $\omega = 2\pi$. If it's really a sine wave (meaning that it started out heading positive with a value of 0 at time 0 or $t = 0$), then we know that the time in seconds, multiplied by the angular frequency will give us the phase of the sine wave because we rotate 2π radians every second.

This is true for any frequency, so if we know the time t in seconds, then we can find the instantaneous phase using Equation 3.6.

$$\varphi = \omega t \quad (3.6)$$

Usually, you'll just see this notated as ωt as in $\sin(\omega t)$.

3.1.11 Negative Frequency

Back in Section 1.4 we looked at how two wheels rotating at the same speed (or frequency) but in opposite directions will look exactly the same if we look at them from only one angle. This was our big excuse for getting into the whole concept of complex numbers – without both the sine and cosine components, we can only know the speed of rotation (frequency) and diameter (amplitude) of the sine wave. In other words, we'll never know the direction of rotation.

As we walk through the world listening to sinusoidal waves, we only get one signal for each sine wave – we don't get a sine and cosine component, just a pressure wave that changes in time. We can measure the frequency and the amplitude, but not the direction of rotation. In other words, the frequency that we're looking at might be either positive or negative, depending on which direction the imaginary wheel is turning.

Here's another way to think of this. Take a Slinky, stretch it out, and look at it from the side. If you didn't have the benefit of perspective, you wouldn't

be able to tell if the Slinky was coiled clockwise or counterclockwise from left to right. One is positive frequency, the other is the negative equivalent.

In the real world, this doesn't really matter too much, but as we'll see later on, when you're doing things like digital filtering, you need to worry about such things.

3.1.12 Speed of Sound

Pay attention during any thunder and lightning storm and you'll be able to figure out that sound travels slower than light. Since the lightning and the thunder occur simultaneously and since the light flash arrives at you earlier than the clap of thunder (unless you're extremely unlucky...) then this must be true. In fact, the speed of sound, abbreviated c is around 344 m/s although it changes with temperature, pressure and humidity.

Note that we're talking about the speed of the wavefront – not the velocity of the air molecules. This latter velocity is dependent on the waveform, as well as its frequency and the amplitude.

The equation we normally use for c in metres per second is

$$c = 332 + (0.6 * t) \quad (3.7)$$

t is the temperature in °C

There is a small deviation of c with frequency shown in Table 3.1, though this is small and therefore generally ignored

Frequency	Deviation
100 (Hz)	-30 ppm
200 (Hz)	-10 ppm
400 (Hz)	-3 ppm
1.25 (kHz)	0 ppm
4 (kHz)	+5 ppm
10 (kHz)	+10 ppm

Table 3.1: Deviation in the speed of sound with frequency. ??

Changes in humidity change the value of c as is seen in Table 3.2.

The difference at a humidity level of 100% of 0.33% is bordering on our ability to detect a pitch shift.

Also – in case you were wondering, “ppm” stands for “parts per million.” It's just like “percent” really, except that you divide by 1000000 instead of

Humidity	Deviation
0%	0 ppm
20%	+415 ppm
40%	+1136 ppm
60%	+1860 ppm
80%	+ 2590 ppm
100%	+3320 ppm

Table 3.2: Deviation in the speed of sound with air humidity levels. ??

100 so it's useful for really small numbers. Therefore 1000 ppm is $\frac{1000}{1000000} = 0.001 = 0.1\%$.

3.1.13 Wavelength

Let's say that you're standing outside, whistling a perfect 1 kHz sine tone. The moment you start whistling, the first wave – the wavefront – is moving away from you at a speed of 344 m/s. This means that exactly one second after you started whistling, the wavefront is 344 m away from you. At exactly that same moment, you are starting to whistle your 1001st cycle (because you're whistling 1000 cycles per second). If we could stop time and look at the sound wave in the air at that moment, we would see the 1000 cycles that you just whistled sitting in the air taking up 344 m. Therefore you have 1000 cycles for every 344 m. Since we know this, we can calculate the length of one wave by dividing the speed of sound by the frequency – in this case, $344/1000 = 34.4$ cm per wave in the air. This is known as the *wavelength*

The wavelength (abbreviated λ) is the distance from a point on a *periodic* (a fancy word meaning 'repeating') waveform to the next identical point. (i.e. crest to crest, or positive zero-crossing to positive zero crossing)

Equation 3.8 is used to calculate the wavelength, measured in metres.

$$\lambda = \frac{c}{f} \quad (3.8)$$

3.1.14 Acoustic Wavenumber

The wavelength of a sinusoidal acoustic wave is a measure of how many metres long a single wave is. We could think of this relationship between frequency and space in a different way. We can also measure the number of radians our wheel turns in one metre – in other words, the amount of

phase change of the waveform per metre. This value is called the *acoustic wavenumber* of the sound wave and is abbreviated k_0 or sometimes, just k . It's measured in radians per metre and is calculated using Equation 3.9.

$$k_0 = \frac{\omega}{c} \quad (3.9)$$

Note that you will see this under a couple of different names – *wave number*, *wavenumber* and *acoustic wavenumber* will show up in different places to mean the same thing. The problem is that there are a couple of different definitions of the term “wavenumber” so you're best to use the proper term “acoustic wavenumber.”

3.1.15 Wave Addition and Subtraction

Go throw a rock in the water on a really calm lake. The result will be a bunch of high and low water levels that expand out from the point where the rock landed. The highs are slightly above the water level that existed before the rock hit, the lows are lower. This is analogous to the high and low pressures that are coming out of a clarinet, being respectively higher and lower than the equilibrium pressure that existed before the clarinet was brought into the room.

Now go and do the same thing out on the ocean as the waves are rolling past. The ripples that you create will cause the bigger wave to rise and fall on a small scale. This is essentially the same as what was happening on the calm lake, but now, the level of equilibrium is changing.

How do we find the final water level? We simply add the two levels together, making sure to pay attention to whether we should be adding a positive value (higher water level) or negative value (lower water level.)

Let's put two small omnidirectional (that is, they radiate sound equally in all directions) loudspeakers, about 34 cm apart in the front of a room. Let's also take a sine wave generator set to produce a 500 Hz sine wave and send it to both speakers simultaneously. What happens in the room?

Constructive Interference

If you're equidistant from the two speakers, then you'll be receiving the same part of the pressure wave at the same time. So, if you're getting the high point in the wave from one speaker, you're getting a high pressure from the second speaker as well.

Likewise, if you're getting a low pressure from one speaker, you're also receiving a low pressure from the other.

The end result of this overlap is that you get twice the pressure difference between the high and low points in your wave. This is because the two waves are interfering with each other *constructively*. This happens because the two have a phase relationship of 0° at your position.

Essentially all we're doing is adding two simultaneous points from the first two graphs and winding up with the bottom graph.

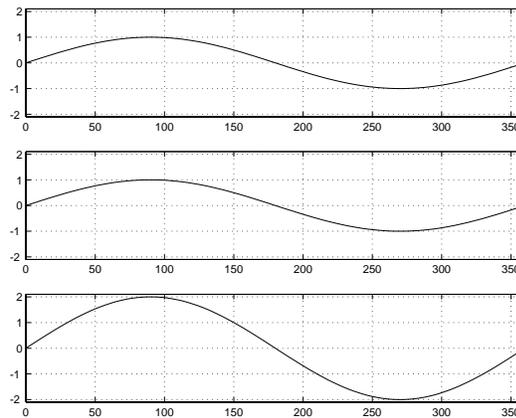


Figure 3.7: The top two plots are the individual signals from two loudspeakers in time measured at a position equidistant to both loudspeakers. The bottom plot is the resulting summed signal.

Destructive Interference

What happens if you're standing on a line with the two loudspeakers, so that the more distant speaker is 34 cm farther away than the closer one.

Now, we have to consider the wavelength of the sound being produced. A 500 Hz sine tone has a wavelength of roughly 68 cm. Therefore, half of a wavelength is 34 cm, or the distance between the two loudspeakers.

This means that the sound from the farther loudspeaker is arriving at your position $1/2$ of a cycle late. In other words, you're getting a high pressure from the closer speaker as you get a low pressure from the farther speaker.

The end result of this effect is that you hear nothing (this is not really true for reasons that we'll talk about later) because the two pressure levels are always opposite each other.

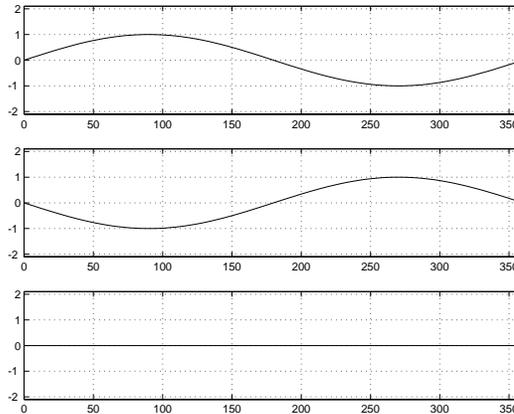


Figure 3.8: The top two plots are the individual signals from two loudspeakers in time measured at a position where one loudspeaker is half a wavelength farther away than the other. The bottom plot is the resulting summed signal.

Beating, Sum and Difference Tones

The discussion of constructive and destructive interference above assumed that the tones coming out of the two loudspeakers have exactly matching frequencies. What happens if this is not the case?

If the two frequencies (let's call them f_1 and f_2 where $f_2 > f_1$) are different then the resulting pressure looks like a periodic wave whose amplitude is being modulated periodically as is shown in Figure 3.9.

The big question is: what does this sound like? The answer to this question is “it depends on how far apart the frequencies are...”

If the frequencies are close together:

First and foremost, you're going to hear the two sine waves of two frequencies, f_1 and f_2 .

Interestingly, you'll also hear beats at a rate equal to the lower frequency subtracted from the higher frequency. For example, if the two tones are at 440 and 444 Hz, you'll hear the two notes beating 4 times per second (or $f_2 - f_1$).

This is the way we tune instruments with each other. If we have two flutes play two A 440's at the same time (with no vibrato), then we should hear no beating. If there's beating, the flutes are out of tune.

If the frequencies are far apart:

First and foremost, you're going to hear the two sine waves of two frequencies f_1 and f_2 .

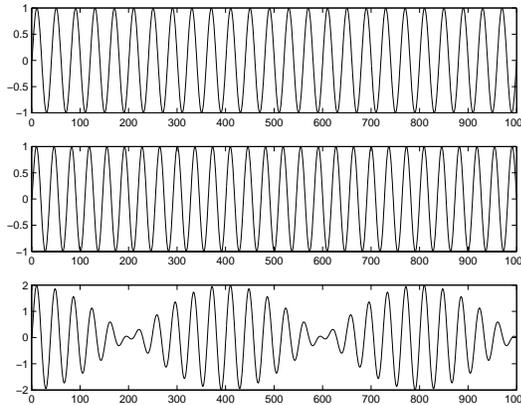


Figure 3.9: The top two plots are sinusoidal waves with slightly different frequencies, f_1 and f_2 . The bottom plot is the sum of the top two. Notice that the modulation in the amplitude of the result is periodic with a “beat frequency” of $f_2 - f_1$

Secondly, you’ll hear a note whose frequency is equal to the difference between the two frequencies being played, $f_2 - f_1$

Thirdly, you’ll hear other tones whose frequencies have the following mathematical relationships with the two tones being played. Although there is some argument between different people, the list below is in order of most to least predominant apparent *difference tones*, *resultant tones* or *combination tones*.

$2f_2 - f_1$, $3f_1 - f_1$, $2f_1 - f_2$, $2f_2 - 2f_1$, $3f_2 - f_1$, $f_2 + f_1$, $2f_2 + f_1$, and so on...

This is a result of a number of effects.

If you’re doing an experiment using two tone generators and a loudspeaker, then the effect is likely a product of the speaker called *intermodulation distortion*. In this case, the combination tones are actually being generated by the driver. We’ll talk about this later.

If you’re using two loudspeakers (or two instruments) then there is some argument as to where the extra tones actually exist. Some arguments say that the tones are in the air, some say that the tones are generated at the eardrum. The most interesting arguments say that the tones are generated in the brain. The proof for this lies in an experiment where different tones are applied to each ear separately (using headphones). In this case, some listeners still hear the combination tones (this is an effect called *binaural beating*).

3.1.16 Time vs. Frequency

We said earlier that the upper harmonics of a periodic waveform are multiples of the first harmonic. Therefore, if I have a complex, but periodic waveform, with a fundamental of 100 Hz, the actual harmonic content is 100 Hz, 200 Hz, 300 Hz, 400 Hz and so on up to ∞ .

Let's assume that the fundamental is lowered to 1 Hz – we're now dealing with an object that is being struck 1 time each second. The fundamental is 1 Hz, so the upper harmonics are 2 Hz, 3 Hz, 4 Hz, 5 Hz and so on up to ∞ .

If we keep slowing down the fundamental to 1 single strike, then the harmonic content is all frequencies up to infinity. Therefore it takes all frequencies sounding simultaneously with the correct phase relationships to create a single click.

If we were to graph this relationship, it would be Figure 3.10, where the two graphs essentially show the same information.

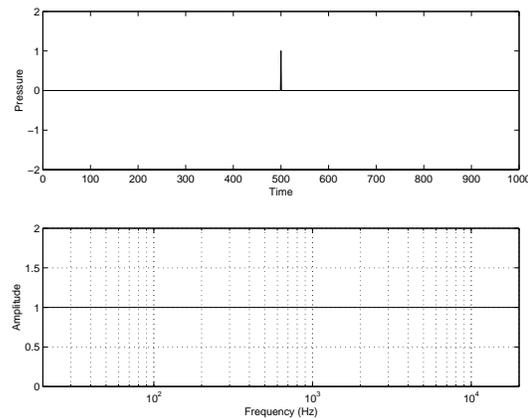


Figure 3.10: Two graphs showing exactly the same information. An infinitely short amplitude spike in the time domain is equivalent of all frequencies being present at that moment in time.

3.1.17 Noise Spectra

The theory explained in Section 3.1.16 that the combination of all frequencies results in a single click relies on an important point that we didn't talk about – relative phase. The click can only happen if all of the phases of the harmonics are aligned properly – if not, then things tend to go awry... If we

have all frequencies with random relative amplitude and phase, the result is noise in its various incarnations.

There is an official document defining four types of noise. The specifications for white, pink, blue and black noise are all found in The Federal Standard 1037C Telecommunications: Glossary of Telecommunication Terms. (I got the definitions from Rane's online dictionary of audio terms at <http://www.rane.com>.)

White Noise

White noise is defined as a noise that has equal amount of energy per frequency. This means that if you could measure the amount of energy between 100 Hz and 200 Hz it would equal the amount of energy between 1000 Hz and 1100 Hz. Because all frequencies have equal level, we call the noise white – just like light that contains all frequencies (colours) equally is white light.

This sounds “bright” to us because we hear pitch in octaves. 1 octave is a doubling of frequency, therefore 100 Hz – 200 Hz is an octave, but 1000 Hz – 2000 Hz (not 1000 Hz – 1100 Hz) is also an octave. Since white noise contains equal energy per Hz, there's ten times as much energy in the 1 kHz octave than in the 100 Hz octave.

Pink Noise

Pink noise is noise that has an equal amount of energy per octave. This means that there is less energy per Hz as you go up in frequency (in fact, there is a power loss of 50% (or a drop of 3.01 dB) each time you go up an octave)

This is used because it sounds relatively “equal” in distribution across frequency bands to us.

Another way of defining this noise is that the power of each frequency f is proportional to $\frac{1}{f}$.

Blue Noise

Blue noise is noise that is the opposite of pink noise in that it doubles the amount of power each time you go up 1 octave. You'll virtually never see it (or hear it for that matter...).

Another way of defining this noise is that the power of each frequency f is proportional to the frequency.

Red Noise (aka Brown Noise or Popcorn Noise)

Red Noise is used when pink noise isn't low-end-heavy enough for you. For example, in cases where you want to use noise to simulate road noise in the interior of a car, then you want a lot of low-frequency information. You'll also see it used in oceanography. In the case of red noise, there is a 6.02 dB drop in power for every increase in frequency of 1 octave. (In other words, the power is proportional to $\frac{1}{f^2}$)

Purple Noise

Purple Noise is to blue noise as red noise is to pink. It increases in power by 6.02 dB for every increase in frequency of 1 octave. (In other words, the power is proportional to f^2 .)

Black Noise

This is an odd case. It is essentially silence with the occasional randomly-spaced spike.

3.1.18 Amplitude vs. Distance

There is an obvious relationship between amplitude and distance – they are inversely proportional. That is to say, the farther away you get, the lower the amplitude. Why?

Let's go back to throwing rocks into a lake. You throw in the rock and it produces a wave in the water. This wave can be considered as a manifestation of an energy transfer from the rock to the water. All of the energy is given to the wave from the rock at the moment of impact – after that, the wave maintains (theoretically) that energy.

The important thing to notice, though, is that the wave expands as it travels out into the lake. Its circumference gets bigger as it travels horizontally (as its radius gets bigger...) Therefore the wave is "longer" (if you're measuring around the circumference). The total amount of energy in the wave, however, has not changed (actually it has gotten a little smaller due to friction, but we're ignoring that effect...) therefore the same amount of energy has to be shared across a longer wavefront. This causes the height (and depth) of the wave to shrink as it expands.

What's the mathematical relationship between the increasing circumference and the increasing radius? Well, the radius is travelling at the constant speed, determined by the density of the water and gravity and other things

like the colour of your left shoe... We know from high school that the circumference is equal to the radius multiplied by about 6.28 (also known as 2π). The graph in Figure 3.11 shows the relationship between the radius and the circumference. You can see that the latter grows much more quickly than the former. What this means is that as the radius slowly expands out from the point of impact, the energy is getting shared between a “length” of the wave that is growing far faster (note that, if we double the radius, we double the circumference).

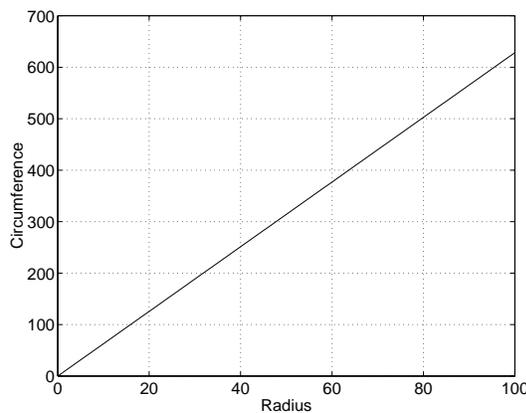


Figure 3.11: The relationship between the circumference of a circle and its radius.

The same holds true with pressure waves expanding from a loudspeaker into a room. The only real difference is that the energy is expanding into 3 dimensions rather than 2, so the surface area of the spherical wavefront (the 3-D version of the circumference of the circular wave on the lake...) increases much more rapidly than the 2-dimensional counterpart. The equation used to find the surface of a sphere is $4\pi R^2$ where R is the radius. As you can see in Figure 3.12, the surface area of the sphere is already at 1200 units squared when the radius has only expanded to 10 units. The result of this in real life is that the energy appears to be dissipating at a rate of 6.02 dB per doubling of distance. (when we double the radius, we increase the surface area of the sphere fourfold.) Of course, all of this assumes that the wavefront doesn't hit anything like a wall or the floor or you...

3.1.19 Free Field

Imagine that you're suspended in infinite space with no walls, ceiling or floor anywhere in sight. If you make a noise, the wavefront of the sound is free

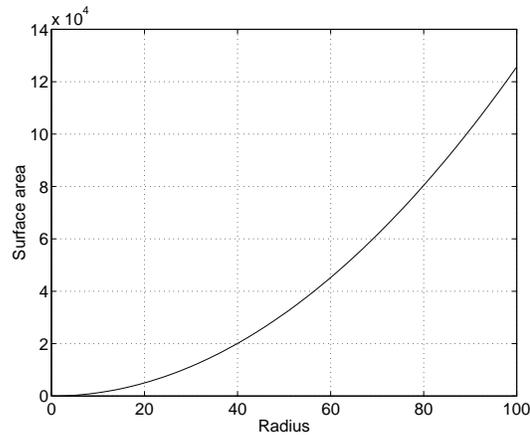


Figure 3.12: The relationship between the surface area of a sphere and its radius.

to move away from you forever, without ever encountering any surface. No reflections or diffraction at all – forever. This space is a theoretical idea known as a *free field* because the wavefront is free expand.

If you put a microphone in this free field, the wavefront from a single sound source would come from a single direction. This seems obvious, but I only mention it to compare with the next section.

For a visual analogy of what we’re talking about, imagine that you’re floating in space and the only thing you can see is a single star. There are at least three things that you’d notice about this odd situation. Firstly, the star doesn’t appear to be very bright, because most of its energy is going in a different direction than towards you. Secondly, you’d notice that everything but the star is very, very dark. Finally, you’d notice that shadows are very distinct and also very, very dark.

3.1.20 Diffuse Field

Now imagine that you’re in a space that is the most reverberant room you’ve ever been in. You clap your hands and the reverb goes on until sometime next Tuesday. (If you’d like to hear what such as space sounds like, run out and buy a copy of the recording of Stuart Dempster and his crowd of trombonists playing in the Sistrin Chapel in Seattle.[[Dempster, 1995](#)]) Anyways, if you were able to keep a record of every reflection in the reverb tail, keeping track of the direction it came from, you’d find that they come from everywhere. They don’t come from everywhere simultaneously – but

if you wait long enough, you'll get a wavefront from every possible direction at some time.

If we consider this in terms of probability, then we can say that, in this theoretical space, sound waves have an equal probability of coming from any direction at any given moment. This is essentially the definition of a *diffuse field*.

For a visual example of this, look out the window of a plane as you're flying through a cloud on a really sunny day. The light from the sun bounces off of all the little particles in the cloud, so, from your perspective, it essentially comes from everywhere. This causes a couple of weird sensations. Firstly, there are no shadows – this is because the light is coming from everywhere so nothing can shadow anything else. Secondly, you have a very difficult time determining distance. Unless you can see the wing of the plane, you have no idea how far away you're actually able to see. This the same reason why people have car accidents in blinding snowstorms. They drive because they think they can see ahead much further than they're really able to.

3.1.21 Acoustic Impedance

Before we look at how rooms behave when you make noise in them, we have to begin by looking at the concept of *acoustic impedance*. Earlier, we saw how sound is transmitted through air by moving molecules bumping up against each other. One air molecule moves and therefore moves the air molecules sitting next to it. In other words, we're talking about energy being transferred from one molecule to another. The ease with which this energy is transferred between molecules is measured by the difference in the acoustic impedances of the two molecules. I'll explain.

We've already seen that sound is essentially a change in pressure over time. If we have a static barometric pressure and then we apply a new pressure to the air molecules, then we change their displacement (we move them) and create a molecular velocity. So far, we've looked at the relationship between the displacement and the velocity of the air molecules, but we haven't looked at how both of these relate to the pressure applied to get the whole thing moving. In the case of a pendulum (a weight hanging on the end of a stick that's free to swing back and forth), the greater the force applied to it, the more it moves and the faster it will go – the higher the pressure, the greater the displacement and the higher the velocity. The same is true of the air molecules – the higher the pressure, the greater the displacement and the higher the velocity. However, the one thing we're ignoring is how hard it is to get the pendulum (or the molecules) moving. If we apply the same force

to two different pendulums, one light one and one heavy one, then we'll get two different maximum displacements and velocities as a result. Essentially, the heavier pendulum is harder to move, so we don't move it as far.

The issue that we're now discussing is how much the pendulum *impedes* your attempts to move it. The same is true of molecules moved by a sound wave. Air molecules are like a light pendulum – they're relatively easy to move. On the other hand, if we were to put a loudspeaker in poured concrete and play a tune, it would be much harder for the speaker to move the concrete molecules – therefore they wouldn't move as far with the same pressure applied by the loudspeaker. There would still be a sound wave going through the concrete (just as the heavy pendulum would move – just not very much) but it wouldn't be very loud.

The measurement of how much velocity results from a given amount of pressure is an indication of how hard it is to move the molecules – in other words, how much the molecules impede the transfer of energy. The higher the impedance, the lower the velocity for a given amount of pressure. This can be seen in Equation 3.10 which is true only for the free field situation.

$$z = \frac{p}{u} \quad (3.10)$$

where z is the acoustic impedance in acoustic ohms (abbreviated Ω).

As you can see in this equation, z is proportional to p and inversely proportional to u . This means that if the impedance goes up and the pressure stays the same, then the velocity will go down.

In the specific case of unbounded *plane waves* (waves with a flat wavefront – not curved like the ones we've been discussing so far), this ratio is also equal to the product of the volume density of the medium, ρ_o and the speed of wave propagation c as is shown in Equation 3.11 [Olson, 1957]. This value z_o is known as the *specific acoustic impedance* or *characteristic impedance* of the medium.

$$z_o = \rho_o c \quad (3.11)$$

Acoustic Resistance and Reactance

Now, here's where things get a little ugly... And I apologize in advance for the mess I'm about to make of all this.

Acoustic impedance is actually the combination of two constituent components called *acoustic resistance* and *acoustic reactance* (just like electrical impedance is the combination of resistance and reactance). Let's think back

to the situation where you have a sound wave going through one medium (say, air for example...) into another medium (like concrete). As we've already seen, the difference in the acoustic impedances of the two media will determine how much of the sound wave gets reflected back into the air and how much will be transmitted into the concrete. What we haven't looked at yet is whether or not there will be a phase change in the reflected and the transmitted pressure waves. It's important here to point out that I'm not talking about a polarity change – I'm talking about a phase change.

Let's just consider the transmitted pressure wave for a while. If there is no change in the phase of the wave as a result of it hitting the second medium, then we can say that the second medium has only an acoustic resistance. If the acoustic impedance has no reactance component (remember it only has two components) then there will be no phase shift between the incident pressure wave and the transmitted pressure wave.

On the other hand, let's say that there is suddenly a 90° delay in the transmitted wave when compared to the incident wave (yes, this can happen – particularly in cases where the second medium is flexible and bends when the sound wave hits it). In this particular case, then the second medium has only an acoustic reactance.

So, an acoustic resistance causes a 0° phase shift and an acoustic reactance causes a 90° phase shift. By now, this should be reminding you of Section 1.5.13. Remember from this chapter that a sinusoidal wave with any magnitude and phase shift can be expressed using a real (0°) and an imaginary (90°) component? The magnitude of the waveform depends on the values of the real and the imaginary components, and the phase shift is determined by the relative values of the two components.

This same system exists for impedance. In fact, you will often see books saying that impedance is a complex value containing both a real and an imaginary component. In other words, there will be a phase shift between the incident wave and the transmitted and reflected waves. This phase shift can be calculated if you know the relative balance of the real component (the acoustic resistance) and the imaginary component (the acoustic reactance) of the acoustic impedance.

If the concepts of acoustic impedance, resistance and reactance are a little confusing, don't worry for now. Go back and read Sections 2.1 and 2.4. If it still doesn't make sense after that, then you can worry. To quote Telly Monster from Sesame Street, "Don't worry, be happy. Or worry, and then be happy."

3.1.22 Power

So far, we've looked at a number of different ways to measure the level of a sound. We've seen the pressure, the particle displacement and velocity and some associated measurements like the SPL. These are all good ways to get an idea of how loud a sound is at a specific point in space, but they are all limited to that one point. All of these measurements tell you how loud a sound is at the point of the receiver, but they don't tell you much about how loud the sound source itself is. If this doesn't sound logical, think about the light radiated from a light bulb – if you measure that light from a distance, you can only tell how bright the light is where you measure it, you can't tell the wattage of the bulb (a measure of how powerful it is).

We've already seen that the particle velocity is proportional to the pressure applied to the particles by the sound source. The higher the pressure, the greater the velocity. However, we've also seen that, the greater the acoustic impedance, the lower the particle velocity for the same amount of pressure. This means that, if we have a medium with a higher acoustic impedance, we'll have to apply more pressure to get the same particle velocity as we would with a lower impedance. Think about the pendulums again – if we have a heavy and a light one and we want them to have the same velocity, we'll have to push harder on the heavy one to get it to move as fast as the light one. In other words, we'll have to do more *work* to get the heavy one to move as fast.

Scientists typically don't like work – otherwise they would have gotten a job in construction instead... As a result, they even use a different word to express work – specifically, they talk about how much work can be done using *power*. The more power you have in a device, the more work it can do. This can be seen from day to day in the way light bulbs are rated. The amount of work that they do (how much light and heat they give off) is expressed in how much power they use when they're turned on. This electrical power rating (expressed in *Watts*) will be discussed in Section [2.1.4](#).

In the case of acoustics, the amount of *work* that is done by the sound source is proportional to the pressure and the particle velocity – the more pressure and/or the more velocity, the more work you had to do to achieve it. Therefore the *acoustic power* measured at a specific point in space can be calculated using Equation ???. Remember that the change in pressure is a result of the work that is done – you put power into the system and you get a change in power as an output.

PUT ACOUSTIC POWER EQUATION HERE

This equation is moderately useful in that it tells us how much work is being done to move out measurement device (a microphone diaphragm), but it still has a couple of problems. Firstly, it is still a measurement of a single point in space, so we can only see the power received, not the total power radiated by the sound source. Another problem with power measurements is that they can't give you a negative value. This is because a positive pressure produces a positive velocity and when the two are multiplied we get a positive power. A negative pressure multiplied by a negative velocity also equals a positive power. This really makes intuitive sense since it's impossible to have a negative amount of work, which is why we need power and pressure measurements in many cases – we need the latter to find out what's going on on the negative side of the stasis pressure.

3.1.23 Intensity

In theory, we can think of the sound power at a single point in space as we did in the previous section. In reality, we cannot measure this, because we don't have any microphones to measure a point that small. Microphones for measuring acoustic fields are pretty small, with diameters on the order of millimeters, but they're not infinitely small. As a result, if we oversimplify a little bit for now, the microphone is giving us an output which is essentially the sum of all of the pressures applied to its diaphragm. If the face of the diaphragm is perpendicular to the direction of travel of the wavefront, then we can say that the microphone is giving us an indication of the *intensity* of the sound wave. Huh?

Well, the intensity of a sound wave is the measure of all of the sound power distributed over a given area normal (perpendicular) to the direction of propagation. For example, let's think about a sound wave as a sphere expanding outwards from the sound source. When the sphere is at the sound source, it has the same amount of power as was radiated by the source, all packed into a small surface area. If we ignore any absorption in the air, as the sphere expands (because the wavefront moves away from the sound source in all directions) the same power is contained in the bigger surface area. Although the sphere gets bigger over time, the total power contained in it never changes.

If we did the same thought experiment, but only considered an angular slice of the sphere - say 45° by 45° , then the same rule would hold true. As the sphere expands, the amount of power contained in the 45° by 45° slice would remain the same, even though its total surface area would increase.

Now, let's think of it a different way. Instead of thinking of the whole

sphere, or an angular slice of it, let's think about a fixed surface area such as 1 cm^2 on the sphere. As the wavefront moves away from the sound source and the sphere expands, the fixed surface area becomes a smaller and smaller component of the total surface area of the sphere. Since the total power distributed over the sphere doesn't change, then the amount of power contained in our little 1 cm^2 gets less and less, proportional to the ratio of the area to the total surface area of the sphere.

If the surface area that we're talking about is part of the sphere expanding from the sound source (in other words, if it's perpendicular to the direction of propagation of the wavefront) then the total sum of power in the area is what is called the sound intensity.

This is why sound appears to get quieter as we move further away from the sound source. Since your eardrum doesn't change in surface area, as you get further from a sound source, it has less intensity – there is less total sound power on the surface of your eardrum because your eardrum is smaller compared to the size of the sphere radiating from the source.

3.2 Acoustic Reflection and Absorption

Imagine that you're an air molecule – and that all of your friends are as well. You all stand in single file front to back and each of you puts your hands on the shoulders of the person in front of you. You're at the back of the line. If you push the person in front of you forwards, then he'll fall forwards, pushing the person in front of him who falls forward and pushes the person in front of her and so on and so on to the front of the line. Now, imagine that the person at the front of the line has his hands on a concrete wall. When the person behind him pushes him forward, he pushes the wall and winds up pushing back on the person who pushed him. That person now falls *backwards*, pushing the person behind who does the same and so on and so on. In the long run, you push someone forward, this causes a chain reaction that ends in you getting pushed.

Imagine the same scenario, but this time, the person at the front of the line is *not* in front of a concrete wall. You push the guy in front of you... the chain reaction happens and eventually, the person in the front of the line falls forward on his face. This now *pulls* the person behind him who falls forward and pulls the person behind her and so on and so on. As you can see, the eventual effect of the initial cause is now different. You push someone forward, a chain reaction happens that ends in you getting pulled.

Remembering that you and your friends are all air molecules, the analogy extends further. The difficulty you have pushing the person in front of you is the acoustic impedance. If you are as easy to push as the person in front, then the difference in impedance between you is 0, therefore there is no reflection. The same is true in air molecules – two air molecules side by side have the same impedance, so there is no reflection from the second one.

The person at the front of the line has a very difficult time pushing the concrete wall just like an air molecule can't push a concrete wall because the acoustic impedance of concrete is higher than that of air. Therefore, if you send a high pressure wave at a reflecting surface with a higher acoustic impedance than the transmission medium, you get a high pressure wave back. You push, and eventually, you get pushed.

When the person at the front of the line has nothing to push, it's the same as an air molecule on a reflecting surface that has a lower acoustic impedance than the transmission medium. If you send a high pressure wave at that reflecting surface, you will get a low pressure wave back. You push, and eventually you get pulled. Of course, the only question that is probably remaining is "how can you have a *lower* acoustic impedance than air?" We'll get to that in a later section.

3.2.1 Impedance Boundaries

We normally think of sound waves as bouncing off of walls in the same way that balls bounce off the cushions on a pool table. This works for the most simple of cases, where you have only one wall and it's made of perfectly flat, polished concrete or marble, but in all other situations, it's best to know a little more about what's going on.

In order to look at this, we have to look at the pressure wave in a slightly different way. So far we've been saying that a compression is a positive (or high) pressure and a rarefaction is a negative (or low) pressure. What we haven't been considering is the direction of the movement of the molecules. For example, in the previous section, we looked at the analogy of a bunch of your friends pushing each other around. One of the little things we skipped over was the details around when you're pushed and when you're pulled... Hopefully, you didn't notice this at the time, but now it's time to deal with the confusion.

Very early in this chapter, we looked at the displacement of the air molecules from their stasis position. Since the molecules move back and forth, the displacement can be a positive or negative number.

We've also looked at the instantaneous excess pressure, p and the peak maximum pressure, P . These two measurements are indications of the distance between the air molecules.

One of the things to note here is that the polarities of the pressure and the displacement are independent. For example, you can have a negative displacement with a positive or a negative pressure. In other words, the molecules can move towards you and either bunch together or pull apart. Using the analogy above, a high pressure means that you are closer to your adjacent friend, regardless of the direction that you moved to get closer...

So, let's create a

EXPLAIN THE CONCEPT OF \mathbf{p} INSTEAD OF \mathbf{P} .

$$\mathbf{p} = \mathbf{P}e^{i(\omega t + kx)} \quad (3.12)$$

MORE EXPLANATION

$$\mathbf{p} = \mathbf{P}(\cos(\omega t + kx) + i\sin(\omega t + kx)) \quad (3.13)$$

MORE EXPLANATION HERE

A sound wave propagates through a medium such as air until it meets a change in acoustic impedance. This change in impedance could be the result of a new medium like a concrete wall, or it might be a less obvious

reason such as a change in air temperature or a change in room shape. We'll look at the more obscure instances a little later. When the wavefront of a propagating sound encounters a change in the propagation medium and the direction of wave propagation is perpendicular to the boundary of the two media, the pressure in the wave is divided between two resulting wavefronts: the *transmitted sound*, which passes into the new medium, and the *reflected sound*, which is transmitted back into its medium of origin as is shown in Figure ?? and expressed in Equation 3.14 ??.

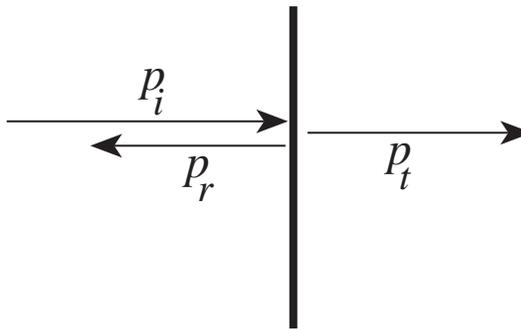


Figure 3.13: The relationship between the incident, transmitted and reflected pressure waves assuming that all rays are perpendicular to the boundary.

$$\mathbf{p}_t = \mathbf{p}_i + \mathbf{p}_r \quad (3.14)$$

where \mathbf{p}_i is the incident pressure in the first medium, \mathbf{p}_r is the reflected pressure in the first medium and \mathbf{p}_t is the pressure transmitted into the second medium, *all measured at the boundary of the two media*.

This equation should look a little weird at first – intuition says that the energy in the incident pressure should be equal to the sum of the reflected and transmitted pressures, and you'd be right if you thought this. Notice, however, that Equation 3.14 uses small \mathbf{p} 's instead of capitals.

FINISH THIS OFF!!

Similar to Equation 3.14, the difference between the incident and reflected particle velocities equals the transmitted particle velocity as is shown in Equation 3.15 [Kinsler and Frey, 1982].

$$\mathbf{u}_t = \mathbf{u}_i - \mathbf{u}_r \quad (3.15)$$

As a result, we can combine Equations 3.10, 3.14 and 3.15 to produce Equation 3.16 [Kinsler and Frey, 1982].

$$\mathbf{z}_t = \frac{\mathbf{P}_i + \mathbf{P}_r}{\mathbf{u}_i - \mathbf{u}_r} = \frac{\mathbf{P}_t}{\mathbf{u}_t} \quad (3.16)$$

where \mathbf{z}_t is the acoustic impedance of the second medium.

3.2.2 Reflection and Transmission Coefficients

The ratios of reflected and transmitted pressures to the incident pressure are frequently expressed as the *pressure reflection coefficient*, \mathbf{R} , and *pressure transmission coefficient*, \mathbf{T} , shown in Equations 3.17 and 3.18 [Kinsler and Frey, 1982].

$$\mathbf{R} = \frac{\mathbf{P}_r}{\mathbf{P}_i} \quad (3.17)$$

$$\mathbf{T} = \frac{\mathbf{P}_t}{\mathbf{P}_i} \quad (3.18)$$

What use are these? Well, let's say that you have a sound wave hitting a wall with a reflection coefficient of $\mathbf{R} = 1$. This then means that $\mathbf{P}_r = \mathbf{P}_i$, which is a mathematical way of saying that all of the sound will bounce back off the wall. Also because of Equation 3.14, this also means that none of the sound will be transmitted into the wall (because $\mathbf{P}_r = 0$ and therefore $\mathbf{T} = 0$), so you don't get angry neighbours. On the other hand, if $\mathbf{R} = 0.5$, then $\mathbf{P}_r = \frac{\mathbf{P}_i}{2}$ which in turn means that $\mathbf{P}_t = \frac{\mathbf{P}_i}{2}$ (and therefore that $\mathbf{T} = 0.5$) and you might be sending some sound next door... although we would have to do a little more math to really decide whether that was indeed the case.

Note that the pressure reflection coefficient can either be a positive number or a negative number. If \mathbf{R} is a positive number then the pressure of the reflection will have the same polarity as the incident wave, however, if \mathbf{R} is negative, then the pressures of the incident and reflected waves will have opposite polarities. THINK BACK TO THE EXAMPLE WITH YOUR FRIENDS...

3.2.3 Absorption Coefficient

So far we have assumed that the media we're talking about that are being used to transmit the sound, whether air or something else, are *adiabatic*. This is a fancy word that means that they don't convert any of the sound power into heat. This, unfortunately is not the case for any substance. All sound transmission media will cause some of the energy in the sound wave to be converted into heat, and therefore it will appear that the substance

has *absorbed* some of the sound. The amount of absorption depends on the characteristics of the material, but a good rule of thumb is that when the material is made of a lot of changes in density, then you're going to get more absorption than in a material with a constant density. For example, air has pretty much the same density all over a room, therefore you don't get much sound energy absorbed by air. Fibreglas insulation, on the other hand is made up of millions of bits of glass and pockets of air, resulting in a large number of big changes in acoustic impedance through the material. The result is that the insulation converts most of the sound energy into heat. A good illustration of this is a rumour that I once heard about an experiment that was done in Sweden some years ago. Apparently, someone tried to measure the sound pressure level of a jet engine in a large anechoic chamber which is a room that is covered in absorbent material so that there are no reflections off any of the walls. Of course, since the walls are absorbent, then this means that they convert sound into heat. The sound of the jet engine was so loud that it caused the absorptive panels to melt! Remember that this was not because the engine made heat, but that it made noise.

Usually, a material's *absorption coefficient*, α is found by measuring the amount of energy reflected off it.

CHECK THIS AND FINISH IT OFF

$$\alpha = \frac{\text{absorbed energy}}{\text{incident energy}} \quad (3.19)$$

Air Absorption

In practice, for lower frequencies, no energy will be lost in the propagation through air. However, for shorter wavelengths, there is an increasing attenuation due to viscothermal losses (meaning losses due to energy being converted into heat) in the medium. These losses in air are on the order of 0.1 dB per metre at 10 kHz as is shown in Figure 3.14.

Usually, we can ignore this effect, since we're usually pretty close to sound sources. The only times that you might want to consider this is when the sound has travelled a very long distance which, in practice, means either a sound source that's really far away outdoors, or a reflection that has bounced around a big room a lot before it finally gets to the listening position.

3.2.4 Comb Filtering Caused by a Specular Reflection

NOT YET WRITTEN

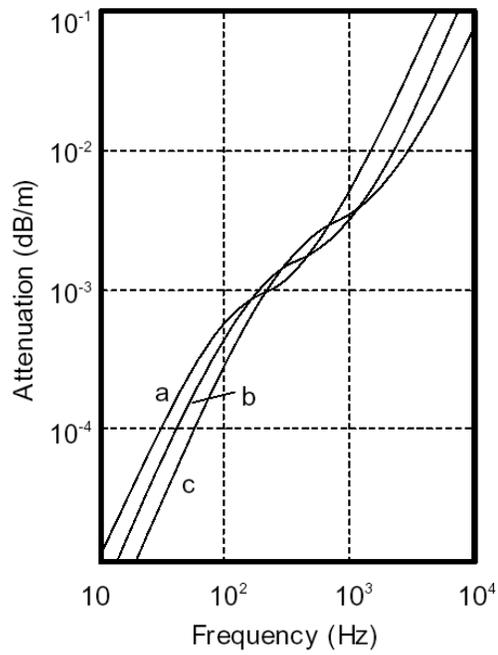


Figure 3.14: Attenuation resulting from air absorption due to propagation distance vs. frequency for various relative humidity levels (a) 20 %; (b) 40 %; (c) 80 % [Kuttruff, 1991]

3.2.5 Huygen's Wave Theory

NOT YET WRITTEN

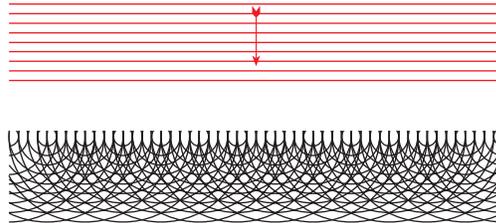


Figure 3.15: The top diagram is a simplified representation of a plane wave travelling from top to bottom of the page. The bottom diagram shows a large number of closely-spaced point sources emitting the same frequency. Notice that the interference pattern of the multiple sources looks very similar to the plane wave. In fact, the more sources you have, and the closer together they are, the closer the more the two results will be.

3.3 Specular and Diffused Reflections

For this section, you will have to imagine a very strange thing – a single reflecting surface in a free field. Imagine that you’re floating in space (except that space is full of air so that you can breathe and sound can be transmitted) next to a wall that extends out to infinity in all directions. It may be easier to think of you standing outdoors on a concrete floor that goes out to infinity in all directions. Just remember – we’re not talking about a wall inside a room yet... just the wall itself. Room acoustics comes later.

3.3.1 Specular Reflections

The discussion in Section 3.2.1 assumes that the wave propagation is *normal*, or perpendicular, to the surface boundary. In most instances, however, the angle of incidence – an angle subtended by a normal to the boundary (a line perpendicular to the surface and intersecting the point of reflection) and the incident sound ray – is an oblique angle. If the reflective surface is large and flat relative to the wavelength of the reflected sound, there exists a simple relationship between the angle of incidence and the angle of reflection, subtended by the reflected ray of sound and the normal to the reflective surface. *Snells law* describes this relationship as is shown in Figure 3.16 and Equation 3.20 [Isaacs, 1990].

$$\sin(\vartheta_i) = \sin(\vartheta_r) \quad (3.20)$$

and therefore, in most cases:

$$\vartheta_i = \vartheta_r \quad (3.21)$$

This is exactly the same as the light that bounces off a mirror. The light hits the mirror and then is reflected off at an angle that is equal to the angle of incidence. As a result, the reflections looks like a light bulb that appears to be behind the mirror. There is one interesting thing to note here – the point on the mirror where the light is reflected is dependent on the locations of the light, the mirror and the viewer. If the viewer moves, then the location of the reflection does as well. If you don’t believe me, go get a light and a mirror and see for yourself.

Since this type of reflection is most commonly investigated as it applies to visual media and thus reflected light, it is usually considered only in the spatial domain as is shown in the above diagram. The study of specular

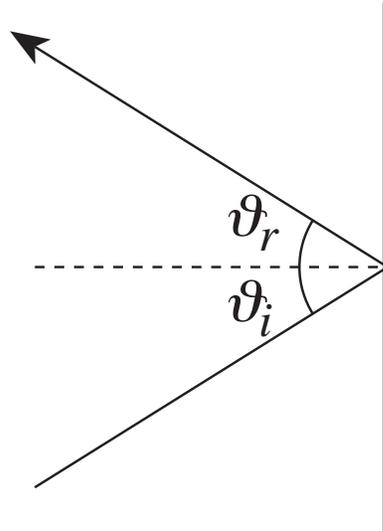


Figure 3.16: Relationship between the angles of incidence and reflection in the case of a specular reflector.

reflections in acoustic environments also requires that we consider the response in the time domain as well. This is not an issue in visual media since the speed of light is effectively infinite in human perception. If the surface is a perfect specular reflector with an infinite impedance, then the reflected pressure wave is an exact copy of the incident pressure wave. As a result, its impulse response is equivalent to a simple delay with an attenuation determined by the propagation distance of the reflection as is shown in Figure 3.17.

3.3.2 Diffused Reflections

If the surface is irregular, then Snells Law as stated above does not apply. Instead of acting as a perfect mirror, be it for light or sound, the surface scatters the incident pressure in multiple directions. If we use the example of a light bulb placed close to a white painted wall, the brightest point on the reflecting surface is independent of the location of the viewer. This is substantially different from the case of a specular reflector such as a polished mirror in which the brightest point, the location of the reflection of the light bulb, would move along the mirrors surface with movements of the viewer. *Lamberts Law* describes this relationship and states that, in a perfectly

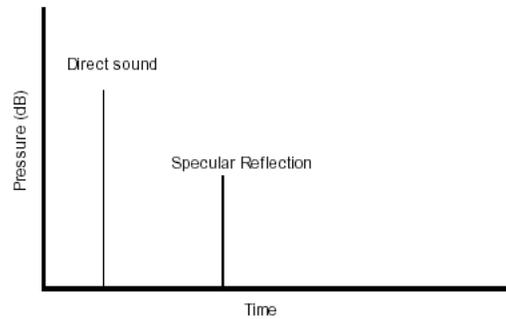


Figure 3.17: Impulse response of direct sound and specular reflection. Note that the Time is referenced to the moment when the impulse is emitted by the sound source, hence the delay in the time of arrival of the initial direct sound.

diffusing reflector, the intensity is proportional to the cosine of the angle of incidence as is shown in Figure 3.18 and Equation 3.22 [Isaacs, 1990].

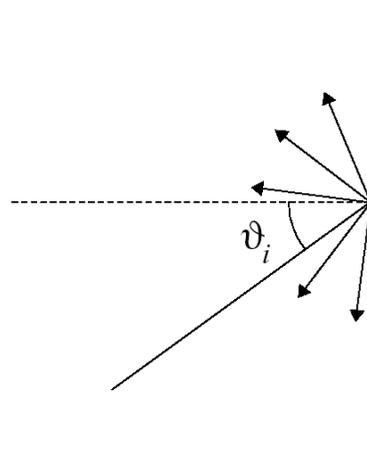


Figure 3.18: Relationship between the angles of incidence and reflection in the case of a diffusive reflector.

$$I_r \propto I_i \cos(\vartheta_i) \quad (3.22)$$

where I_r and I_i are the intensities of the reflected and incident sound waves respectively.

Note that, in this case of a perfectly diffusing reflector, the “bright point”

is the point on the surface of the reflector where the wavefront hits perpendicular to the surface. This means that it is irrelevant where the viewer is located. This can be seen in Figure 3.19 which shows a reflecting surface that is partly diffuse and specular. In this case, if the viewer were to move, the smaller specular reflection would move, but the diffuse reflection would not.



Figure 3.19: Photograph of a wall surface which exhibits both specular and diffusive reflective properties. Note that there are two bright spots. The small area on the right is the specular reflection, the larger area in the centre is the diffuse component.

There are a number of physical characteristics of diffused reflections that differ substantially from their specular counterparts. This is due to the fact that, whereas the received reflection from a specular reflector originates from a single point on the surface, the reflection from a diffusive reflector is distributed over a larger area as is shown in Figure 3.20.

Dalenbäck [Svensson, 1994] lists the results of this distribution in the spatial, temporal and frequency domains as the following:

1. Non-specular regions are covered
2. temporal smearing and amplitude smoothing
3. reception angle smear
4. directivity smear
5. frequency content in reflection is affected
6. creation of a more uniform reverberant field

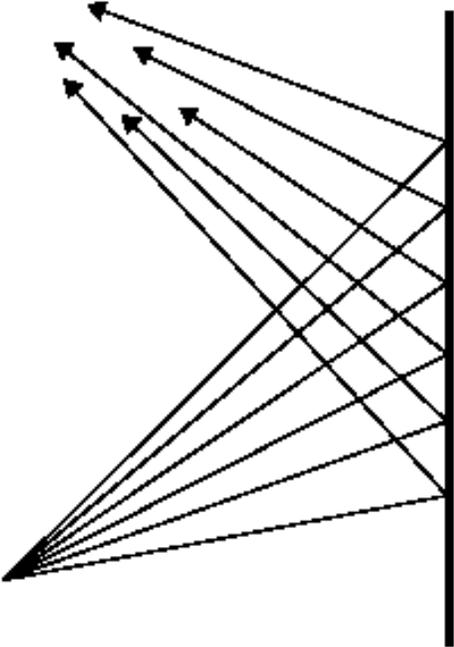


Figure 3.20: Diffused reflection showing spatial distribution of the reflected power for a single receiver.

The first issue will be discussed below. The second, third and fourth points are the product of the fact that the received reflection is distributed over the surface of the reflector. This results in multiple propagation distances for a single reflection as well as multiple angles and reflection locations. Since the reflection is distributed over both space and time at the listening position, there is an effect on the frequency content. Whereas, in the case of a perfect specular reflector, the frequency components of the resulting reflection form an identical copy of the original sound source, a diffusive reflector will modify those frequency characteristics according to the particular geometry of the surface. Finally, since the reflections are more widely distributed over the surfaces of the enclosure, the reverberant field approaches a perfectly diffuse field more rapidly.

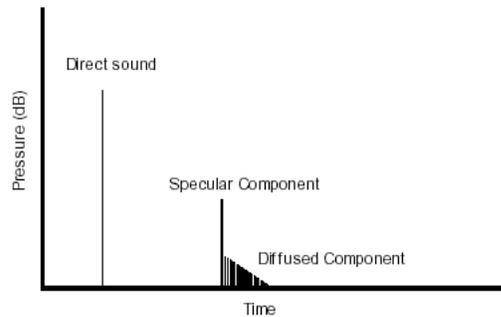


Figure 3.21: Impulse response of direct sound as well as the specular and simplified diffused reflection components. Note that the Time is referenced to the moment when the impulse is emitted by the sound source, hence the delay in the time of arrival of the initial direct sound.

Surface Types

The relative balance of the specular and diffused components of a reflection off a given surface are determined by the characteristics of that surface on a physical scale on the order of the wavelength of the acoustic signal. Although a specular reflection is the result of a wave reflecting off a flat, non-absorptive material, a non-specular reflection can be caused by a number of surface characteristics such as irregularities in the shape or absorption coefficient (and therefore acoustic impedance). In order to evaluate the specific qualities of assorted diffusion properties, various surface characteristics are discussed.

Irregular Surfaces

The natural world is comprised of very few specular reflectors for light waves even fewer for acoustic signals. Until the development of artificial structures, reflecting surfaces were, in almost all cases, irregularly-shaped (with the possible exception of the surface of a very calm body of water). As a result, natural acoustic reflections are almost always diffused to some extent. Early structures were built using simple construction techniques and resulted in flat surfaces and therefore specular reflections.

For approximately 3000 years, and up until the turn of the 20th century, architectural trends tended to favour florid styles, including widespread use of various structural and decorative elements such as fluted pillars, entablatures, mouldings, and carvings. These random and periodic surface irregularities resulted in more diffused reflections according to the size, shape and absorptive characteristics of the various surfaces. The rise of the International Style in the early 1900s [Nuttgens, 1997] saw the disappearance of these largely irregular surfaces and the increasing use of expansive, flat surfaces of concrete, glass and other acoustically reflective materials. This stylistic move was later reinforced by the economic advantages of these design and construction techniques.

Maximum length sequence diffusers

The link between diffused reflections and better-sounding acoustics has resulted in much research in the past 30 years on how to construct diffusive surfaces with predictable results. This continues to be an extremely popular topic at current conferences in audio and acoustics with a great deal of the work continuing on the breakthroughs of Schroeder.

In his 1975 paper, Schroeder outlined a method of designing surface irregularities based on maximum length sequences (MLS) [Golomb, 1967] which result in the diffusion of a specific frequency band. This method relies on the creation of a surface comprised of a series of reflection coefficients alternating between +1 and -1 in a predetermined periodic pattern.

Consider a sound wave entering the mouth of a well cut into the wall from the concert hall as shown in Figure 3.22.

Assuming that the bottom of the well has a reflection coefficient of 1, the reflection returns to the entrance of the well having propagated a distance equalling twice its depth d_n , and therefore undergoing a shift in phase relative to the sound entering the well. The magnitude of this shift is dependent on the relationship between the wavelength and the depth according to Equation 3.23.

$$\varphi = 4\pi \frac{d_n}{\lambda} \quad (3.23)$$

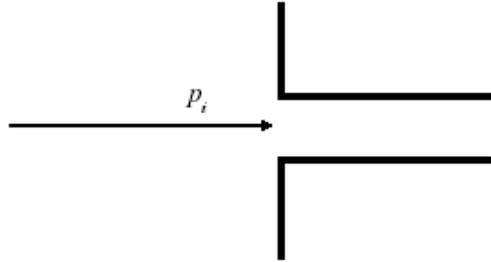


Figure 3.22: Pressure wave incident upon the mouth of a well.

where φ is the phase shift in radians, d_n is the depth of the well and λ is the wavelength of the incident sound wave.

Therefore, if $\lambda = 4d_n$, then the reflection will exit the well having undergone a phase shift of π radians. According to Schroeder, this implies that the well can be considered to have a reflective coefficient of -1 for that particular frequency, however this assumption will be expanded to include other frequencies in the following section.

Using an MLS, the particular required sequence of positive and negative reflection coefficients can be calculated, resulting in a sequence such as the following, for $N=15$:

+ + + - - - + - - + + - + - +

This is then implemented as a series of individually separated wells cut into the reflecting surface as is shown in Figure 3.23. Although the depth of the wells is dependent on a single so-called design wavelength denoted λ_o , in practice it has been found that the bandwidth of diffused frequencies ranges from one-half octave below to one half octave above this frequency [Schroeder, 1975]. For frequencies far below this bandwidth, the signal is typically assumed to be unaffected. For example, consider a case where the depth of the wells is equal to one half the wavelength of the incident sound wave. In this case, the wells now exhibit a reflective coefficient of +1; exactly the opposite of their intended effect, rendering the surface a flat and therefore specular reflector.

The advantage of using a diffusive surface geometry based on maximum length sequences lies in the fact that the power spectrum of the sequence is flat except for a minor dip at DC [Schroeder, 1975]. This permits the acoustical designer to specify a surface that maintains the sound energy in the room through reflection while maintaining a low interaural cross correlation

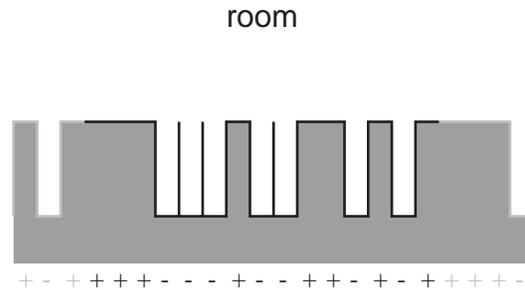


Figure 3.23: MLS Diffuser showing the relationship between the individual wells cut into the wall and the MLS sequence.

(IACC) through predictable diffusion characteristics which do not impose a resonant characteristic on the reflection. The principal disadvantage of the MLS-based diffusion scheme is that it is specific to a relatively narrow frequency band, thus making it impractical for wide-band diffusion.

Schroeder Diffusers

The goal became to find a surface geometry which would permit designers the predictability of diffusion from MLS diffusers with a wider bandwidth. The new system was again introduced by Schroeder in 1979 in a paper describing the implementation of the *quadratic residue diffuser* or *Schroeder diffuser* [Schroeder, 1979] – a device which has since been widely accepted as one of the de facto standards for easily creating diffusive surfaces. Rather than relying on alternating reflecting coefficient patterns, this method considers the wall to be a flat surface with varying impedance according to location. This is accomplished using wells of various specific depths arranged in a periodic sequence based on residues of a quadratic function as in Equation 3.24 [Schroeder, 1979].

$$s_n = n^2, \text{mod}(N) \quad (3.24)$$

where s_n is the sequence of relative depths of the wells, n is a number in the sequence of non-negative consecutive integers (0, 1, 2, 3 ...) denoting the well number, and N is a non-negative odd prime number.

If you're uncomfortable with the concept of the modulo function, just think of it as the remainder. For example, $5, \text{mod}(3) = 2$ because $\frac{5}{3} = 1$ with a remainder of 2. It's the remainder that we're looking for.

For example, for modulo 17, the series is

0, 1, 4, 9, 16, 8, 2, 15, 13, 13, 15, 2, 8, 16, 9, 4, 1, 0, 1, 4, 9, 16, 8, 2, 15

... As may be evident from the representation of this series in Figure 3.24, the pattern is repeating and symmetrical around $n = 0$ and $\frac{N}{2}$.

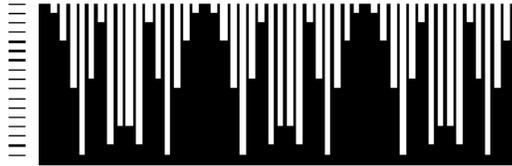


Figure 3.24: Schroeder diffuser for $N = 17$ following the relative depths listed above. Note that this diagram shows 3 repetitions of the period of the sequence.

The actual depths of the wells are dependent on the design wavelength of the diffuser. In order to calculate these depths, Schroeder suggests Equation 3.25.

$$d_n = s_n \frac{\lambda_o}{2N} \quad (3.25)$$

where d_n is the depth of well n and λ_o is the *design wavelength* [Schroeder, 1979].

The widths of these wells w should be constant (meaning that they should all be the same) and small compared to the design wavelength (no greater than $\frac{\lambda_o}{2}$; Schroeder suggests $0.137\lambda_o$). Note that the result of Equation 3.25 is to make the median well depth equal to one-quarter of the design wavelength. Since this arrangement has wells of varying depths, the resulting bandwidth of diffused sound is increased substantially over the MLS diffuser, ranging approximately from one-half octave below the design frequency up to a limit imposed by $\lambda > \frac{\lambda_o}{N}$ and, more significantly, $\lambda > 2w$ [Schroeder, 1979].

The result of this sequence of wells is an apparently flat reflecting surface with a varying and periodic impedance corresponding to the impedance at the mouth of each well. This surface has the interesting property that, for the frequency band mentioned above, the reflections will be scattered to propagate along predictable angles with very small differences in relative amplitude.

3.3.3 Reading List

3.4 Diffraction of Sound Waves

NOT YET WRITTEN

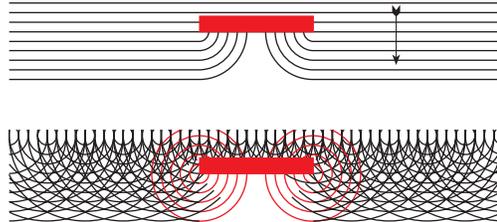


Figure 3.25: An overly-simplified diagram showing how an obstruction (the red block) will shadow a plane wave on one side. Also shown is an overly-simplified example of diffraction, shown as the waves centered at the corners of the obstruction. Omitted are the reflections off the obstruction, and the diffraction off the top two corners.

3.4.1 Reading List

3.5 Struck and Plucked Strings

3.5.1 Travelling waves and reflections

Go find a long piece of rope, tie one end of it to a solid object like a fence rail and pull it tight so that you look something like Figure 3.26



Figure 3.26: You, a rope and a fence post before anything happens.

Then, with a flick of your wrist, you quickly move your end of the rope up and back down to where you started. If you did this properly, then the rope will have a bump in it as is shown in Figure 3.27. This bump will move quickly down the rope towards the other end.



Figure 3.27: You, a rope and a fence post just after you have flicked your wrist to put a bump in the rope.

When the bump hits the fence post, it can't move it because fence posts are harder to move than rope molecules. Since the fence post is at the end of the rope, we say that it *terminates* the rope. The word *termination* is one that we use a lot in this book, both in terms of acoustic as well as electronics. All it means is the end of the system – the rope, the air, the wire... whatever the “system” is. So, an acoustician would say that the rope is *terminated with a high impedance* at the fence post end.

Remember back to the analogy in Section 3.2.2 where the person in the front was pushing on a concrete wall. You pushed the person ahead of you and you wound up getting pushed in the opposite direction that you pushed

in the first place. The same is true here. When the wave on the rope reflects off a higher impedance, then the reflected wave does the same thing. You pull the rope up, and the reflection pulls you down. This end result is shown in Figure 3.28.



Figure 3.28: You, a rope and a fence post just after the wave has reflected off the fence post (the high impedance termination).

3.5.2 Standing waves (aka Resonant Frequencies)

Take a look at a guitar string from the side as is shown in the top diagram in Figure 3.29. It's a long, thin piece of metal that's pulled tight and, on each end it's held up by a piece of plastic called the *bridge* at the bottom and the *nut* at the top. Since the nut and the bridge are firmly attached to heavy guitar bits, they're harder to move than the string, therefore we can say that the string is terminated at both ends with a high impedance.

Let's look at what happens when you pluck a perfect string from the the centre point in slow motion. Before you pluck the string, it's sitting at the equilibrium position as shown in the top diagram in Figure 3.29.

You grab a point on the string, and pull it up so that it looks like the second diagram in Figure 3.29. Then you let go...

One way to think of this is as follows: The string wants to get back to its equilibrium position, so it tries to flatten out. When it does get to the equilibrium position, however, it has some momentum, so it passes that point and keeps going in the opposite direction until it winds up on the other side as far as it was pulled in the first place. If we think of a single molecule at the centre of the string where you plucked, then the behaviour is exactly like a simple pendulum. At any other point on the string, the molecules were moved by the adjacent molecules, so they also behave like pendulums that weren't pulled as far away from equilibrium as the centre one.

So, in total, the string can be seen as an infinite number of pendulums, all connected together in a line, just as is explained in Huygens theory.

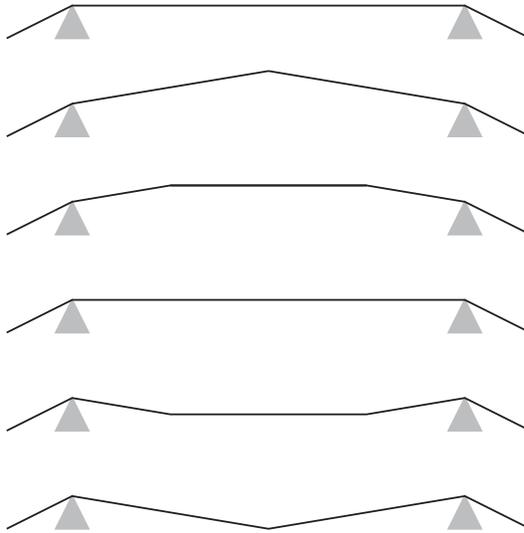


Figure 3.29: A frame-by-frame diagram showing what happens to a guitar string when you pluck it.

If we were to pluck the string and wait a bit, then it would settle down into a pretty predictable and regular motion, swinging back and forth looking a bit like a skipping rope being turned by a couple of kids. If we were to make a movie of this movement, and look at a bunch of frames of the film all at the same time, they might look something like Figure 3.30.

In Figure 3.30, we can see that the string swings back and forth, with the point of largest displacement being in the centre of the string, halfway between the two anchored points at either end. Depending on the length, the tension and the mass of the string, it will swing back and forth at some speed (we'll look at how to calculate this a little later...) which will determine the number of times per second it oscillates. That frequency is called the *fundamental resonant frequency* of the string. If it's in the right range (between 20 Hz and 20,000 Hz) then you'll hear this frequency as a musical pitch.

In reality, this is a bit of an oversimplification. The string actually resonates at other frequencies. For example, if you look at Figure 3.31, you'll see a different *mode* of oscillation. Notice that the string still can't move at the two anchored end points, but it now also does not move in the centre. In fact, if you get a skipping rope or a telephone cord and wiggle it back and forth regularly at the right speed, you can get it to do exactly this

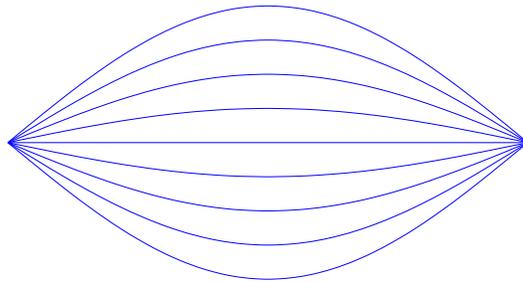


Figure 3.30: The first mode of vibration of a string anchored on both ends. This will vibrate at a frequency f .

pattern. I would highly recommend trying.

A short word here about technical terms. The point on the string that doesn't move is called a *node*. You can have more than one node on a string as we'll see below. The point on the string that has the highest amplitude of movement is called the *antinode*, because it's the evil opposite of the node, I suppose...

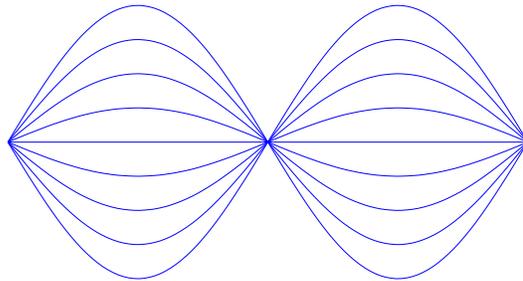


Figure 3.31: The second mode of vibration of a string anchored on both ends. This will vibrate at a frequency $2f$.

One of the interesting things about the mode shown in Figure 3.31 is that its wavelength on the string is exactly half the wavelength of the mode shown in Figure 3.30. As a result, it vibrates back and forth twice as fast and therefore has twice the frequency. Consequently, the pitch of this vibration is exactly one octave higher than the first.

This pattern continues upwards. For example, we have seen modes of vibration with one and with two “bumps” on the string, but we could also have three as is shown in Figure 3.32. The frequency of this mode would be three times the first. This trend continues upwards with an integer number of bumps on the string until you get to infinity.

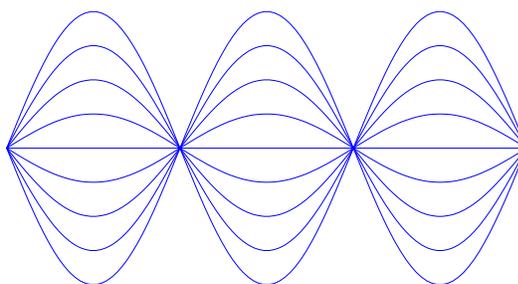


Figure 3.32: The third mode of vibration of a string anchored on both ends. This will vibrate at a frequency $3f$.

Since the string is actually vibrating with all of these modes at the same time, with some relative balance between them, we wind up hearing a fundamental frequency with a number of harmonics. The combined timbre (or sound colour) of the sound of the string is determined by the relative levels of each of these harmonics as they evolve over time. For example, if you listen to the sound of a guitar string, you might notice that it has a very bright sound immediately after it has been plucked, and that the sound gets darker over time. This is because at the start of the sound, there is a relatively high amount of energy in the upper harmonics, but these decay more quickly than the lower ones and the fundamental. Therefore, at the end of the sound, you get only the lowest harmonics and fundamental of the string, and therefore a darker sound quality.

It might be a little difficult to think that the string is moving at a maximum in the middle for some modes of vibration in exactly the same place as it's not moving at all for other modes. If this is confusing, don't worry, you're completely normal. Take a look at Figure 3.33 which might help to alleviate the confusion. Each mode is an independent component that can be considered on its own, but the total movement of the string is the result of the sum of all of them.

One of the neat tricks that you can do on a stringed instrument such as

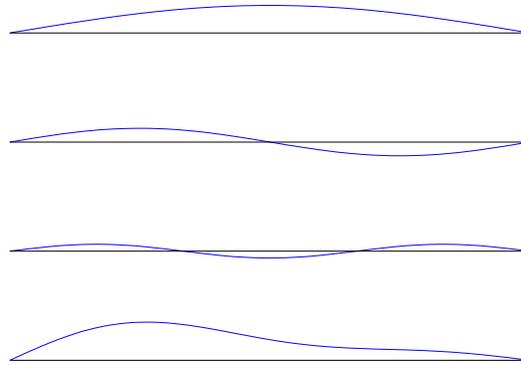


Figure 3.33: The sum of the three first modes of vibration of a string. The top three plots show the first, second and third modes of vibration with relative amplitudes 1, 0.5 and 0.25 respectively. The bottom plot is the sum of the top three.

a violin or guitar is to play these modes selectively. For example, the normal way to play a violin is to clamp the string down to the fingerboard of the instrument with your finger to effectively shorten it. This will produce a higher note if the string is plucked or bowed. However, you could gently touch the string at exactly the halfway point and play. In this case, the string still has the same length as when you're not touching it. However, your finger is preventing the string from moving at one particular point. That point (if your finger is halfway up the string) is supposed to be the point of maximum movement for all of the odd harmonics (which include the fundamental – the 1st harmonic). Since your finger is there, these harmonics can't vibrate at all, so the only modes of vibration that work are the even-numbered harmonics. This means that the second harmonic of the string is the lowest one that's vibrating, and therefore you hear a note an octave higher than normal. If you know any string players, get them to show you this effect. It's particularly good on a cell or double bass because you can actually see the harmonics as a shape of the string.

3.5.3 Impulse Response vs. Resonance

Think back to the beginning of this section when you were playing with the skipping rope. Let's take the same rope and attach it to two fence posts, pulling it tight. In our previous example, you flicked the rope with your wrist, making a bump in it that travelled along the rope and reflected off the opposite end. In our new setup, what we'll do is to tap the rope and

watch what happens. What you'll (hopefully) be able to see is that the bump you created by tapping travels in two opposite directions to the two ends of the rope. These two bumps reflect and return, meeting each other at some point, crossing each other and so on. This process is shown in Figures 3.34 through 3.37.

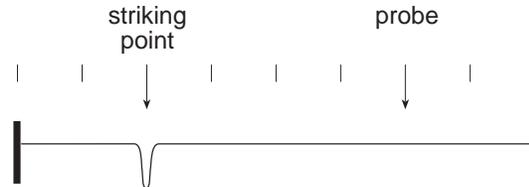


Figure 3.34: A rope tied to two fence posts just after you have tapped it.

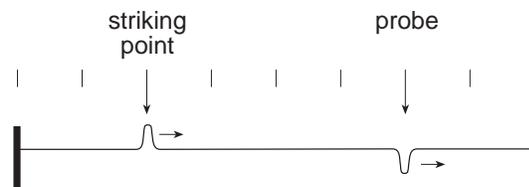


Figure 3.35: A rope tied to two fence posts a little later. Note that the bump you created is travelling in two directions, and each of the two resulting bumps is smaller than the original. The bump on the left has already reflected off the left end of the rope.

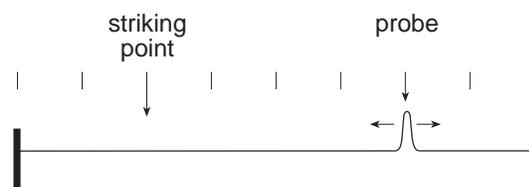


Figure 3.36: The two bumps after they have reflected off the high impedance terminations (the fence posts). When the two bumps meet each other, they appear for an instant to be one bump

Let's assume that you are able to make the bump in the rope infinitely narrow, so that it appears as a spike or an *impulse*. Let's also put a theoretical probe on the rope that measures its vertical movement at a single point over time. We'll also, for the sake of simplicity, put the probe the same distance from one of the fence posts as you are from the other post. This

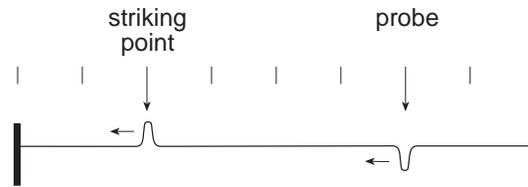


Figure 3.37: After the two reflections have met each other, they continue on in opposite directions. (The right bump has reflected off the right end of the rope.) The process then repeats itself indefinitely, assuming that there are no losses of energy in the system due to things like friction.

is to ensure that the probe is at the point where the two little spikes meet each other to make one big spike. If we graphed the output of the probe over time, it would look like Figure 3.38.

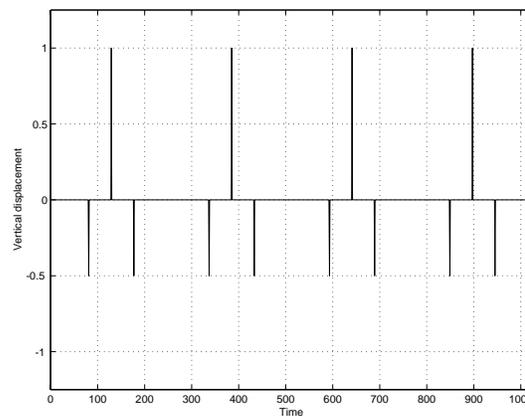


Figure 3.38: The output of the probe showing the vertical movement of the rope over time. This response corresponds directly to Figures 3.34 to 3.37

This graph shows how the rope responds in time when the impulse (an instantaneous change in displacement or pressure which instantaneous) is applied to it. Consequently we call it the *impulse response* of the system. Note that the graph in Figure 3.38 corresponds directly to Figures 3.35 to 3.37 so that you can see the relationship between the displacement at the point where the probe is located on the string and passing time. Note that only the first three spikes correspond to the pictures – after those three have gone by, the whole thing repeats over and over.

As we'll see later in Section 8.2, we are able to do some math on this impulse response to find out what the frequency content of the signal is –

in other words, the harmonic content of the signal. The results of this is shown in Figure 3.39.

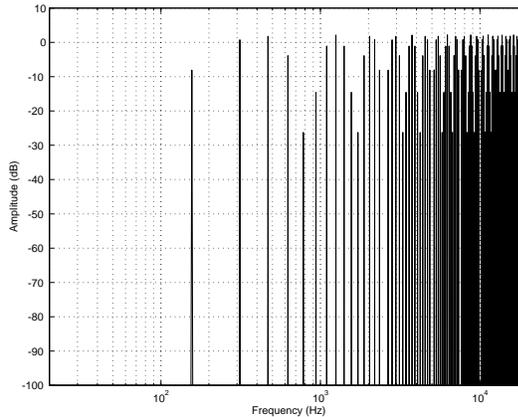


Figure 3.39: The frequency content of the string’s vibrations at the location of the probe.

This graph shows us that we have the fundamental frequency and all its harmonics at various levels up to ∞ Hz. The differences in the levels of the harmonics is due to the relative locations of the striking point and the probe on the string. If we were to move either or both of these locations, then the relative times of arrival of the impulses would change and the balance of the harmonics would change as well. Note that the actual frequencies shown in the graph are completely arbitrary. These will change with the characteristics of the string as we’ll see below in Section 3.5.4.

“So what?” I hear you cry. Well, this tells us the resonant frequencies of the string. Basically, Figure 3.39 (which is a frequency content plot based on the impulse response in time) is the same as the description of the standing wave in Section 3.5.2. Each spike in the graph corresponds to a frequency in the standing wave series.

3.5.4 Tuning the string

If we weren’t able to change the pitch of a string, then many of our musical instruments would sound pretty horrid and we would have very boring music... Luckily, we can change the frequency of the modes of vibration of a string using any combination of three variables.

String Mass

The reason a string vibrates in the first place is because, when you pull it back and let go, it doesn't just swing back to the equilibrium position and stop dead in its tracks. It has momentum and therefore passes where it wants to be, then turns around and swings back again. The heavier the string is, the more momentum it has, so the harder it is to stop. Therefore, it will move slower. (I could make an analogy here about people, but I'll behave...) If it moves slower, then it doesn't vibrate as many times per second, therefore heavier strings vibrate at lower frequencies.

This is why the lower strings on a guitar or a piano have a bigger diameter. This makes them heavier per metre, therefore they vibrate at a lower frequency than a lighter string of the same length. As a result, your piano doesn't have to be hundred of metres long...

String Length

The fundamental mode of vibration of a string results in the string looking like a half-wavelength of a sound wave. In fact, this half-wavelength is exactly that, but the medium is the string itself, not the air. If we make the wavelength longer by lengthening the string, then the frequency is lowered, just as a longer wavelength in air corresponds to a lower frequency.

String Tension

As we saw earlier, the reason the string vibrates is because it's trying to get back to the equilibrium position. The tension on the string – how much it's being stretched – is the force that's trying to pull it back into position. The more you pull (therefore the higher the tension) the faster the string will move, therefore this higher the pitch.

QUESTION TO SELF: IS THERE A SIMPLE EQUATION FOR PREDICTING THE FUNDAMENTAL RESONANT FREQUENCY OF THE STRING?

3.5.5 Harmonics in the real-world

While it's nice to think of resonating strings in this way, that's not really the way things work in the real world, but it's pretty close. If you look at the plots above that show the modes of vibration of a string, you'll notice that the ends start at the correct slope. This is a nice, theoretical model, but it really doesn't reflect the way things really behave. If we zoom in at

the point where the string is terminated (or anchored) on one end, we can see a different story. Take a look at Figure 3.40. The top plot shows the way we've been thinking up to now. The perfect string is securely anchored by a bridge or clamp, and inside the clamp it is free to bend as if it were hinged by a single molecule. of course, this is not the case, particularly with metal strings as we find on most instruments.

The bottom diagram tells a more realistic story. Notice that the string continues on a straight line out of the clamp and then gradually bends into the desired position – there is no fixed point where the bend occurs at an extreme angle.

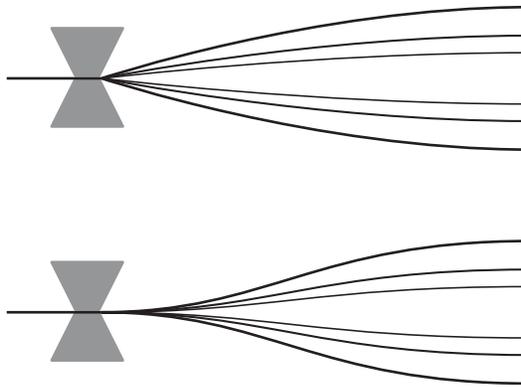


Figure 3.40: Theory vs. reality on a string termination. The top diagram shows the theoretical bend of a string at the anchor point. The bottom diagram shows a more realistic behaviour, particularly for stiff strings as in a piano.

Now think back to the slopes of the string at the anchor point for the various modes of vibration. The higher the frequency of the mode, the shorter the wavelength on the string, and the steeper the slope at the string termination. This means that higher harmonics are trying to bend the string more at the ends, yet there is a small conflict here... there is typically less energy in the higher harmonics, so it's more difficult for them to move the string, let alone bending it more than the fundamental. As a result, we get a strange behaviour as is shown in Figure 3.41

As you can see in this diagram, the lower harmonic is able to bend the string more, closer to the termination than the higher harmonic. This means that the effective length of the string is shorter for higher harmonics than for lower ones. As a result, the higher the harmonic, the more incorrect it is mathematically, and the sharper it is musically speaking. Essentially, the higher the harmonic, the sharper it gets.



Figure 3.41: High and low harmonics on a stiff string. Note that the lower mode (represented as a blue line) has a longer effective length than the higher mode (the red line).

This is why good piano tuners tune by ear rather than using a machine. Let's say you want a Middle C on a piano to sound in tune with the C one octave above it. The fundamental of the higher C is theoretically the same frequency as the second harmonic of the lower one. But, we now know that the harmonic of the lower one is sharper than it should be, mathematically speaking. Therefore, in order to sound in tune, the higher C has to be a little higher than its theoretically correct tuning. If you tune the strings using a tuner, then they will have the correct frequencies for the theoretical world, but not for your piano. You will have to tune the various frequencies a little to high for higher pitches in order to sound correct.

3.5.6 Reading List

The Physics of the Piano Blackham, E. D. Scientific American December 1965

Normal Vibration Frequencies of a Stiff Piano String Fletcher, H. JASA Vol 36, No 1 Jan 1964

Quality of Piano Tones Fletcher H. et al JASA Vol 34 No 6 June 1962

3.6 Waves in a Tube (Woodwind Instruments)

Imagine that you're very, very small and that you're standing at the end of the inside a pipe that is capped at both ends. (Essentially, this is exactly the same as if you were standing at one end of a hallway without doors.) Now, clap your hands. If we were in a free field (described in Section 3.1.19) then the power in your hand clap would be contained in a forever-expanding spherical wavefront, so it would get quieter and quieter as it moves away. However, you're not in a free field, you're trapped in a pipe. As a result, the wave can only expand in one direction – down the pipe. The result of this is that the sound power in your hand clap is trapped, and so, as the wavefront moves down the pipe, it doesn't get any quieter because the wavefront doesn't expand – it's only moving forward.

3.6.1 Waveguides

In theory, if there was no friction against the pipe walls, and there was no absorption in the air, even if pipe were hundreds of kilometers long, the sound of your handclap would be as loud at the other end as it was 1 m down the pipe... (In fact, back in the early days of audio effects units, people used long pieces of hose as delay lines.)

One other thing – as you get further away from the sound source in the pipe, the wavefront becomes flatter and flatter. In fact, in not very much distance at all, it becomes a planewave. This means that if you look at the pressure wave across the pipe, you would see that it is perpendicular with the walls of the pipe. When this happens, the pipe is guiding the wave down its length, so we call the pipe a *waveguide*.

3.6.2 Resonance in a closed pipe

When the pressure wave hits the capped end of the pipe, it reflects off the wall and bounces back in the direction it came in. If we sent a high-pressure wave down the pipe, since the cap has a higher acoustic impedance than the air in the pipe, the reflection will also be a high-pressure wave. This wavefront will travel back down the waveguide (the pipe) until it hits the opposite capped end, and the whole process repeats. This process is shown in Figures 3.42 through 3.45.

If we assume that the change in pressure is instantaneous and that it returns to the equilibrium pressure instantaneously then we are creating an impulse. The response at the microphone over time, therefore, is the impulse

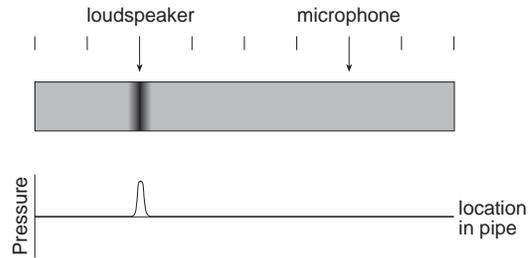


Figure 3.42: A pipe that is closed with a high-impedance cap at both ends. There is a loudspeaker and a microphone inside the pipe at the shown locations. The black level of the diagram shows the pressure inside the pipe – the darker the higher the pressure.

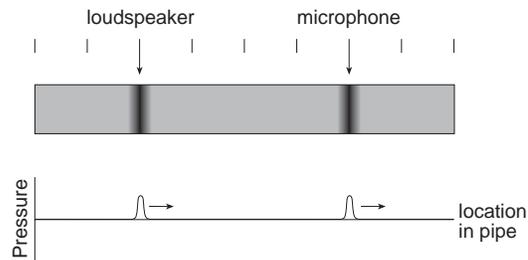


Figure 3.43: The same pipe a little later. Note that the high pressure wave is travelling in two directions, and each of the two resulting bumps is smaller than the original. The wave on the left has already reflected off the left end of the pipe.

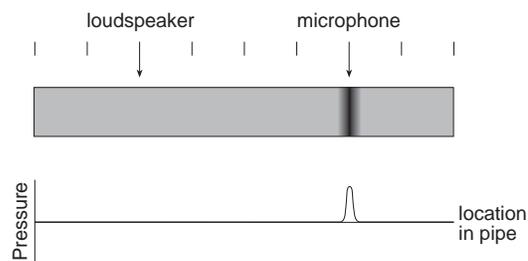


Figure 3.44: The two waves after they have reflected off the high impedance terminations (the pipe caps). When the two waves meet each other, they appear for an instant to be one wave.

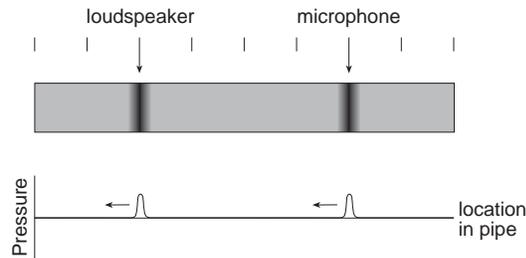


Figure 3.45: After the two reflections have met each other, they continue on in opposite directions. (The right wave has reflected off the right end of the pipe.) The process then repeats itself indefinitely, assuming that there are no losses of energy in the system due to things like friction on the pipe walls or losses in the caps.

response of the closed pipe shown like Figure 3.46. Note that this impulse response corresponds directly to Figures 3.34 through 3.37. Also note that, unlike the struck rope, all pressure wave is always positive if we create a positive pressure to begin with. This is, in part, because we are now looking at air pressure whereas, in the case of the string, we were monitoring vertical displacement. In fact, if we were graphing the molecules displacements in the pipe, we would see the values alternating between positive and negative just as in the impulse response for the string.

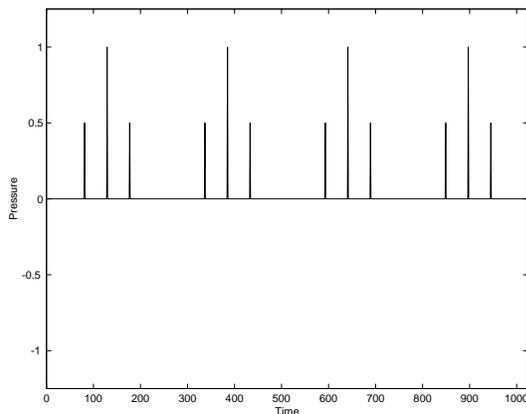


Figure 3.46: The impulse response of a pipe that has a cap with an infinite acoustic impedance at both ends. The sound source that produced the impulse is assumed to be at one of the ends of the pipe, so we only see a single reflection. Note that the impulse does not decay over time since all of the acoustic power is trapped inside the pipe. The first three spikes in this impulse response corresponds directly to Figures 3.35 through 3.37.

Just as with the example of the vibrating string in the previous Chapter, we can convert the impulse response in Figure 3.46 into a plot of the frequency content of the signal as is shown in Figure 3.47. Note that this response is only true at the location of the microphone. If it or the loudspeaker's position were changed, so would the relative balances of the harmonics. However, the frequencies of the harmonics would not change – these are determined by the length of the pipe and the speed of sound as we'll see below.

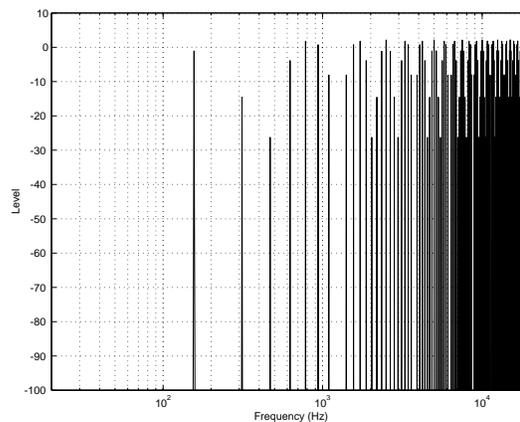


Figure 3.47: The frequency content of the pressure waves in the air inside the pipe at the location of the microphone.

Figure 3.47 tells us that the signal inside the pipe can be decomposed into a series of harmonics, just like we saw on the string in the previous section. The only problem here is that they're a little more difficult to imagine because we're talking about longitudinal waves instead of transverse waves.

FINISH OFF THIS SECTION ON LONGITUDINAL STANDING WAVES
INCLUDE ANIMATION

It is important to note that, in almost every way, this system is identical to a guitar string. The only big difference is that the wave on the guitar string is a transverse wave whereas the pipe has a longitudinal wave, but their basic behaviours are the same.

Of course, it's not very useful to have a musical instrument that is a completely closed pipe – since none of the sound gets out, you probably won't get very large audiences. Then again, you might wind up with the perfect instrument for playing John Cage's *4'33"*...

So, let's let a little sound out of the pipe. We won't change the caps on the two ends, we'll just cut a small rectangular hole in the side of the pipe at one end. To get an idea of what this would look like, take a look at Figure 3.48.

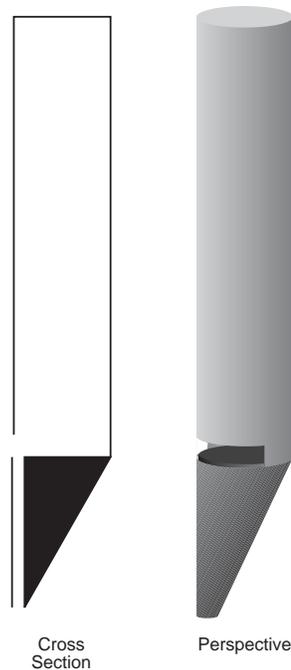


Figure 3.48: A cross section and perspective drawing of a closed organ pipe.

Let's think about what's happening here in slow motion. Air comes into the pipe from the bottom through the small section on the left. It's at a higher pressure than the air inside the pipe, so when it reaches the bottom of the pipe, it sends a high pressure wavefront up to the other end of the pipe. While that's happening, there's still air coming into the pipe. The high pressure wavefront bounces back down the pipe and gets back to where it started, meeting the new air that's coming into the bottom. These two collide and the result is that air gets pushed out the hole on the side of the pipe. This, however, causes a negative pressure wavefront which travels up the pipe, bounces back down and arrives where it started where it sucks air into the hole in the side of the pipe. This causes a high pressure wavefront, etc etc... This whole process repeats itself so many times a second, that we can measure it as an oscillation using a microphone (if the pipe is the right

length – more on this in the next section).

Consider the description above as an observer standing on the outside of the pipe. We just see that there is air getting pushed out and sucked into the hole on the side of the pipe a bunch of times a second. This results in high and low pressure wavefronts radiating outwards from the hole and we hear sound.

Overtone series in a closed pipe

TO BE WRITTEN

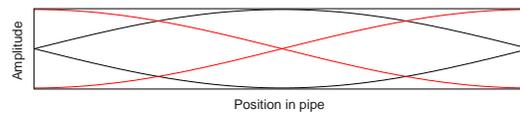


Figure 3.49: The relationship between the position of the probe in a closed pipe (the horizontal axis) and the particle velocity maximum and minimum (in black) as well as the pressure maximum and minimum (in red). This diagram shows the fundamental resonance of the pipe.

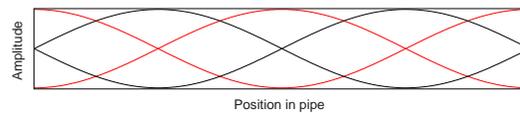


Figure 3.50: The relationship between the position of the probe in a closed pipe (the horizontal axis) and the particle velocity maximum and minimum (in black) as well as the pressure maximum and minimum (in red). This diagram shows the first overtone in the pipe – half the wavelength and therefore twice the frequency of the fundamental.

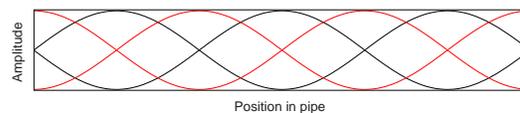


Figure 3.51: The relationship between the position of the probe in a closed pipe (the horizontal axis) and the particle velocity maximum and minimum (in black) as well as the pressure maximum and minimum (in red). This diagram shows the second overtone in the pipe – one third the wavelength and therefore three times the frequency of the fundamental.

Tuning a closed pipe

So, if we know the length of pipe, how can we calculate the resonant frequency? This might seem like an arbitrary question, but it's really important if you've been hired to build a pipe organ. You want to build the pipes the right length in order to provide the correct pitch.

Take a look back at the example of an impulse response of a closed pipe, shown in Figure 3.46. As you can see, the sequence shown consists of three spikes and that sequence is repeated over and over forever (if it's an undamped system). The time it takes for the sequence to repeat itself is the time it takes for the impulse to travel down the length of the pipe, bounce off one end, come all the way back up the pipe, bounce off the other end and to get back to where it started. Therefore the wavelength of the fundamental resonant frequency of a closed pipe is equal to two times the length of the pipe, as is shown in Equation 3.26. Note that this equation is a general one for calculating the wavelengths of all resonant frequencies of the pipe. If you want to find the wavelength of the fundamental, set n to equal 1.

$$\lambda_n = \frac{2L}{n} \quad (3.26)$$

where λ_n is the wavelength of the n th resonant frequency of the closed pipe of length L .

Since the wavelength of the fundamental resonant frequency of a closed pipe is twice the length of the pipe, you'll often hear closed pipes called *half-wavelength resonators*. This is a just geeky way of saying that it's closed at both ends.

If you'd like to calculate the actual frequency that the pipe resonates, then you can just calculate it from the wavelength and the speed of sound using Equation 3.8. What you would wind up with would look like Equation 3.27.

$$f_n = n \frac{c}{2L} \quad (3.27)$$

3.6.3 Open Pipes

NOT YET WRITTEN

Overtone series in a closed pipe

TO BE WRITTEN

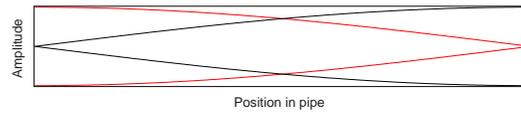


Figure 3.52: The relationship between the position of the probe in an open pipe (the horizontal axis) and the particle velocity maximum and minimum (in black) as well as the pressure maximum and minimum (in red). This diagram shows the fundamental resonance of the pipe.

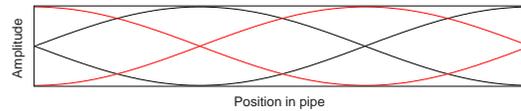


Figure 3.53: The relationship between the position of the probe in an open pipe (the horizontal axis) and the particle velocity maximum and minimum (in black) as well as the pressure maximum and minimum (in red). This diagram shows the first overtone in the pipe – one third the wavelength and therefore three times the frequency of the fundamental.

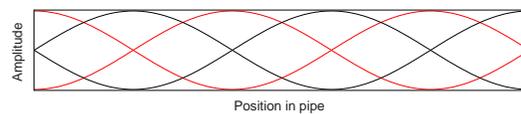


Figure 3.54: The relationship between the position of the probe in a closed pipe (the horizontal axis) and the particle velocity maximum and minimum (in black) as well as the pressure maximum and minimum (in red). This diagram shows the second overtone in the pipe – one fifth the wavelength and therefore five times the frequency of the fundamental.

Tuning a closed pipe**Effective vs. actual length****3.6.4 Reading List**

The Physics of Woodwinds Benade, A. H. Scientific American October 1960

The Acoustics of Orchestral Instruments and of the Organ Richardson,
E. G. Arnold and Co. (1929)

Horns, Strings and Harmony Benade, A. H. Doubleday and Co., Inc.
(1960)

On Woodwind Instrument Bores Benade, A. H. JASA Vol 31 No 2 Feb
1959

3.7 Helmholtz Resonators

Back in Section 3.1.2 we looked at a type of simple harmonic oscillator. This was comprised of a mass on the end of a spring. As we saw, if there is no such thing as friction, if you lift up the mass, it will bounce up and down forever. Also, if we graphed the displacement of the mass over time, we'd see a sinusoidal waveform. If there is such as things as friction (such as air friction), then some of the energy is lost from the system as heat and the oscillator is said to be damped. This means that, eventually, it will stop bouncing.

We could make a similar system using only moving air instead of a mass on a spring. Go find a wine bottle. If it's not empty, then you'll have to empty it (I'll leave it to you to decide exactly how this should be accomplished). If we simplify the shape of the bottle a little bit, it is a tank with an opening into a long neck. The top of the neck is open to the outside world.

There is air in the tank (the bottle) whose pressure can be changed (we can make it higher or lower than the outside pressure) but if we do change it, it will want to get back to the normal pressure. This is essentially the same as the compression of the spring. If we compress the spring, it pushes back to try to be normal. If we pull the spring, it pulls against us to try to be normal. If we compress the air in the bottle, it pushes back to try to be the same as the outside pressure. If we decompress the air, it will pull back, trying to normalize. So, the tank is the spring.

As we saw in Section 3.6.1 on waveguides, if air goes into one end of a tube, then it will push air out the other end. In essence, the air in the tube is one complete thing that can move back and forth inside the tube. Remember that the air in the tube has some mass (just like the mass on the end of the spring) and that one end is connected to the tank (the spring).

Therefore, a wine bottle can be considered to be a simple harmonic oscillator. All we need to do is to get the mass moving back and forth inside the neck of the bottle. We already know how to do this – you blow across the top of the bottle and you'll hear a note.

We can even take the analogy one step further. The system is a damped oscillator (good thing too... otherwise every bottle that you ever blew across would still be making sound... and that would be pretty annoying) because there is friction in the system. The air mass inside the neck of the bottle rubs against the inside of the neck as it moves back and forth. This resulting friction between the air and the neck causes losses in the system, however the wider the neck of the bottle, the less friction per mass, as we'll see in

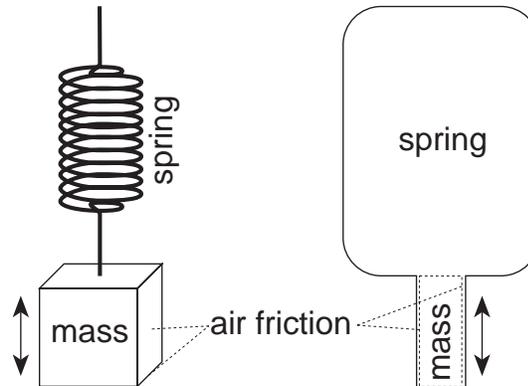


Figure 3.55: A diagram showing two equivalent systems. On the left is a mass supported by a spring, whose oscillation is damped by resistance to its movement caused by air friction. On the right is a mass of air in a tube (the neck of the bottle) pushing and pulling against a spring (the air in the bottle) with the system oscillation damped by the air friction in the bottle neck.

the math later.

This thing that we’ve been looking at is called a wine bottle, but an acoustician would call it a *Helmholtz resonator* named after H. L. F. von Helmholtz. FIND SOME INFORMATION ABOUT HELMHOLTZ AND PUT IT HERE.

What’s so interesting about this device that it warrants its own name other than “wine bottle?” Well, if we were to assume (incorrectly) that a wine bottle was a quarter-wavelength resonator (hey, it’s just a funny-shaped pipe that’s closed on one end, right?) and we were to calculate its fundamental resonant frequency, we’d find that we’d calculate a *really* wrong number. This is because a Helmholtz resonator doesn’t behave like a quarter-wavelength resonator, it has a much lower fundamental resonant frequency. Also, it’s a lot more stable – it’s pretty easy to get a quarter-wavelength resonator to fluke up to one of its harmonics just by plowing a little harder. If this wasn’t true, flute music would be really boring (or flutists would need more fingers). It’s much more difficult to get a beer bottle to give you a large range of notes just by blowing harder.

How do we calculate the fundamental resonant frequency of a Helmholtz resonator? (I knew you were just dying to find out...) This is shown in Equation 3.28

$$\omega_0 = c\sqrt{\frac{S}{L'V}} \quad (3.28)$$

where ω_0 is the fundamental resonant frequency of the resonator in radians per second (to convert to Hz, see Equation 3.5), c is the speed of sound in m/s, S is the cross-sectional area of the neck of the bottle, L' is the *effective length* of the neck (see below for more information on this) and V is the volume of the bottle (not including the neck) [Kinsler and Frey, 1982].

So, as you can see, this isn't too bad to calculate except for the bit about the "effective length" of the neck. How does the effective length compare to the actual length of the neck of the bottle? As we saw in the previous chapter, the acoustical length of an open pipe is not the same as its real measurable length. In fact, it's a bit longer, and the amount by which it's longer is dependent on the frequency and the cross-sectional area of the pipe. This is also true for the neck of a Helmholtz resonator, since it's open to the world on one end.

If the pipe is *unflanged* (meaning that it doesn't flare out and get bigger on the end like a horn), then you can calculate the effective length using Equation 3.29.

$$L' = L + 1.5a \quad (3.29)$$

where L is the actual length of the pipe, and a is the inside radius of the neck [Kinsler and Frey, 1982].

If the pipe is *flanged* (meaning that it does flare out like a horn), then the effective length is calculated using Equation 3.30 [Kinsler and Frey, 1982]

$$L' = L + 1.7a \quad (3.30)$$

Please note that these equations won't give you exactly the right answer, but they'll put you in the ballpark. Things like the actual shape of the bottle and the neck, how the flange is shaped, how the neck meets the bottle... many things will have a contribution to the actual frequency of the resonator.

Is this useful information? Well, consider that you now know that the oscillation frequency is dependent on the mass of the air inside the neck of the bottle. If you make the mass smaller, then the frequency of the oscillation will go up. Therefore, if you stick your finger into the top of a beer bottle and blow across it, you'll get a higher pitch than if your finger wasn't there. The further you stick your finger in, the higher the pitch. I once saw a woman in a bar in Montreal win money in a talent contest by playing "Girl From Ipanema" on a single beer bottle in this manner. Therefore, yes. It is useful information.

3.7.1 Reading List

3.8 Characteristics of a Vibrating Surface

NOT YET WRITTEN

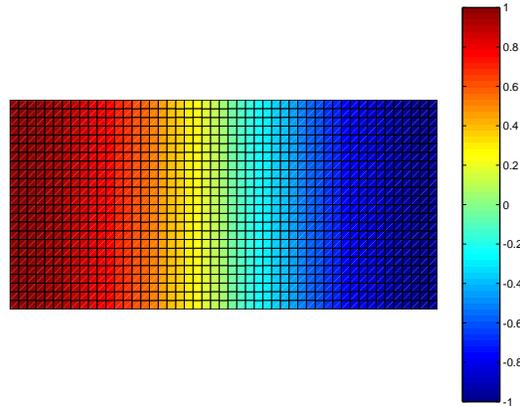


Figure 3.56: The first mode of vibration along the length of a rectangular plate. Note that there is no vibration across its width.

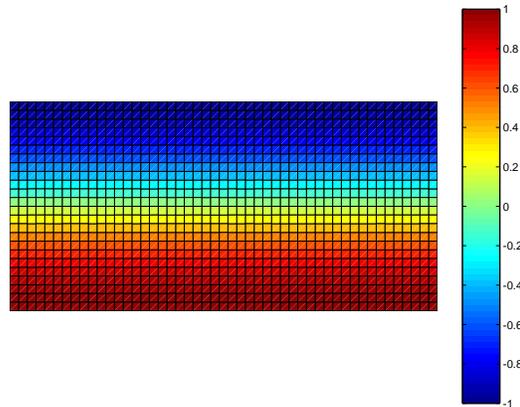


Figure 3.57: The first mode of vibration across the width of a rectangular plate. Note that there is no vibration along its length.

3.8.1 Reading List

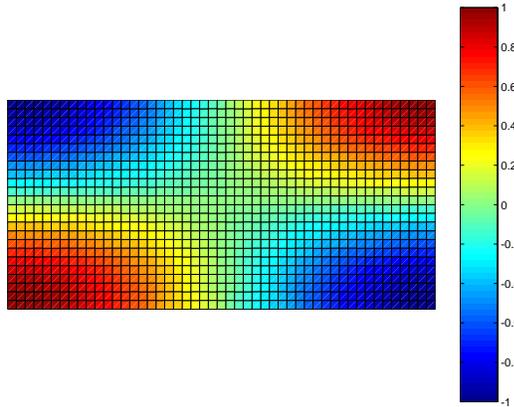


Figure 3.58: The first modes of vibration along the length and width of a rectangular plate.

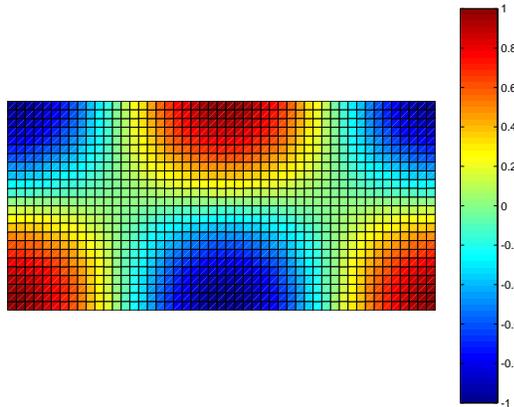


Figure 3.59: The second mode of vibration along the length and the first mode of vibration across the width of a rectangular plate.

3.9 Woodwind Instruments

NOT YET WRITTEN

3.9.1 Reading List

3.10 Brass Instruments

NOT YET WRITTEN

3.10.1 Reading List

The Physics of Brasses Benade, A. H. Scientific American July 1973

Horns, Strings and Harmony Benade, A. H. Doubleday and Co. Inc. (1960)

The Trumpet and the Trombone Bate, P. W. W. Norton and Co., Inc. (1966)

Complete Solutions of the "Webster" Horn Equation Eisner, E. JASA Vol 41 No 4 Pt 2 April 1967

On Plane and Spherical Waves in Horns of Non-uniform Flare Jansson, E. V. and Benade, A. H. Technical Report, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm March 15, 1973

3.11 Bowed Strings

We've already got most of the work done when it comes to bowed strings. The string itself behaves pretty much the same way it does when it's struck. That is to say that it tends to prefer to vibrate at its resonant modal frequencies. The only real question then is: how does the bow transfer what appears to be continuous energy into the string to keep vibrating?

Think of pushing a person on a swing. They're sitting there, stopped, and you give them a little push. This moves them away, then they come back towards you, and just when they're starting to head away from you again, you give them another little push. Every time they get close, you push them away once they've started moving away. Using this technique, you can use just a little energy on each push, in synch with the resonant frequency of the person on the swing to get them going quite high in the air. The trick here is that you're pushing at the right time. You have to add energy to the system when they're not moving too fast (at the bottom of the swing) because that would mean that you have to work hard, but at a time when you're pushing in the same direction that they're moving.

This is basically the way a bow works. Think of the string as the swing and the bow as you. In between the two is a mysterious substance called *rosin* which is essentially just dried tree sap (actually, it's dried turpentine from pine wood as far as I've been able to find out). The important thing about rosin is that it is a substance with a very high *static friction coefficient*, but a very low *dynamic friction coefficient*.

Huh? Well, first let's define what a friction coefficient is. Put a book on a table and then push it so that it slides – then stop pushing. The book stops moving. The reason is friction – the book really doesn't want to move across the table because friction is preventing it from doing so. The amount of friction that's pushing back against you is determined by a bunch of things including the mass of the book, and the smoothness of the two surfaces (the book cover and the table top). This amount of friction is called the *friction coefficient*. The higher the coefficient, the more friction there is and the harder it is to move the book – in essence the more the book “sticks” to the table.

Now we have to look at two different types of friction coefficients – static and dynamic. Just before you start to push the book across the table, it's just sitting there, stopped (in other words, it's static – i.e. it's not moving). After you've started pushing the book, and it's moving across the table, it's probably easier to keep it going than it was to get it started. Therefore the dynamic friction coefficient (the amount of friction there is when the object

is moving) is lower than the static friction coefficient (the amount of friction there is when the object is stopped).

So, going back to the statement that rosin has a very high static friction coefficient, but a very low dynamic friction coefficient, think about what happens when you start bowing a string.

1. You put the bow on the string and nothing is moving yet.
2. You push the bow, but it has rosin on it which has a very high static friction coefficient – it sticks to the string, so it pulls the string along with it.
3. The string, however, is under tension, and it wants to return to the place where it started. As it gets pulled further and further away, it is pulling back more and more.
4. Finally, we reach a point where the tension on the string is pulling back more than the rosin can hold, so the string lets go and starts to slide back. This sliding happens very easily because rosin has a very low dynamic friction coefficient.
5. So, the string slides back to where it came from in the opposite direction to that in which the bow is moving. It passes the equilibrium position and moves back too far and therefore starts to slow down.
6. Once it gets back as far as it's going to go, it turns around and heads back towards the equilibrium position, in the same direction of travel as the bow...
7. At some point a moment later, the string and the bow are moving at the same speed in the same direction, therefore they're stopped relative to each other. Remember that the rosin has a high static friction coefficient, so the string sticks to the bow and the whole process repeats itself.

In many ways this system is exactly the same as pushing a person on a swing. You feed a little energy into the oscillation on every cycle and therefore keep the whole thing going with only a minimum of energy expended. The ultimate reason so little energy is expended is because you are smart about when you inject it into the system. In the case of the bow and the string, the timing of the injection looks after itself.

3.11.1 Reading List

The Physics of Violins Hutchins, C. M. Scientific American November 1962

The Mechanical Action of Violins Saunders, F. A. JASA Vol 9 No 2 Oct 1937

Regarding the Sound Quality of Violins and a Scientific Basis for Violin Construction Meinel, H. JASA Vol 29 No 7 July 1957

Subharmonics and Plate Tap Tones in Violin Acoustics Hutchins, C. M. et al JASA Vol 32 No 11 Nov 1960

On the Mechanical Theory of Vibrations of Bowed Strings and of Musical Instruments of the Violin Family Raman, C. V. Indian Association for the Cultivation of Science (1918)

The Bowed String and the Player Schelleng, J. C. JASA Vol 53 No 1 Jan 1973

The Physics of the Bowed String Schelleng, J. C. Scientific American January 1974

3.12 Percussion Instruments

NOT YET WRITTEN

3.12.1 Reading List

3.13 Vocal Sound Production

NOT YET WRITTEN

3.13.1 Reading List

The Acoustics of the Singing Voice Sundberg, J. Scientific American March 1977

Singing : The Mechanism and the Technic Vannard, W. Carl Fischer, Inc. (1967)

Towards an Integrated Physiologic-Acoustic Theory of Vocal Registers Large, J. NATS Bulletin Vol 28, No. 3 Feb/Mar 1972

Articulatory Interpretation of the "Singing Formant" Sundberg, J. JASA Vol 55, No 4 April 1974

3.14 Directional Characteristics of Instruments

NOT YET WRITTEN

3.14.1 Reading List

3.15 Pitch and Tuning Systems

3.15.1 Introduction

Go back – way back to a time just before harmony. when we had only a melody line, people would sing a tune made up of notes that were probably taught to them by someone else singing. Then people thought that it would be a really neat idea to add more notes at the same time – harmony! The problem was deciding which notes sounded good together (*consonant*) and which sounded bad (*dissonant*). Some people sat down and decided mathematically that some sounds ought to be consonant or dissonant while other people decided that they ought to use their ears.

For the most part, they both won. It turns out that, if you play two frequencies at the same time, you’ll like the combinations that have mathematically simple relationships – and there’s a reason which we talked about in a previous chapter – beating.

If I play an “A” below Middle C on a piano, the sound includes the fundamental 220 Hz, as well as all of the upper harmonics at various amplitudes. So we’re hearing 220 Hz, 440 Hz, 660 Hz, 880 Hz, 1100 Hz, 1320 Hz, and so on up to infinity. (Actually, the exact frequencies are a little different as we saw in Section 3.5.5.)

If I play another note which has a frequency that is relatively close to the fundamental of the 220 Hz I hear beating between the two fundamentals at a rate of $f_1 - f_2$. The same thing will happen if I play a note with a fundamental which is close to one of the harmonics of the 220 Hz (hence $2f_1 - f_2$ and $3f_1 - f_2$ and so on...).

For example, if I play a 220 Hz tone and a 445 Hz tone, I’ll hear beating, because the 445 Hz is only 5 Hz away from one of the harmonics of the 220 Hz tone. This will sound “dissonant” because, basically, we don’t like intervals that “beat.”

If I play a 440 Hz tone with the 220 Hz tone, there is no beating, because 440 Hz happens to be one of the harmonics of the 220 Hz tone. If there’s no beating then we think that it’s consonant.

Therefore, if I wanted to create a system of tuning the notes in a scale, I would do it using formulae for calculating the frequencies of the various notes that ensured that, at least for my most-used chords, the intervals in the chords would not beat. That is to say that when I play notes simultaneously, the various fundamentals have simple mathematical relationships.

3.15.2 Pythagorean Temperament

The philosophy described above resulted in a tuning system called Pythagorean Temperament which was based on only 2 intervals. The idea was that, the smaller the mathematical relationship between two frequencies, the better, so the two smallest possible relationships were used to create an entire system. Those frequency ratios were 1:2 and 2:3.

Exactly what does this mean? Well, essentially what we're saying when we talk about an interval of 1: 2 is that we have two frequencies where one is twice the other. For example, 100 Hz and 200 Hz. ($100:200 = 1:2$). As we have seen previously, this is known as an "octave."

The second ratio of 2:3 is what we nowadays call a "Perfect Fifth." In this interval, the two notes can be seen as harmonics of a common fundamental. Since the frequencies have a ratio of 2:3 we can consider them to be the 2nd and 3rd harmonics of another frequency 1 octave below the lower of the two notes.

Using only these two ratios (or intervals) and a little math, we can tune all the notes in a diatonic scale as follows:

If we start on a note which we'll call the "tonic" at, oh, 220 Hz, we tune another note to have a 2:3 ratio with it. 2:3 is the same as 220:330, so the second note will be 330 Hz.

We can then repeat this using the 330 Hz note and tuning another note with it... If we wish to stay within the same octave as our tonic, however, we'll have to tune up by a ratio of 2:3 and then transpose that down an octave (2:1). In other words, we'd go up a Perfect 5th and down an octave, creating a Major Second from the Tonic. Another way to do this is simply to tune down a Perfect Fourth (the inversion of a Perfect Fifth). This would result, however in a different ratio, being 3:4.

How did we arrive at this ratio? Well, we'll start by doing it empirically. If we started on a 330 Hz tone, and went up a Perfect Fifth to 495 Hz and then down an octave, we'd be at 247.5 Hz. $247.5:330 = 3:4$.

If we were to do this mathematically, we'd say that our first ratio is 2:3, the second ratio is 2:1. These two ratios must be multiplied (when we add intervals, we multiply the frequency ratios which multiply just like fractions) so $2:3 * 2:1 = 4:3$. Therefore, in order to go down a fourth, the ratio is 4:3, up is 3:4.

In order to tune the other notes in the diatonic scale, we keep going, up a Perfect 5th, down a Perfect Fourth (or up 2:3, down 4:3) and so on until we have done 3 ascending fifths and 2 descending fourths. (or, without octave transpositions, 5 ascending fifths). That gets us 6 of the 7 notes we require

(being, in chronological order, the Tonic, Fifth, Second, Sixth, and Third in a “Major” scale. The Fourth of the scale is achieved simply by calculating the 3:4 ratio with the Tonic.

It may be interesting to note that the first 5 notes we tune are the “common” pentatonic scale.

We could use the same system to get all of the chromatic notes by merely repeating the procedure for a total of 6 ascending fifths and 5 descending fourths (or 11 ascending fifths). This will present us with a problem, however.

If we do complete the system, starting at the tonic and calculating the other 11 notes in a chromatic scale, we’ll end up with what is known as one *wolf fifth*.

What’s a wolf fifth? Well, if we kept going around through the system until the 12th note, we ought to end up an octave from where we started – the problem is that we wind up a little too sharp. So, we tune the octave above the tonic to the tonic and put up with it being “out of tune” with the note a fifth below.

In fact, it’s wiser to put this this wolf fifth somewhere else in the scale – such as in an interval less likely to be played than the tonic and the fourth of the scale.

Let’s do a little math for a moment. If we go up a fifth and down a fourth and so on and so on through to the 12th note, we are actually doing the following equation:

$$f * \frac{3}{2} * \frac{3}{4} \quad (3.31)$$

If we calculated this we’d see that we get the equation

$$f * \frac{531441}{262144} \quad (3.32)$$

instead of the $f * \frac{2}{1}$ that we’d expect for an octave.

Question:

how far away from a “real” octave is the note you wind up with?

Answer:

well, if we transpose the $f * \frac{531441}{262144}$ note down an octave, we get $\frac{1}{2} * f * \frac{531441}{262144}$ or $f * \frac{531441}{524288}$.

In other words, the ratio between the tonic and the note which is an octave below the 12th note in the pythagorean tuning system is 531441 : 524288.

This amount of error is called the *Pythagorean Comma*.

Just for your information, according to the Grove dictionary, “Medieval theorists who discussed intervallic ratios nearly always did so in terms of Pythagorean intonation.” (look up “Pythagorean intonation”)

For an investigation of the relative sizes of intervals within the scale, look at Table 4.1 on page 86 of “The Science of Musical Sound” By Johann Sundberg.

3.15.3 Just Temperament

The Pythagorean system is pretty good if you’re playing music that has a lot of open fifths and fourths (as Pythagorean music seems to have had, according to Grout) but what if you like a good old major chord – how does that sound? Well, one way to decide is to look at the intervals. To achieve a major third above the tonic in Pythagorean temperament, we go up 2 fifths and down 2 fourths (although not in that order) therefore, the frequency is

$$f * \frac{3}{2} * \frac{3}{4} * \frac{3}{2} * \frac{3}{4} \quad (3.33)$$

or

$$f * \frac{81}{64} \quad (3.34)$$

This ratio of 81:64 doesn’t use very small numbers, and, in fact, we do hear considerably more beating than we’d like. So, once upon a time, a “better” tuning system was invented to accomodate people who liked intervals other than fourths and fifths.

The basic idea of this system, called *Just Temperament* (or *pure temperament* or *just intonation*) is that the notes in the diatonic scale have a relatively simple mathematical relationship with the tonic. In order to do this, you have to be fairly flexible about what you call “simple” (Pythagoras never used a bigger number than 4, remember) but the musical benefits outweigh the mathematical ones.

In just temperament we use the ratios shown in Table 3.3:

These are all fairly simple mathematical relationships, and note that the fourth and fifth are identical to that of the Pythagorean system.

This system has a much better intonation (i.e. less beating) when you plunk down a I, IV or V triad because the ratios of the frequencies of the

| Musical Interval | Frequency ratio of two notes |
|------------------|------------------------------|
| Tonic | 1:1 |
| 2nd | 9:8 |
| 3rd | 5:4 |
| 4th | 4:3 |
| 5th | 3:2 |
| 6th | 5:3 |
| 7th | 15:8 |
| Octave | 2:1 |

Table 3.3: A list of the ratios of frequencies of notes in a just temperament scale.

three notes in each of the chords have the simple relationships 4:5:6. (i.e. 100 Hz, 125 Hz and 150 Hz)

The system isn't perfect, however, because intervals that should be the same actually have different sizes. For example,

- The seconds between the tonic and the 2nd, the 4th and 5th, and the 6th and 7th have a ratio of 9:8
- The second between the 2nd and 3rd, and the 5th and 6th have a ratio of 10:9

The implications of this are intonation problems when you stray too far from your tonic key. Thus a keyboard instrument tuned using Just Intonation must be tuned for a specific key and you're not permitted to transpose or modulate without losing your audience or your sanity or both. Of course, the problem only occurs in instruments with fixed tunings for notes (such as keyboard and fretted string instruments). Everyone else can compensate on the fly.

The major third and the tonic in Just Intonation have a frequency ratio of 5:4. The major third and the tonic in Pythagorean Intonation have a ratio of 64:81. The difference between these two intervals is

$$64:81 - 5:4 = 80:81$$

This is the amount of error in the major third in Pythagorean Tuning and is called the *syntonic comma*.

3.15.4 Meantone Temperaments

Meantone Temperaments were attempts to make the Just system “better.” The idea was that you fudge a little with a couple of notes in the scale to make different keys more palatable on the instrument. There were a number of attempts at this by people like Werckmeister, Valotti, Ramos, and de Caus, each with a different system that attempted to improve on the other. (Although Werckmeister was quite happy to use another system called “Equal Temperament”)

These are well discussed in chapters 18.3 – 18.5 of the book “Music Acoustics” by Donald Hall

3.15.5 Equal Temperament

The problem with all of the above systems is that an instrument tuned using either of them was limited to the number of keys it could play in. This is simply because of the problem mentioned with Just Intonation, different intervals that should be the same size, are not.

The simple solution to this problem is to decide on a single size for the smallest possible interval (in our case, the semitone) and just multiply to get all other intervals.

This was probably first suggested by a Sicilian abbot named Girolamo Roselli who, in turn convinced Frescobaldi to use it (but not without the “aid of ‘frequent and gratuitous beverages’” (Grove Dictionary under “Temperaments”))

So, we have to divide an octave into semitones, or 12 equal divisions. How do we do this? Well, we need to find a ratio that can multiply by itself 12 times to arrive at twice the number we started with. (i.e. adding 12 semitones gives you an octave.) How do we do this?

We know that $3 \times 3 = 9$ so we say that $3^2 = 9$ therefore, going backwards, “the square root of $9 = 3$.” What this final equation says is “the number, when multiplied by itself once gives 9 is 3.” We want to find the number which, when multiplied by itself 12 times equals 2 (since we’re going up an octave) so, the number we’re looking for is “the twelfth root of 2.” (or about 1:1.06) That’s it.

Therefore, given any frequency f , the note one semitone up is $f \times$ the 12th root of 2. If you want to go up 2 semitones, you have to multiply by the 12th root of 2 twice or :

$$f \times \sqrt[12]{2} \times \sqrt[12]{2} \tag{3.35}$$

or

$$f * \sqrt[12]{2}^2 \quad (3.36)$$

So, in order to go up any number of semitones, we simply do the following equation :

$$f * \sqrt[12]{2}^x \quad (3.37)$$

where x is the number of semitones

The advantage of this is that you can play in any key on one instrument. The disadvantage is that every key is “out of tune.” But, they’re all equally out of tune, so we have gotten used to it. Sort of like we have gotten used to eating fast food, despite the fact that it tastes pretty wretched, sometimes even bordering on rancid...

To get an intuitive idea of the fact that equal temperament intervals are out of tune, even if they don’t sound like it most of the time, take a look at Figures 3.60 and 3.61

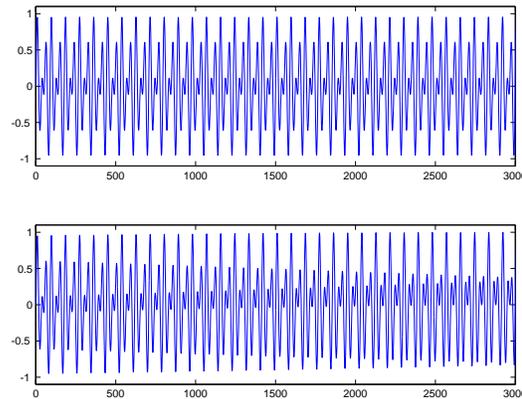


Figure 3.60: The time response of a perfect fifth played with sine waves. In both cases, the root is 1 kHz. The X axis shows time in samples at a 44.1 kHz sampling rate. The top plot shows a just temperament perfect fifth, with the frequencies 1 kHz and 1.5 kHz. The bottom plot shows an equal temperament fifth, with the frequencies 1 kHz and 1.49830707687668 kHz. Notice that the bottom plot modulates in time. If I had plotted more time, it would be evident that the modulation is periodic.

3.15.6 Cents

There are some people who want a better way of dividing up the octave. Basically, some people just aren’t satisfied with 12 equal divisions, so they

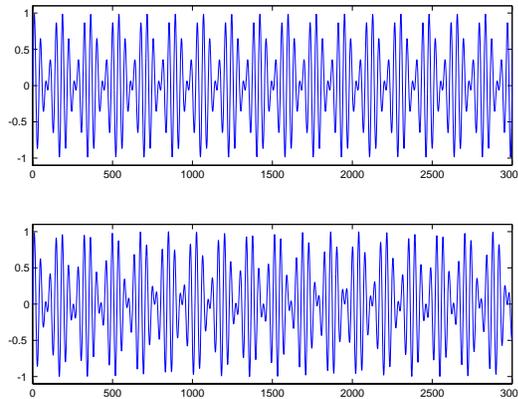


Figure 3.61: The time response of a major third played with sine waves. In both cases, the root is 1 kHz. The X axis shows time in samples at a 44.1 kHz sampling rate. The top plot shows a just temperament major third, with the frequencies 1 kHz and 1.25 kHz. The bottom plot shows an equal temperament third, with the frequencies 1 kHz and 1.2592104989487 kHz. Notice that the bottom plot modulates in time. If I had plotted more time, it would be evident that the modulation is periodic.

divided up the semitone into 100 equal parts and called them *cents*. Since a cent is 1/100 of a semitone, it's an interval which, when multiplied by itself 1200 times (12 semitones 100 cents) makes an octave, therefore the interval is the 1200th root of 2 (or 1:1.00058).

Therefore, 1 cent above 440 Hz is

$$440 * \sqrt[1200]{2} = 440.254Hz. \quad (3.38)$$

We can use cents to compare tuning systems. Remember that 100 cents is 1 semitone, 200 cents is 2 semitones, and so on.

There is a good comparison in cents of various tuning systems in “Musical Acoustics” by Donald Hall. (p. 453)

3.15.7 Suggested Reading List

Tuning and Temperament, A Historical Survey Barbour, J. M.

3.16 Room Acoustics – Reflections, Resonance and Reverberation

Let’s ignore all the stuff we talked about in Section 3.2 for a minute. We’ll pretend that we didn’t talk about acoustical impedance and that we didn’t look at all sorts of nasty-looking equations. We’ll pretend for this chapter, when it comes to a sound wave reflecting off a wall, that all we know is Snell’s Law described in Section 3.3.1. So, for now, we’ll think that all walls are mirrors and that the angle of incidence equals the angle of reflection.

3.16.1 Early Reflections

Let’s consider a rectangular room with one sound source and you in it as is shown in Figure 3.62. We’ll also ignore the fact that the room has a ceiling or a floor... One very convenient way to consider the way sound moves in this room is to not try to understand everything about it – but instead to be a complete egoist and ask “what sound gets to me?”

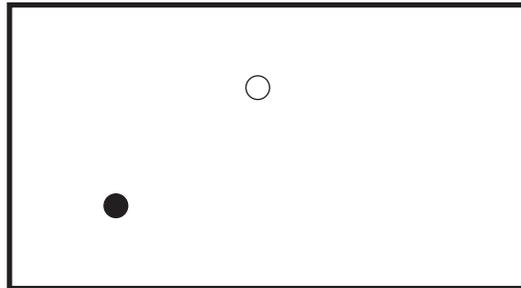


Figure 3.62: A sound source (black dot) and a listener (white dot) in a room (Black rectangle)

As we discussed earlier, when the source makes a sound, the wavefront expands spherically in all directions, but we’re not thinking that way at the moment. All you care about is you, so we can say that the sound travels from the source to you in a straight line. It takes a little while for the sound to get to you, according to the speed of sound and the distance between you and the source. This sound that you get first is usually called the *direct sound*, because it travels directly from the sound source to you. This path that the wave travels is shown in Figure 3.63 as a straight red line. In addition, if you were replaced by an omnidirectional microphone, we could think of the impulse response (explained in Section ??) as being a single copy of the sound arriving a little later than when it was emitted, and slightly lower in

level because the sound had to travel some distance. This impulse response is shown in Figure 3.64.

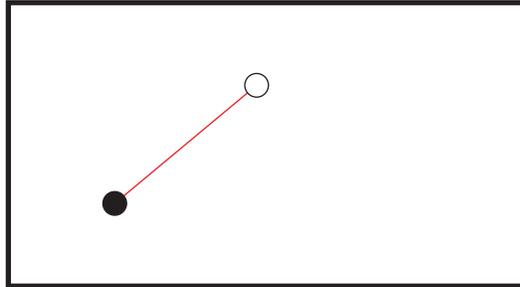


Figure 3.63: The direct sound (red line) travelling from the source to the receiver.

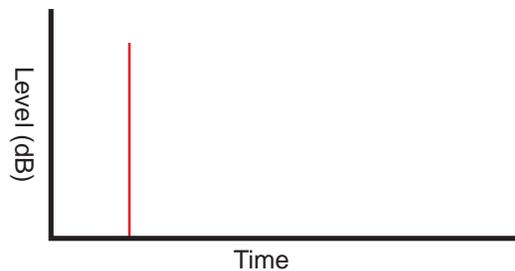


Figure 3.64: The impulse response of the direct sound.

Of course, the sound is really travelling out in all directions, which means that a lot of it is heading towards the walls of the room instead of heading towards you. As a result, there is a ray of sound that travels from the sound source, bounces off of one wall (remember Snell's Law) and comes straight to you. Of course, this will happen with all four walls – a single reflection from each wall reaches you a little while after the direct sound and probably at different times according to the distance travelled. These are called *first-order reflections* because they contain only a single bounce off a surface. They're shown as the blue lines in Figure 3.65 with the impulse response shown in Figure 3.66.

We also get situations where the sound wave bounces off two different walls before the sound reaches you, resulting in second-order reflections. In our perfectly rectangular room, there will be two types of these second-order reflections. In the first, the two walls that are reflecting are parallel

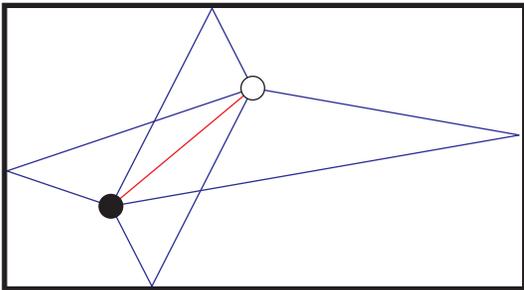


Figure 3.65: The first-order reflections (blue lines) travelling from the source to the receiver.

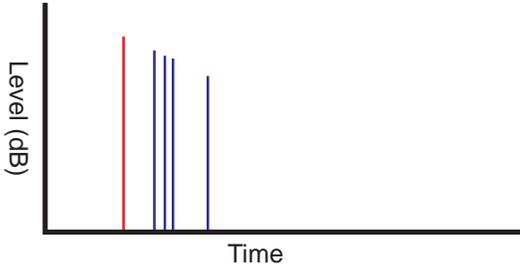


Figure 3.66: The impulse response of the first-order reflections (blue lines).

and opposite to each other. In the second, the two walls are adjacent and perpendicular. These are shown as the green lines in Figure 3.67 and the impulse response in Figure 3.68. Note in the impulse response that it's possible for a second-order reflection to arrive earlier than a first-order reflection, particularly if you are in a long rectangular room. For example, if you're sitting in the front row of a big concert hall, it's possible that you get a second-order reflection off the stage and side walls before you get a first-order reflection off the wall in the back behind the audience. The moral of the story here is that the order of reflection is only a general indicator of its order of arrival.

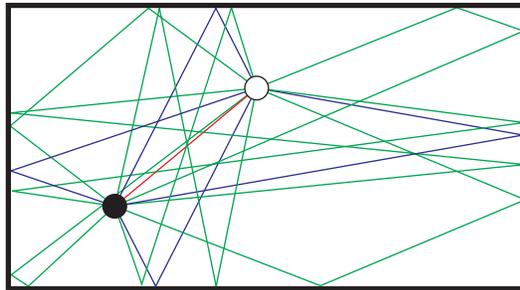


Figure 3.67: The second-order reflections (green lines) travelling from the source to the receiver.

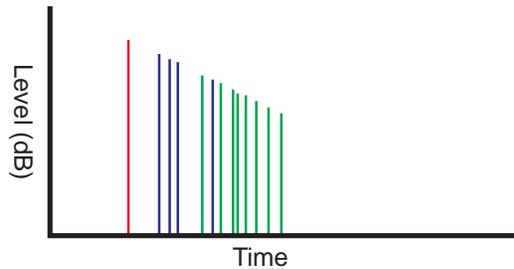


Figure 3.68: The impulse response of the second-order reflections (green lines).

3.16.2 Reverberation

If the walls were perfect reflectors and there was no such thing as sound absorption in air, this series of more and more reflections would continue forever. However, there is a little energy in the sound wave lost in the air,

and in the wall, so eventually, the reflections get quieter and quieter as they reach a higher and higher order until eventually, there is nothing.

Let's say that your sound source is a person clapping their hands once – a sound with a very fast attack and decay. The first thing you hear is the direct sound, then the early reflections. These are probably separated enough in time and space that your brain can interpret them as separate events. Be careful about what I mean by this previous sentence. I do not necessarily mean that you will hear the direct and earlier reflections as separate hand claps (although if the room is big enough you might...) Instead, I mean that your brain uses these discrete components in the sound that arrives at the listening position to determine a bunch of information about the sound source and the room. We'll talk about that more later.

If we consider higher and higher orders of reflections, then we get more and more reflections per second as time goes by. For example, in our rectangular, two-dimensional room, there are 4 first-order reflections, 8 second-order reflections, 12 third-order reflections and so on and so on. These will pile up on each other very quickly and just become a complete mess of sound that apparently comes from almost everywhere all at the same time (actually, you will start to approach a diffuse field situation). When the reflections get very dense, we typically call the collection of all of them *reverberation* or *reverb*. Essentially, reverberation is what you have when there are too many reflections to think about. So, instead of trying to calculate every single reflection coming in from every direction at every time, we just give up and start talking about the *statistical properties* of the room's acoustics. So, you won't hear about a 57th order reflection coming in a predicted time. Instead, you'll hear about the probability of a reflection coming from a certain direction at a given time. (This is sort of the same as trying to predict the weather. Nobody will tell you that it will definitely rain tomorrow starting at 2:34 in the afternoon. Instead, they'll say that there is a 70% chance of rain. Hiding behind statistics helps you to avoid being incorrect...)

One immediately obvious thing about reverberation in a real room is that it takes a little while for it to die away or decay. So then the question is, how do we measure the reverberation time? Well, typically we have to oversimplify everything we do in audio, so one way to oversimplify this measurement is to just worry about one frequency. What we'll do is to get a loudspeaker that's emitting a sine tone with a constant level – therefore, just a single frequency. Also, we'll put a microphone somewhere in the room and look at its output level on a decibel scale. If we leave the speaker on for a while, the sound pressure level at the microphone will stabilize and stay constant. Then we turn off the sine wave, and the reverberation will decay

down to nothing. Interestingly, if the room is behaving according to theory, if we plot the output of the microphone over time on a decibel scale, then the decay of the reverberation will be a straight line as is shown in Figure 3.69.

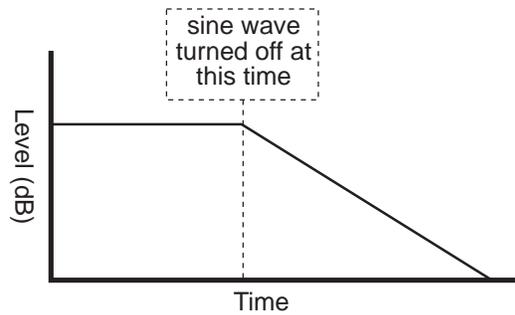


Figure 3.69: Sound pressure level vs. time for a single sine wave in a room. The sine wave was turned on once upon a time – long ago enough that the SPL in the room has stabilized at the microphone’s position. Note that, when the tone is turned off, the decay in the room is linear (a straight line) on a decibel scale.

The amount of time it takes the reverberation to decay a total of 60 decibels is what we call the *reverberation time* of the room, abbreviated RT_{60} .

Sabine Equation

Once upon a time (actually, around the year 1900), a guy named *Wallace Clement Sabine* did some experiments and some math and figured out that we can arrive at an equation to predict the reverberation time of a room if we know a couple of things about it.

Let’s consider that the more absorptive the surfaces in the room, the more energy we lose in the walls, so the faster the reverberation will decay. Also, the more surface area there is (i.e. the bigger the walls, floor and ceiling) the more area there is to absorb sound, therefore the reverb will decay faster. So, the average absorption coefficient (see Section ??) and the surface area will be inversely proportional to the reverb time.

Also consider, however, that the bigger the room, the longer the sound will travel before it hits anything to reflect (or absorb) it. Therefore the bigger the room volume, the longer the reverb time.

Thinking about these three issues, and after doing some experiments with a stopwatch, Sabine settled on Equation 3.39:

$$RT_{60} = \frac{55.26V}{Ac} \quad (3.39)$$

Where c is the speed of sound in the room and A is the total sound absorption by the room which can be calculated using Equation 3.40.

$$A = S\bar{\alpha} \quad (3.40)$$

Where S is the total surface area of the room and $\bar{\alpha}$ is the “average value of the statistical absorption coefficient.” [Morfey, 2001]

Eyring Equation

There is second famous equation that is used to calculate the reverberation time of a room, developed by C. F. Eyring in 1930. This equation is very similar to the Sabine equation, in fact you can use Equation 3.39 as the main equation. You just have to change the value of A using Equation 3.41:

$$A = S \ln \left(\frac{1}{1 - \bar{\alpha}} \right) \quad (3.41)$$

Note that some books will call this the *Norris-Eyring Equation* [Morfey, 2001].

3.16.3 Resonance and Room Modes

I lied. All of the description of reflections and reverberation that I talked about above only applies to high frequencies. Low frequencies behave completely differently in a room. Remember back to Section ?? on woodwind instruments that, as soon as you have a pipe that is closed on both ends, it will have a natural resonance at a frequency that is determined by its length. Since the pipe is closed on both ends, then the fundamental resonant frequency has a wavelength equal to twice the length of the pipe. All we need to do to make the pipe resonate at that frequency and its harmonics is to put a sound source in there that has any energy at the resonant frequencies. Since, as we saw in Section ??, an impulse contains all frequencies, if we snap our fingers inside the pipe, it will ring.

Now, let’s change the scale a bit and consider a closed pipe the length of a room. This pipe will still resonate at its fundamental frequency and its harmonics, but these will be very low because the pipe will be very long.

If we wanted to calculate the fundamental resonant frequency of this pipe of length L (equal to the Length of the room), we just need to find the frequency with a wavelength of $2L$ as is shown in Equation 3.42

$$f = \frac{c}{2L} \quad (3.42)$$

We also know that the harmonics of this frequency f will resonate as well. These are easy to calculate by just multiplying f by an integer – so the first harmonic will be $1f$, the second harmonic will be $2f$ and so on.

It's worth your while to compare Equation 3.42 to Equation 3.27 back in the section on resonance in closed pipes. You'll notice that both equations are the same.

Axial (One-dimensional) Modes

The interesting thing is that a room behaves in exactly the same way. We can think of a room with a length of L as a closed pipe with a length of L . In fact, I'm not even saying that a room is like a closed pipe – I'm saying that a room *is* a closed pipe. Therefore the room will resonate at a frequency with a wavelength of $2L$ and its harmonics. This can be calculated using Equation 3.43 which is exactly the same as Equation ??, but written slightly differently, and with a p added for the harmonic number.

$$f = \frac{c p}{2L} \quad (3.43)$$

This behaviour doesn't just apply to the length of the room. It also applies to the width and length – therefore the room is actually behaving as three pipes of lengths L , W and H (for Length, Width and Height) at the same time. These three fundamental frequencies (and their harmonics) will all resonate independently of each other.

There are a couple of interesting things to discuss when it comes to axial (one-dimensional) room modes.

Firstly, just as with the resonating closed pipe, there is the issue of the relationship between the particle pressure and the particle velocity. If we look at the particles adjacent to the wall, we see that these molecules cannot move, therefore the amplitude of their velocity wave is 0, and the amplitude of the pressure wave is at its maximum. Conversely, at the centre of the room, the amplitude of the velocity wave is maximum, and the amplitude of the pressure wave is 0.

FIGURE HERE?

Secondly, there's the issue of phase. Remember back to the discussion on closed pipes that we can consider them to be waveguides. This means that the sound energy at one end of the pipe gets out the other end without being attenuated because it can't expand in three dimensions – it can only travel in one. Also, it means that the pressure wave is in phase at any cross section of the pipe. At the resonant frequency of the room, this is also true. If you could freeze time and actually see the pressure wave in the room, you would see that the pressure wave at the resonant frequency is in phase across the room. So, if you have a loudspeaker in one corner of the room playing a sine wave at the fundamental resonance of the length of the room, then the sound wave is not expanding outwards as a sphere from the loudspeaker. Instead, it's travelling back and forth in the room as a plane wave.

FIGURE HERE?

Tangential (Two-dimensional) Modes

The axial room modes can be thought of in exactly the same way as a standing wave in a pipe or on a string. In all of these cases, the resonance is limited to a single dimension. However, a room has more than one dimension. There is also the issue of resonances in two-dimensions, known as *tangential* room modes. These behave in exactly the same way as a rectangular resonating plate (assuming of course that we're talking about a rectangular room).

FINISH THIS OFF

$$f = \frac{c}{2} \sqrt{\left(\frac{p}{L}\right)^2 + \left(\frac{q}{W}\right)^2} \quad (3.44)$$

FINISH THIS OFF

Oblique (Three-dimensional) Modes

FINISH THIS OFF

$$f = \frac{c}{2} \sqrt{\left(\frac{p}{L}\right)^2 + \left(\frac{q}{W}\right)^2 + \left(\frac{r}{H}\right)^2} \quad (3.45)$$

Equation 3.45 can also be used as the master equation for calculating any type of room mode – axial, tangential or oblique. For example, let's say that you wanted to calculate the 2nd harmonic of the axial mode for the width of the room. You set q to equal 2 and set p and r to 0. This winds up making Equation 3.45 exactly the same as Equation 3.43 because the 0's make the L and H components go away.

Coupling

sound source couples to the mode

FINISH THIS OFF

We also have to consider how well the mode couples to the receiver.

FINISH THIS OFF

Modal Density and Overlap

FINISH THIS OFF

3.16.4 Schroeder frequency

So, now we've seen that, at high frequencies, we worry about reflections and statistical behaviour of the room. At low frequencies, we worry about room modes, since they'll ring longer than the reverberation and be the dominant characteristic in the room's behaviour. The question that should be sitting in the back of your head is "what's the crossover frequency between the low and the high?"

This frequency is known as the *Schroeder frequency* or the *Schroeder large-room frequency* and is defined as the frequency where the modes start bunching so closely together that they no longer are seen as resonant peaks. In most definitions, you'll see people talking about *modal density* which is just a measure of the number of resonant modes in a given bandwidth. (This number increases with frequency.)

As a result, a room can be considered in the same way that we think of two-way speakers – the low end has a modal behaviour, the high end is statistical and the crossover is at the Schroeder frequency. This frequency can be calculated using Equation 3.46.

$$f_{min} \approx c \frac{4}{\sqrt{\pi A}} - \frac{S}{16V} \quad (3.46)$$

where f_{min} is the Schroeder frequency, A is the room absorption calculated using Equation 3.40, S is the surface area of the boundaries and V is the room's volume.

3.16.5 Room Radius (aka Critical Distance)

If you go way back in this book, you'll remember that we mentioned that, for every doubling of distance from a sound source, you get a 6 dB drop in sound pressure level. This rule is true, but only if you're in a free field like

an anechoic chamber or at the top of a pole outdoors. What happens when you're in a real room? This is where things get a little strange.

If you're close to the sound source, then the direct sound is very loud compared to all of the reflections, reverberation and modes. As a result, you can sort of pretend that you're in a free field, so, as you move away from the source, you lose 6 dB per doubling of distance, as if you were normal. However, if you're far away from the sound source, the total summed power coming into your SPL meter from the reflections, reverberation and room modes is greater than the direct sound. When this is the case, then no matter where you go in the room, you'll get the same reading on your SPL meter.

Of course, this means that there must be some middle distance where the direct sound's level is equal to the reflected sound pressure level. This distance is known as the *room radius* or the *critical distance*.

This is a fairly easy thing to measure. Send a sine tone out of a loudspeaker and measure its level with an SPL meter while you're standing fairly close to it. Back away from the speaker and keep looking at your SPL meter. It should drop as you get farther and farther away, but it will reach some minimum level and not drop below that, no matter how much farther away you get. The distance from the loudspeaker at which this minimum value is reached is the room radius.

Of course, if you don't want to be bothered to measure this, you could always calculate it using Equation 3.47 []. Notice that the value is dependent on the volume of the room, V and the reverberation time.

$$r_h = 0.1 \sqrt{\frac{V}{\pi RT_{60}}} \quad (3.47)$$

3.16.6 Reading List

3.17 Sound Transmission

I've lived most of my adult life in apartment buildings in Montreal and Ottawa. In 12 years, I lived in 10 different apartments in those two cities, and as a result, I feel that I am a qualified expert in sound proofing... or, more to the point, the lack of it.

DEFINE SOUND TRANSMISSION INDEX

3.17.1 Airborne sound

The easiest way to introduce airborne sound transmission is to break it up into high frequency and low frequency behaviour. This is because, in most situations, these two frequency bands are transmitted very differently between two spaces.

High-frequency Transmission

NOT WRITTEN YET

Low-frequency Transmission

NOT WRITTEN YET

3.17.2 Impact noise

NOT WRITTEN YET

3.17.3 Reading List

3.18 Desirable Characteristics of Rooms and Halls

NOT YET WRITTEN

3.18.1 Reading List

Acoustical Designing in Architecture Knudsen, V. O. and Harris, C. M. John Wiley and Sons, Inc. (1950)

Acoustics, Noise and Buildings Parkin, P. H. and Humphreys, H. R. Faber and Faber Ltd (1958)

Architectural Acoustics Knudsen, V. O. John Wiley and Sons, Inc. (1932)

Music, Acoustics and Architecture Beranek, L. L. John Wiley and Sons, Inc. (1962)

Architectural Acoustics Knudsen, V. O. Scientific American November 1963

Chapter 4

Psychoacoustics and perception

4.1 What's the difference?

NOT YET WRITTEN

4.2 How your ears work

NOT YET WRITTEN

4.3 Human Response Characteristics

4.3.1 Introduction

Our ability to perceive things using any of our senses is limited by two things:

- physical limitations and
- the brain's ability to process the information.

Physical limitations determine the absolute boundaries of range for things like hearing and sight. For example, there is some maximum frequency (which may or may not be something about 20 kHz, depending on who you ask and how old your are...) above which we cannot hear sound. This ceiling is set by the physical construction of the ear and its internal components.

The brain's ability to process information is a little tougher to analyze. For example, we'll talk about a thing called "psychoacoustic masking" which basically says that if you are presented with a loud sound and a soft sound simultaneously, you won't "hear" the soft sound (for example, if I whisper something to you at a Motorhead concert, chances are you won't know that I exist...). Your ear is actually receiving both sounds, but your brain is ignoring one of them.

4.3.2 Frequency Range

We said earlier that the limits on human frequency response are about 20 Hz in the low frequency range and 20 kHz in the upper end. This has been disputed recently by some people who say that, even though tests show that you cannot hear a single sine wave at, say, 25 kHz, you are able to perceive the effects a 25 kHz harmonic would have on the timbre of a violin. This subject provides audio people with something to argue about when they've agreed about everything else...

4.3.3 Dynamic Range

The dynamic range of your hearing is determined by two limits called the *threshold of hearing* and the *threshold of pain*.

The threshold of hearing is the quietest sound that you are able to hear, specified at 1 kHz. This value is $20 \mu\text{Pa}$ or $20 * 10^{-6}$ Pascals. Just to give you an idea of how quiet this is, the sound of blood rushing in your head is louder to you than this. Also, at $20 \mu\text{Pa}$ of sound pressure level, the hair cells

inside your inner ear are moving back and forth with a total peak-to-peak displacement that is less than the diameter of a hydrogen atom \square .

Note that the reference for calculating sound pressure level in dBspl is $20 \mu\text{Pa}$, therefore, a 1 kHz sine tone at the threshold of hearing has a level of 0 dBspl.

One important thing to remember is that the threshold of hearing is not the same sound pressure level at all frequencies, but we'll talk about this later.

The threshold of pain is a sound pressure level that is so loud that it causes you to be in pain. This level is somewhere around 200 Pa, depending on which book you read and how masochistic you are. This means that the threshold of pain is around 140 dBspl. This is very loud.

So, based on these two numbers, we can calculate that the human hearing system has a total dynamic range of about 140 dB.

4.4 Loudness

4.4.1 Equal loudness contours

Back in 1933, a couple of researchers by the name of Fletcher and Munson decided to gather some information about how we perceive different frequencies at different amplitudes. What they came up with was a bunch of lines we now call “Equal Loudness Contours” or the “Fletcher and Munson Curves”.

These curves indicate two important pieces of information.

1. the threshold of hearing at different frequencies
2. the apparent levels of equal loudness at different frequencies

Threshold of hearing at different frequencies

We said earlier that the softest sound that we are able to hear is equivalent to a pressure variation of $20 \mu \text{ Pa}$. This is true, but only at one frequency (1 kHz). If we change the frequency, we would find that we are less sensitive to soft sounds and therefore require a higher amplitude to hear the tone. The bottom line of the equal loudness contours indicates the amplitude (in dBspl) we would require to hear a sound in the absence of all other sound. Although this plot has a specific shape, there are two basic trends to point out.

We are less sensitive to high frequencies and low frequencies than to mid-range frequencies

The frequency band in which our hearing is most sensitive is approximately 1 kHz – 5 kHz. This is an interesting area for two reasons. Firstly, the bulk of our speech (specifically consonant sounds) relies on information in this frequency range (although it’s like that the speech evolved to capitalize on the sensitive frequency range). Secondly, the anthropologically-minded will be interested to note that the sound of a snapping twig has lots of information which is smack in the middle of the 1 kHz – 5 kHz range. This is a useful characteristic when you look like lunch to a large-toothed animal that’s sneaking up behind you...

Apparent levels of equal loudness at different frequencies

If I were to play you sine waves at 100 Hz and 1 kHz and then I were to ask you to set the loudness of the two tones so that they were the same, you would set the lower pitch to be louder (if we were to measure it in dBspl)

than the higher pitch. The same effect would happen if we tried it with a 1 kHz tone and a 10 kHz tone, except that the 10 kHz would now be the louder of the two (even though they sound the same to you). Again, there are two interesting things to note about this effect.

The curve of equal loudness has virtually the same shape as the absolute threshold curve, even at other amplitudes

The curve tends to flatten out when the volume goes up. What does this mean? Firstly, when you turn down the stereo, you are less sensitive to low and high frequencies (compared to the mid-range frequencies) than when the stereo was turned up. Therefore the balance changes (particularly in the low end). If the level is low, then you'll think that you hear less bass. This is why there's a LOUDNESS switch on your stereo. It boosts the bass to compensate for your low-level Fletcher and Munson curves... Secondly, things sound better when they're louder. This is because there's a "better balance" in your hearing perception than when they're at a lower level. This is why the salesperson at the stereo store will crank up the volume when you're buying speakers... they sound good that way... everything does.

4.4.2 Phons (also known as "Phones")

Most of the world measures things in dBspl. This is a good thing because it's a valid measurement of pressure referenced to some fixed amount of pressure. As Fletcher and Munson discovered, though, those numbers have fairly little to do with how loud things sound to us... So someone decided that it would be a really good idea to come up with a system which was related to dBspl but "fixed" according to how we perceive things.

The system they came up with (which is virtually never used – but you'll stumble on it in old textbooks) measures the amplitude of sounds in something called phons.

Here's how to find a value in phons for a given measured amplitude.

1. Measure the amplitude of the sound in dBspl
2. Check the frequency of the sound in Hz
3. Plot the intersection of the two values on the chart of the Fletcher and Munson curves
4. Find the nearest curve contour and check what the value of that curve is at 1 kHz
5. That value is the amplitude of the sound in phons

The idea is that all sounds along a single Fletcher and Munson contour have the same apparent loudness level, and therefore are given the same value in phons.

Note: I got an email from Bert Noeth, a professor teaching sound and acoustics in Belgium who tells me that, in Europe, the lines of equal intensity are called “isophons.”

4.4.3 Sones

There is another frequency-dependent amplitude measurement called the Sone – but you’ll never see it except in definitions and psychoacoustics tests, so I’ll just say “they exist” and if you want more information, check out the textbook. We won’t speak of them again...

4.5 Weighting Curves

Let's say that you're hired to go measure the level of background noise in an office building. So, you wait until everyone has gone home, you set up your brand-new and very expensive sound pressure level meter and you'll find out that the noise level is really high – something like 90 dBspl or more.

This is a very strange number, because it doesn't *sound* like the background noise is 90 dBspl... so why is the meter giving you such a high number? The answer lies in the quality of your meter's microphone. Basically, the meter can hear better than you can – particularly at very low frequencies. You see, the air conditioning system in an office building makes a lot of noise at very low frequencies, but as we saw earlier, you aren't very good at hearing very low frequencies.

The result is that the sound pressure level meter is giving you a very accurate reading, but it's pretty useless at representing what you hear. So, how do we fix this? Easy! We just make the meter's "hearing" as bad as yours.

So, what we have to do is to introduce a filter in between the microphone output and the measuring part of the meter. This filter should simulate your hearing abilities.

There's a problem, however. As we saw in Section 4.4.1, the "EQ curve" of your hearing changes with level. Remember, the louder the sound, the flatter your personal frequency response. This means that we're going to need a different filter in our sound pressure level meter according to the sound pressure of the signal that we're measuring.

The filter that we use to simulate human hearing is called a *weighting filter* because it applies different weights (or levels of importance) to different frequencies. The frequency response characteristics of the filter is usually called a *weighting curve*.

There are three standard weighting curves, although we typically only use two of them in most situations. These three curves are shown in Figure ?? and are called the *A-weighting*, *B-weighting*, and *C-weighting* curves.

SHOW WEIGHTING CURVES FREQUENCY RESPONSES

As can be seen in Figure ??, the A-weighting curve has the most attenuation in the low and high frequency bands. This corresponds to your naturally poor hearing abilities at low sound pressure levels. Therefore, you should use the A-weighting curve when you are measuring signals with a sound pressure level of XXX dBspl or less.

FINISH OFF THE DESCRIPTIONS OF THE C and B WEIGHTING CURVES.

There is an important lesson to learn here – beware! Notice that the A-weighting curve has a great deal of attenuation in the low and high frequencies. Therefore, if you have a piece of equipment that is noisy, and you want to make its specifications look better than they really are, you can use an A-weighting curve to reduce the noise level. Manufacturers who want to make their gear have better specifications will use an A-weighting curve to improve the looks of their noise specifications.

You may also see instances where people use an A-weighting curve to measure acoustical noise floors even when the sound pressure level of the noise is much higher than XXX dBspl. This really doesn't make much sense, since the frequency response of your hearing is better than an A-weighting filter at higher levels. Again, this is used to make you believe that things aren't as loud as they appear to be.

4.6 Masking

Let's go back to the bottom curve on the Fletcher and Munson graphs. This contour tells us the absolute minimum of our abilities to perceive sound. If we measure a sound to have a pressure that, when plotted on the same graph is below the line, then we can't hear it. Sometime after Fletcher and Munson, however, there was a discovery that this curve was not an absolute. In fact, it moves around quite a bit depending on what you're hearing at the time. It tends to move up and surround sounds that you are hearing. I'll explain.

If you play a relatively loud 1 kHz sine wave, the spectrum of the tone will be plotted on a Frequency vs. Amplitude graph as a vertical line.

If you then play another sine tone (in addition to the first) with a close frequency to 1 kHz but a comparatively low amplitude, you won't hear it.

If you then raise the amplitude of the second tone until you can hear it, plot that point on a graph, and repeat the process with other nearby frequencies, you'll wind up with a graph that looks like this:

This is telling us that, if I play a sine tone at 1 kHz with an amplitude of about 60 dBspl, I will not be able to hear any simultaneously sounding tone which, when plotted on the same graph, is below the dotted line.

This is the simultaneous masking curve. Others can be plotted for forwards and backward masking which occur when the two tones happen at different times.

4.7 Localization

Lord Rayleigh to Wightman and Kistler

How do you localize sound? For example, if you close your eyes and you hear a sound and you point and say “the sound came from over there,” how did you know? And, how good are you at it?

Well, you have two general things to sort out

1. which direction is the sound coming from?
2. how far away is it?

We determine the direction of the sound using a couple of basic components that rely on the fact that we have two ears

1. Which ear got the sound first?
2. Which ear is the sound louder in?
3. What is the timbre of the sound
4. How does it change if I turn my head a little bit?

The first thing you rely on is the interaural (a fancy word meaning “between the two ears” give or take) time of arrival of the sound. If the right ear hears the sound first, then the sound is on your right, if the left ear hears the sound first, then the sound is on your left. If the two ears get the sound simultaneously, then the sound is directly ahead, or above, or below or behind you.

The next thing you rely on is the interaural amplitude difference. If the sound is louder in the right ear, then the sound is probably on your right. Interestingly, if the sound is louder in your right ear, but arrives in your left ear first, then your brain decides that the interaural time of arrival is the more important cue and basically ignores the amplitude information.

You also have these things sticking out of your head which most people call their ears but are really called your pinnae (1 pinna, 2 pinnae). These things bounce sound around inside them differently depending on which direction the sound is coming from. They tend to block really high frequencies coming from the rear (because they stick out a bit...) so rear sources sound “darker” than front sources. These are of a little more help when you turn your head back and forth a bit (which you do involuntarily anyway...)

Once upon a time, a guy named Lord Rayleigh wrote a book called “The Theory of Sound.” He said that the brain uses the phase difference of low frequencies to sort out where things are, whereas for high frequencies, the amplitude differences between the two ears are used. This is a pretty good estimation, although there’s a couple of people by the names of Wightman and Kistler in the States working for NASA doing a lot of research in the matter.

Research in this field is a big thing these days because of all the companies trying to make virtual reality machines.

4.7.1 Cone of Confusion

Exactly how good are you at the localization of sound sources? Well, if the sound is in front of you, you’ll be accurate to within about 2° or so. If the sound is directly to one side, you’ll be up to about 7° off, and if the sound is above you, you’re really bad at it... typical errors for localizing sources above your head are about $14^\circ - 20^\circ$.

Why? Well, probably because, anthropologically speaking, more stuff was attacking from the ground than from the sky. You’re better equipped if you can hear where the sabre-toothed tiger is rather than the large now-extinct human-eating flying penguin...

If you play a sound for someone on one side of their head and asked them to point at the direction of the sound source, they would typically point in the wrong direction, but be pointing in the correct angle. That is, if the sound is 5° front of centre, some people will point 5° rear of centre, Some people will point 5° above centre and so on. If you made a diagram of the incorrect (and correct) guesses, you’d wind up drawing a cone sticking out of the test subject’s ear. This is called the “cone of confusion.”

4.8 Precedence Effect

Stand about, oh, 10 m from a large flat wall outdoors and clap your hands. You should hear an echo. It'll be pretty close behind the clap (in fact it ought to be about 60 ms behind the clap...) but it's late enough that you can hear it. Now, go closer to the wall, and keep clapping as you go. How close will you be when you can no longer hear a distinct echo? (not one that goes "echo.....echo" but one that at least appears to have a sound coming from the wall which is separate from the one at your hands...)

It turns out that you'll be about 5 m away. You see, there is this small time window of about 30 ms or so where, if the echo is outside the window, you hear it as a distinct sound all to itself; if the echo is within the window, you hear it as a component of the direct sound.

It also turns out that you have a predisposition to hear the direction of the combined sound as originating near the earlier of the two sounds when they're less than 30 ms apart.

This localization tendency is called the *precedence effect* or the *Haas effect* [].

Okay, okay so I've oversimplified a bit. Let's be a little more specific. The actual time window is dependent on the signal. For very transient sounds, the window is much shorter than for sustained sounds. For example, you're more likely to hear a distinct echo of a xylophone than a cello. (And, in case you're wondering, a duck's quack does echo...) So the upper limit of the Haas window is between 5 ms and 40 ms, depending on the signal.

Also, if the delay time is very short (less than about 1 ms), then you will perceive the reflection as having a timbral effect known as a *comb filter* (explained in Section ??).

In between these two times, the precedence effect is in action, and its principal function is to aid you in determining either the distance to the reflecting surface, the distance to the sound source, or both.

4.8.1 Suggested Reading List

4.9 Distance Perception

Go outdoors, close your eyes and stand there and listen for a while. Pay attention to how far away things sound and try not to be distracted by how far away you know that they are. Chances are that you'll start noticing that everything sounds really close. You can tell what direction things are coming from, but you won't really know how far away they are. This is because you probably aren't getting enough information to tell you the distance to the sound source. So, the question becomes, "what are the cues that you need to determine the distance to the sound source?"

There are a number of cues that we rely on to figure out how far away a sound source is. These are, in no particular order...

- reflection patterns
- direct-to-reverberant ratio
- sound pressure level
- high-frequency content

4.9.1 Reflection patterns

One of the most important cues that you get when it comes to determining the distance to a sound source lies in the pattern of reflections that arrive after the direct sound. Both the level and time of arrival relationships between these reflections and the direct sound tell you not only how far away the sound source is, but where it is located relative to walls around you, and how big the room is.

Go to an anechoic chamber (or a frozen pond with snow on it...). Take a couple of speakers and put them directly in front of you, aimed at your head with one about 2 m away and the other 4 m distant. Then, make the two speakers the same apparent level at your listening position using the amplifier gains. If you switch back and forth between the two speakers you will not be able to tell which is which – this is because the lack of reflections in the anechoic chamber rob you of your only cues that give you this information.

4.9.2 Direct-to-reverberant ratio

Anyone who has used a digital reverberation unit knows that the way to make things sound farther away in the mix is to increase the level of the

reverb relative to the dry sound. This is essentially creating the same effect we have in real life. We've seen in Section ?? that, as you get farther and farther away from a sound source in a room, the direct sound gets quieter and quieter, but the energy from the room's reflections – the reverberation – stays the same. Therefore the direct-to-reverberant level ratio gets smaller and smaller.

4.9.3 Sound pressure level

If you know how loud the sound source is normally, then you can tell how far away it is based on how loud it is. This is particularly true of sound sources with which we are very familiar like human speech, for example.

Of course, if you have never heard the sound before, or if you really don't know how loud it is (like the sound of a jet engine, for example) then the sound's loudness won't help you tell how far away it is.

4.9.4 High-frequency content

As we saw in Section 3.2.3, air has a tendency to absorb high-frequency information. Over short distances, this absorption is minimal, and therefore not detectable by humans. However, over long distances, the absorption becomes significant enough for us to hear the difference. Of course, we are again assuming that we are familiar with the sound source. If you don't know the normal frequency balance of the sound, then you won't know that the high end has been rolled off due to distance.

4.9.5 Suggested Reading List

4.10 Hearing Loss and Tinnitus

NOT YET WRITTEN

See <http://www.cccd.edu/faq/tinnitus.html> for info

4.10.1 Suggested Reading List

4.11 Perception of sound qualities

NOT YET WRITTEN

4.11.1 Suggested Reading List

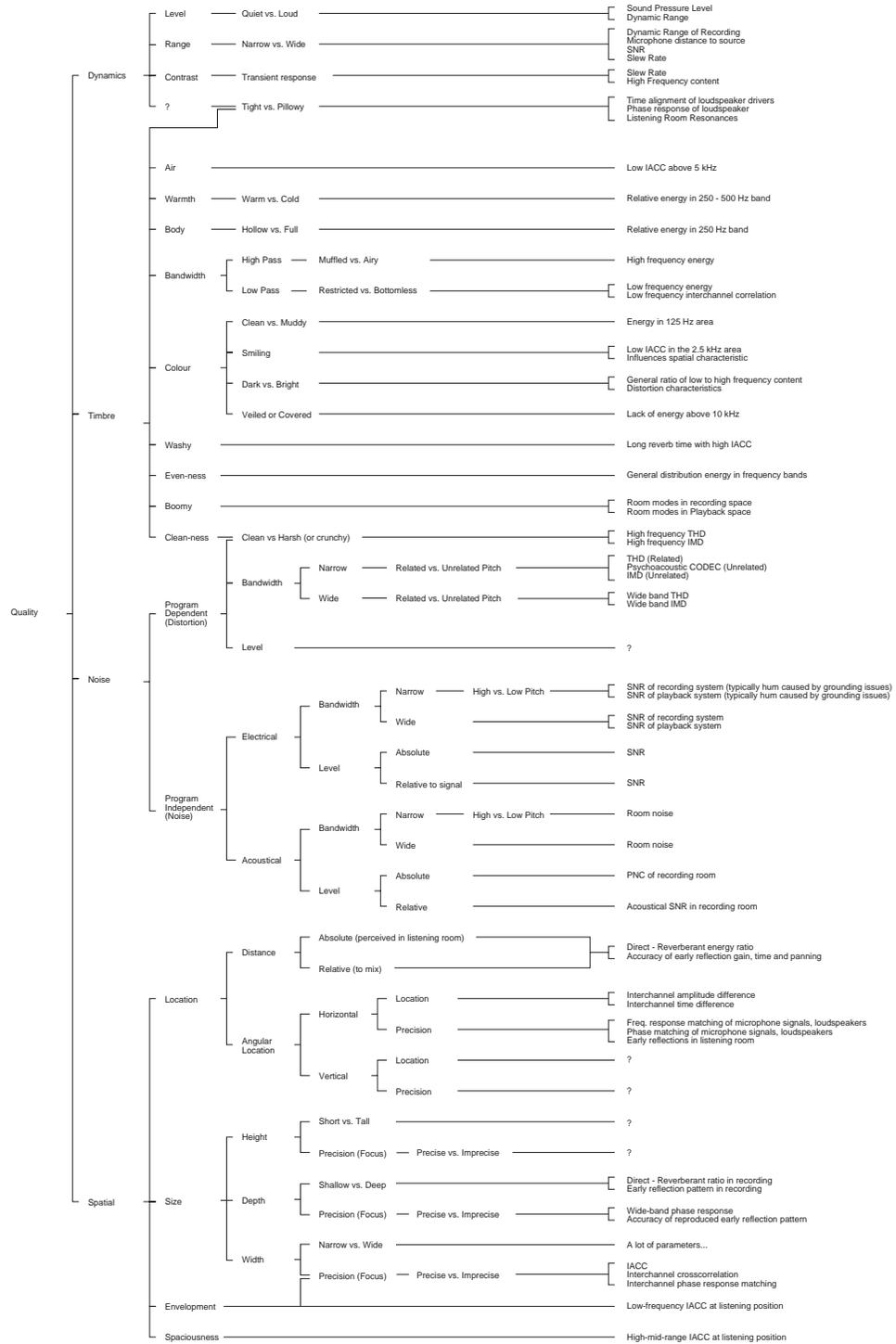


Figure 4.1: My own personal, unproven and un-researched list of descriptions of sound qualities, possibly even bordering on perceptual attributes and how I think they correlate with physical measurements. Please be wary of quoting this list – it's just a map of what is in my head.

Chapter 5

Electroacoustics

5.1 Filters and Equalizers

Thanks to George Massenburg at GML Inc. (www.massenburg.com) for his kind permission to use include chapter which was originally written as part of a manual for one of their products.

5.1.1 Introduction

Once upon a time, in the days before audio was digital, when you made a long-distance phone call, there was an actual physical connection made between the wire running out of your phone and the phone at the other end. This caused a big problem in signal quality because a lot of high-frequency components of the signal would get attenuated along the way. Consequently, booster circuits were made to help make the relative levels of the various frequencies equal. As a result, these circuits became known as *equalizers*. Nowadays, of course, we don't need to use equalizers to fix the quality of long-distance phone calls, but we do use them to customize the relative balance of various frequencies in an audio signal.

In order to look at equalizers and their smaller cousins, *filters*, we're going to have to look at their *frequency response* curves. This is a description of how the output level of the circuit compares to the input for various frequencies. We assume that the input level is our reference, sitting at 0 dB and the output is compared to this, so if the signal is louder at the output, we get values greater than 0 dB. If it's quieter at the output, then we get negative values at the output.

5.1.2 Filters

Before diving straight in and talking about how equalizers behave, we'll start with the basics and look at four different types of filters. Just like a coffee filter keeps coffee grinds trapped while allowing coffee to flow through, an audio filter lets some frequencies pass through unaffected while reducing the level of others.

Low-pass Filter

One of the conceptually simplest filters is known as a *low-pass filter* because it allows low frequencies to pass through it. The question, of course, is “how low is low?” The answer lies in a single frequency known as the *cutoff frequency* or f_c . This is the frequency where the output of the filter is 3.01 dB lower than the maximum output for any frequency (although we normally round this off to -3 dB which is why it's usually called the *3 dB down point*). “What's so special about -3 dB?” I hear you cry. This particular number is chosen because -3 dB is the level where the signal is at one half the power of a signal at 0 dB. So, if the filter has no additional gain incorporated into it, then the cutoff frequency is the one where the output is exactly one half the power of the input. (Which explains why some people call it the *half-power point*.)

As frequencies get higher and higher, they are attenuated more and more. This results in a *slope* in the frequency response graph which can be calculated by knowing the amount of extra attenuation for a given change in frequency. Typically, this slope is specified in decibels per octave. Since the higher we go, the more we attenuate in a low pass filter, this value will always be negative.

The slope of the filter is determined by its *order*. If we oversimplify just a little, a first-order low-pass filter will have a slope of -6.02 dB per octave above its cutoff frequency (usually rounded to -6 dB/oct). If we want to be technically correct about this, then we have to be a little more specific about where we finally reach this slope. Take a look at the frequency response plot in Figure 5.1. Notice that the graph has a nice gradual transition from a slope of 0 (a horizontal line) in the really low frequencies to a slope of -6 dB/oct in the really high frequencies. In the area around the cutoff frequency, however, the slope is changing. If we want to be really accurate, then we have to say that the slope of the frequency response is really 0 for frequencies less than one tenth of the cutoff frequency. In other words, for frequencies more than one *decade* below the cutoff frequency. Similarly, the

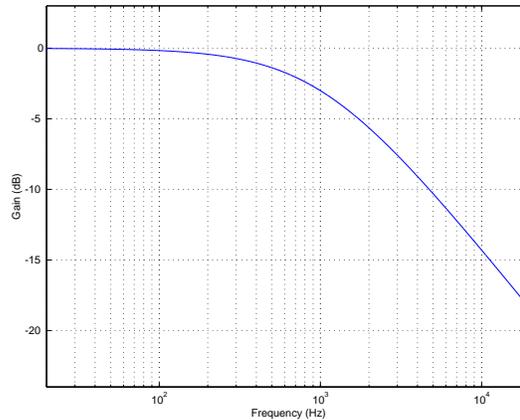


Figure 5.1: The frequency response of a first-order low pass filter with a cutoff frequency of 1 kHz. Note that the cutoff frequency is where the response has dropped in level by 3 dB. The slope can be calculated by dividing the drop in level by the change in frequency that corresponds to that particular drop.

slope of the frequency response is really -6.02 dB/oct for frequencies more than one decade above the cutoff frequency.

If we have a higher-order filter, the cutoff frequency is still the one where the output drops by 3 dB, however the slope changes to a value of $-6.02n \text{ dB/oct}$, where n is the order of the filter. For example, if you have a 3rd-order filter, then the slope is

$$\text{slope} = \text{order} * -6.02\text{dB/oct} \quad (5.1)$$

$$= 3 * -6.02\text{dB/oct} \quad (5.2)$$

$$= -18.06\text{dB/oct} \quad (5.3)$$

High-pass Filter

A high-pass filter is essentially exactly the same as a low-pass filter, however, it permits high frequencies to pass through while attenuating low frequencies as can be seen in Figure 5.2. Just like in the previous section, the cutoff frequency is where the output has a level of -3.01 dB but now the slope below the cutoff frequency is positive because we get louder as we increase in frequency. Just like the low-pass filter, the slope of the high-pass filter is dependent on the order of the filter and can be calculated using the equation $6.02n \text{ dB/oct}$, where n is the order of the filter.

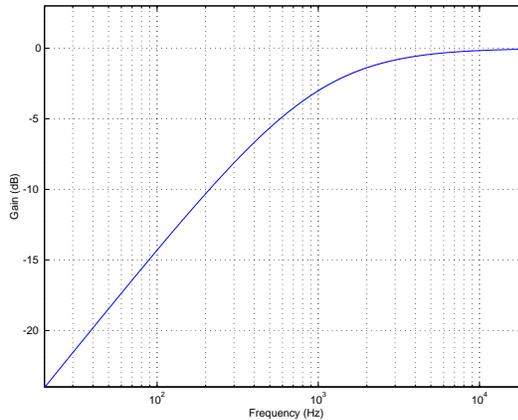


Figure 5.2: The frequency response of a first-order high pass filter with a cutoff frequency of 1 kHz.

Remember as well that the slope only applies to frequencies that are at least one decade away from the cutoff frequency.

Band-pass Filter

Let's take a signal and send it through a high-pass filter and a low-pass filter in series, so the output of one feeds into the input of the other. Let's also assume for a moment that the two cutoff frequencies are more than a decade apart.

The result of this probably won't hold any surprises. The high-pass filter will attenuate the low frequencies, allowing the higher frequencies to pass through. The low-pass filter will attenuate the high frequencies, allowing the lower frequencies to pass through. The result is that the high and low frequencies are attenuated, with a middle band (called the *passband*) that's allowed to pass relatively unaffected.

Bandwidth

This resulting system is called a *bandpass filter* and it has a couple of specifications that we should have a look at. The first is the width of the passband. This *bandwidth* is calculated using the difference two cutoff frequencies which we'll label f_{c1} for the lower one and f_{c2} for the higher one. Consequently, the bandwidth is calculated using the equation:

$$BW = f_{c2} - f_{c1} \quad (5.4)$$

So, using the example of the filter frequency response shown in Figure 4, the bandwidth is $10,000 \text{ Hz} - 20 \text{ Hz} = 9980 \text{ Hz}$.

Centre Frequency

We can also calculate the middle of the passband using these two frequencies. It's not quite so simple as we'd like, however. Unfortunately, it's not just the frequency that's half-way between the low and high frequency cutoff's. This is because frequency specifications don't really correspond to the way we hear things. Humans don't usually talk about frequency – they talk about pitches and notes. They say things like “Middle C” instead of “262 Hz.” They also say things like “one octave” or “one semitone” instead of things like “a bandwidth of 262 Hz.”

Consider that, if we play the A below Middle C on a well-tuned piano, we'll hear a note with a fundamental of 220 Hz. The octave above that is 440 Hz and the octave above that is 880 Hz. This means that the bandwidth of the first of these two octaves is 220 Hz (it's $440 \text{ Hz} - 220 \text{ Hz}$), but the bandwidth of the second octave is 440 Hz ($880 \text{ Hz} - 440 \text{ Hz}$). Despite the fact that they have different bandwidths, we hear them each as one octave, and we hear the 440 Hz note as being half-way between the other two notes. So, how do we calculate this? We have to find what's known as the *geometric mean* of the two frequencies. This can be found using the equation

$$f_{centre} = \sqrt{f_{c1}f_{c2}} \quad (5.5)$$

Q

Let's say that you want to build a bandpass filter with a bandwidth of one octave. This isn't difficult if you know the centre frequency and if it's never going to change. For example, if the centre frequency was 440 Hz, and the bandwidth was one octave wide, then the cutoff frequencies would be 311 Hz and 622 Hz (we won't worry too much about how I arrived at these numbers). What happens if we leave the bandwidth the same at 311 Hz, but change the centre frequency to 880 Hz? The result is that the bandwidth is now no longer an octave wide – it's one half of an octave. So, we have to link the bandwidth with the centre frequency so that we can describe it in

terms of a fixed musical interval. This is done using what is known as the *quality* or Q of the filter, calculated using the equation:

$$Q = \frac{f_{centre}}{BW} \quad (5.6)$$

Now, instead of talking about the bandwidth of the filter, we can use the Q which gives us an idea of the width of the filter in musical terms. This is because, as we increase the centre frequency, we have to increase the bandwidth proportionately to maintain the same Q . Notice however, that if we maintain a centre frequency, the smaller the bandwidth gets, the bigger the Q becomes, so if you're used to talking in terms of musical intervals, you have to think backwards. A big Q is a smaller interval as can be seen in the plot of a number of different Q 's in Figure 5.8.

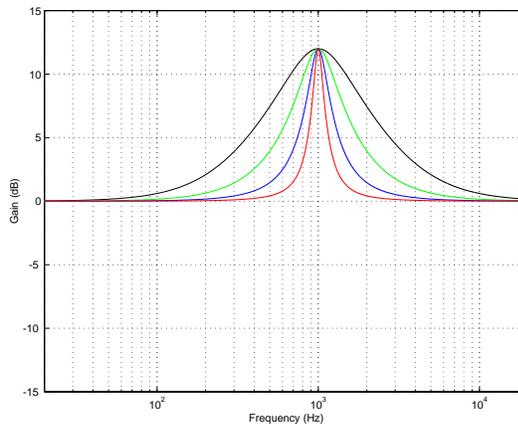


Figure 5.3: The frequency responses of various bandpass filters with different Q 's and a matched centre frequency of 1 kHz.

Notice in Figure 5.8 that you can have a very high Q , and therefore a very narrow bandwidth for a bandpass filter. All of the definitions still hold, however. The cutoff frequencies are still the points where we're 3 dB lower than the maximum value and the bandwidth is still the distance in Hertz between these two points and so on...

Band-reject Filter

Although bandpass filters are very useful at accentuating a small band of frequencies while attenuating others, sometimes we want to do the opposite.

We want to attenuate a small band of frequencies while leaving the rest alone. This can be accomplished using a *band-reject filter* (also known as a *bandstop filter*) which, as its name implies, rejects (or usually just attenuates) a band of frequencies without affecting the surrounding material. As can be seen in Figure 5.4, this winds up looking very similar to a bandpass filter drawn upside down.

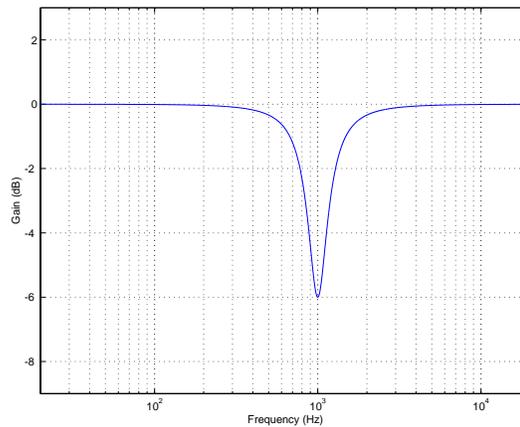


Figure 5.4: The frequency response of a band-reject filter with a centre frequency of 1 kHz.

The thing to be careful of when describing band-reject filters is the fact that cutoff frequencies are still defined as the points where we've dropped in level by 3 dB. Therefore, we don't really get an intuitive idea of how much we drop at the centre frequency. Looking at Figure 5.4 we can see that, although the band-reject filter looks just like the bandpass filter upside down, the bandwidth is quite different. This is a fairly important point to remember a little later on in the section on symmetry.

Notch Filter

There is a special breed of band-reject filter that is designed to have almost infinite attenuation at a single frequency, leaving all others intact. This, of course is impossible, but we can come close. If we have a band-reject filter with a very high Q , the result is a frequency response like the one shown in Figure 5.5. The shape is basically a flat frequency response with a narrow, deep notch at one frequency – hence the name *notch filter*

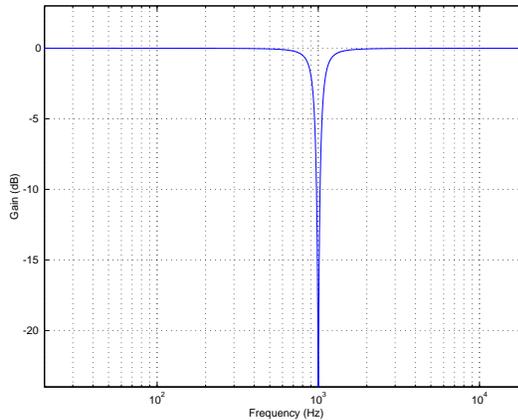


Figure 5.5: The frequency response of a notch filter with a centre frequency of 1 kHz.

5.1.3 Equalizers

Unlike its counterpart from the days of long-distance phone calls, a modern equalizer is a device that is capable of attenuating and boosting frequencies according to the desire and expertise of the user. There are four basic types of equalizers, but we'll have to talk about a couple of issues before getting into the nitty-gritty.

An equalizer typically consists of a collection of filters, each of which permits you to control one or more of three things: the *gain*, *centre frequency* and *Q* of the filter. There are some minor differences in these filters from the ones we discussed above, but we'll sort that out before moving on. Also, the filters in the equalizer may be connected in *parallel* or in *series*, depending on the type of equalizer and the manufacturer.

To begin with, as we'll see, a filter in an equalizer comes in three basic models, the *bandpass*, and the *band reject*, which are typically chosen by the user by manipulating the gain of the filter. On a decibel scale, positive gain results in a bandpass, whereas negative gain produces a band reject. In addition, there is the *shelving filter* which is a variation on the highpass and low pass filters.

The principal difference between filters in an equalizer and the filters defined in Section 1 is that, in a typical equalizer, instead of attenuating all frequencies outside the passband, the filter typically leaves them at a gain of 0 dB. An example of this can be seen in the plot of an equalizer's bandpass filter in Figure 5.6. Notice now that, rather than attenuating all

unwanted frequencies, the filter is applying a known gain to the passband. The further away you get from the passband, the less the signal is affected. Notice, however, that we still measure the bandwidth using the two points that are 3 dB down from the peak of the curve.

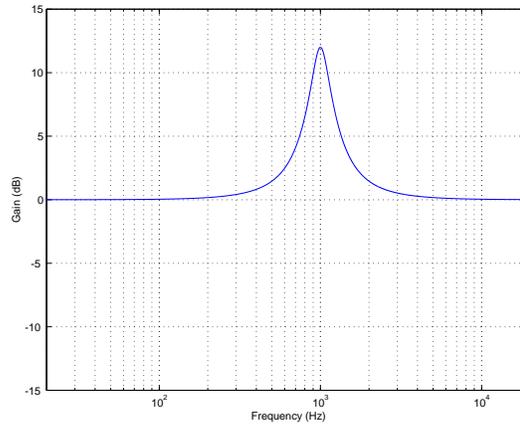


Figure 5.6: The frequency response of a bandpass filter with a centre frequency of 1 kHz, a Q of 4, and a gain of 12 dB in a typical equalizer.

Filter symmetry

Constant Q Filter

Let's look at the frequency response of a filter with a centre frequency of 1 kHz, a Q of 4 and a varying amount of boost or cut. If we plot these responses on the same graph, they look like Figure 5.11.

Notice that, although these two curves have “matching” parameters, they do not have the same shape. This is because the bandwidth (and therefore the Q) of a filter is measured using its 3 dB down point – not the point that's 3 dB away from the peak or dip in the curve. Since the measurement is not symmetrical, the curves are not symmetrical. This is true of any filter where the Q is kept constant and gain is modified. If you compare a boost of any amount with a cut of the same amount, you'll always get two different curves. This is what is known as a *constant Q* filter because the Q is kept as a constant. The result is called an *asymmetrical filter* (or *non-symmetrical filter*) because a matching boost and cut are not mirror images of each other.

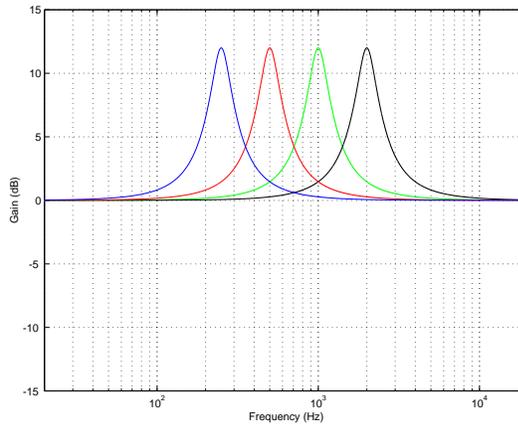


Figure 5.7: The frequency responses of bandpass filters with a various centre frequencies, a Q of 4, and a gain of 12 dB in a typical equalizer. Blue $f_c = 250$ Hz. Red $f_c = 500$ Hz. Green $f_c = 1000$ Hz. Black $f_c = 2000$ Hz.

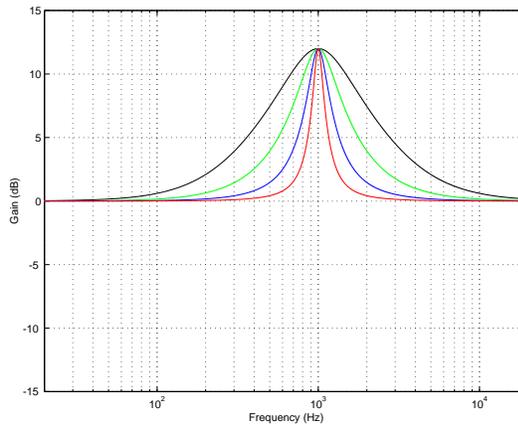


Figure 5.8: The frequency responses of bandpass filters with a centre frequency of 1 kHz, various Q's, and a gain of 12 dB in a typical equalizer. Black Q = 1. Green Q = 2. Blue Q = 4. Red Q = 8.

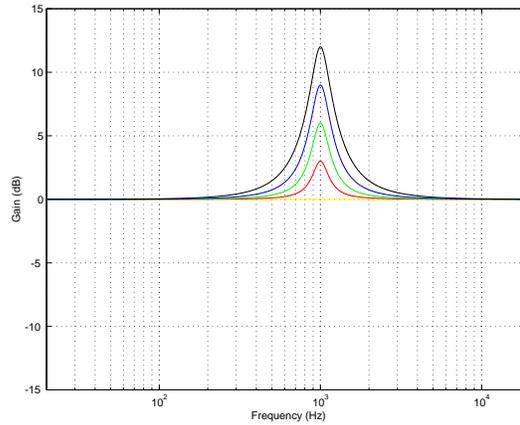


Figure 5.9: The frequency responses of bandpass filters with a centre frequency of 1 kHz, a Q of 4, and various gains from 0 dB to 12 dB in a typical equalizer. Yellow gain = 0 dB. Red gain = 3 dB. Green gain = 6 dB. Blue gain = 9 dB. Black gain = 12 dB.

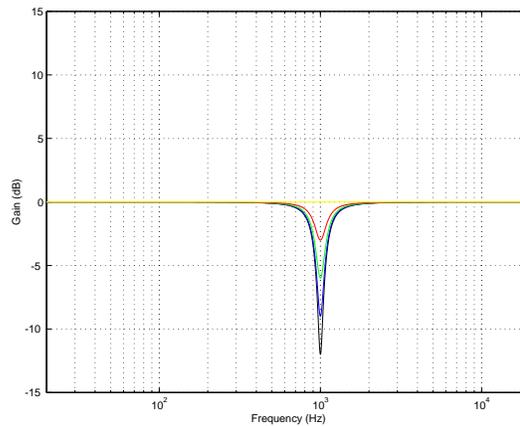


Figure 5.10: The frequency responses of bandpass filters with a centre frequency of 1 kHz, a Q of 4, and various gains from -12 dB to 0 dB in a typical equalizer. Yellow gain = 0 dB. Red gain = -3 dB. Green gain = -6 dB. Blue gain = -9 dB. Black gain = -12 dB.

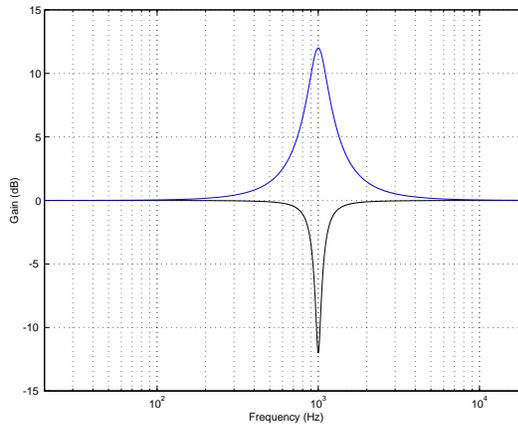


Figure 5.11: The frequency responses of two filters, each with a centre frequency of 1 kHz, and a Q of 4. The Blue curve shows a gain of 12 dB, the black curve, a gain of -12 dB.

There are advantages and disadvantages to this type of filter. The primary advantage is that you can have a very selective cut if you're trying to eliminate a single frequency, simply by increasing the Q. The primary disadvantage is that you cannot undo what you have done. This statement is explained in the following section.

Reciprocal Peak/Dip Filter

Instead of building a filter where the cut and boost always maintain a constant Q, let's set about to build a filter that is *symmetrical* – that is to say that a matching boost and cut at the same centre frequency would result in the same shape. The nice thing about this design is that, if you take two such filters and connect them in series and set their parameters to be the same but opposite gains (for example, both with a centre frequency of 1 kHz and a Q of 2, but one has a boost of 6 dB and the other has a cut of 6 dB) then they'll cancel each other out and your output will be identical to your input. This also applies if you've equalized something while recording – assuming that you live in a perfect world, if you remember your original settings on the recorded EQ curve, you can undo what you've done by duplicating the settings and inverting the gain.

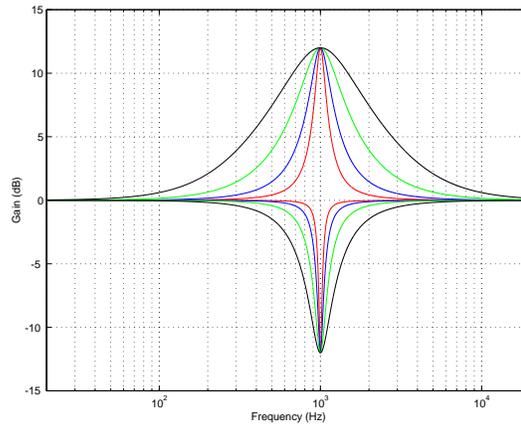


Figure 5.12: The frequency responses of various constant Q filters, all with a centre frequency of 1 kHz, gains of either 12 dB or -12 dB (depending on whether it's a boost or a cut) and various Q's. Black Q = 1. Green Q = 2. Blue Q = 4. Red Q = 8.

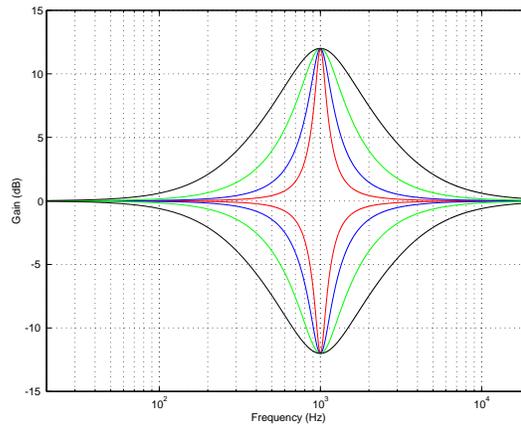


Figure 5.13: The frequency responses of various reciprocal peak/dip filters, all with a centre frequency of 1 kHz, gains of either 12 dB or -12 dB (depending on whether it's a boost or a cut) and various boost Q's. Black Q = 1. Green Q = 2. Blue Q = 4. Red Q = 8.

Parallel vs. Series Filters

Let's take two reciprocal peak/dip filters, each set with a Q of 2 and a gain of 6 dB. The only difference between them is that one has a centre frequency of 1 kHz and the other has a centre frequency of 1.2 kHz. If we use both of these filters on the same signal simultaneously, we can achieve two very different resulting frequency responses, depending on how they're connected.

If the two filters are connected in series (it doesn't matter what order we connect them in), then the frequency band that overlaps in the boosted portion of the two filters' responses will be boosted twice. In other words, the signal goes through the first filter and is amplified, after which it goes through the second filter and the amplified signal is boosted further. This arrangement is also known as a circuit made of *combining filters*.

If we connect the two filters in parallel, however, a different situation occurs. Now each filter boosts the original signal independently, and the two resulting signals are added, producing a small increase in level, but not as significant as in the case of the series connection. This arrangement is also known as a circuit made of *non-combining filters*.

The primary advantage to having filters in connected in series rather than in parallel lies in possibility of increased gain or attenuation. For example, if you have two filters in series, each with a boost of 12 dB and with matched centre frequencies, the total resulting gain applied to the signal is 24 dB (because a gain of 12 dB from the second filter is applied to a signal that already has a gain of 12 dB from the first filter). If the same two filters were connected in parallel, the total gain would be only 18 dB. (This is because a the addition of two identical signals results in a doubling of level which corresponds to an additional gain of only 6 dB.)

The main disadvantage to having filters connected in series rather than in parallel is the fact that you can occasionally result in frequency bands being boosted more than you're intuitively aware. For example, looking at Figure 17, we can see that, based on the centre frequencies of the two filters, we would expect to have two peaks in the total frequency response at 1 kHz and 1.2 kHz. The actual result, as can be seen, is a single large peak between the two expected centre frequencies. Also, it should be noted that a group of non-combining filters will have a ripple in their output frequency response.

Shelving Filter

The nice thing about high pass and low pass filters is that you can reduce (or eliminate) things you don't want like low-frequency noise from air conditioners, for example. But, what if you want to boost all your low frequencies instead of cutting all your high's? This is when a *shelving filter* comes in handy. The response curve of shelving filters most closely resemble their high- and low-pass filter counterparts with a minor difference. As their name suggests, the curve of these filters level out at a specified frequency called the *stop frequency*. In addition, there is a second defining frequency called the *turnover frequency* which is the frequency at which the response is 3 dB above or below 0 dB. This is illustrated in Figure 20.

The *transition ratio* is sort of analogous to the *order* of the filter and is calculated using the turnover and stop frequencies as shown below.

$$R_T = \frac{f_{stop}}{f_{turnover}} \quad (5.7)$$

where R_T is the transition ratio.

The closer the transition ratio is to 1, the greater the slope of the transition in gain from the unaffected to the affected frequency ranges.

These filters are available as high- and low-frequency shelving units, boosting high and low frequencies respectively. In addition, they typically have a symmetrical response. If the transition ratio is less than 1, then the filter is a low shelving filter. If the transition ratio is greater than 1, then the filter is a high shelving filter.

The disadvantage of these components lies in their potential to boost frequencies above and below the audible audio range causing at the least wasted amplifier power and interference from local AM radio signals, and at the worst, loudspeaker damage. For example, if you use a high shelf filter with a stop frequency of 10 kHz to increase the level of the high end by 12 dB to brighten things up a bit, you will probably also wind up boosting signals above your hearing range. In a typical case, this may cause some unpredictable signals from your tweeter due to increased intermodulation distortion of signals you can't even hear. To reduce these unwanted effects, super sonic and subsonic signals can be attenuated using a low pass or high pass filter respectively outside the audio band. Using a peaking filter at the appropriate frequency instead of a filter with a shelving response can avoid the problem altogether.

The most common application of this equalizer is the tone controls on home sound systems. These bass and treble controls generally have a maxi-

imum slope of 6 dB per octave and reciprocal characteristics. They are also frequently seen on equalizer modules on small mixing consoles.

Graphic Equalizer

Graphic equalizers are seen just about everywhere these days, primarily because they're intuitive to use. In fact, they are probably the most-used piece of signal processing equipment in recording. The name "graphic equalizer" comes from the fact that the device is made up of a number of filters with centre frequencies that are regularly spaced, each with a slider used for gain control. The result is that the arrangement of the sliders gives a graphic representation of the frequency response of the equalizer. The most common frequency resolutions available are one-octave, two-third-octave and one-third-octave, although resolutions as fine as one-twelfth-octave exist. The sliders on most graphic equalizers use *ISO standardized band center frequencies*. They virtually always employ reciprocal peak/dip filters wired in parallel. As a result, when two adjacent bands are boosted, there remains a comparatively large dip between the two peaks. This proves to be a great disadvantage when attempting to boost a frequency between two center frequencies. Drastically excessive amounts of boost may be required at the band centers in order to properly adjust the desired frequency. This problem is eliminated in graphic EQ's using the much-less-common combining filters. In this system, the filter banks are wired in series, thus adjacent bands have a cumulative effect. Consequently, in order to boost a frequency between two center frequencies, the given filters need only be boosted a minimal amount to result in a higher-boosted mid-frequency.

Virtually all graphic equalizers have fixed frequencies and a fixed Q. This makes them simple to use and quick to adjust, however they are generally a compromise. Although quite suitable for general purposes, in situations where a specific frequency or bandwidth adjustment is required, they will prove to be inaccurate.

Paragraphic Equalizer

One attempt to overcome the limitations of the graphic equalizer is the *paragraphic equalizer*. This is a graphic equalizer with fine frequency adjustment on each slider. This gives the user the ability to sweep the center frequency of each filter somewhat, thus giving greater control over the frequency response of the system.

Sweep Filters

These equalizers are most commonly found on the input stages of mixing consoles. They are generally used where more control is required over the signal than is available with graphic equalizers, yet space limitations restrict the sheer number of potentiometers available. Typically, the equalizer section on a console input strip will have one or two sweep filters in addition to low and a high shelf filters with fixed turnover frequencies. The frequencies of the mid-range filters are usually reciprocal peak/dip filters with an adjustable (or *sweepable*) center frequencies and fixed Q's.

The advantage of this configuration is a relatively versatile equalizer with a minimum of knobs, precisely what is needed on an overcrowded mixer panel. The obvious disadvantage is its lack of adjustment on the bandwidth, a problem that is solved with a parametric equalizer.

Parametric Equalizer

A parametric equalizer is one that allow the user to control the gain, centre frequency and Q of each filter. In addition, these three parameters are independent – that is to say that adjusting one of the parameters will have no effect on the other two. They are typically comprised of combining filters and will have either reciprocal peak/dip or constant-Q filters. (Check your manual to see which you have – it makes a huge difference!) In order to give the user a wider amount of control over the signal, the frequency ranges of the filters in a parametric equalizer typically overlap, making it possible to apply gain or attenuation to the same centre frequency using at least two filters.

The obvious advantage of using a parametric equalizer lies in the detail and versatility of control afforded by the user. This comes at a price, however – it unfortunately takes much time and practice to master the use of a parametric equalizer.

Semi-parametric equalizer

A less expensive variation on the true parametric equalizer is the *semi-parametric* or *quasi-parametric equalizer*. From the front panel, this device appears to be identical to its bigger cousin, however, there is a significant difference between the two. Whereas in a true parametric equalizer, the three parameters are independent, in a semi-parametric equalizer, they are not. As a result, changing the value of one parameter will cause at least one, if not both, of the other two parameters to change unexpectedly. As a result,

although these devices are less expensive than a true parametric, they are less trustworthy and therefore less functional in real working situations.

5.1.4 Summary

5.1.5 Phase response

So far, we've only been looking at the frequency response of a filter or equalizer. In other words, we've been looking at what the magnitude of the output of the filter would be if we send sine tones through it. If the filter has a gain of 6 dB at a certain frequency, then if we feed it a sine tone at that frequency, then the amplitude of the output will be 2 times the amplitude of the input (because a gain of 2 is the same as an increase of 6 dB). What we haven't looked at so far is any shift in phase (also known as *phase distortion*) that might be incurred by the filtering process. Any time there is a change in the frequency response in the signal, then there is an associated change in phase response that you may or may not want to worry about. That phase response is typically expressed as a shift (in degrees) for a given frequency. Positive phase shifts mean that the signal is delayed in phase whereas negative phase shifts indicate that the output is ahead of the input.

“The output is *ahead* of the input!?” I hear you cry. “How can the output be ahead of the input? Unless you've got one of those new digital filters that can see into the near future...” Well, it's actually not as strange as it sounds. The thing to remember here is that we're talking about a sine wave – so don't think about using an equalizer to help your drummer get ahead of the beat... It doesn't mean that the whole signal comes out earlier than it went in. This is because we're not talking about negative delay – it's negative phase.

Let's look at a typical example. Figure 24 below shows the phase response for a typical filter found in an average equalizer with a frequency response shown in Figure 10. Note that some frequencies have a negative phase shift while others have a positive phase shift. If you're looking really carefully, you may notice a relationship between the slope of the frequency response and the polarity of the phase response – but if you didn't notice this, don't worry...

Minimum phase

While it's true that a change in frequency response of a signal necessarily implies that there is a change in its phase, you don't have to have the

| Category | Graphic | | Sweep | Parametric | |
|-----------------------------------|---------|-------------|-------|-----------------|-----------------|
| | Graphic | Paragraphic | | Semi-parametric | True Parametric |
| Control Gain | Y | Y | Y | Y | Y |
| Centre Frequency | N | Y | Y | Y | Y |
| Q | N | N | N | Y | Y |
| Shelving Filter? | N | N | Y | Optional | Optional |
| Combining / Non-combining | N | N | N | Depends | C |
| Reciprocal peak/dip or Constant Q | R p/d | R p/d | R p/d | Typically R p/d | Depends |

Table 5.1: Summary of the typical control characteristics on the various types of equalizers. "Depends" means that it depends on the manufacturer and model.

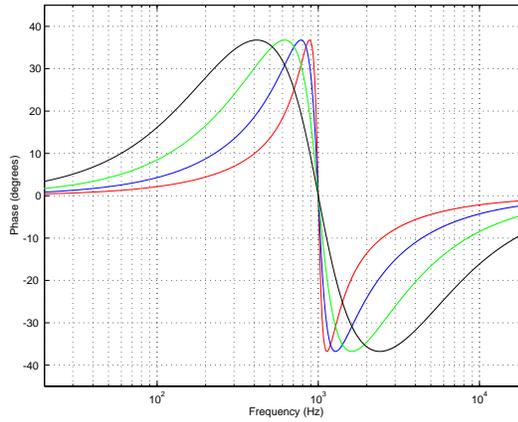


Figure 5.14: The phase responses of bandpass filters with a centre frequency of 1 kHz, various Q 's, and a gain of 12 dB in a typical equalizer. Black $Q = 1$. Green $Q = 2$. Blue $Q = 4$. Red $Q = 8$. (Compare these curves to the plot in Figure 10)

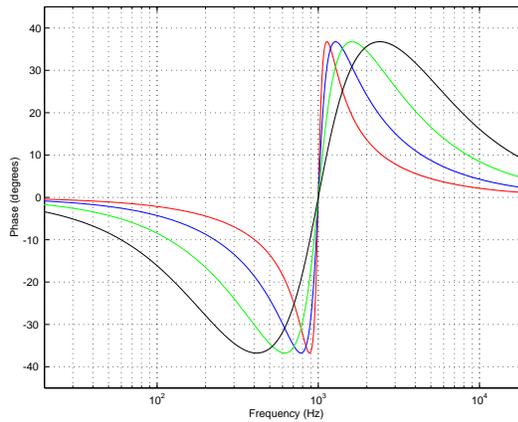


Figure 5.15: The phase responses of bandpass filters with a centre frequency of 1 kHz, various Q 's, and a gain of -12 dB in a typical equalizer. Black $Q = 1$. Green $Q = 2$. Blue $Q = 4$. Red $Q = 8$. (Compare these curves to the plot in Figure 5.14)

same phase shift for the same frequency response change. In fact, different manufacturers can build two filters with centre frequencies of 1 kHz, gains of 12 dB and Q's of 4. Although the frequency responses of the two filters will be identical, their phase responses can be very different.

You may occasionally hear the term *minimum phase* to describe a filter. This is a filter that has the frequency response that you want, and incurs the smallest (hence “minimum”) shift in phase to achieve that frequency response.

Two things to remember about minimum phase filters: 1) Just because they have the minimum possible phase shift doesn't necessarily imply that they sound the best. 2) A minimum phase filter can be “undone” – that is to say that if you put your signal through a minimum phase filter, it is possible to find a second minimum phase filter that will reverse all the effects of the first, giving you exactly the signal you started with.

Linear Phase

If you plot the phase response of a filter for all frequencies, chances are you'll get a smooth, fancy-looking curve like the ones in Figure 24. Some filters, on the other hand, have a phase response plot that's a straight line if you graph the response on a linear frequency scale (instead of a log scale like we normally do...). This line usually slopes upwards so the higher the frequency, the bigger the phase change. In fact, this would be exactly the phase response of a straight delay line – the higher the frequency, the more of a phase shift that's incurred by a fixed delay time. If the delay time is 0, then the straight line is a horizontal one at 0° for all frequencies.

Any filter whose phase response is a straight line is called a *linear phase filter*. Be careful not to jump to the conclusion that, because it's a linear phase filter, it's better than anything else. While there are situations where such a filter is useful, they work well in all situations to correct all problems. Different intentions require different filter characteristics.

Ringling

The phase response of a filter is typically strongly related to its Q. The higher the Q (and therefore the smaller the bandwidth) the greater the change in phase around the centre frequency. This can be seen in Figure 24 above. Notice that, the higher the Q, the higher the slope of the phase response at the centre frequency of the filter. When the slope of the phase response of a filter gets very steep (in other words, when the Q of the filter is very

high) an interesting thing called *ringing* happens. This is an effect where the filter starts to oscillate at its centre frequency for a length of time after the input signal stops. The higher the Q , the longer the filter will ring, and therefore the more audible the effect will be. In the extreme cases, if the Q of the filter is 0, then there is no ringing (but the bandwidth is infinity and you have a flat frequency response – so it's not a very useful filter...). If the Q of the filter is infinity, then the filter becomes a sine wave generator.

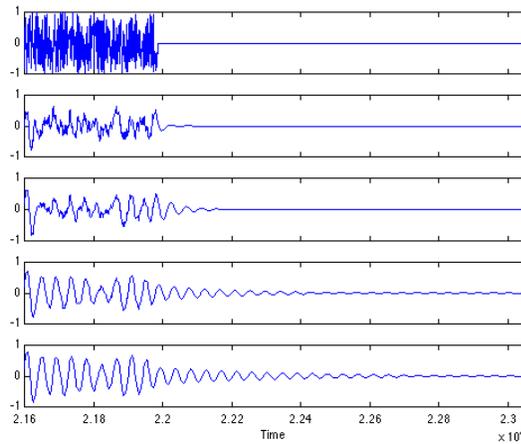


Figure 5.16: Ringing caused by minimum phase bandpass filters with centre frequencies of 1 kHz and various Q 's. The input signal is white noise, abruptly cut to digital zero as is shown in the top plot. There are at least three things to note: 1) The higher the Q , the longer the filter will ring at the centre frequency after the input signal has stopped. 2) The higher the Q , the more the output signal approaches a sine wave at the centre frequency. 3) Even a filter with a Q as low as 1 rings – although this will likely not be audible due to psychoacoustic masking effects.

5.1.6 Applications

All this information is great – but why and how would you use an equalizer? Well, there are lots of different reasons FINISH THIS OFF

5.1.7 Spectral sculpting

This is probably the most obvious use for an equalizer. You have a lead vocal that sounds too bright so you want to cut back the high frequency content. Or you want to get bump up the low mid range of a piano to warm it up a bit. This is the primary intention of the tone controls on the cheapest ghetto blaster through to the best equalizer in the recording studio.

It's virtually impossible to give a list of "tips and tricks" in this category, because every instrument and every microphone in every recording situation will be different. There are time when you'll want to use an equalizer to compensate for deficiencies in the signal because you couldn't afford a better mic for that particular gig. On the other hand there may be occasions where you have the most expensive microphone in the world on a particular instrument and it still needs a little tweaking to fix it up. There are, however, a couple of good rules to follow when you're in this game.

First of all – don't forget that you can use an equalizer to cut as easily well as boost. Consider a situation where you have a signal that has too much bass – there are two possible ways to correct the problem. You could increase the mids and highs to balance, or you could turn down the bass. There are as many situations where one of these is the correct answer as there are situations where the other answer is more appropriate. Try both unless you're in a *really* big hurry.

Second of all – don't touch the equalizer before you've heard what you're tweaking. I often notice when I go to a restaurant that there are a huge number of people who put salt and pepper on their meal before they've even tasted a single morsel. Doesn't make much sense... Hand them a plate full of salt and they'll still shake salt on it before raising a hand to touch their fork. The same goes for equalization. Equalize to fix a problem that you can hear – not because you found a great EQ curve that worked great on kick drum at the last session.

Thirdly – don't overdo it. Or at least, overdo it to see how it sounds when it's overdone, then bring it back. Again, back to a restaurant analogy – you know that you're in a restaurant that knows how to cook steak when there's a disclaimer on the menu that says something to the effect of "We are not responsible for steaks ordered well done." Everything in moderation – unless, of course, you're intending to plow straight through the fields of moderation and into the barn of excess.

Fourthly, there's a number of general descriptions that indicate problems that can be fixed, or at least tamed with equalization. For example, when someone says that the sound is "muddy," you could probably clean this up by reducing the area around 125 – 250 Hz with a low-Q filter. The table below gives a number of basic examples, but there are plenty more – ask around...

One last trick here applies when you hear a resonant frequency sticking out, and you want to get rid of it, but you just don't know what the exact frequency is. You know that you need to use a filter to reduce a frequency – but finding it is going to be the problem. The trick is to search and destroy

| <i>Symptom description</i> | <i>Possible remedy</i> |
|----------------------------|-------------------------------|
| Bright | Reduce high frequency shelf |
| Dark, veiled, covered | Increase high frequency shelf |
| Harsh, crunchy | Reduce 3 – 5 kHz region |
| Muddy, thick | Reduce 125 – 250 Hz region |
| Lacks body or warmth | Increase 250 – 500 Hz |
| Hollow | Reduce 500 Hz region |

Table 5.2: Some possible spectral solutions to general comments about the sound quality

by making the problem worse. Set a filter to *boost* instead of cutting a frequency band with a fairly high *Q*. Then, sweep the frequency of the filter until the resonance sticks out more than it normally does. You can then fine tune the centre frequency of the filter so that the problem is as bad as you can make it, then turn the boost back to a cut.

5.1.8 Loudness

Although we rarely like to admit it, we humans aren't perfect. This is true in many respects, but for the purposes of this discussion, we'll concentrate specifically on our abilities to hear things. Unfortunately, our ears don't have the same frequency response at all listening levels. At very high listening levels, we have a relatively flat frequency response, but as the level drops, so does our sensitivity to high and low frequencies. As a result, if you mix a tune at a very high listening level and then reduce the level, it will appear to lack low end and high end. Similarly, if you mix at a low level and turn it up, you'll tend to get more low end and high end.

One possible use for an equalizer is to compensate for the perceived lack of information in extreme frequency ranges at low listening levels. Essentially, when you turn down the monitor levels, you can use an equalizer to increase the levels of the low and high frequency content to compensate for deficiencies in the human hearing mechanism. This filtering is identical to that which is engaged when you press the "loudness" button on most home stereo systems. Of course, the danger with such equalization is that you don't know what frequency ranges to alter, and how much to alter them – so it is not recommendable to do such compensation when you're mixing, only when you're at home listening to something that's already been mixed.

5.1.9 Noise Reduction

It's possible in some specific cases to use equalization to reduce noise in recordings, but you have to be aware of the damage that you're inflicting on some other parts of the signal.

High-frequency Noise (Hiss)

Let's say that you've got a recording of an electric bass on a really noisy analog tape deck. Since most of the perceivable noise is going to be high-frequency stuff and since most of the signal that you're interested in is going to be low-frequency stuff, all you need to do is to roll off the high end to reduce the noise. Of course, this is the best of all possible worlds. It's more likely that you're going to be coping with a signal that has some high-frequency content (like your lead vocals, for example...) so if you start rolling off the high end too much, you start losing a lot of brightness and sparkle from your signal, possibly making the end result worse than you started. If you're using equalization to reduce noise levels, don't forget to occasionally hit the "bypass" switch of the equalizer once in a while to hear the original. You may find when you refresh your memory that you've gone a little too far in your attempts to make things better.

Low-frequency Noise (Rumble)

Almost every console in the world has a little button on every input strip that has a symbol that looks like a little ramp with the slope on the left. This is a high-pass filter that is typically a second-order filter with a cutoff frequency around 100 Hz or so, depending on the manufacturer and the year it was built. The reason that filter is there is to help the recording or sound reinforcement engineer get rid of low-frequency noise like "stage rumble" or microphone handling noise. In actual fact, this filter won't eliminate all of your problems, but it will certainly reduce them. Remember that most signals don't go below 100 Hz (this is about an octave and a half below middle C on a piano) so you probably don't need everything that comes from the microphone in this frequency range – in fact, chances are, unless you're recording pipe organ, electric bass or space shuttle launches, you won't need nearly as much as you think below 100 Hz.

Hummmmmmm...

There are many reasons, forgivable and unforgivable, why you may wind up with an unwanted hum in your recording. Perhaps you work with a poorly-installed system. Perhaps your recording took place under a buzzing streetlamp. Whatever the reason, you get a single frequency (and perhaps a number of its harmonics) singing all the way through your recording. The nice thing about this situation is that, most of the time, the hum is at a predictable frequency (depending on where you live, it's likely a multiple of either 50 Hz or 60 Hz) and that frequency never changes. Therefore, in order to reduce, or even eliminate this hum, you need a very narrow band-reject filter with a lot of attenuation. Just the sort of job for a notch filter. The drawback is that you also attenuate any of the music that happens to be at or very near the notch centre frequency, so you may have to reach a compromise between eliminating the hum and having too detrimental of an effect on your signal.

5.1.10 Dynamic Equalization

A dynamic equalizer is one which automatically changes its frequency response according to characteristics of the signal passing through it. You won't find many single devices what fit this description, but you can create a system that behaves differently for different input signals if you add a compressor to the rack. This is easily accomplished today with digital multi-band compressors which have multiple compressors fed by what could be considered a crossover network similar to that used in loudspeakers.

Dynamic enhancement

Take your signal and, using filters, divide it into two bands with a crossover frequency at around 5 kHz. Compress the higher band using a fast attack and release time, and adjust the output level of the compressor so that when the signal is at a peak level, the output of the compressor summed with the lower frequency band results in a flat frequency response. When the signal level drops, the low frequency band will be reduced more than the high frequency band and a form of high-frequency enhancement will result.

Dynamic Presence

In order to add a sensation of "presence" to the signal, use the technique described in Section 3.4.1 but compress the frequency band in the 2 kHz to

5 kHz range instead of all high frequencies.

De-Essing

There are many instances where a close-mic technique is used to record a narrator and the result is a signal that emphasizes the sibilant material in the signal – in particular the “s” sound. Since the problem is due to an excess of high frequency, one option to fix the issue could be to simply roll off high frequency content using a low-pass filter or a high-frequency shelf. However, this will have the effect of dulling all other material in the speech, removing not only the “s’s” but all brightness in the signal. The goal, therefore, is to reduce the gain of the signal when the letter “s” is spoken. This can be accomplished using an equalizer and a compressor with a *side chain*. In this case, the input signal is routed to the inputs of the equalizer and the compressor in parallel. The equalizer is set to boost high frequencies (thus making the “s’s” even louder...) and its output is fed to the side chain input of the compressor. The compression parameters are then set so that the signal is not normally compressed, however, when the “s” is spoken, the higher output level from the equalizer in the side chain triggers compression on the signal. The output of the compressor has therefore been “de-essed” or reduced in sibilance.

Although it seems counterintuitive, don’t forget that, in order to reduce the level of the high frequencies in the output of the compressor, you have to increase the level of the high frequencies at the output of the equalizer in this case.

Pop-reduction

A similar problem to de-essing is the “pop” that occurs when a singer’s plosive sounds (p’s and b’s) cause a thump at the diaphragm of the microphone. There is a resulting overload in the low frequency component of the signal that can be eliminated using the same technique described in Section 3.4.3 where the low frequencies (250 Hz and below) are boosted in the equalizer instead of the high frequency components.

5.1.11 Further reading

What is a filter? – from Julius O. Smith’s substantial website.

5.2 Compressors, Limiters, Expanders and Gates

5.2.1 What a compressor does.

So you're out for a drive in your car, listening to some classical music played by an orchestra on your car's CD player. The piece starts off very quietly, so you turn up the volume because you really love this part of the piece and you want to hear it over the noise of your engine. Then, as the music goes on, it gets louder and louder because that's what the composer wanted. The problem is that you've got the stereo turned up to hear the quiet sections, so these new loud sections are REALLY loud – so you turn down your stereo. Then, the piece gets quiet again, so you turn up the stereo to compensate.

What you are doing is to manipulate something called the “dynamic range” of the piece. In this case, the dynamic range is the difference in level between the softest and the loudest parts of the piece (assuming that you're not mucking about with the volume knob). By fiddling with your stereo, you're making the soft sounds louder and the loud sounds softer, and therefore *compressing* the dynamics. The music still appears to have quiet sections and loud sections, but they're not as different as they were without your fiddling.

In essence, this is what a compressor does – at the most basic level, it makes loud sounds softer and soft sounds louder so that the music going through it has a smaller (or compressed) dynamic range. Of course, I'm oversimplifying, but we'll straighten that out.

Let's look at the *gain response* of an ideal piece of wire. This can be shown as a *transfer function* as seen in Figure 1.

Now, let's look at the gain response for a simple device that behaves as an oversimplified compressor. Let's say that, for a sine wave coming in at 0 dBV (1 V_{rms}, remember?) the device has a gain of 1 (or output=input). Let's also say that, for every 2 dB increase in level at the input, the gain of this device is reduced by 1 dB – so, if the input level goes up by 2 dB, the output only goes up by 1 dB (because it's been reduced by 1 dB, right?) Also, if the level at the input goes down by 2 dB, the gain of the device comes up by 1 dB, so a 2 dB drop in level at the input only results in a 1 dB drop in level at the output. This generally makes the soft sounds louder than when they went in, the loud sounds softer than when they went in, and anything at 0 dBV come out at exactly the same level as it goes in.

If we compare the *change* in level at the input to the change in level at the output, we have a comparison between the original dynamic range and the new one. This comparison is expressed as a ratio of the change in input

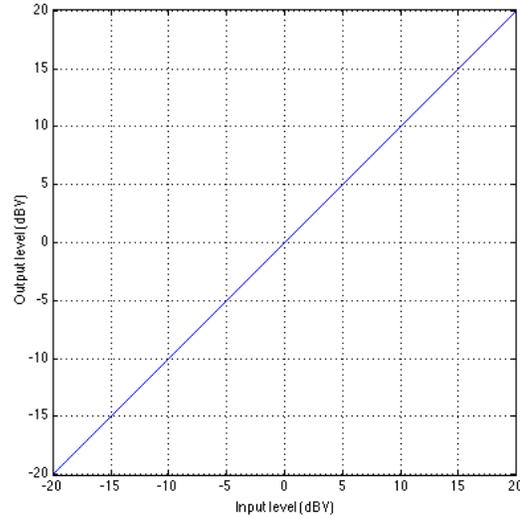


Figure 5.17: The gain response or transfer function of a device with a gain of 1 for all input levels. Essentially, output = input.

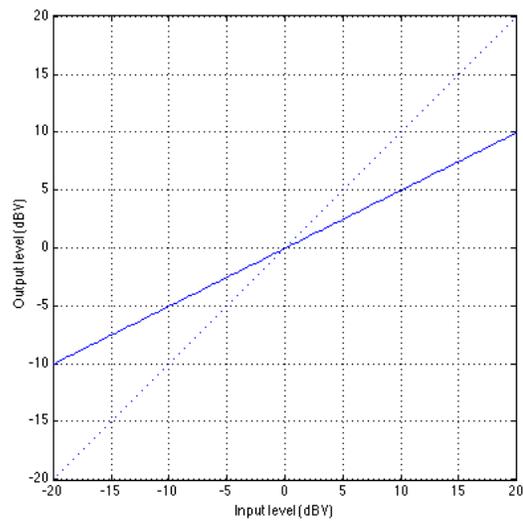


Figure 5.18: The gain response (or transfer function) of a device with a different gain for different input levels. Note that a 2 dB rise in level at the input results in a 1 dB rise in level at the output.

level in decibels to change in output level in decibels. So, if the output goes up 2 dB for every 1 dB increase in level at the input, then we have a 2:1 *compression ratio*. The higher the compression ratio, the greater the effect on the dynamic range.

Notice in Figure 2 that there is one input level (in this case, 0 dBV) that results in a gain of 1 – that is to say that the output is equal to the input. That input level is known as the *rotation point* of the compressor. The reason for this name isn't immediately obvious in Figure 2, but if we take a look at a number of different compression ratios plotted on the same graph as in Figure 3, then the reason becomes clear.

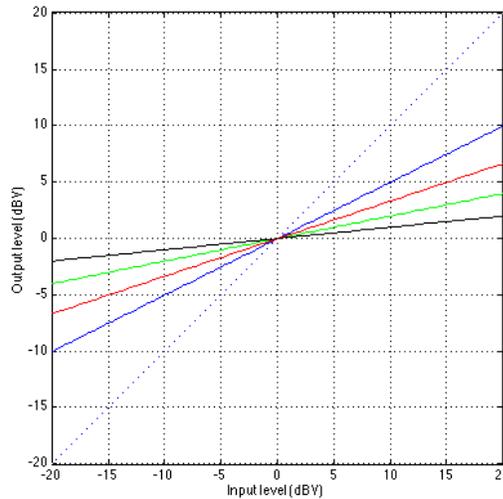


Figure 5.19: The gain response of various compression ratios with the same rotation point (at 0 dBV). Blue = 2:1 compression ratio, red = 3:1, green = 5:1, black = 10:1.

Normally, a compressor doesn't really behave in the way that's seen in any of the above diagrams. If we go back to thinking about listening to the stereo in the car, we actually leave the volume knob alone most of the time, and only turn it down during the really loud parts. This is the way we want the compressor to behave. We'd like to leave the gain at one level (let's say, at 1) for most of the program material, but if things get really loud, we'll start turning down the gain to avoid letting things get out of hand. The gain response of such a device is shown in Figure 4.

The level where we change from being a *linear gain device* (meaning that the gain of the device is the same for all input levels) to being a compressor is called the *threshold*. Below the threshold, the device applies the same

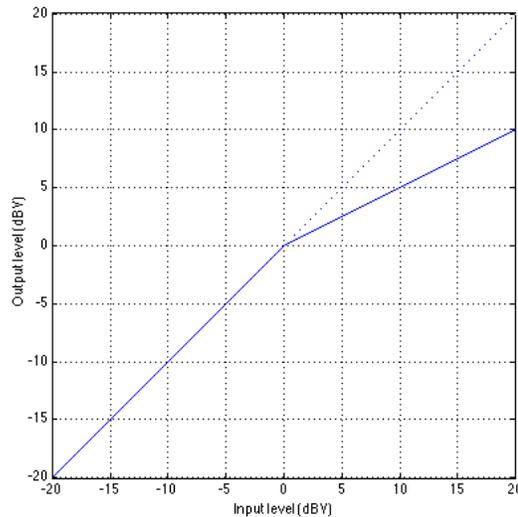


Figure 5.20: A device which exhibits unity gain for input signals with a level of less than 0 dBV and a compression of 2:1 for input signals with a level of greater than 0 dBV.

gain to all signal levels. Above the threshold, the device changes its gain according to the input level. This sudden bend in the transfer function at the threshold is called the *knee* in the response.

In the case of the plot shown in Figure 4, the rotation point of the compressor is the same as the threshold. This is not necessarily the case, however. If we look at Figure 5, we can see an example of a curve where this is illustrated.

This device applies a gain of 5 dB to all signals below the threshold, so an input level of -20 dBV results in an output of -15 dBV and an input at -10 dBV results in an output of -5 dBV. Notice that the threshold is still at 0 dBV (because it is the input level over which the device changes its behaviour). However, now the rotation point is at 10 dBV.

Let's look at an example of a compressor with a gain of 1 below threshold, a threshold at 0 dBV and different compression ratios. The various curves for such a device are shown in Figure 6 below. Notice that, below the threshold, there is no difference in any of the curves. Above the threshold, however, the various compression ratios result in very different behaviours.

There are two basic “styles” in compressor design when it comes to the threshold. Some manufacturers like to give the user control over the threshold level itself, allowing them to change the level at which the compressor

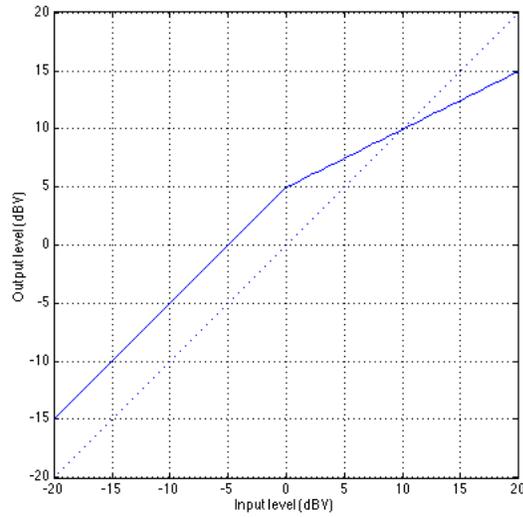


Figure 5.21: An example of a device where the threshold is not the rotation point. The threshold is 0 dBV and the rotation point is 10 dBV.

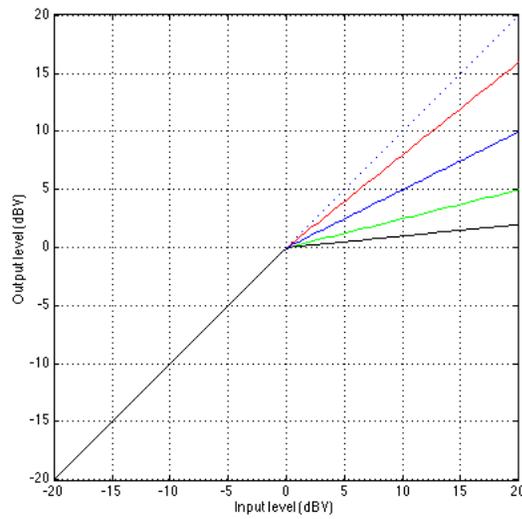


Figure 5.22: A plot showing a number of curves representing various settings of the compression ratio with a unity gain below threshold and a threshold of 0 dBV. red = 1.25:1, blue = 2:1, green = 4:1, black = 10:1.

“kicks in.” This type of compressor typically has a unity gain below threshold, although this isn’t always the case. Take a look at Figure 7. This shows a number of curves for a device with a compression ratio of 2:1, unity gain below threshold and an adjustable threshold level.

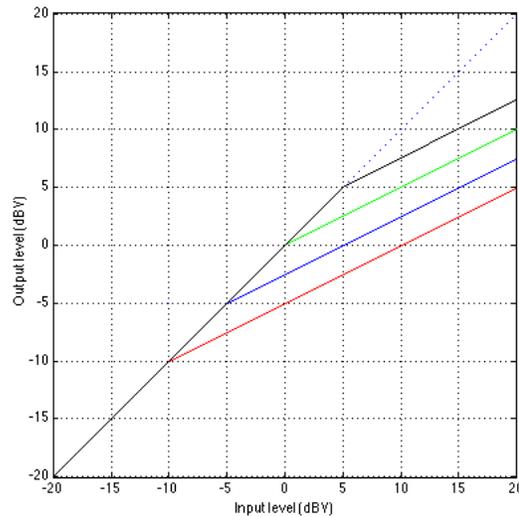


Figure 5.23: A plot showing a number of curves representing various settings of the threshold with a unity gain below threshold and a compression ratio of 2:1. red threshold = -10 dBV, blue threshold = -5 dBV, green threshold = 0 dBV, black threshold = 5 dBV.

The advantage of this design is that the bulk of the signal, which is typically below the threshold, remains unchanged – by changing the threshold level, we’re simply changing the level at which we start compressing. This makes the device fairly intuitive to use, but not necessarily a good design for the final sound quality.

Let’s think about the response of this device (with a 2:1 compression ratio). If the threshold is turned up to 12 dBV, then any signal coming in that’s less than 12 dBV will go out unchanged. If the input signal has a level of 20 dBV, then the output will be 16 dBV, because the input went 8 dB above threshold and the compression ratio is 2:1, so the output goes up 4 dB.

If the threshold is turned down to -12 dBV, then any signal coming in that’s less than -12 dBV will go out unchanged. If the input signal has a level of 20 dBV, then the output will be 4 dBV, because the input went 32 dB above threshold and the compression ratio is 2:1, so the output goes up 16 dB.

So what? Well, as you can see from Figure 7, changing the compression ratio will affect the output level of the loud stuff by an amount that's determined by the relationship between the threshold and the compression ratio.

Consider for a moment how a compressor will be used in a recording situation: we use the compressor to reduce the dynamic range of the louder parts of the signal. As a result, we can increase the overall level of the output of the compressor before going to tape. This is because the spikes in the signal are less scary and we can therefore get closer to the maximum input level of the recording device. As a result, when we compress, we typically have a tendency to increase the input level of the device that follows the compressor. Don't forget, however, that the compressor itself is adding noise to the signal, so when we boost the input of the next device in the audio chain, we're increasing not only the noise of the signal itself, but the noise of the compressor as well. How can we reduce or eliminate this problem? Use compressor design philosophy number 2...

Instead of giving the user control over the threshold, some compressor designers opt to have a fixed threshold and a variable *gain before compression*. This has a slightly different effect on the signal.

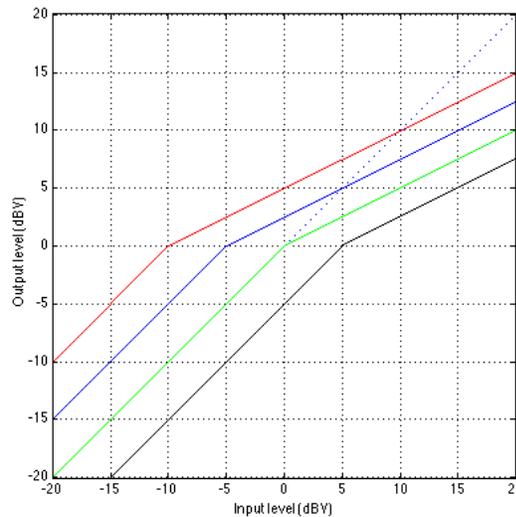


Figure 5.24: A plot showing a number of curves representing various settings of the gain before compression with a fixed threshold. The compression ratio in this example is 2:1. The threshold is fixed at 0 dBV, however, this value does not directly correspond to the input signal level as in Figure 7. The red curve has a gain of 10 dB, blue = 5 dB, green = 0 dB, black = -5 dB.

Let's look at the implications of this configuration using the response in Figure 8 which has a fixed threshold of 0 dBV. If we look at the green curve with a gain of 0 dB, then signals coming in are not amplified or attenuated before hitting the threshold detector. Therefore, signals lower than 0 dBV at the input will be unaffected by the device (because they aren't being compressed and the gain is 0 dB). Signals greater than 0 dBV will be compressed at a 2:1 compression ratio.

Now, let's look at the blue curve. The low-level signals have a constant 5 dB gain applied to them – therefore a signal coming in a -20 dBV comes out at -15 dBV. An input level of -15 dBV results in an output of -10 dBV. If the input level is -5 dBV, a gain of 5 dB is applied and the result of the signal hitting the threshold detector is 0 dBV – the level of the threshold. Signals above this -5 dBV level (at the input) will be compressed.

If we just consider things in the theoretical world, applying a 5 dB gain before compression (with a threshold fixed at 0 dBV) results in the same signal that we'd get if we didn't change the gain before compression, reduced the threshold to -5 dBV and then raised the output gain of the compressor by 5 dB. In the practical world, however, we are reducing our noise level by applying the gain before compression, since we aren't amplifying the noise of the compressor itself.

There's at least one manufacturer that takes this idea one step further. Let's say that you have the output of a compressor being sent to the input of a recording device. If the compressor has a variable threshold and you're looking at the record levels, then the more you turn down the threshold, the lower the signal going into the recording device gets. This can be seen by looking at the graph in Figure 7 comparing the output levels of an input signal with a level of 20 dBV. Therefore, the more we turn down the threshold on the compressor, the more we're going to turn up the input level on the recorder.

Take the same situation but use a compressor with a variable gain before compression. In this case, the more we turn up the gain before compression, the higher the output is going to get. Now, if we turn up the gain before compression, we are going to turn down the input level to the recorder to make sure that things don't get out of hand.

What would be nice is to have a system where all this gain compensation is done for you. So, using the example of a compressor with gain before compression: we turn up the gain before compression by some amount, but at the same time, the compressor turns down its output to make sure that the compressed part of the signal doesn't get any louder. In the case where the compression ratio is 2:1, if we turn up the gain before compression by 10

dB, then the output has to be turned down by 5 dB to make this happen. The output attenuation in dB is equal to the gain before compression (in dB) divided by the compression ratio.

What would this response look like? It's shown in Figure 9. As you can see, changes in the gain before compression are compensated so that the output for a signal above the threshold is always the same, so we don't have to fiddle with the input level of the next device in the chain.

If we were to do the same thing using a compressor with a variable threshold, then we'd have to boost the signal at the output, thus increasing the apparent noise floor of the compressor and making it sound as bad as it is...

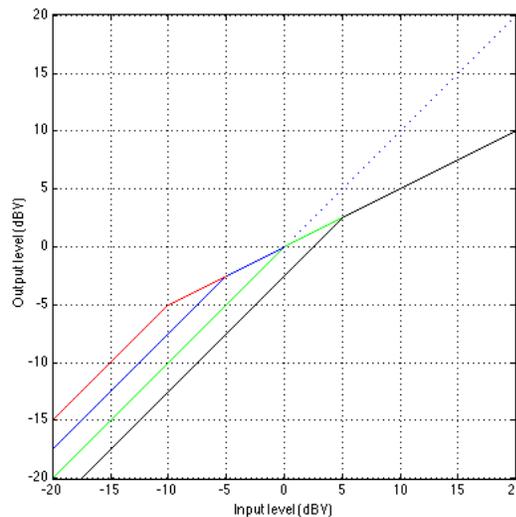


Figure 5.25: The gain response curves for various settings on a compressor with a magic output gain stage that compensates for changes in either the threshold or the gain before compression stage so that you don't have to.

As you can see from Figure 9, the advantage of this system is that adjustments in the gain before compression (or the threshold) don't have any affect on how the loud stuff behaves – if you're past the threshold, you get the same output for the same input.

Compressor gain characteristics

So far we've been looking at the relationship between the output level and the input level of a compressor. Let's look at this relationship in a different

way by considering the gain of the compressor for various signals.

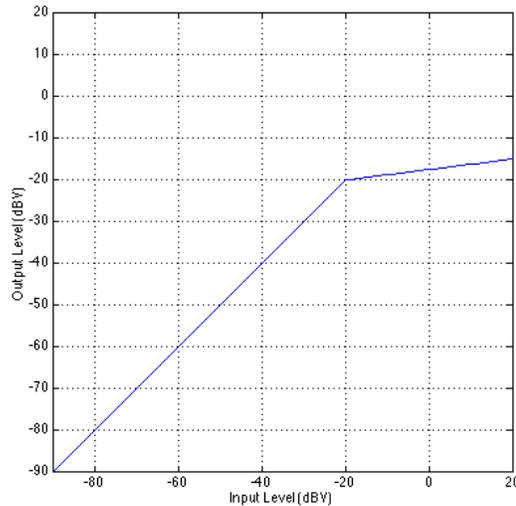


Figure 5.26: The transfer function of a compressor with a gain before compression of 0 dB, a threshold at -20 dBV and a compression ratio of 8:1.

Figure 10 shows the level of the output of a compressor with a given threshold and compression ratio. As we would expect, below the threshold, the output is the same as the input, therefore the gain for input signals with a level of less than -20 dBV in this case is 0 dB – unity gain. For signals above this threshold, the higher the level gets the more the compressor reduces the gain – in fact, in this case, for every 8 dB increase in the input level, the output increases by only 1 dB, therefore the compressor reduces its gain by 7 dB for every 8 dB increase. If we look at this gain vs. the input level, we have a response that is shown in Figure 11.

Notice that Figure 11 plots the gain in decibels vs. the input level in dBV. The result of this comparison is that the gain reduction above the threshold appears to be a linear change with an increase in level. This response could be plotted somewhat differently as is shown in Figure 12.

You'll now notice that there is a rather dramatic change in gain just above the threshold for signals that increase in level just a bit. The result of this is an audible gain change for signals that hover around the threshold – an artifact called *pumping*. This is an issue that we'll deal with a little later.

Let's now consider this same issue for a number of different compression ratios. Figures 13, 14 and 15 show the relationships of 4 different compres-

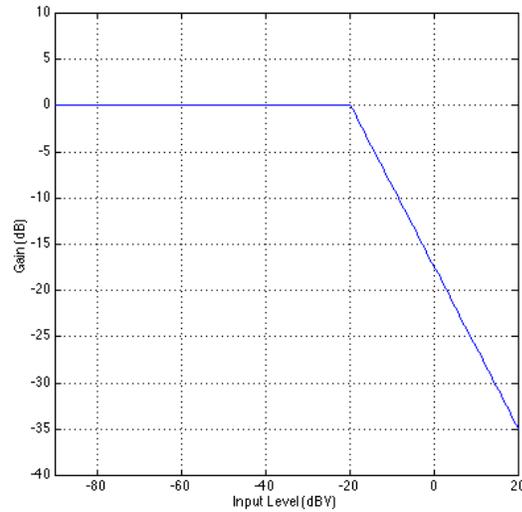


Figure 5.27: The gain vs. input response of a compressor with a gain before compression of 0 dB, a threshold at -20 dBV and a compression ratio of 8:1.

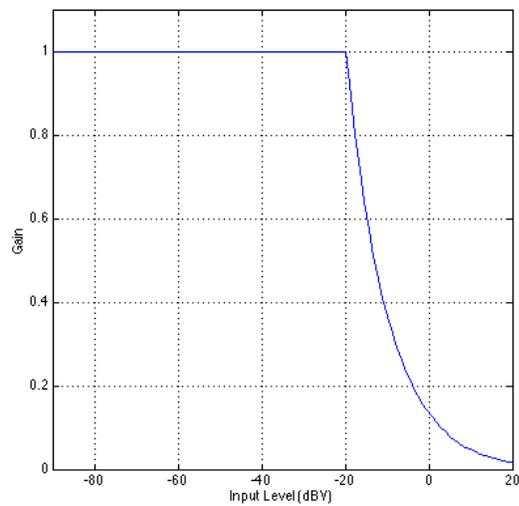


Figure 5.28: The gain vs. input response of a compressor with a gain before compression of 0 dB, a threshold at -20 dBV and a compression ratio of 8:1. Notice that the gain is not plotted in decibels in this case. In effect, Figures 10, 11 and 12 show the same information.

sion ratios with the same thresholds and gains before compression to give you an idea of the change in the gain of the compressor for various ratios.

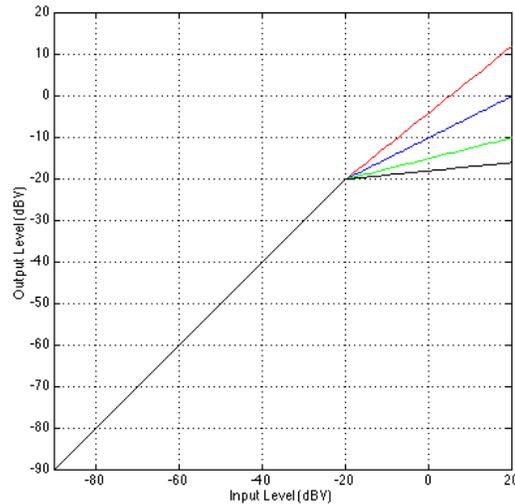


Figure 5.29: The transfer function of a compressor with a gain before compression of 0 dB, a threshold at -20 dBV. Four different compression ratios are shown: red = 1.25:1, blue = 2:1, green = 4:1, black = 10:1.

Soft Knee Compressors

There is a simple solution to the problem of the pumping caused by the sudden change in gain when the signal level crosses the threshold. Since the problem is caused by the fact that the gain change is sudden because the knee in the response curve is a sharp corner, the solution is to soften the sharp corner into a gradual bend. This response is called a *soft knee* for obvious reasons.

Signal level detection

So far, we've been looking at a device that alters its gain according to the input level, but we've been talking in terms of the input level being measured in dBV – therefore, we're thinking of the signal level in V_{RMS} . In fact, there are two types of level detection available – compressors can either respond to the *RMS* value of the input signal, or the *peak* value of the input signal. In fact, some compressors give you the option of selecting some combination of the two instead of just selecting one or the other.

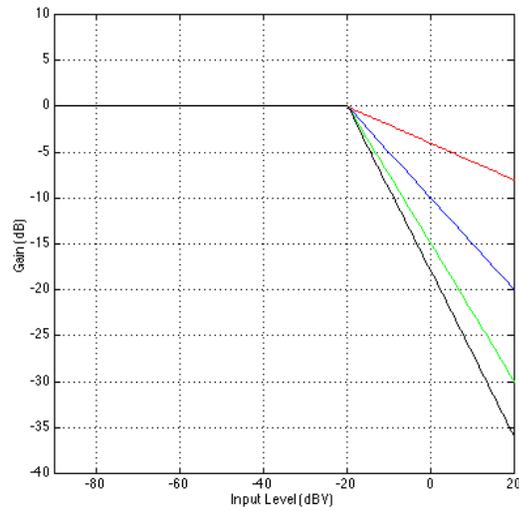


Figure 5.30: The gain vs. input response of a compressor with a gain before compression of 0 dB, a threshold at -20 dBV. Four different compression ratios are shown: red = 1.25:1, blue = 2:1, green = 4:1, black = 10:1.

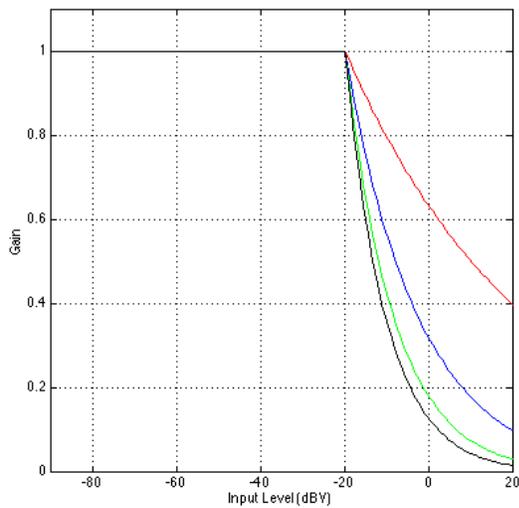


Figure 5.31: The gain vs. input response of a compressor with a gain before compression of 0 dB, a threshold at -20 dBV. Four different compression ratios are shown: red = 1.25:1, blue = 2:1, green = 4:1, black = 10:1.

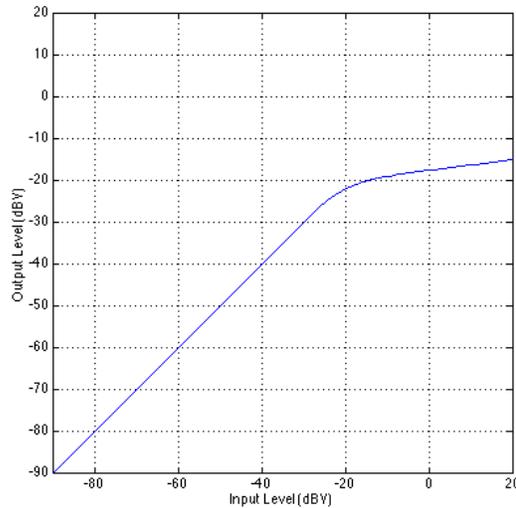


Figure 5.32: The gain vs. input level plot for a soft knee compressor with a gain before compression of 0 dB, a threshold at -20 dBV and a compression ratio of 8:1. Compare this plot to the one in Figure 10.

Probably the simplest signal detection method is the RMS option. As we'll see later, the signal that is input to the device goes to two circuits: one is the circuit that changes the gain of the signal and sends it out the output of the device. The second, known as the control path determines the RMS level of the signal and outputs a control signal that changes the gain of the first circuit. In this case, the speed at which the control circuit can respond to changes in level depends on the time constant of the RMS detector built into it. For more info on time constants of RMS measurements, see Section 2.1.6. The thing to remember is that an RMS measurement is an average of the signal over a given period of time, therefore the detection system needs a little time to respond to the change in the signal. Also, remember that if the time constant of the RMS detection is long, then a short, high level transient will get through the system without it even knowing that it happened.

If you'd like your compressor to respond a little more quickly to the changes in signal level, you can typically choose to have it determine its gain based on the peak level of the signal rather than the RMS value. In reality, the compressor is not continuously looking at the instantaneous level of the voltage at the input – it's usually got a circuit built in that looks at a smoothed version of the absolute value of the signal. Almost all compres-

sors these days give you the option to switch between a peak and an RMS detection circuit.

On high-end units, you can have your detection circuit respond to some mix of the simultaneous peak and RMS values of the input level. Remember from Chapter 2.1.6 that the ratio of the peak to the RMS is called the *crest factor*. This ratio of peak/RMS can either be written as a value from 0 to something big, or it may be converted into a dB scale. Remember that, if the crest factor is near 0 (or -infinity dB), then the RMS value is much greater than the peak value and therefore the compressor is responding to the RMS of the signal level. If the crest factor is a big number (or a smaller number in dB – but much greater than -infinity), then the compressor is responding to the peak value of the input level.

Time Response: Attack and Release

Now that we're talking about the RMS and the smoothed peak of the signal, we have to start considering what time it is. Up to now, we've been only looking at the output level or the gain of the compressor based on a static input level. We're assuming that the only thing we're sending through the unit is a steady-state sine tone. Of course, this is pretty boring to listen to, but if we're going to look at real-world signals, then the behaviour of the compressor gets pretty complicated.

Let's start by considering a signal that's quiet to begin with and suddenly gets louder. For the purposes of this discussion, we'll simulate this with a pulse-modulated sine wave like the one shown in Figure 17.

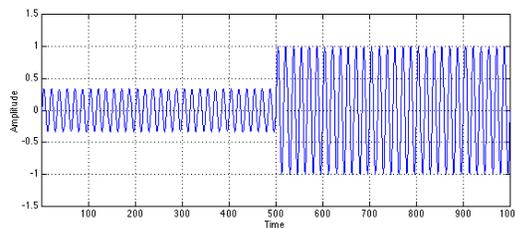


Figure 5.33: A sine wave that is suddenly increased in level from a peak value of 0.33 to a peak value of 1.

Unfortunately, a real-world compressor cannot respond instantaneously to this sudden change in level. In order to be able to do this, the unit would have to be able to see into the future to know what the new peak value of the signal will be before we actually hit that peak. (In fact, some digital

compressors can do this by delaying the signal and turning the present into the past and the future into the present, but we'll pretend that this isn't happening for now...).

Let's say that we have a compressor with a gain before compression of 0 dB and a threshold that's set to a level that's higher than the lower-level signal in Figure 17, but lower than the higher-level signal. So, the first part of the signal, the quiet part, won't be compressed and the later, louder part will. Therefore the compressor will have to have a gain of 1 (or 0 dB) for the quiet signal and then a reduced gain for the louder signal.

Since the compressor can't see into the future, it will respond somewhat slowly to the sudden change in level. In fact, most compressors allow you to control the speed with which the gain change happens. This is called the *attack time* of the compressor. Looking at Figure 18, we can see that the compressor has a sudden awareness of the new level (at Time = 500) but it then settles gradually to the new gain for the higher signal level. This raises a question – the gain starts changing at a known time, but, as you can see in Figure 18, it approaches the final gain forever without really reaching it. The question that's raised is “what is the time of the attack time?” In other words, if I say that the compressor has an attack time of 200 ms, then what is the relationship between that amount of time and the gain applied by the compressor. The answer to this question is found in the chapter on capacitors. Remember that, in a simple RC circuit, the capacitor charges to a new voltage level at a rate determined by the *time constant* which is the product of the resistance and the capacitance. After 1 time constant, the capacitor has charged to 63 % of the voltage being applied to the circuit. After 5 time constants, the capacitor has charged to over 99 % of the voltage, and we consider it to have reached its destination. The same numbers apply to compressors. In the case of an attack time of 200 ms, then after 200 ms has passed, the gain of the compressor will be at 63 % of the final gain level. After 5 times the attack time (in this case, 1 second) we can consider the device to have reached its final gain level. (In fact, it never reaches it, it just gets closer and closer and closer forever...)

What is the result of the attack time on the output of the compressor? This actually is pretty interesting. Take a look at Figure 19 showing the output of a compressor that has the signal in Figure 17 sent into it and responding with the gain in Figure 18. Notice that the lower-level signal goes out exactly as it went in. We would expect this because the gain of the compressor for that portion of the signal is 1. Then the signal suddenly increases to a new level. Since the compressor detection circuit takes a little while to figure out that the signal has gotten louder, the initial new loud

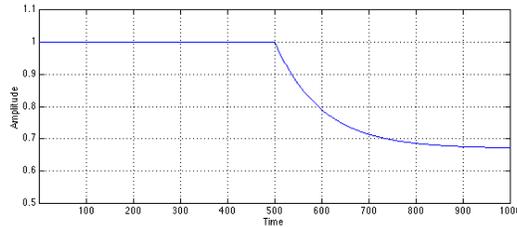


Figure 5.34: The change in gain over time for a sudden increase in signal level going from a signal that's lower than the threshold to one that's higher. This is called the *attack time* of the compressor. (Notice that this looks just like the response of a capacitor being charged to a new voltage level. This is not a coincidence.)

signal gets through, almost unchanged. As we get further and further into the new level in time, however, the gain settles to the new value and the signal is compressed as we would expect. The interesting thing to note here is that a portion of the high-level signal gets through the compressor. The result is that we've created a signal that sounds like more of a transient than the input. This is somewhat contrary to the way most people tend to think that a compressor behaves. The common belief is that a compressor will control all of your high-level signals, thus reducing your dynamic range – but this is not exactly the case as we can see in this example. In fact, it may be possible that the perceived dynamic range is greater than the original because of the accents on the transient material in the signal.

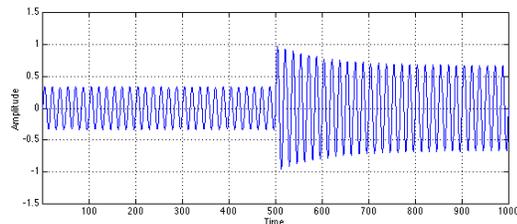


Figure 5.35: The output of a compressor which is fed the signal shown in Figure 17 and responds with the gain shown in Figure 18.

Similarly, what happens when the signals decreases in level from one that is being compressed to one that is lower than the threshold? Again, it takes some time for the compressor's detection circuit to realize that the level has changed and therefore responds slowly to fast changes. This response time is called the *release time* of the compressor. (Note that the release time is

measured in the same way as the attack time – it’s the amount of time it takes the compressor to get to 63% of its intended gain.)

For example, we’ll assume that the signal in Figure 20 is being fed into a compressor. We’ll also assume that the higher-level signal is above the compression threshold and the lower-level signal is lower than the threshold.

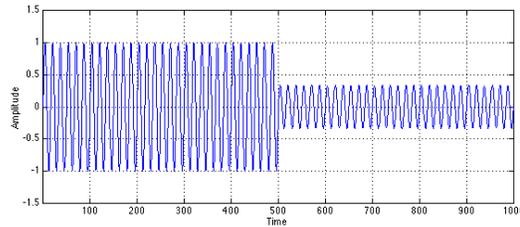


Figure 5.36: A sine wave that is suddenly decreased in level from a peak value of 1 to a peak value of 0.33.

This signal will result in a gain reduction for the first part of the signal and no gain reduction for the latter part, however, the release time of the compressor results in a transition time from these two states as is shown in Figure 21.

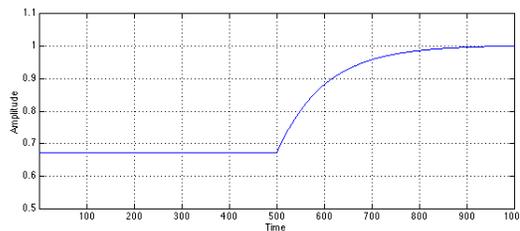


Figure 5.37: The change in gain over time for a sudden decrease in signal level going from a signal that’s higher than the threshold to one that’s lower. This is called the release time of the compressor. (Notice that this looks just like the response of a capacitor being charged to a new voltage level. This is not a coincidence.)

Again, the result of this gain response curve is somewhat interesting. The output of the compressor will start with a gain-reduced version of the louder signal. When the signal drops to the lower level, however, the compressor is still reducing the gain for a while, therefore we wind up with a compressed signal that’s below the threshold – a signal that normally wouldn’t be compressed. As the compressor figures out that the signal has dropped, it releases its gain to return to a unity gain, resulting in an output

signal shown in Figure 22.

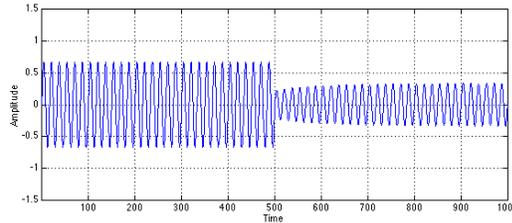


Figure 5.38: The output of a compressor which is fed the signal shown in Figure 20 and responds with the gain shown in Figure 21.

Just for comparison purposes, Figures 23 and 24 show a number of different attack and release times.

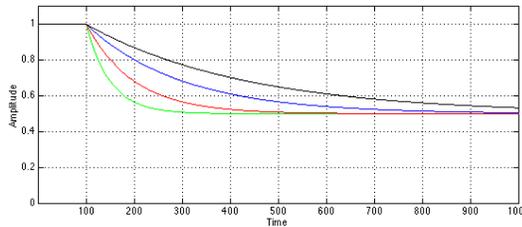


Figure 5.39: Four different attack times for compressors with the same thresholds and compression ratios.

One last thing to discuss is a small issue in low-cost RMS-based compressors. In these machines, the attack and release times of the compressor are determined by the time constant of the RMS detection circuit. Therefore, the attack and release times are identical (normally, we call them “symmetrical”) and not adjustable. Check your manual to see if, by going into RMS mode, you’re defeating the attack time and release time controls.

5.2.2 How compressors compress

On its simplest level, a compressor can be thought of as a device which controls its gain based on the incoming signal. In order to do this, it takes the incoming audio and sends it in two directions, along the *audio path*, which is where the signal goes in, gets modified and comes out the output; and the *control path*, (also known as a *side chain*) where the signal comes in,

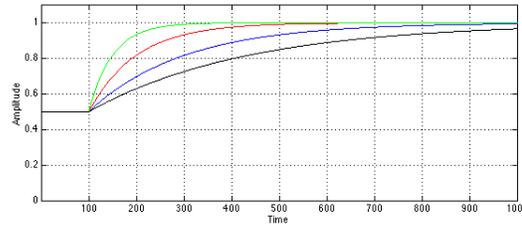


Figure 5.40: Four different release times for compressors with the same thresholds and compression ratios.

gets analysed and converted into a different signal which is used to control the gain of the audio path.

As a result, we can think of a basic compressor as is shown in the block diagram below.

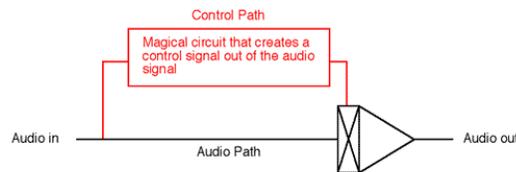


Figure 5.41: INSERT CAPTION

Notice that the input gets split in two directions right away, going to the two different paths.

At the heart of the audio path is a device we haven't seen before – it's drawn in block diagrams (and sometimes in schematics) as a triangle (so we know right away it's an amplifier of some kind) attached to a box with an "X" through it on the left. This device is called a *voltage controlled amplifier* or *VCA*. It has one audio input on the left, one audio output on the right and a *control voltage* (or *CV*) input on the top. The amplifier has a gain which is determined by the level of the control voltage. This gain is typically applied to the current through the VCA, not the voltage – this is a new concept as well... but we'll get to that later.

If you go to the VCA store and buy a VCA, you'll find out that it has an interesting characteristic. Usually, it will have a logarithmic change in gain for a linear change in voltage at the control voltage input. For example, one particular VCA from THAT corporation has a gain of 0 dB (so input =

output) if the CV is at 0 V. If you increase the CV by 6 mV, then the gain of the VCA goes down by 1 dB.

So, for that particular VCA, we could make the following table which we'll use later on.

| Control Voltage (mV) | Gain of Audio signal (dB) |
|----------------------|---------------------------|
| -12 | +2 |
| -6 | +1 |
| 0 | 0 |
| +6 | -1 |
| +12 | -2 |

Table 5.3: INSERT CAPTION HERE

The only problem with the schematic so far is that the VCA is a current amplifier not a voltage amplifier. Since we prefer to think in terms of voltage most of the time, we'll need to convert the voltage signal that we're feeding into the compressor into a current signal of the same shape. This is done by sticking the VCA in the middle of an inverting amplifier circuit as shown below:

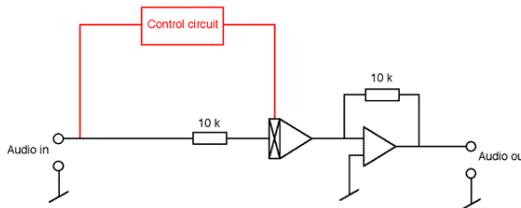


Figure 5.42: INSERT CAPTION

Here's an explanation of why we have to build the circuit like this. Remember back to the first stuff on op amps – one of the results of the feedback loop is that the voltage level at the negative input leg **MUST** be the same as the positive input leg. If this wasn't the case, then the huge gain of the op amp would result in a clipped output. So, we call the voltage level at the negative input “virtual ground” because it has the same voltage level as ground, but there's really no direct connection to ground. If we assume that the VCA has an impedance through it of 0Ω (a pretty safe assumption), then the voltage level at the signal input of the VCA is also the same as ground. Therefore the current through the resistor on the left in the above schematic

is equal to the Audio in voltage divided by the resistor value. Now, if we assume that the VCA has a gain of 0 dB, then the current coming out of it equals the current going into it. We also happen to remember that the input impedance of the op amp is infinite, therefore all the current on the wire coming out of the VCA must go through the resistor in the feedback loop of the op amp. This results in a voltage drop across it equal to the current multiplied by the resistance.

Let's use an example. If the Audio in is 1 Vrms, then the current through the input resistor on the left is 0.1 mA rms. That current goes through the VCA (which we'll assume for now has a gain of 0 dB) and continues on through the feedback resistor. Since the current is 0.1 mA rms and the resistor is 10 k Ω , then the voltage drop across it is 1 Vrms. Therefore, the Audio out is 1 Vrms but opposite in polarity to the input. This is exactly the same as if the VCA was not there.

Now take a case where the VCA has a gain of +6 dB for some reason. The voltage level at its input is 0 V (virtual ground) so the current through the input resistor is still 0.1 mA rms (for a 1 V rms input signal). That current gets multiplied by 2 in the VCA (because it has a gain of +6 dB) making it 0.2 mA rms. This all goes through the feedback resistor which results in a voltage drop of 10k * 0.2 mA rms = 2 V rms (and opposite in polarity). Ah hah! The gain applied to the current by the VCA now shows up as a gain on the voltage at the output. Now all we need to do is to build the control circuitry and we have a compressor...

The control circuitry needs a number of things: a level detector, compression ratio adjustment, threshold, threshold level adjustment and some other stuff that we'll get to. We'll take them one at a time. For now, let's assume that we have a compressor which is only looking at the RMS level to determine its compression characteristics.

RMS Detector

This is an easy one. You buy a chip called an RMS detector. There's more stuff to do once you buy that chip to make it happy, but you can just follow the manufacturer's specifications on that one. This chip will give you a DC voltage output which is determined by the logarithmic level of an AC input. For example, using a THAT Corp. chip again... The audio input of the chip is measured relative to 0.316 V rms (which happens to be -10 dBV). If the RMS level of the audio input of the chip is 0.316 V rms, then the output of the chip is 0 V DC. If you increase the level of the input signal by 1 dB, the output level goes up by 6 mV. Conversely, if the input level goes down by 1

dB, then the output goes down by 6 mV. An important thing to notice here is that a logarithmic change in the input level results in a linear change in the output voltage. So, we build a table for future reference again:

| Input level (dBV) | Output Voltage (mV) |
|-------------------|---------------------|
| -8 | +12 |
| -9 | +6 |
| -10 | 0 |
| -11 | -6 |
| -12 | -12 |

Table 5.4: INSERT CAPTION HERE

Now, what would happen if we took the output from this RMS detector and connected it directly to the control voltage input of the VCA like in the diagram below?

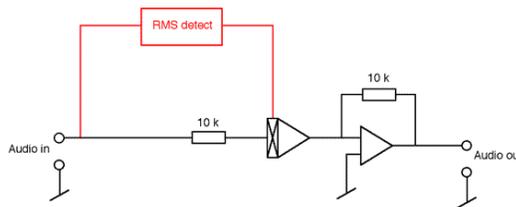


Figure 5.43: INSERT CAPTION

Well, if the input level to the whole circuit was -10 dBV, then the RMS detector would output a 0 V control voltage to the CV input of the VCA. This would cause it to have a gain of 0 dB and its output would be -10 dBV. BUT, if the input was -11 dBV, then the RMS detector output would be -6 mV making the VCA gain go up by 1 dB, raising the output level to -10 dBV. Hmmmmm... If the input level was -9 dBV, then the RMS detector's output goes to 6 mV and the VCA gain goes to -1 dB, so the output is -10 dBV. Essentially, no matter what the input level was, the output level would always be the same. That's a compression ratio of infinity : 1.

Although the circuit above would indeed compress with a ratio of infinity : 1, that's not terribly useful to us for a number of reasons. Let's talk about how to reduce the compression ratio. Take a look at this circuit.

If the potentiometer has a linear scale, and the wiper is half-way up the pot, then the voltage at the wiper is one half the voltage applied to the top

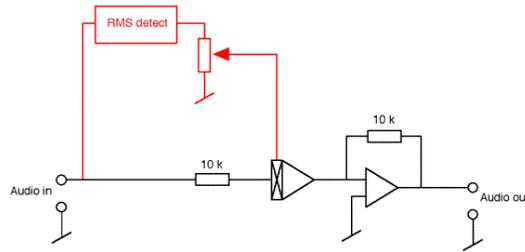


Figure 5.44: INSERT CAPTION

of the pot. This means, in turn, that the voltage applied to the CV input of the VCA is one half the voltage output from the RMS detector. How does this affect us? Well, if the input level of the circuit is -10 dBV, then the RMS detector outputs 0 V, the wiper on the pot is at 0 V and the gain of the VCA is 0 dB, therefore the output level is -10 dBV. If, however, the input level goes up by 1 dB (to -9 dBV), then the RMS detector output goes up by 6 mV, the wiper on the pot goes up by 3 mV, therefore the gain of the VCA goes down by 0.5 dB and the output level is -9.5 dB. So, for a 2 dB change in level at the input, we get a 1 dB change in level at the output – in other words, a $2:1$ compression ratio.

If we put the pot at another location, we change the ratio of the voltage at the top of the pot (which is dependent on the input level to the RMS detector) to the gain (which is controlled by the wiper voltage). So, we have a variable compression ratio from $1:1$ (no compression) to infinity: 1 (complete limiting) and a rotation point at -10 dBV. This is moderately useful, but real compressors have a threshold. So – how do we make this happen?

Threshold

The first thing we'll need to make a threshold detection circuit is a way of looking at the signal and dividing it into a low voltage area (in which nothing gets out of the circuit) and a high area (in which the output = the input to the circuit). We already looked at how to do this in a rather crude fashion – it's called a half-wave rectifier. Since the voltage of the wiper on the pot is going positive and negative as the input signal goes up and down respectively, all we need to do is rectify the signal after the wiper so that none of the negative voltage gets through to the CV input of the VCA. That

way, when the signal is low, the gain of the VCA will be 0 dB, leaving the signal unaffected. When the signal goes positive, the rectifier lets the signal through, the gain of the VCA goes down and the compressor compresses.

One way to do this would simply be to put a diode in the circuit pointing away from the wiper. This wouldn't work very well because the diode would need the 0.7 V difference across it to turn on in the first place. Also, the turn-on voltage of the diode is a little sloppy, so we wouldn't know exactly what the threshold was (but we'll come back to this later). what we need then, is something called a *precision rectifier* – a circuit that looks like a perfect diode. This is pretty easy to build with a couple of diodes and an op amp as is shown in the circuit below.

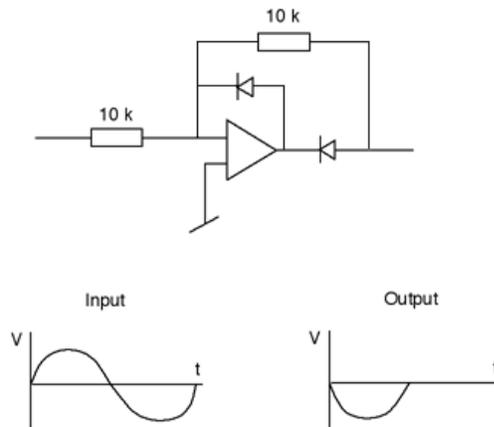


Figure 5.45: INSERT CAPTION

Notice that the circuit has two effects – the first is that it is a half-wave rectifier, so only the positive half of the input gets through. The second is that it is an inverting amplifier, so the output is opposite in polarity to the input – therefore, in order to get things back in the right polarity, we'll have to flip the polarity once again with a second inverting amplifier with unity gain.

If we add this circuit between the wiper and the VCA CV input like the diagram shown below, what will happen?

Now, if the input level is -10 dBV or lower, the output of the RMS detector is 0 V or lower. This will result in the output of the half-wave rectifier being 0 V. This will be multiplied by -1 in the polarity inversion,

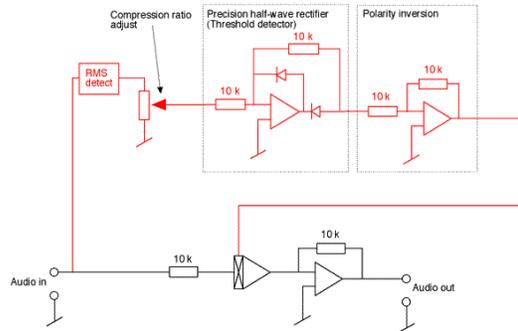


Figure 5.46: INSERT CAPTION

resulting in a level of 0 V at the CV input of the VCA. This means that if the input signal is -10 dBV or lower, there is no gain change. If, however, the input level goes above -10 dBV, then the output of the RMS detector goes up. This gets through the rectifier and comes out multiplied by -1, so for every increase in 1 dB above -10 dBV at the input, the output of the rectifier goes DOWN by an amount determined by the position of the wiper. This is multiplied by -1 again at the polarity inversion and sent to the CV input of the VCA causing a gain change. So, we have a threshold at -10 dBV at the input. But, what if we wanted to change the threshold level?

In order to change the threshold level, we have to trick the threshold detection circuit into thinking that the signal has reached the threshold before it really has. Remember that the output of the RMS detection circuit (and therefore the wiper on the pot) is DC (well, technically speaking, it varies if the input signal's RMS level varies, but we'll say it's DC for now). So, we need to mix some DC with this level to give the input to the rectification circuit an additional boost. For example, up until now, the threshold is -10 dBV at the input because that's where the output of the RMS detector crosses from negative to positive voltage. If we wanted to make the threshold -20 dBV, then we'd need to find out the output of the RMS detector if the signal was -20 dBV (that would be -60 mV because it's 6 mV per dB and 10 dB below -10 dBV) and add that much DC voltage to the signal before sending it into the rectification stage. There are a couple of ways to do this, but one efficient way is to combine an inverting mixer circuit with the half-wave rectifier that's already there.

The threshold level adjustment is just a controllable DC level which is mixed with the DC level coming out of the RMS detector. One important

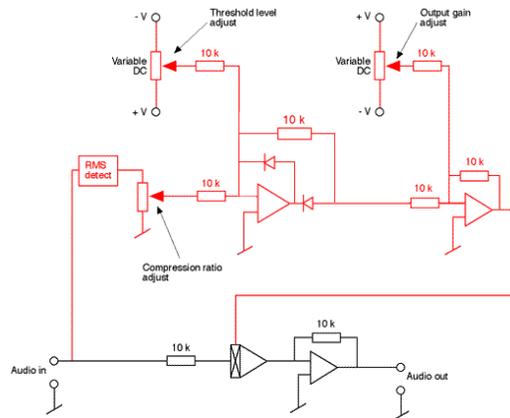


Figure 5.48: INSERT CAPTION

it's basically the same as an output level knob. In fact, it IS an output level knob.

Everything I've said here is basically a lead-up to the pdf file below from THAT Corp. It's a good introduction to how a simple RMS-based compressor works. It includes all the response graphs that I left out here, and goes a little further to explain how to include a soft knee for your circuit. Definitely recommended reading if you're planning on learning more about these things...

5.2.3 Suggested Reading List

Basic Compressor/Limiter Design – from THAT Corporation www.thatcorp.com/datashts/an100a.pdf
 Users Manual for the GML 8900 Dynamic Range Controller www.gmlinc.com/8900.pdf

5.3 Analog Tape

NOT YET WRITTEN

5.3.1 Suggested Reading List

5.4 Noise Sources

5.4.1 Introduction

Noise can very basically be defined as anything in the audio that we don't want.

There are two basic sources of noise in an audio chain, internal noise and noise from external sources.

Internal noise is inherent in the equipment, we can either mask it with better gain structures (i.e. more gain earlier in the chain rather than later... You want to be almost clipping the output of your mic preamplifier and attenuating as you go through the remainder of your gear. This way, you're not boosting the noise of all previous equipment, except of course for the self-noise of the microphones) or just buy better equipment...

Noise from external sources is caused by *Electromagnetic Interference*, also known as *EMI*, requires 3 simultaneous things...

1. Source of electromagnetic noise
2. Transmission medium in which the noise can propagate (usually a piece of wire, for example)
3. A receiver sensitive to the noise

5.4.2 EMI Transmission

EMI has four possible means of transmission: common impedance coupling, electrical field coupling, magnetic field coupling and electromagnetic radiation.

Common Impedance Coupling

This occurs when there is a shared wire between the source and the receiver.

Let's say, for example, that two units, a microphone preamplifier and a food processor, are both connected to the same power bar. This means that, from the power bar to the earth, the two devices are making use of the same wires for their references to ground. This means that, in the case of this shared ground wire, they are coupled by a common impedance to the earth (the impedance of the ground wire from the power bar to the earth). If one of the two devices (the food processor maybe...) is a source of noise in the form of alternating current, and therefore voltage (because the wire has some impedance and $V=IR$) at the ground connection in the powerbar, then the "ground" voltage at the microphone preamp will modulate. If this is supposed to be a DC 0 V forever relative to the rest of the studio, and it

isn't, then the mic pre will output the audio signal plus the noise coupled from the common impedance.

Electrical field coupling

This is determined by the capacitance between the source and the receiver or their transmission media.

Once upon a time, we looked at the construction of a capacitor as being two metal plates side by side but not touching each other. If we push electrons into one of the plates, we'll repel electrons out of the other plate and we appear to have "current" flowing through the capacitor. The higher the rate of change of the current flowing in and out of the plate, the easier it is to move current in and out of the other plate.

Consider that if we take any two pieces of metal and place them side by side without touching, we're going to create a capacitor. This is true of two wires side by side inside a mic cable, or two wires resting next to each other on the floor or so on. If we send a high frequency through one of the wires, and we have some small capacitance between that wire and the "receiver" wire, we'll get some signal appearing on the latter.

The level of this noise is proportional to:

1. The area that the source and receiver share (how big the plates are, or in this case, how long the wires are side by side)
2. The frequency of the noise
3. The amplitude of the noise voltage (note that this is "voltage")
4. The permittivity of the medium (dielectric) between the two

The level of the noise is inversely proportional to

1. the square of the distance between the sender and the receiver (or in some cases their connected wires)

Magnetic field coupling

This is determined by the mutual inductance between the source and receiver.

Remember back to the chapter where we talked about the right hand rule and how, when we send AC through a wire, we generate a pulsing magnetic field around it whose direction is dependent on the direction of the current and whose amplitude (or distance from the wire) is proportional to the level of the current. If we place another wire in this moving magnetic field, we will induce a current in the second wire – which is how a transformer works.

Although this concept is great for transformers, it's a problem when we have microphone cables sitting next to high-current AC cables. If there's lots of current through the AC cable at 60 Hz, and we place the mic cable in the resulting generated magnetic field, then we will induce a current in the mic cable which is then amplified by the mic preamplifier to an audible level. This is bad, but there are a number of ways to avoid it as we'll see.

The level of this noise is proportional to:

1. The loop area of the receiver (therefore it's best not to create a loop with your mic cables)
2. The frequency of source
3. The current of source
4. The permeability of the medium between them

The level of this noise is inversely proportional to:

1. the square of the distance between them (so you should keep mic cables away from AC cables! – and if they have to cross, cross them at a right angle to each other to minimize the effects)

Electromagnetic radiation

This occurs when the source and receiver are at least 1/6th of a wavelength apart (therefore the receiver is in the far field – where the wavefront is a plane and the ratio of the electrostatic to the electromagnetic field strengths is constant)

An example of noise caused by electromagnetic radiation is RFI (Radio Frequency Interference) caused by radio transmitters, CB etc.

5.4.3 Suggested Reading List

www.engineeringharmonics.com

Grounding and Shielding for Sound and Video Philip Giddings

5.5 Reducing Noise – Shielding, Balancing and Grounding

How do we get rid of the noise caused by the four sources described in Chapter 5.4? There are three ways: shielding, balancing and grounding

5.5.1 Shielding

This is the first line of defense against outside noise caused by high-frequency electrical field and magnetic field coupling as well as electromagnetic radiation. The theory is that the shielding wire, foil or conduit will prevent the bulk of the noise coming in from the outside.

It works by relying on two properties

1. Reflection back to the outside world where it can't do any harm... (and, to a small extent, re-reflection within the shield, but this is a VERY small extent)
2. Absorption – where the energy is absorbed by the shield and sent to ground.

The effectiveness of the shield is dependent on its:

1. Thickness – the thinner the shield the less effective. This is particularly true of low-frequency noise... Aluminum foil shield works well at rejecting up to 90 dB at frequencies above 30 MHz, but it's inadequate at fending off low-frequency magnetic fields (in fact it's practically transparent below 1 kHz), We rely on balancing and differential amplifiers to get rid of these.
2. Conductivity – the shield must be able to sink all stray currents to the ground plane more easily than anything else.
3. Continuity – we cannot break the shield. It must be continuous around the signal paths, otherwise the noise will leak in like water into a hole in a boat. Don't forget that the holes in your equipment for cooling, potentiometers and so on are breaks in the continuity. General guideline: keep the diameter of your holes at less than 1/20 of the wavelength of the highest frequency you're worried about to ensure at least 20 dB of attenuation. Most high-frequency noise problems are caused by openings in the shield material.

5.5.2 Balanced transmission lines

It is commonly believed even at the highest levels of the audio world that a balanced signal and a *differential* or *symmetrical* signal are the same thing.

This is not the case. A differential (or symmetrical) signal is one where one channel of audio is sent as two voltages on two wires, which are usually twisted together. These two signals are identical in every respect with the exception that they are opposite in polarity. These signals are known by such names as “inverting and non-inverting” or “Live and Return” – the “L” and “R” in XLR (the X is for eXternal – the ground). They are received by a differential amplifier which subtracts the return from the live and produces a single signal with a gain of 6 dB (since a signal minus its negative self is the same as 2 times the signal and therefore 6 dB louder). The theoretical benefit of using this system is that any noise that is received on the transmission cables between the source and the receiver is (theoretically) identical on both wires. When these two versions of the noise arrive at the receiver’s differential amplifier, they are theoretically eliminated since we are subtracting the signal from a copy of itself. This is what is known as the Common Mode Rejection done by the differential input. The ability of the amplifier to reject the common signals (or mode) is measured as a ratio between the output and one input leg of the differential amplifier and is therefore called the Common Mode Rejection Ratio (CMRR).

Having said all that, I want to come back to the fact that I used the word “theoretical” a little too often in the last paragraph. The amount and quality of the noise on those two transmission lines (the live and the return) in the so-called “balanced” wire is dependent on a number of things.

1. The proximity to the noise source. This is what is causing the noise to wind up on the two wires in the first place. If the source of the noise is quite near to the receiving wire (for example, in the case of a high-voltage/current AC cable sitting next to a microphone cable) then the closer wire within our “balanced” pair will receive a higher level of noise than the more distant wire. Remember that this is inversely proportional to the square of the distance, so it can cause a major problem if the AC and mic cables are sitting side by side. The simplest way to avoid this difference in the noise on the two wires is to wrap them together. This ensures that, over the length of the cable, the two internal wires average out to being equally close to the AC cable and therefore we pick up the same amount of noise – therefore the differential amplifier will cancel it.

2. The termination impedance of the two wires. In fact, technically speaking, a balanced transmission line is one where the impedance between each of the two wires and ground is identical for each end of the transmission. Therefore the impedance between live and ground is identical to the impedance between return and ground at the output of the sending device and at the input of the receiving device. This is not to say that the in-

put and output impedances are matched. They are not. If the termination impedances are mismatched, then the noise on each of the wires will be different and the differential amplifier will not be subtracting a signal from a copy of itself – therefore the noise will get through. Some manufacturers are aware of this and save themselves some money while still providing you with a balanced output. Mackie consoles, for example, drive the signal on the tip of their 1/4" balanced outputs, but only put a resistor between the ring and ground (the sleeve) on the same output connector. This is still a balanced output despite the fact that there is no signal on the ring because the impedance between the tip and ground matches the impedance between the ring and ground (they're careful about what resistor they put in there...)

5.5.3 Grounding

The grounding of audio equipment is there for one primary purpose: to keep you alive. If something goes horribly wrong inside one of those devices and winds up connecting the 120 V AC from the wall to the box (chassis) itself, and you come along and touch the front panel while standing in a pool of water, YOU are the path to ground. This is bad. So, the manufacturers put a third pin on their AC cables which is connected to the chassis on the equipment end, and the third pin in the wall socket.

Let's look at how the wiring inside a wall socket is connected to begin with. Take a look at Figures 5.49 to 5.53.



Figure 5.49: A typical North American electrical outlet showing the locations of the two spade connections and the third, round ground pin. Note that the orange cable contains three independent conductors, each with a different coloured insulator.

The third pin in the wall socket is called the *ground bus* and is connected to the electrical breaker box somewhere in the facility. All of the ground busses connect to a primary ground point somewhere in the building. This is the point at which the building makes contact with the earth through a

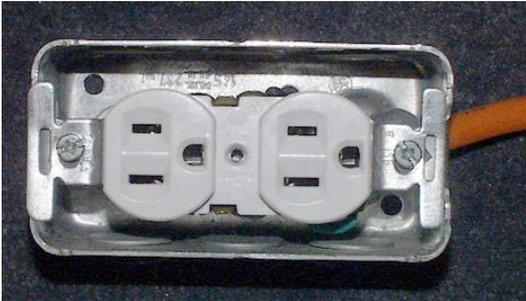


Figure 5.50: The same outlet as is shown in Figure 5.49 with the safety faceplate removed.



Figure 5.51: PUT CAPTION HERE

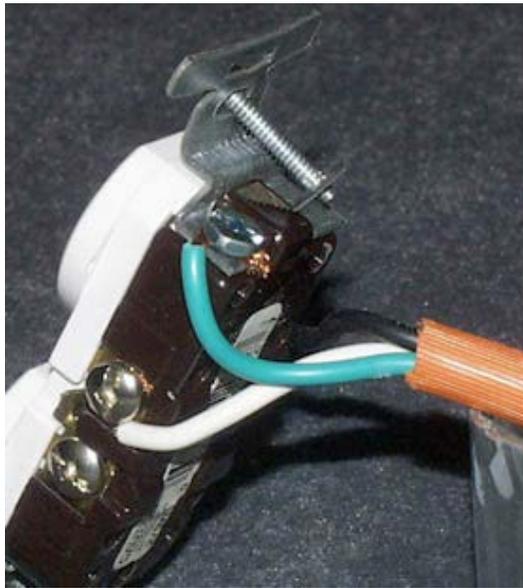


Figure 5.52: The beginnings of the inside of the outlet showing the connection of the white and green wires to the socket. The white wire is at 0 V and is connected in parallel through the brass plate on the side of the socket to the two larger spades. The green wire is also at 0 V and is connected in parallel to the round safety ground pin as well as the box that houses the socket. The third lack wire which is at $120 V_{RMS}$ is connected to the socket on the opposite side and cannot be seen in this photograph.



Figure 5.53: The socket completely removed from the housing.

spike or piling called the grounding electrode. The wires which connect these grounds together MUST be heavy-gauge (and therefore very low impedance) in order to ensure that they have a MUCH lower impedance than you when you and it are a parallel connection to ground. The lower this impedance, the less current will flow through you if something goes wrong.

MUCH MORE TO COME!

5.5.4 Suggested Reading List

Grounding and Shielding for Sound and Video by Philip Giddings www.engineerharmonics.com

Hum and Buzz in Unbalanced Interconnect Systems by Bill Whitlock

www.jensen-transformers.com/an/an004.pdf

Sound System Interconnection www.rane.com/note110.html

Grounding www.trinitysoundcompany.com/grounding.html

Ground loop problems and how to get rid of them www.hut.fi/Misc/Electronics/docs/groundlo

Considerations in Grounding and Shielding Audio Devices www.rane.com/pdf/groundin.pdf

A Clean Audio Installation Guide – from Benchmark Media www.benchmarkmedia.com/appno
a/caig

Fundamentals of Studio Grounding, Richard Majestic: Broadcast Engineering, April 1992

The Proper Use of Grounding and Shielding, Philip Giddings: Sound and Video Contractor, September 20, 1995

| Conductor Out From | Low Dynamic Range
(< 60 dB) | Med DR
(60 to 80 dB)
Low EMI | Med DR
(60 to 80 dB)
High EMI | High DR
(> 80 dB)
Low EMI | High DR
(> 80 dB)
High EMI |
|----------------------------|--------------------------------|------------------------------------|-------------------------------------|---------------------------------|----------------------------------|
| Ground Electrode | 6 | 2 | 00 | 00 | 0000 |
| Master Bus | 10 | 8 | 6 | 4 | 0 |
| Local Bus | 14 | 12 | 12* | 12* | 10* |
| Max. Resist. for any cable | 0.5Ω | 0.1Ω | 0.01Ω | 0.001Ω | 0.0001Ω |

Table 5.5: Suggested technical ground conductor sizes. This table is from "Audio Systems: Design and Installation" by Philip Giddings (Focal Press, 1990, ISBN 0-240-80286-1). If you're installing or maintaining a studio, you should own this book. (*Do not share ground conductors – run individual branch grounds. In all cases the ground conductor must not be smaller than the neutral conductor of the panel it services.)

5.6 Microphones – Transducer type

5.6.1 Introduction

A microphone is one of a small number of devices used in audio that can be called a *transducer*. Generally speaking, a transducer is any device that converts one kind of energy into another (for example, electrical energy into mechanical energy). In the case of microphones, we are converting mechanical energy (the movement of the air particles due to sound waves) into electrical energy (the output of the microphone). In order to choose and use your microphones effectively for various purposes, you should know a little about how this magical transformation occurs.

5.6.2 Dynamic Microphones

Back in the chapter on induction and transformers, we talked about an interesting relationship between magnetism and current. If you have a magnetic field, which is comprised of what we call magnetic lines of force all “pointing” in the same direction, and you move a piece of wire through it so that the wire cuts the lines of force perpendicularly, then you’ll generate a current in the wire. (If this is coming as a surprise, then you should read Chapters 2.6 and 2.7.)

Dynamic microphones rely on this principal. Somewhere inside the microphone, there’s a piece of metal or wire that’s sitting in a strong magnetic field. When a sound wave hits the microphone, it pushes and pulls on a membrane called a *diaphragm* that will, in turn, move back and forth proportionally to some component of the particle movement (either the pressure or the velocity – but we’ll talk more about that later). The movement of the diaphragm causes the piece of metal to move in the magnetic field, thus producing current that is representational of the movement itself. The result is an electrical representation of the sound wave where the electrical energy is actually converted from mechanical energy of the air molecules.

One important thing to note here is the fact that the current that is generated by the metal moving in the magnetic field is proportional to the velocity at which it’s moving. The faster it moves, the bigger the current. As a result, you’ll often hear dynamic microphones referred to as *velocity* microphones. The problem with this name is that some people like using the term “velocity microphone” to mean something completely unrelated – and as a result, people get very confused when you go from book to book and see the term in multiple places with multiple meanings. For a further

discussion on this topic, see the section on Pressure Gradient microphones in Section .

Ribbon Dynamic Microphones

The simplest design of dynamic transducer we can make is where the diaphragm *is* the piece of metal that's moving in the magnetic field. Take a strip of aluminium a couple of μm (micrometers) thick, 2 to 4 mm wide and a couple of centimeters long and bend it so that it's corrugated (see Figure 5.54) to make it a little stiff across the width. This will be the diaphragm of the microphone. It's nice and light, so it moves very easily when the sound wave hits it. Now we'll support the microphone from the top and bottom and hang it in between the North and South poles of a strong magnet as shown in Figure 5.54.

Referring to the construction in Figure 5.54: if a sound wave with a positive pressure hits the front of the diaphragm, it moves backwards and generates a current that goes up the length of the aluminium (you can double check this using the right hand rule described in Chapter 2.6). Therefore, if we connect wires that run from the top and bottom of the diaphragm out to a preamplifier, we'll get a signal.

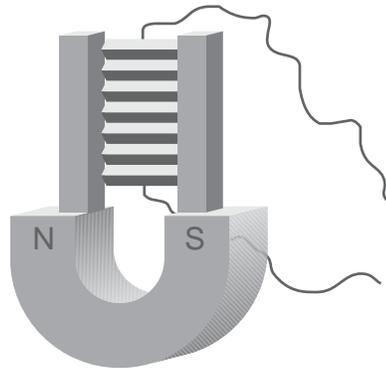


Figure 5.54: The construction of a simple ribbon microphone. The diaphragm is the corrugated (folded) foil placed between the two poles of the magnet.

There are a couple of small problems with this design. Firstly, the current that's generated by one little strip of aluminium that's getting pushed back and forth by a sound wave will be *very* small. So small that a typical microphone preamplifier won't have enough gain to bring the signal up to a useable level. Secondly, consider that the impedance of a strip of aluminium

a couple of centimeters long will be very small, which is fine, except that the input of the microphone preamp is expecting to “see” an impedance which is at least around 200Ω or so. Luckily, we can fix both of these problems in one step by adding a transformer to the microphone.

The output wires from the diaphragm are connected to the primary coil of a transformer that steps up the voltage to the secondary coil. The result of this is that the output of the microphone is increased proportionally to the turns ratio of the transformer, and the apparent impedance of the diaphragm is increased proportionally to the square of the turns ratio. (See Section 2.7 of the electronics section if this doesn’t make sense.) So, by adding a small transformer inside the body of the microphone, we kill both birds with one stone. In fact, there is a third dead bird lying around here as well – we can also use the transformer to balance the output signal by including a centre tap on the secondary coil and making it the ground connection for the mic’s output. (See Chapter 5.5 for a discussion on balancing if you’re not sure about this.)

That’s pretty much it for the basic design of a ribbon condenser microphone – different manufacturers will use different designs for their magnet and ribbon assembly. There is an advantage and a couple of disadvantages in this design that we should discuss at this point. Firstly, the advantage: since the diaphragm in a ribbon microphone is a very small piece of aluminium, it is very light, and therefore very easy to move quickly. As a result, ribbon microphones have a good high-frequency response characteristic (and therefore a good transient response). On the contrary, there are a number of disadvantages to using ribbon microphones. Firstly, you have to remember that the diaphragm is a very thin and relatively fragile strip of aluminium. you cannot throw a ribbon microphone around in a road case and expect it to work the next day – they’re just too easily broken. Since the output of the diaphragm is proportional to the its velocity, and since that velocity is proportional to frequency, the ribbon has a very poor low-frequency response. There’s also the issue of noise: since the ribbon itself doesn’t have a large output, it must be boosted in level a great deal, therefore increasing the noise floor as well. The cost of ribbon microphones is moderately high (although not insane) because of the rather delicate construction. Finally, as we’ll see a little later, ribbon microphones are particularly susceptible to low-frequency noises caused by handling and breath noise.

Moving Coil Dynamic Microphones

In the chapter on induction, we talked about ways to increase the efficiency of the transfer of mechanical energy into electrical energy. The easiest way to do this is to take your wire that's moving in the magnetic field and turn it into a coil. The result of this is that the individual turns in the coil reinforce each other producing more current.

This same principal can be applied to a dynamic microphone. If we replace the single ribbon with a coil of copper wire sitting in the gap of a carefully constructed magnet, we'll generate a lot more current with the same amount of movement. Take a look at Figures 5.55 and 5.55.

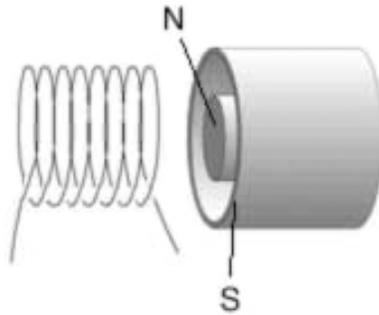


Figure 5.55: An exploded view of a coil of wire with a diameter carefully chosen to fit in the circular gap of a permanent magnet.

Now, when the coil is moved in and out of the magnet, a current is generated that is proportional to the velocity of the movement. How do we create this movement? We glue the front of the coil on to a diaphragm made of plastic as is shown in the cross section in Figure 5.57.

Pressure changes caused by sound waves hitting the front of the diaphragm push and pull it, moving the coil in and out of the gap. This causes the wire in coil to cut perpendicularly through the magnetic lines of force, thus generating a current that is substantially greater than that produced by the ribbon in a ribbon microphone.

This signal still need to be boosted, and the impedance of the coil isn't high enough for us to simply take the wire connected to the coil and connect it to the microphone's output. Therefore, we use a step-up transformer again, just as we did in the case of the ribbon mic, to increase the signal strength, increase the output impedance to around 200Ω , and to provide a balanced output.

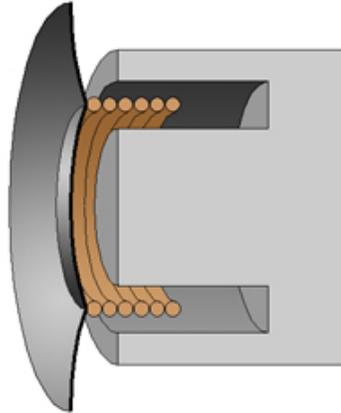


Figure 5.56: A cross section of the same device when assembled. Note that the front of the coil of wire is attached to the inside of the diaphragm.



Figure 5.57: A moving coil dynamic microphone with the protection grid removed. The “front” of the microphone shows a second protective layer made of mesh and hard plastic. The diaphragm and assembly are below this.



Figure 5.58: The underside of the diaphragm showing the copper coil glued to the back of the diaphragm. This coil fits inside the circular gap in the magnet. See Figure 3a for part labels.

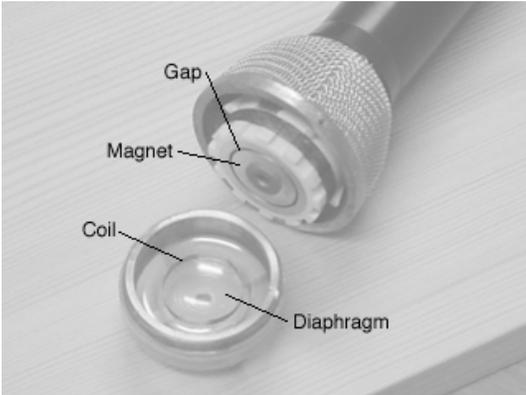


Figure 5.59: The same photograph as Figure 2a with the various parts labeled.

There are a number of advantages and disadvantages to using moving coil microphones. One of the biggest advantages is the rugged construction of these devices. For the most part, moving coil microphones border on being indestructible – in fact, it's almost difficult to break one without intentionally doing so. This is why you'll see them in road cases of touring setups – they can withstand a great deal of abuse. Secondly, since there are so many of these devices in production, and because they have a fairly simple design, the costs are quite affordable. On the side of the disadvantages, you have to consider that the coil in a moving coil microphone is relatively heavy and difficult to move quickly. As a result, it's difficult to get a good high frequency response from such a microphone. Similarly, since the output of the coil is dependent on its velocity, very low frequencies will result in little output as well.

5.6.3 Condenser Microphones

The goal in a microphone is to turn the movement of a diaphragm into a change in electrical potential. As we saw in dynamic microphones, this can be done using induction however, there is another way.

DC Polarized Condenser Microphones

Electret Condenser Microphones

5.6.4 RF Condenser Microphones

5.6.5 Suggested Reading List

5.7 Microphones – Directional Characteristics

5.7.1 Introduction

If you take a look at any book of science experiments for kids you'll find a project that teaches children about barometric pressure. You take a coffee can and stretch a balloon over the open end, sealing it up with rubber bands. Then you use some tape to stick a drinking straw on the balloon so that it hangs over the end. The whole contraption should look like the one shown in Figure 1.

So, what is this thing? Believe it or not, it's a barometer. If the seal on the ballon is tight, then no air can escape from the can. As a result, if the barometric pressure outdoors goes down (when the weather is rainy), then the pressure inside the can is relatively high (relative to the outside, that is...) so the balloon swells up in the middle and the straw points down as is shown in Figure 2.

On a sunny day, the barometric pressure goes up and the balloon is pushed into the can, so the straw points up as can be seen in Figure 3.

This little barometer tells us a great deal about how a microphone works. There are two big things to remember about this device:

1. The displacement of the balloon is caused by the difference in air pressure on each side of it. In the case of the coffee can, the balloon moves to try and make the pressure inside the can the same as the outside of the can. The balloon always moves away from the side with the higher air pressure.
2. In the case of the sealed can, it doesn't matter which direction the pressure outside the can is coming from. On a low-pressure day, the balloon pushes out of the can, and this is true whether the can is rightside up, on its side, or even upside down.

5.7.2 Pressure Transducers

The barometer in the introduction responds to large changes in barometric pressure on the order of 2 or 3 Pascals over long periods of time on the order of hours or days. However, if we made a miniature copy of the coffee can and the balloon, we're have a device that would respond much more quickly to much smaller changes in pressure. In fact, if we were to call the miniaturized balloon a *diaphragm* and make it about 1 cm in diameter or so,

it would be perfect for responding to changes in pressure caused by passing sound waves instead of weather systems.

So, let's consider what we have. A small can, sealed on the front by a very thin diaphragm. This device is built so that the diaphragm moves in and out of the can with changes in pressure between 20 micropascals and 2 pascals (give or take...) and frequencies up to about 20 kHz or so. There's just one problem: the diaphragm moves a distance that is proportionate to the amplitude of the pressure change in the air, so if the barometric pressure goes down on a rainy day, the diaphragm will get stretched out and will probably tear. So, to prevent this from happening, we'll drill a very small hole called a *capillary tube* in the back of the can for very long term changes in the pressure. If you'd like to build one, the construction diagrams are shown in Figure 4.

The biologically-minded reader may be interested to note that this is essentially the construction of the human ear. The diaphragm is your eardrum, the canister is your head (or at least a small cavity behind your eardrum inside your head) and the capillary tube is your eustachian tube that connects the back of your eardrum to your mouth. When you undergo a wide change in air pressure in a longer period of time (like when you're taking off in an airplane, for example), your eardrum is pushed out and "pops" just like the diaphragm would be. And, like the capillary tube, the eustachian tube lets the new pressure "equalize" on the back of the eardrum – therefore, by yawning or swallowing, you put your eardrum back where it belongs.

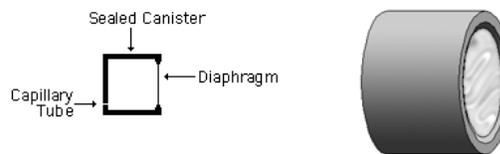


Figure 5.60: The construction of a miniature coffee can barometer.

How will this system behave? Remember that, just like the coffee can barometer, the back of the diaphragm is effectively sealed so if the pressure outside the can is high, the diaphragm will get pushed into the can. If the pressure outside the can is low, then the diaphragm will get pulled out of the can. This will be true no matter where the change in pressure originated. (The capillary tube is of a small enough diameter that fast pressure changes in the audio range don't make it through into the can, so we can consider

the can to be completely sealed.)

What we have is called a *pressure transducer*. Remember that a transducer is any device that converts one kind of energy into another. In the case of a microphone, we're converting mechanical energy (the movement of the diaphragm) into electrical energy (the change in voltage and/or current at the output). How that conversion actually takes place is dealt with in a previous chapter – what we're concerned about in this chapter is the pressure part.

A perfect pressure transducer responds identically to a change in air pressure originating in any direction, and therefore arriving at the diaphragm from any *angle of incidence*. Every microphone has a “front,” “side” and “back,” but because we're trying to be a little more precise about things (translation: because we're geeks) we break this down further into angles. So, directly in front of the microphone is considered to be an angle of incidence of 0° . We can rotate around from there to 90° on the side and 180° at the back as is shown in Figure 5.

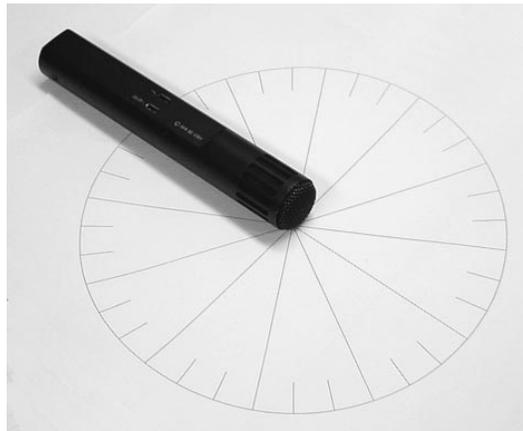


Figure 5.61: A microphone showing various angles of incidence (full marking every 30° , small markings every 10°). Note that we can consider the rotational angle in any plane whereas this photo only indicates the horizontal plane.

We can make a graph of the way in which a perfect pressure transducer will respond to the pressure changes by graphing its *sensitivity*. This is a word used for the *gain* of a microphone caused by the angle of incidence of the incoming sound (although, as we saw earlier, other issues are included in the sensitivity as well). Remember in the case of a perfect pressure transducer, the sensitivity will be the same regardless of the angle of incidence,

so if we consider that the gain for a sound source that's *on-axis* or with an angle of incidence of 0° is normalized to a value of 1, then all other angles of incidence will be 1 as well. This can be plotted on a cartesian X-Y graph as is shown in Figure 6. The equation below can be used to calculate the sensitivity for a pressure transducer. (Okay, okay, it's not much of an equation – for any angle, the sensitivity is 1...)

$$S_P = 1 \quad (5.8)$$

where S_P is the sensitivity of a pressure transducer.

For any angle, you can just multiply the pressure by the sensitivity for that angle of incidence to find the voltage output.

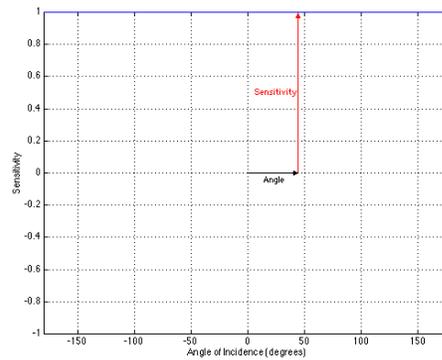


Figure 5.62: A Cartesian plot of the sensitivity of a perfect pressure transducer normalized to the on-axis response.

Most people like to see this in a little more intuitive graph called a *polar plot* shown in Figure 7. In this kind of graph, the sensitivity is graphed as a radius from the centre of the graph at a given angle of rotation.

One thing to note here: most books plot their polar plots with 0° pointing directly upwards (towards 12 o'clock). Technically speaking, this is incorrect – a proper polar plot starts with 0° on the right side (towards 3 o'clock). This is the system that I'll be using for all polar plots in this book.

Just for the sake of having a version of the plots that look nice and clean, Figures 8 and 9 are duplicates of Figures 6 and 7.

Most people don't call these microphones "pressure transducers" – because the microphone is equally sensitive to all sound sources regardless of direction they're normally called *omnidirectional* microphones. Some people shorten this even further and call them *omni's*.

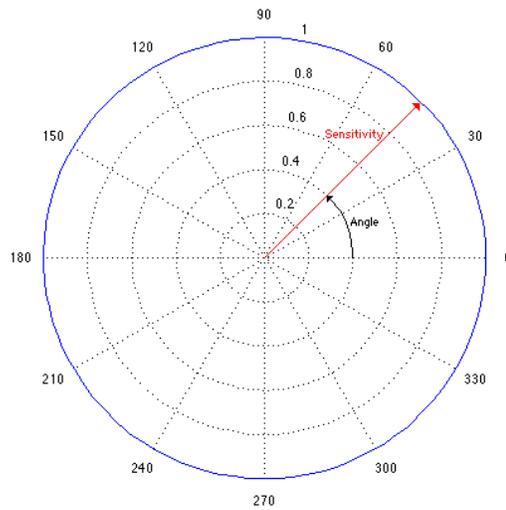


Figure 5.63: A polar plot of the sensitivity of a perfect pressure transducer normalized to the on-axis response. Note that this plot shows the same information as the plot in Figure 6.

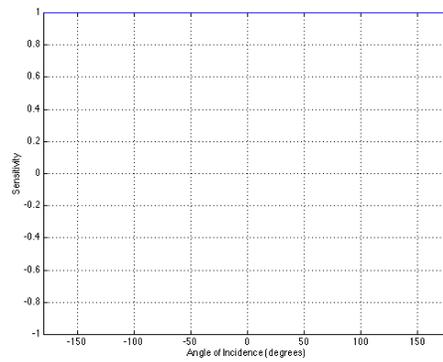


Figure 5.64: Cartesian plot of the sensitivity of a perfect pressure transducer normalized to the on-axis response.

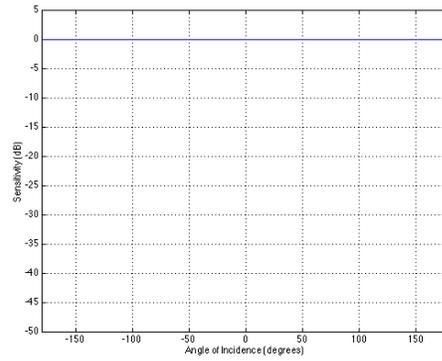


Figure 5.65: Cartesian plot of the sensitivity (in dB referenced to the on-axis sensitivity) of a perfect pressure transducer normalized to the on-axis response.

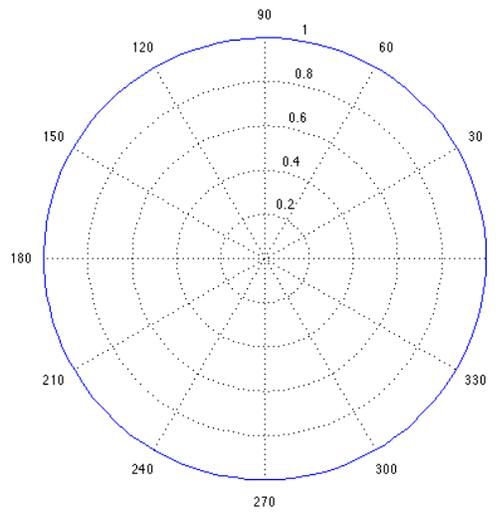


Figure 5.66: Polar plot of the sensitivity of a perfect pressure transducer normalized to the on-axis response. Note that this plot shows the same information as the plot in Figure 8.

5.7.3 Pressure Gradient Transducers

What happens if the diaphragm is held up in mid-air without being sealed on either side? Figure 9 shows just such a system where the diaphragm is supported by a ring and is exposed to the outside world on both sides.



Figure 5.67: The construction of a diaphragm that's open on both sides.

Let's assume for the purposes of this discussion that a movement of the diaphragm to the left of the resting position somehow magically results in a positive voltage at the output of this microphone (for more info on how this miracle actually occurs, read Section 5.6.) Therefore if the diaphragm moves in the opposite direction, the voltage at the output will be negative. Let's also assume that the side of the diaphragm facing the right is called the "front" of the microphone.

If there's a sound source producing a high pressure at the front of the diaphragm, then the diaphragm is pushed backwards and the voltage output is positive. Positive pressure causes positive voltage. If the sound source stays at the front and produces a low pressure, then the diaphragm is pulled frontwards and the resulting voltage output is negative. Negative pressure causes negative voltage. Therefore there is a positive relationship between the pressure at the front of the diaphragm and the voltage output – meaning that, the polarity of the voltage at the output is the same as the pressure at the front of the microphone.

What happens if the sound source is at the rear of the microphone at an angle of incidence of 180° ? Now, a positive pressure pushes on the diaphragm from the rear and causes it to move towards the front of the microphone. Remember from two paragraphs back that this causes a negative voltage output. Positive pressure causes negative voltage. If the source is in the rear and the pressure is negative, then the diaphragm is pulled towards the

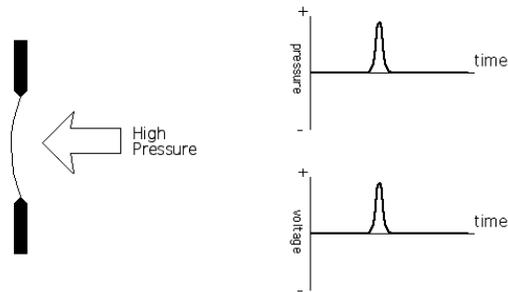


Figure 5.68: A positive pressure at the front of the microphone moves the diaphragm towards the back and causes a positive voltage at the output.

rear of the microphone, resulting in a positive voltage output. Now, we have a situation where there is a negative relationship between the pressure at the rear of the microphone and the voltage output – the polarity of the voltage at the output is opposite to the pressure at the rear.

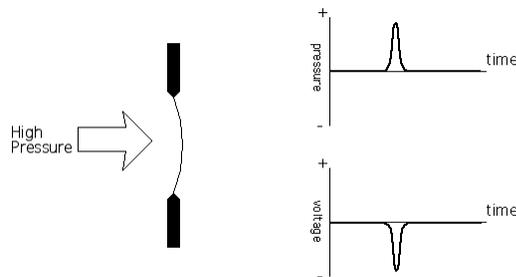


Figure 5.69: A positive pressure at the back of the microphone moves the diaphragm towards the front and causes a negative voltage at the output.

What happens when the sound source is at an angle of incidence of 90° – directly to one side of the microphone? If the source produces a high pressure, then this reaches both sides of the diaphragm equally and therefore the diaphragm doesn't move. The result is that the output voltage is 0 – there is no output. The same will be true if the pressure is negative because the two low-pressure areas on either side of the microphone will be pulling equally on the diaphragm.

This phenomenon of the sound source outputting a high or low pressure with no voltage at the output of the microphone shows exactly what's hap-

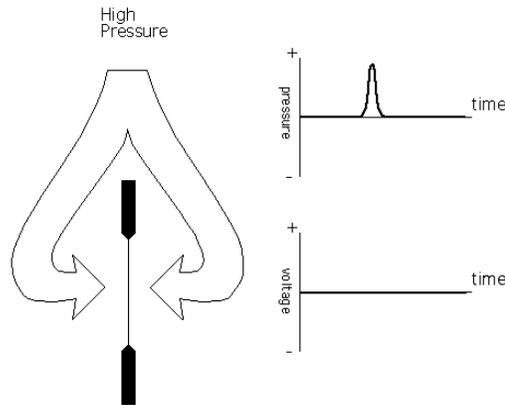


Figure 5.70: A positive pressure at the side of the microphone causes no movement in the diaphragm and causes 0 volts at the output.

pening in this microphone. The movement of the diaphragm (and therefore the output of the microphone) is dependent on the *difference* in pressure on the two sides of the diaphragm. If the pressure on the two sides is the same, the difference is 0 and therefore the output is 0. The bigger the difference in pressure, the bigger the voltage output. Another word for “difference” is “gradient” and thus this design is called a *Pressure Gradient Transducer*.

We know the sensitivity of the microphone at four angles of incidence – at 0° the sensitivity is 1 just like a Pressure Transducer. At 180° , the sensitivity is -1. The voltage waveform will look like the pressure waveform, but it will be upside down – inverted in polarity because it’s multiplied by -1. At 90° and 270° the sensitivity will be 0 – no matter what the pressure is, the voltage output will be 0.

The question now is, what happens at all the other angles? Well, it might be already obvious. There’s a simple function that converts angles to a number where 0° corresponds to a value of 1, 90° to 0, 180° to -1 and 270° to 0 again. The function is called a cosine – it turns out that the sensitivity of this construction of microphone is the cosine of the angle of incidence as is shown in Figure 13. So, the equation for calculating the sensitivity is:

$$S_G = \cos(\alpha) \quad (5.9)$$

where S_G is the sensitivity of a pressure gradient transducer and α is the angle of incidence.

Again, most people prefer to see this in a polar plot as is shown in Figure

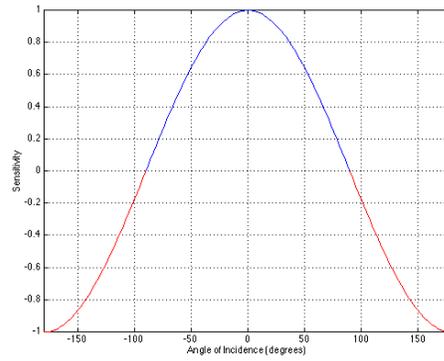


Figure 5.71: Cartesian plot of the sensitivity of a pressure gradient transducer. Note that the negative polarity lobe has been highlighted in red.

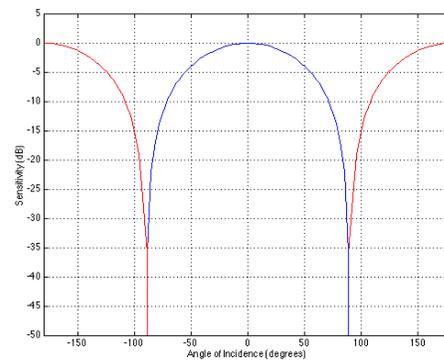


Figure 5.72: Cartesian plot of the sensitivity (in dB referenced to the on-axis sensitivity) of a pressure gradient transducer. Note that the negative polarity lobe has been highlighted in red.

15, however, what most people don't know is that they're not really looking at an accurate polar plot of the sensitivity. In this case, we're looking at a polar plot of the absolute value of the cosine of the angle of incidence – if this doesn't make sense, don't worry too much about it.

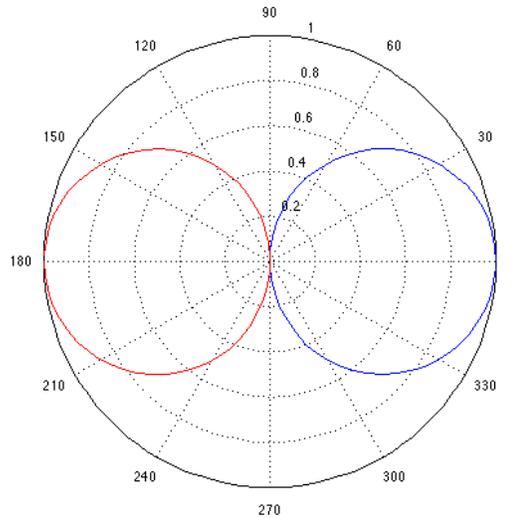


Figure 5.73: Polar plot of the absolute value of the sensitivity of a pressure gradient transducer. Blue indicates positive polarity, red indicates negative polarity. Note that this plot shows the same information as the plot in Figure 14.

Notice that in the graph in Figure 15, the front half of the plot is in blue while the rear half is red. This is to indicate the polarity of the sensitivity, so at an angle of 180° , the radius of the plot is 1, but because the plot at that angle is red, it's -1. At 30° , the radius is 0.5 and because it's blue, then it's positive.

Just like pressure transducers are normally called omnidirectional microphones, pressure gradient transducers are usually called either *bidirectional* microphones (because they they're sensitive in two directions – the front and back) or *figure eight* microphones (because the polar pattern looks like the number 8).

There's one thing that we should get out of the way right now. Many people see the figure 8 pattern of a bidirectional microphone and jump to the assumption that the mic has two outputs – one for the front and one for the back. This is not the case. The microphone has one output and one output only. The sound picked up by the front and rear *lobes* is essentially

mixed acoustically and output as a single signal. You cannot separate the two lobes to give you independent outputs.

5.7.4 Combinations of Pressure and Pressure Gradient

It is possible to create a microphone that has some combination of both a pressure component and a pressure gradient component. For example, look at the diagram in Figure 16. This shows a microphone where the diaphragm is not entirely sealed in a can as in the pressure transducer design, but it's not completely open as in the pressure gradient design.



Figure 5.74: A microphone that is one half pressure transducer and one half pressure gradient design. Note that if you build this microphone it probably will not work properly – this is an approximate drawing for conceptual purposes. The significant things to note here are the vents that allow some of the pressure changes from the outside world into the back of the diaphragm. Note, however, that the back of the diaphragm is not completely exposed to the outside world.

In this case, the path to the back of the diaphragm from the outside world is the same length as the path to the front of the diaphragm when the sound source is at 180° – not 90° as in a pure pressure gradient transducer. This then means that there will be no output when the sound source is at the rear of the microphone. In this case, the sensitivity pattern is created by creating a mixture of 50 percent Pressure and 50 percent Pressure Gradient. Therefore, we're multiplying the two pure sensitivity patterns by 0.5 and adding them together. This results in the pattern shown in Figure 17 – notice the similarity between this pattern and the perfect pressure gradient sensitivity pattern – it's just a cosine wave that's been offset by enough to eliminate the negative components.

If we plot this sensitivity pattern on a polar plot, we get the graph shown in Figure 18. Notice that this pattern looks somewhat like a heart shape, so it's normally called a *cardioid* pattern (“cardio” meaning “heart” as in “cardio-vascular” or “cardio-pulmonary”)

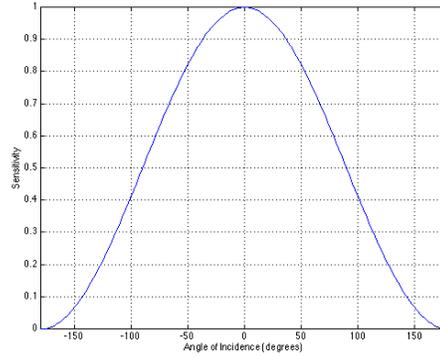


Figure 5.75: Cartesian plot of the sensitivity pattern of a microphone that is one half Pressure and one half Pressure Gradient transducer.

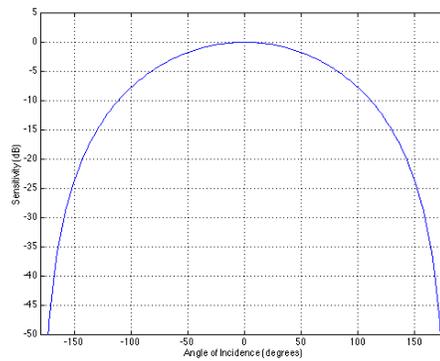


Figure 5.76: Cartesian plot of the sensitivity (in dB referenced to the on-axis sensitivity) of a microphone that is one half Pressure and one half Pressure Gradient transducer.

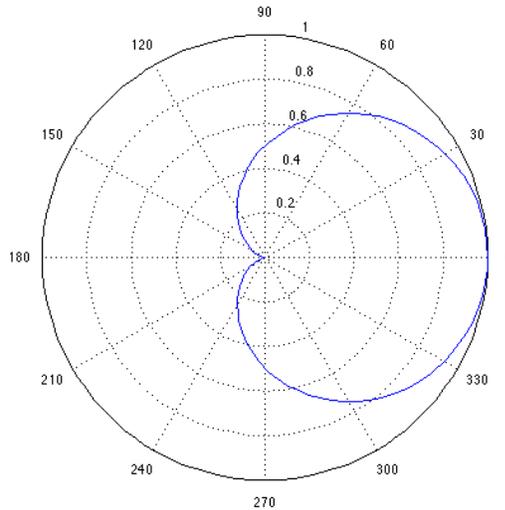


Figure 5.77: Polar plot of the sensitivity pattern of a cardioid microphone (one half Pressure and one half Pressure Gradient transducer). Note that this plot shows the same information as the plot in Figure 16. A good rule of thumb to remember about this polar pattern is that the sensitivity is 0.5 (or -6 dB) at 90°.

5.7.5 General Sensitivity Equation

We can now develop a general equation for calculating the sensitivity pattern of a microphone that contains both Pressure and Pressure Gradient components as follows:

$$S = P + G * \cos(\alpha) \quad (5.10)$$

where S is the sensitivity of the microphone, P is the Pressure component, G is the Pressure Gradient component, α is the angle of incidence and where $P + G = 1$.

For example, for a microphone that is 50 percent Pressure and 50 percent Pressure Gradient, the sensitivity equation would be:

$$S = P + G * \cos(\alpha) \quad (5.11)$$

$$S = 0.5 + 0.5 * \cos(\alpha) \quad (5.12)$$

This sensitivity equation can then be used to create any polar pattern between a perfect pressure transducer and a perfect pressure gradient. All we

need to do is to decide how much of each we want to add in the equation. For a perfect omnidirectional microphone, we make $P=1$ and $PG=0$. Therefore the microphone is a 100 percent pressure transducer and 0 percent pressure gradient transducer. There are five “standard” polar patterns, although one of these is actually two different standards, depending on the manufacturer. The five most commonly-seen polar patterns are:

| Polar Pattern | P | G |
|-----------------|-------|-------|
| Omnidirectional | 1 | 0 |
| Subcardioid | 0.75 | 0.25 |
| Cardioid | 0.5 | 0.5 |
| Supercardioid | 0.333 | 0.666 |
| Hypercardioid | 0.25 | 0.75 |
| Bidirectional | 0 | 1 |

Table 5.6: INSERT CAPTION HERE

What do these polar patterns look like? We’ve already seen the omnidirectional, cardioid and bidirectional patterns. The others are as follows.

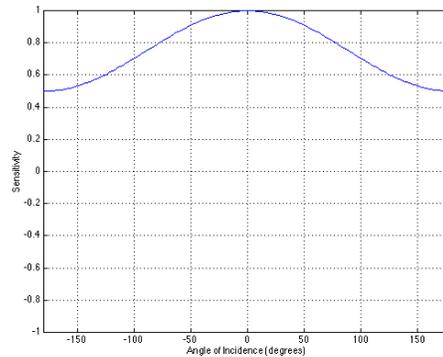


Figure 5.78: Cartesian plot of the sensitivity of a subcardioid microphone.

Just to compare the relationship between the various directional patterns, we can look at all of them on the same plot. This gets a little complicated if they’re all on the same polar plot – just because things get crowded, but if we see them on the same Cartesian plot (see the graphs above for the corresponding polar plots) then we can see that all of the simple directional patterns are basically the same.

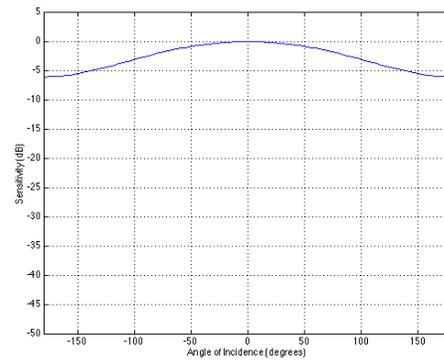


Figure 5.79: Cartesian plot of the sensitivity (in dB referenced to the on-axis sensitivity) of a subcardioid microphone.

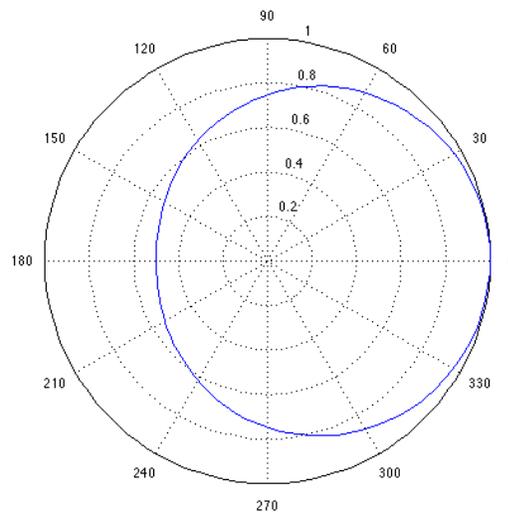


Figure 5.80: Polar plot of a subcardioid microphone. Notice that the maximum attenuation of 0.5 (or -6.02 dB) is at the rear of the microphone at 180°.

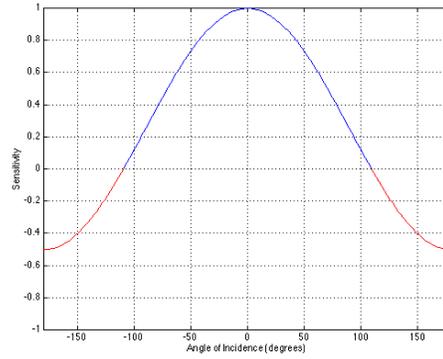


Figure 5.81: Cartesian plot of a hypercardioid microphone using the values $P=0.25$ and $PG=0.75$.

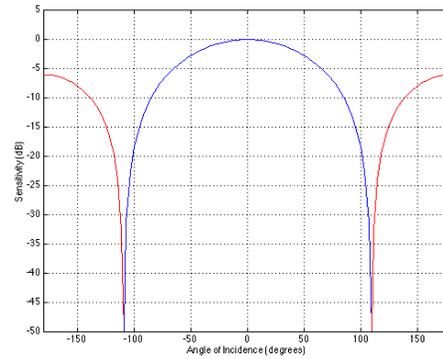


Figure 5.82: Cartesian plot of the sensitivity (in dB referenced to the on-axis sensitivity) of a hypercardioid microphone using the values $P=0.25$ and $PG=0.75$. Note that the negative polarity lobe has been highlighted in red.

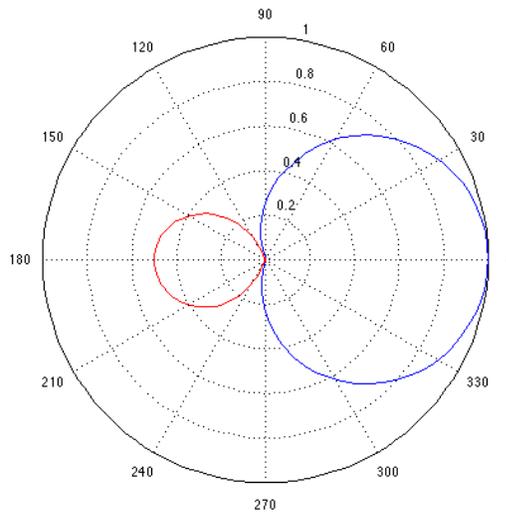


Figure 5.83: Polar plot of a hypercardioid microphone using the values $P=0.25$ and $PG=0.75$. Notice that the maximum attenuation of 0 (or -infinity dB) is at about 109° .

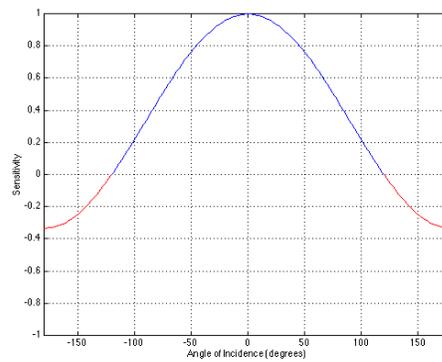


Figure 5.84: Cartesian plot of a supercardioid microphone using the values $P=0.333$ and $PG=0.666$

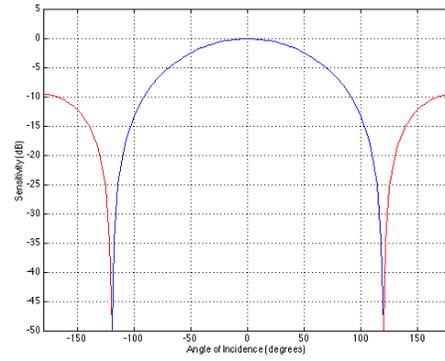


Figure 5.85: Cartesian plot of the sensitivity (in dB referenced to the on-axis sensitivity) of a supercardioid microphone using the values $P=0.333$ and $P_G=0.666$. Note that the negative polarity lobe has been highlighted in red.

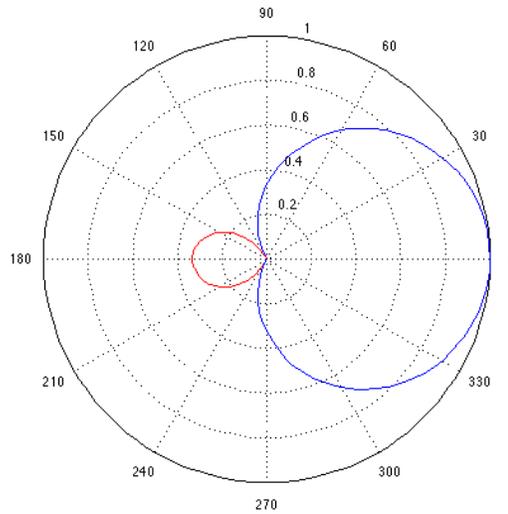


Figure 5.86: Polar plot of a supercardioid microphone using the values $P=0.333$ and $P_G=0.666$. Notice that the maximum attenuation of 0 (or -infinity dB) is at 120° .

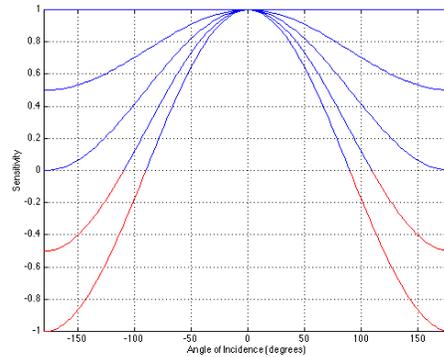


Figure 5.87: Most of the standard polar patterns on one Cartesian plot. From top to bottom, these are omnidirectional, subcardioid, cardioid, hypercardioid, and bidirectional. Note that red sections of the plot point out the fact that the sensitivity is negative polarity.

One of the interesting things that becomes obvious in this plot is the relationship between the angle of incidence where the sensitivity is 0 – sometimes called the *null* because there is no output – and the mixture of the Pressure and Pressure Gradient components. All mixtures between omnidirectional and cardioid have no null because there is no angle of incidence that results in no output. The cardioid microphone has a single null at 180° , or, directly to the rear of the microphone. As we increase the mixture to have more and more Pressure Gradient component, the null splits into two symmetrical points on the polar plot that move around from the rear of the microphone to the sides until, when the transducer is a perfect bidirectional, the nulls are at 90° and 270° .

5.7.6 Do-It-Yourself Polar Patterns

If you go to your local microphone store and buy a “normal” single-diaphragm cardioid microphone like a B & K 4011 (don’t worry if you’re surprised that there might be something other than a microphone with a single diaphragm... we’ll talk about that later) the manufacturer has built the device so that it’s the appropriate mixture of Pressure and Pressure Gradient. Consider, however, that if you have a perfect omnidirectional microphone and a perfect bidirectional microphone, then you could strap them together, mix their outputs electrically in a run-of-the-mill mixing console, and, assuming that everything was perfectly aligned, you’d be able to make your own cardioid. In fact, if the two real microphones were exactly matched, you could

make any polar pattern you wanted just by modifying the relative levels of the two signals.

Mathematically speaking, the output of the omnidirectional microphone is the Pressure component and the output of the Bidirectional Microphone is the Pressure Gradient component. The two are just added in the mixer so you're fulfilling the standard sensitivity equation:

$$S = P + G * \cos(\alpha) \quad (5.13)$$

where P is the gain applied to the omnidirectional microphone and G is the gain applied to the bidirectional microphone.

Also, let's say that you have two cardioid microphones, but that you put them in a back-to-back configuration where the two are pointing 180° away from each other. Let's look at this pair mathematically. We'll call microphone 1 the one pointing "forwards" and microphone 2 the second microphone pointing 180° away. Note that we're also assuming for a moment that the gain applied to both microphones is the same.

$$S_{TOTAL} = (0.5 + 0.5 * \cos(\alpha)) + (0.5 + 0.5 * \cos(\alpha + 180)) \quad (5.14)$$

$$S_{TOTAL} = 0.5 + 0.5 + 0.5 * (\cos(\alpha) + \cos(\alpha + 180)) \quad (5.15)$$

$$S_{TOTAL} = 1 + 0.5 * (\cos(\alpha) + \cos(\alpha + 180)) \quad (5.16)$$

Now, consider that the cosine of every angle is the opposite polarity to the cosine of the same angle + 180°. In other words:

$$\cos(\alpha) = -1 * \cos(\alpha + 180^\circ) \quad (5.17)$$

Consider therefore that the cosine of any angle added to the cosine of the same angle + 180° will equal 0. In other words:

$$\cos(\alpha) + \cos(\alpha + 180^\circ) = 0 \quad (5.18)$$

Let's go back to the equation that describes the back to back cardioids:

$$S_{TOTAL} = 1 + 0.5 * (\cos(\alpha) + \cos(\alpha + 180)) \quad (5.19)$$

We now know that the two cosines cancel each other, therefore the equation simplifies to:

$$S_{TOTAL} = 1 + 0.5 * (0) \quad (5.20)$$

$$S_{TOTAL} = 1 \quad (5.21)$$

Therefore, the result is an omnidirectional microphone. This result is possibly easier to understand intuitively if we look at graphs of the sensitivity patterns as is shown in Figures 26 and 27.

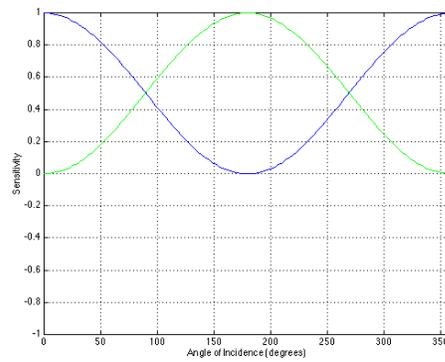


Figure 5.88: Cartesian plot of the sensitivity patterns of two cardioid microphones aimed 180° apart. The blue plot is the forward-facing cardioid, the green is the rear-facing cardioid. Note that, if summed, the resulting output would be 1 for any angle.

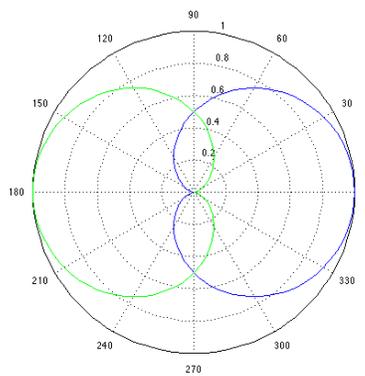


Figure 5.89: Polar plot of the sensitivity patterns of two cardioid microphones aimed 180° apart. Note that, if summed, the resulting output would be 1 for any angle.

Similarly, if we inverted the polarity of the rear-facing microphone, the resulting mixed output (if the gains applied to the two cardioids were equal) would be a bidirectional microphone. The equation for this mixture would be:

$$S_{TOTAL} = (0.5 + 0.5 * \cos(\alpha)) + -1 * (0.5 + 0.5 * \cos(\alpha + 180)) \quad (5.22)$$

$$S_{TOTAL} = 0.5 + 0.5 * \cos(\alpha) - -0.5 - -0.5 * \cos(\alpha + 180)) \quad (5.23)$$

$$S_{TOTAL} = 0.5 - -0.5 + 0.5 * \cos(\alpha) - -0.5 * \cos(\alpha + 180) \quad (5.24)$$

$$S_{TOTAL} = 0.5 * \cos(\alpha) - -0.5 * \cos(\alpha + 180) \quad (5.25)$$

$$S_{TOTAL} = \cos(\alpha) \quad (5.26)$$

So, as you can see, not only is it possible to create any microphone polar pattern using the summed outputs of a bidirectional and an omnidirectional microphone, it can be accomplished using two back-to-back cardioids as well. Of course, we're still assuming at this point that we're living in a perfect world where all transducers are matched – but we'll stay in that world for now...

5.7.7 The Influence of Polar Pattern on Frequency Response Pressure Transducers

Remember that a pressure transducer is basically a sealed can, just like the coffee can barometer. Therefore, any change in pressure in the outside world results in the displacement of the diaphragm. High pressure pushes the diaphragm in, low pressure pulls it out. Unless the change in pressure is extremely slow with a period on the order of hours (which we obviously will not hear as a sound wave – and which leaks through the capillary tube) then the displacement of the diaphragm is dependent on the pressure, regardless of frequency. Therefore a perfect pressure transducer will respond to all frequencies similarly. This is to say that, if a pressure wave arriving at the diaphragm is kept at the same peak pressure value, but varied in frequency, then the output of the microphone will be a voltage waveform that changes in frequency but does not change in peak voltage output.

A graph of this would look like Figure 28.

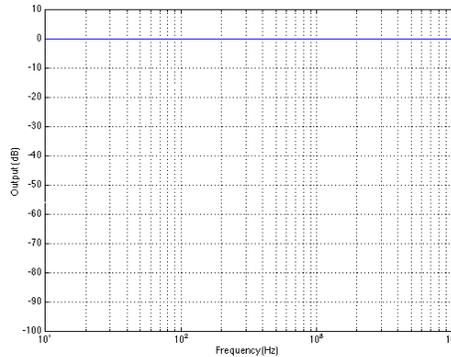


Figure 5.90: The frequency response of a perfect Pressure transducer. Note that all frequencies have equal output assuming that the peak value of the pressure wave is the same at all frequencies.

Pressure Gradient Transducers

The behaviour of a Pressure Gradient transducer is somewhat different because the incoming pressure wave reaches both sides of the diaphragm. Remember that the lower the frequency, the longer the wavelength. Also, consider that, if a sound source is on-axis to the transducer, then there is a pathlength difference between the pressure wave hitting the front and the rear of the diaphragm. That pathlength difference remains at a constant delay time regardless of frequency, therefore, the lower the frequency the more alike the pressures at the front and rear of the diaphragm because the phase difference is smaller with lower frequencies. The delay is constant and short because the diaphragm is typically small.

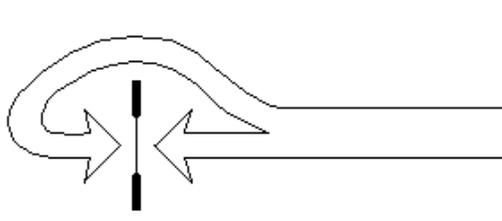


Figure 5.91: A diagram of a Pressure Gradient transducer showing the two paths to the front and rear of the diaphragm from a source on axis.

Since the sensitivity at the rear of the diaphragm has a negative polarity and the front has a positive polarity, then the result is that the pressure at

the rear is subtracted from the front.

With this in mind, let's start at a frequency of 0 Hz and work our way upwards. At 0 Hz, then the pressure at the rear of the diaphragm equals the pressure at the front, therefore the diaphragm does not move and there is no output. (Note that, right away, we're looking at a different beast than the perfect Pressure transducer. Without the capillary tube, the pressure transducer would give us an output with a 0 Hz pressure applied to it.)

As we increase in frequency, the phase difference in the pressure wave at the front and rear of the diaphragm increases. Therefore, there is less and less cancellation at the diaphragm and we get more and more output. In fact, we get a doubling of output for every doubling of frequency – in other words, we have a slope of +6 dB per octave.

Eventually, we get to a frequency where the pressure at the rear of the microphone is 180° later than the pressure at the front. Therefore, if the pressure at the front of the microphone is high and pushing the diaphragm in, then the pressure at the rear is low and pulling the diaphragm in. At this frequency, we have constructive interference and an increased output by 6 dB.

If we increase the frequency further, then the phase difference between the front and rear increases and we start approaching a delay of 360° . At that frequency (which will be twice the frequency where we had +6 dB output) we will have no output at all – therefore a level of – infinity dB.

As the frequency increases, we result in a common pattern of peaks and valleys shown in Figure 30. Because this pattern looks like a hair comb, it's called a *comb filter*.

The frequencies of the peaks and valleys in the comb filter frequency response are determined by the distance between the front and the rear of the diaphragm. This distance, in turn, is principally determined by the diameter of the diaphragm. The smaller the diameter, the shorter the delay and the higher the frequency of the lowest-frequency peak.

Most manufacturers build their bidirectional microphones so that the lowest frequency peak in the frequency response is higher than the range of normal audio. Therefore, the “standard” frequency response of a bidirectional microphone starts an output of 0 at 0 Hz and doubles for every doubling of frequency to a maximum output that is somewhere around or above 20 kHz.

This, of course, is a problem. We don't want a microphone that has a rising frequency response, so we have to fix it. How? Well, we just build the diaphragm so that it has a natural resonance down in the low frequency range. This means that, if you thump the diaphragm like a drum head, it

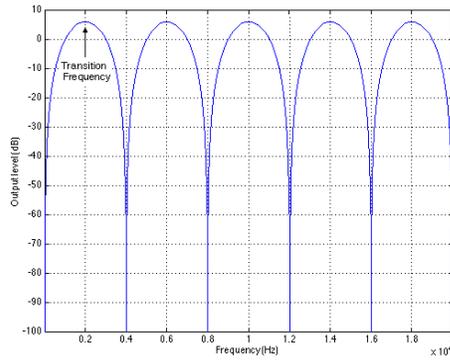


Figure 5.92: A linear plot of a comb filter caused by the interference of the pressures at the front and rear of a Pressure Gradient transducer. The harmonic relationship between the peaks and dips in the frequency response is evident in this plot.

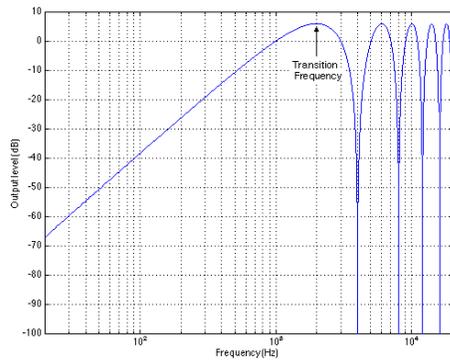


Figure 5.93: A semi-logarithmic plot of a comb filter caused by the interference of the pressures at the front and rear of a Pressure Gradient transducer. The 6 dB/octave rise in the response up to the lowest-frequency peak is evident in this plot.

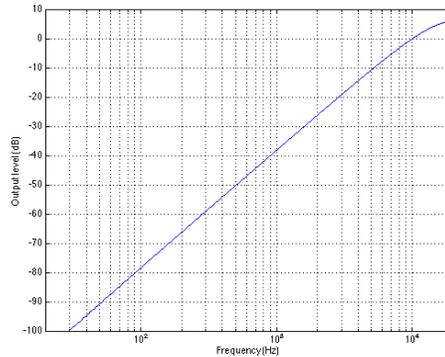


Figure 5.94: The output of a Pressure Gradient transducer whose design ensures that the entire audio range lies below the lowest-frequency peak in the frequency response.

will ring at a very low note. The higher the frequency, the further you get from the peak of the resonance. This resonance acts like a filter that has a gain that increases by 6 dB for every halving of frequency. Therefore, the lower the frequency, the higher the gain. This counteracts the rising natural slope of the diaphragm's output and produces a theoretically flat frequency response. The only problem with this is that, at very low frequencies, there is almost no output to speak of, so we have to have enormous gain and the resulting output is basically nothing but noise.

The moral of this story is that Pressure Gradient microphones have no very low frequency output. Also, keep in mind that any microphone with a Pressure Gradient component will have a similar response. Therefore, if you want to record program material with low frequency content, you have to stick with omnidirectional microphones.

5.7.8 Proximity Effect

Most microphones that have a pressure gradient component have a correction filter built in to fix the low frequency problems of the natural response of the diaphragm. In many cases, this correction works very well, however, there is a specific case where the filter actually makes things worse.

Consider that a pressure gradient microphone has a naturally rising frequency response because the incoming pressure wave arrives at the front as well as at the rear of the diaphragm. Pressure microphones have a naturally flat frequency response because the rear of the diaphragm is sealed from the

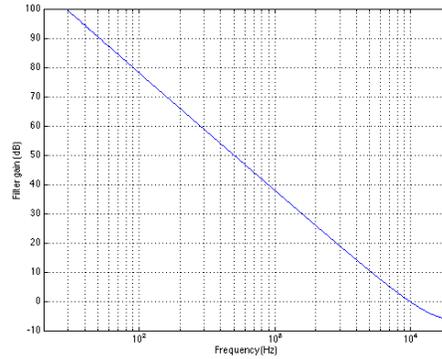


Figure 5.95: The blue plot shows the gain response of a theoretical filter required to “fix” the frequency response of the transducer shown in Figure 32. Note the extremely high gain required in the low frequency range. The red plot shows the gain achieved by making the diaphragm naturally resonant at a low frequency. Note that there is a bottom limit to the benefits of the resonance.

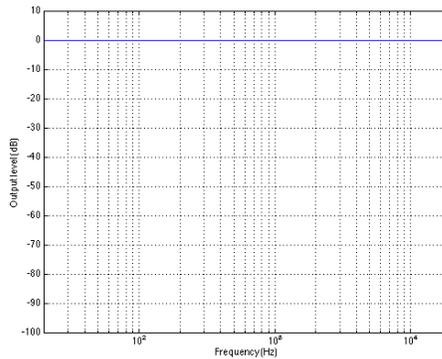


Figure 5.96: The blue plot shows the result of the frequency response of the output of the transducer shown in Figure 32 filtered using the theoretical (blue) gain response plotted in Figure 33. Note that this is a theoretical result that does not take real life into account... The red plot shows a more likely scenario where the extra gain provided by the resonance doesn't extend all the way down to 0 Hz.

outside world. Also, consider a little rule of thumb that says that, in a free, unbounded space, the pressure of a sound wave is reduced by half for every doubling of distance. The implication of this rule is that, if you're very close to a sound source, a small change in distance will result in a large change in sound level. At a greater distance from the sound source, the same change in distance will result in a smaller change in level. For example, if you're 1 cm from the sound source, moving away by 1 cm will cut the level by half, a drop of 6 dB. If you're 1 m from the sound source, moving away by 1 cm will have a negligible effect on the sound level. So what?

Imagine that you have a pressure gradient microphone that is placed very close to a sound source. Consider that the distance from the sound source (say, a singer's mouth...) to the front of the diaphragm will be on the order of millimeters. At the same time, the distance to the rear of the diaphragm will be comparatively very far – possibly 4 to 8 times the distance to the singer's lips. Therefore there is a very large drop in pressure for the sound wave arriving at the rear of the diaphragm. The result is that the rear of the diaphragm is effectively sealed from the outside world by virtue of the fact that the sound pressure level at that side of the diaphragm is much lower than that at the front. Consequently, the natural frequency response becomes more like a pressure transducer than a pressure gradient transducer.

What's the problem? Well, remember that the microphone has a filter that boosts the low end built into it to correct for problems in the natural frequency response – problems that don't exist when the microphone is close to the sound source. As a result, when the microphone is very close to the source, there is a boost in the low frequencies because the correction filter is applied to a now naturally flat frequency response. This boost in the low end is called *proximity effect* because it is caused by the microphone being in close proximity to the sound source.

There are a number of microphones that rely on the proximity effect to boost the low frequency components of the signal. These are typically sold as vocal mic's such as the Shure SM58. If you measure the frequency response of such a microphone from 1 m away, then you'll notice that there is almost no low-end output. However, in typical usage, there is plenty of low end. Why? Because, in typical usage, the microphone is stuffed in the singer's mouth – therefore there's lots of low end because of proximity effect.

Remember, when the microphone has a pressure gradient component, the frequency response is partially dependent on the distance to the diaphragm. Also remember that, for some microphones, you have to be placed close to the source to get a reasonably low frequency response, whereas other

microphones in the same location will have a boosted low frequency response.

5.7.9 Acceptance Angle

As we saw in a previous section, the bandwidth of a filter is determined by the frequency band limited by the points where the signal is 3 dB lower than the maximum output of the filter. Microphones have a spatial equivalent called the *acceptance angle*. This is the frontal angle of the microphone where the sensitivity is within 3 dB of the on-axis response. This angle will vary with polar pattern.

In the case of an omnidirectional, all angles of incidence have a sensitivity of 0 dB relative to the on-axis response of the microphone. Consequently, the acceptance angle is $\pm 180^\circ$ because the sensitivity never drops below -3 dB relative to the on-axis sensitivity.

A subcardioid, on the other hand, has a sensitivity that drops below -3 dB when the angle of incidence of the sound source is outside the acceptance angle of $\pm 99.9^\circ$. A cardioid has an acceptance angle of $\pm 65.5^\circ$, a hypercardioid has an acceptance angle of $\pm 52.4^\circ$, and a bidirectional has an acceptance angle of $\pm 45.0^\circ$.

| Polar Pattern (P : G) | Acceptance Angle |
|-------------------------------|------------------|
| Omnidirectional (1 : 0) | $\pm 180^\circ$ |
| Subcardioid (0.75 : 0.25) | $\pm 99.9^\circ$ |
| Cardioid (0.5 : 0.5) | $\pm 65.5^\circ$ |
| Supercardioid (0.375 : 0.625) | $\pm 57.9^\circ$ |
| Hypercardioid (0.25 : 0.75) | $\pm 52.4^\circ$ |
| Bidirectional (0 : 1) | $\pm 45.0^\circ$ |

Table 5.7: Acceptance Angles for various microphone polar patterns.

5.7.10 Random-Energy Response (RER)

Think about an omnidirectional microphone in a diffuse field (the concept of a diffuse field is explained in Section ??). The omni is equally sensitive to all sounds coming from all directions, giving it some output level. If you put a cardioid microphone in exactly the same place, you wouldn't get as much output from it because, although it's as sensitive to on-axis sounds as the omni, all other directions will be attenuated in comparison.

Since a diffuse field is comprised of random signals coming from random directions, we call the theoretical power output of a microphone in a diffuse

field the *Random-Energy Response* or *RER*. Note that this measurement is of the *power* output of the microphone.

The easiest way to get an intuitive understanding of the RER of a given polar pattern is that it is simply the square of the surface area of a three-dimensional plot of the pattern. The reason we square the surface area is that we are looking at the power of the output which, as we saw in Section ??, is the square of the signal.

The RER of any polar pattern can be calculated using Equation 5.27.

$$RER = \int_0^\pi \int_0^{2\pi} S^2 \sin \alpha d\phi d\alpha \quad (5.27)$$

where S is the sensitivity of the microphone, α is the angle of rotation around the microphone's "equator" and ϕ is the angle of rotation around the microphones axis. These two angles are shown in the explanation of spherical coordinates in Section ??.

If you're having some difficulties grasping the intricacies of Equation 5.27, don't panic. Double integrals aren't something we see every day. We know from Section ?? that, because we're dealing with integrals, then we must be looking for the area of some shape. So far so good. (The area we're looking for is the surface area of the three-dimensional plot of the polar pattern.)

FINISH THIS OFF

There are a couple of good rules of thumb to remember when it comes to RER.

1. An omni has the greatest sum of sensitivities to sounds from all directions, therefore it has the highest RER of all polar patterns.
2. A cardioid and a bidirectional both have the same RER.
3. A hypercardioid has the lowest RER of all first-order gradient polar patterns.

5.7.11 Random-Energy Efficiency (REE)

It's a little difficult to remember the strange numbers that the RER equation comes up with, so we rarely bother. Instead, it's really more interesting to see how the various microphone polar patterns compare to each other. So, what we do is call the RER of an omnidirectional the "reference" and then look at how the other polar patterns' RER's compare to it.

| Polar Pattern (P : G) | RER |
|-------------------------------|--------------------|
| Omnidirectional (1 : 0) | 4π |
| Subcardioid (0.75 : 0.25) | $\frac{7\pi}{3}$ |
| Cardioid (0.5 : 0.5) | $\frac{4\pi}{3}$ |
| Supercardioid (0.375 : 0.625) | $\frac{13\pi}{12}$ |
| Hypercardioid (0.25 : 0.75) | π |
| Bidirectional (0 : 1) | $\frac{4\pi}{3}$ |

Table 5.8: Random Energy Responses for various microphone polar patterns.

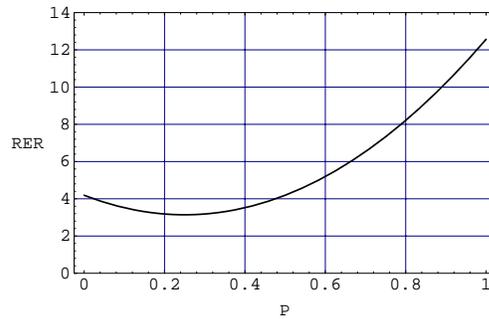


Figure 5.97: Random Energy Response vs. the Pressure component, P in the microphone.

This relationship is called the *Random-Energy Efficiency*, abbreviated *REE* of the microphone, and is calculated using Equation 5.28.

$$REE = \frac{RER}{RER_{omni}} \quad (5.28)$$

So, as we can see, all we're doing is calculating the ratio of the microphone's RER to that of an omni. As a result, the lower the RER, the lower the REE.

This value can be expressed either as a linear value, or it can be calculated in decibels using Equation 5.29.

$$REE_{dB} = 10 \log(REE) \quad (5.29)$$

Notice in Equation 5.29 that we're multiplying by 10 instead of the usual 20. This is because the REE is a power measurement and, as we saw in Section ??, for power you multiply by 10 instead of 20.

In the practical world, this value gives you an indication of the relative outputs of microphone (assuming that they have similar electrical sensitivities, explained in Section ??) when you put them in a reverberant field. Let's say that you have an omnidirectional microphone in the back of a concert hall to pick up some of the swimmy reverberation sound. If you replace it with a hypercardioid in the same location, you'll have to crank up the gain of the hypercardioid by 6 dB to get the same output as the omni because the REE of a hypercardioid is - 6 dB.

| Polar Pattern (P : G) | REE | REE (dB) |
|-------------------------------|-----------------|----------|
| Omnidirectional (1 : 0) | 1 | 0dB |
| Subcardioid (0.75 : 0.25) | $\frac{7}{12}$ | -2.34dB |
| Cardioid (0.5 : 0.5) | $\frac{1}{3}$ | -4.77dB |
| Supercardioid (0.375 : 0.625) | $\frac{13}{48}$ | -5.67dB |
| Hypercardioid (0.25 : 0.75) | $\frac{1}{4}$ | -6.02dB |
| Bidirectional (0 : 1) | $\frac{1}{3}$ | -4.77dB |

Table 5.9: Random Energy Efficiency for various microphone polar patterns.

5.7.12 Directivity Factor (DRF)

Of course, usually microphones (at least for classical recordings in reverberant spaces...) are not just placed far away in the back of the hall. Then again, they're not stuck up the musicians'... uh... *down* the musicians'

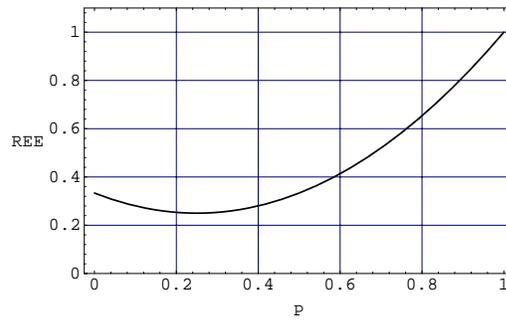


Figure 5.98: Random Energy Efficiency vs. the Pressure component, P in the microphone.

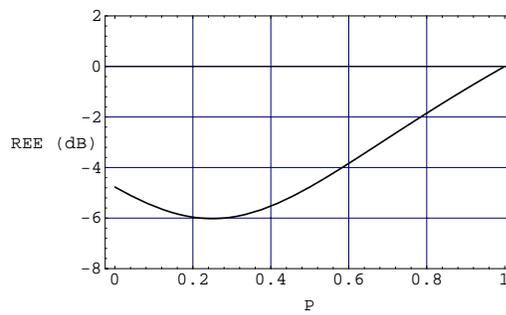


Figure 5.99: Random Energy Response on a decibel scale vs. the Pressure component, P in the microphone.

throats, either. They're somewhere in between where they're getting a little direct sound, possibly on-axis, and some diffuse, reverberant sound. So, one of the characteristics we're interested in is the relationship between these two signals. This specification is called the *Directivity Factor* (the *DRF*) of the microphone. It is the ratio of the response of the microphone in a diffuse field to the response to a free-field source with the same intensity as the diffuse field signal, located on -axis to the microphone. In essence, this is a measure of the direct-to-reverberant ratio of the microphone's polar pattern.

Since

1. the imaginary free field source has the same intensity as the diffuse-field signal, and
2. the power output of the microphone for that signal, on-axis, would be the same as the RER of an omni in a diffuse field...

we can calculate the DRF using Equation 5.31

$$DRF = \frac{1}{REE} \quad (5.30)$$

| Polar Pattern (P : G) | DRF | Decimal equivalent |
|-------------------------------|-----------------|--------------------|
| Omnidirectional (1 : 0) | 1 | 1 |
| Subcardioid (0.75 : 0.25) | $\frac{12}{7}$ | 1.71 |
| Cardioid (0.5 : 0.5) | 3 | 3 |
| Supercardioid (0.375 : 0.625) | $\frac{48}{13}$ | 3.69 |
| Hypercardioid (0.25 : 0.75) | 4 | 4 |
| Bidirectional (0 : 1) | 3 | 3 |

Table 5.10: Directivity Factor for various microphone polar patterns.

5.7.13 Distance Factor (DSF)

We can use the DRF to get an idea of the relative powers of the direct and reverberant signals coming from the microphones. Essentially, it tells us the relative sensitivities of those two signals, but what use is this to us in a panic situation when the orchestra is sitting out there waiting for you to put up the microphones and \$ 1000 per minute is going by while you place the mic's... It's not, really... so we need to translate this number into a useable one in the real world.

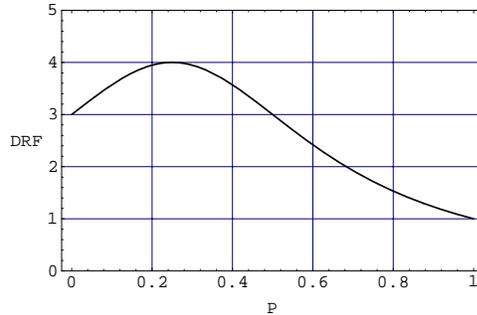


Figure 5.100: Directivity Factor vs. the Pressure component, P in the microphone.

Well, consider a couple of things:

1. if you move away from a sound source in a real room, the direct sound will drop by 6 dB per doubling of distance
2. if you move away from a sound source in a real room, the reverberant sound will not change. This is true, even inside the room radius. The only reason the level drops inside this area is because the direct sound is much louder than the reverberant sound.
3. the relative balance of the direct sound and the reverberant sound is dependent on the DRF of the microphone's polar pattern.

So, we now have a question. If you have a sound source in a reverberant space, and you put an omnidirectional microphone somewhere in front of it, and you want a cardioid to have the same direct-to-reverberant ratio as the omni, where do you put the omni? If you put the two microphones side-by-side, the cardioid will sound closer since it will get the same direct sound, but less reverberant energy than the omni. Therefore, the cardioid must be placed farther away, but how much farther? This is actually quite simple to calculate. All we have to do is to convert the DRF (which is a measurement based on the power output of the microphone) into a *Distance Factor* (or *DSF*). This is done by coming back from the power measurement into an amplitude measurement (because the distance to the sound source is inversely proportional to the relative level of the received direct sound. In other words, if you go twice as far away, you get half the amplitude.) So, we can calculate the DSF using Equation ??.

$$DSF = \sqrt{DRF} \quad (5.31)$$

| Polar Pattern (P : G) | DSF | Decimal equivalent |
|-------------------------------|------------------------|--------------------|
| Omnidirectional (1 : 0) | 1 | 1 |
| Subcardioid (0.75 : 0.25) | $2\sqrt{\frac{3}{7}}$ | 1.31 |
| Cardioid (0.5 : 0.5) | $\sqrt{3}$ | 1.73 |
| Supercardioid (0.375 : 0.625) | $4\sqrt{\frac{3}{13}}$ | 1.92 |
| Hypercardioid (0.25 : 0.75) | 2 | 2 |
| Bidirectional (0 : 1) | $\sqrt{3}$ | 1.73 |

Table 5.11: Distance Factor for various microphone polar patterns.

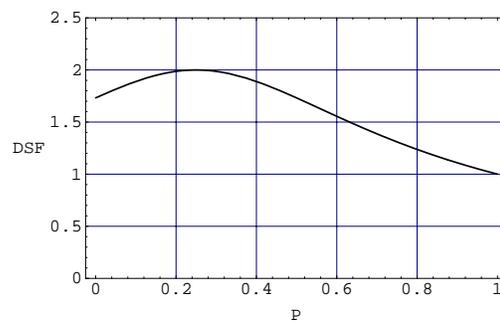


Figure 5.101: Distance Factor vs. the Pressure component, P in the microphone.

So, what does this mean? Have a look at Figure 5.102. All of the microphones in this diagram will have the same direct-to-reverberant outputs. The relative distances to the sound source have been directly taken from Table 5.10 as you can see...

5.7.14 Variable Pattern Microphones

5.7.15 Suggested Reading List

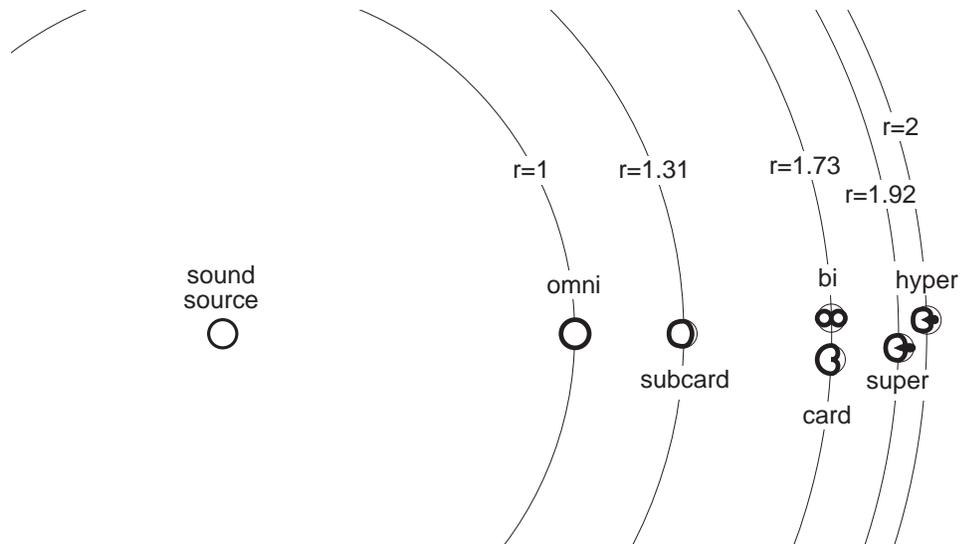


Figure 5.102: Diagram showing the distance factor in practice. In theory, all of the outputs of these microphones at these specific distances from the sound source will all have the same direct-to-reverberant ratios.

5.8 Loudspeakers – Transducer type

Note that this is just an outline at this point. Bear with me while I think about this before I start writing it.

5.8.1 Introduction

Loudspeakers are basically comprised of two things:

1. one or more *drivers* to push and pull the air in the room
2. an *enclosure* (fancy word for “box”) to make the driver sound and look better. This topic is discussed in Section [5.9](#)

We can group the different types of loudspeaker drivers into a number of categories:

1. *Dynamic* (which can further be subdivided into two other types:
 - *Ribbon*
 - *Moving Coil*
2. *Electrostatic*
3. *Servodrive*, which is only found in subwoofers (low-frequency drivers)
4. *Plasma*, which is only found in homes of very rich people

5.8.2 Ribbon Loudspeakers

As we have now seen many times, if you put current through a piece of wire, you generate a magnetic field around it. If that wire is suspended in another magnetic field, then the field that you generate will cause the wire to move. The direction of movement is determined by the polarity of the field that you created using the current in the wire. The velocity of the movement is determined by the strength of the magnetic fields, and therefore the amount of current in the wire.

Ribbon loudspeakers use exactly this principle. We suspend a piece of corrugated aluminum in a magnetic field and connect a lead wire to each end of the ribbon as is shown in Figure [5.103](#).

When we apply a current through the wire and ribbon, we generate a magnetic field, and the ribbon moves. If the current goes positive and

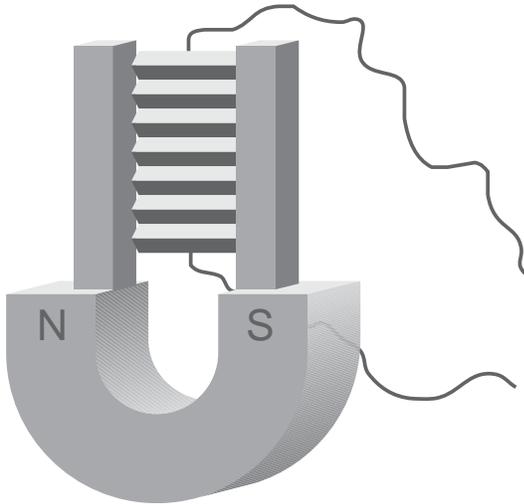


Figure 5.103: When you put current in the wire, a magnetic field is created around it and the ribbon. Therefore the ribbon moves.

negative, then the ribbon moves forwards and backwards. Therefore, we have a loudspeaker driver where the ribbon itself is the diaphragm of the loudspeaker.

This ribbon has a low mass, so it's easy to move quickly (making it appropriate for high frequencies) but it doesn't create a large magnetic field, so it cannot play very loudly.. Also, if you thump the ribbon with your finger, you'll see that it has a very low resonant frequency, mainly because it's loosely suspended. As a result, this is a good driver to use for a tweeter, but it's difficult to make it behave for lower frequencies.

There are advantages and disadvantages to using ribbon loudspeakers:

Advantages

- The ribbon has a low mass, therefore it's good for high frequencies.

Disadvantages

- You can't make it very large, or have a very large excursion (it'll break apart) so it's not good for low frequencies or high sound pressure levels.
- The magnets have to produce a very strong magnetic field (making them very heavy) because the ribbon can't.

- The impedance of the driver is very low (because it's just a piece of aluminum) – so it may be a nasty load for your amplifier, unless you look after this using a transformer.

5.8.3 Moving Coil Loudspeakers

REWRITE ALL OF THIS CHAPTER

Think back to the chapter on electromagnetism and remember the right hand rule. If you put current through a wire, you'll create a magnetic field surrounding it; likewise if you move a wire in a magnetic field, you'll induce a current. A moving coil loudspeaker uses a coil of wire suspended in a stationary magnetic field (compliments of a permanent magnet). If you send current through the coil, it induces a magnetic field around the coil (think of a transformer). Since the coil is suspended, it is free to move, which it does according to the strengths and directions of the two magnetic fields (the permanent one and the induced one). The bigger the current, the bigger the field, therefore the greater the movement.



Figure 5.104:

In order for the system to work well, you need reasonably strong magnetic fields. The easiest way to do this is to use a really strong permanent magnet. You could also improve the *packing density* of the voice coil. This essentially means putting more metal in the same space by changing the cross-section of the wire. The close-ups shown below illustrate how this can be done. The third (and least elegant) method is to add more wire to the coil. We'll talk about why this is a bad idea later.

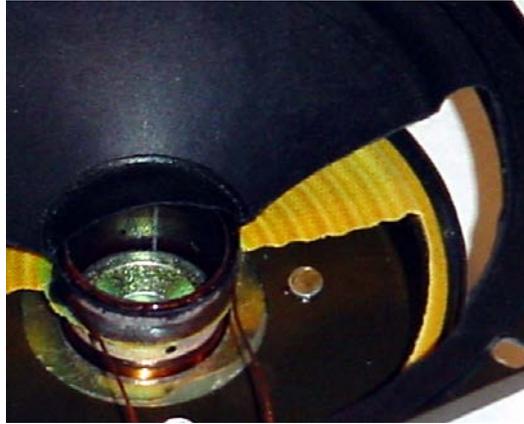


Figure 5.105:

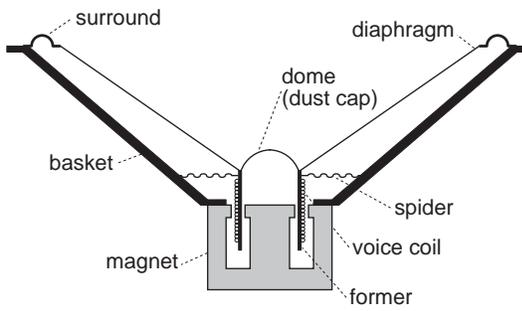


Figure 5.106: A cross section of a simplified model of a moving coil loudspeaker.

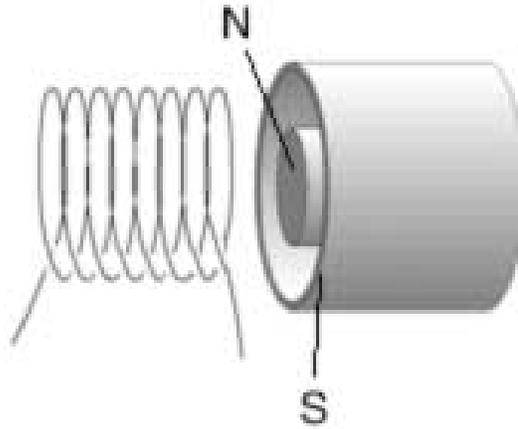


Figure 5.107: A diaphragm is glued to the front (this side) of the coil (called the voice coil. It has two basic purposes: 1) to push the air and 2) to suspend the coil in the magnetic field

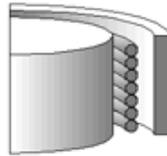


Figure 5.108: Voice coil using wire with a round cross section. This is cheap and easy to make, but less efficient.

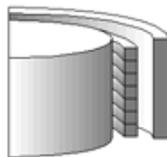


Figure 5.109: Voice coil using wire with a flat cross section. This has greater packing density, producing a stronger magnetic field and is therefore more efficient.

Loudspeakers have to put a great deal of acoustic energy into a room, so they have to push a great deal of air. This can be done in one of two ways:

- use a big diaphragm and move lots of molecules by a little bit (big diaphragm, small excursion)
- use a little diaphragm and move a few molecules by a lot (little diaphragm, big excursion)

In the first case, you have to move a big mass (the diaphragm and the air next to it) by a little, in the second case, you move a little mass by a lot – either way you need to get a lot of energy into the room. If the loudspeaker is inefficient, you'll be throwing away large amounts of energy produced by your power amplifier. This is why you worry about things like packing density of the voice coil. If you try to solve the problem simply by making the voice coil bigger (by adding more wire), you also make it heavier, and therefore harder to move.

There are advantages and disadvantages to using moving coil loudspeakers:

Advantages

- They are pretty rugged (that's to say that they can take a lot of punishment – not that they look nice if they're carpeted...)
- They can make very loud sounds
- They are easy (and therefore cheap) to construct

Disadvantages

- There's a big hunk of metal (the voice coil) that you're trying to move back and forth. In this case, inertia is *not* your friend. The more energy you want to emit (you'll need lots in low frequencies) the bigger the coil – the bigger the heavier – the heavier, the harder to move quickly – the harder to move quickly, the more difficult it is to produce high frequencies. The moral: a driver (single coil and diaphragm without an enclosure) can't effectively produce all frequencies for you. It has to be optimized for a specific frequency range.
- They're heavy because strong permanent magnets weigh a lot. This is especially bad if you're part of a road crew for a rock band.
- They have funny-looking impedance curves – but we'll talk about that later.

5.8.4 Electrostatic Loudspeakers

When we learned how capacitors worked, we talked a bit about *electrostatic attraction*. This is the tendency for oppositely charged particles to attract each other, and similarly charged particles to repel each other. This is what causes electrons in the plates of a capacitor to bunch up when there's a voltage applied to the device. Let's now take this principle a step farther.

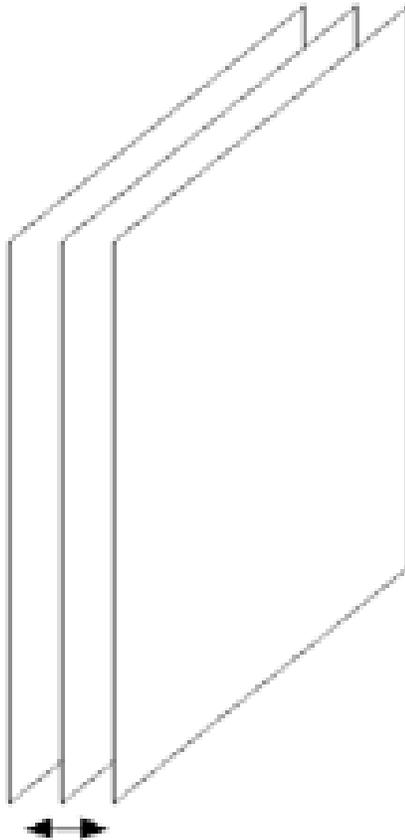


Figure 5.110: Put three conductive plates side by side (just like in a capacitor, but with an extra plate). Have the two outside plates fixed, but the middle one suspended so that it can move (shown by the little arrow at the bottom of the diagram).

Charge the middle plate with a DC *polarizing* voltage. It is now equally attracted to the two outside plates (assuming that their charges are equal).

If we change the charge on the outside plates such that they are different

from each other, the inside plate will move towards the plate of more opposite polarity.

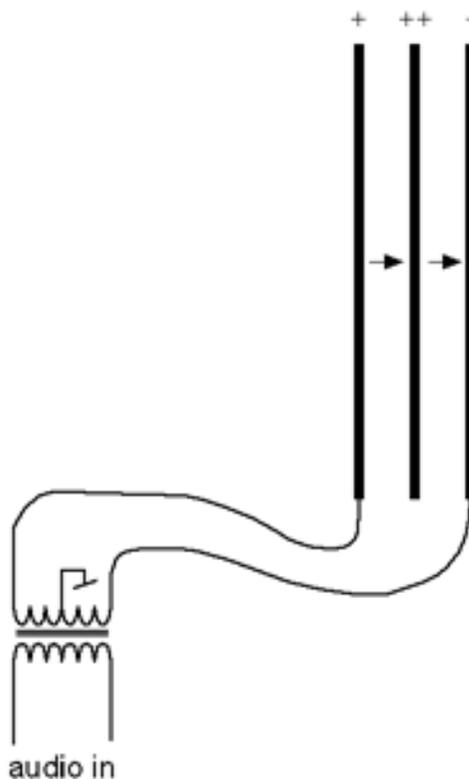


Figure 5.111: If the middle plate is very positive and the outside plates are slightly positive and slightly negative relative to each other, the middle plate will be attracted to (and be pulled towards) the more negative plate while it is repelled (and therefore pushed away from) the more positive plate. This system is known as push-pull for obvious reasons.

If we then perforate the plates (or better yet, make them out of a metal mesh) the movement of the inside plate will cause air pressure changes that radiate through the holes into the room.

There are (as expected) advantages and disadvantages to this system.

Disadvantages

- In order for this system to work, you need enough charge to create an adequate attraction and repulsion. This requires one of two things:

- a very big polarizing voltage on the middle plate (on the order of 5000 V). Most electrostatics use a high polarizing voltage (hence the necessity in most models to plug them into your wall and your amplifier).

- a very small gap between the plates. If the gap is too small, you can't have a very big excursion of the diaphragm, therefore you can't produce enough low frequency energy without a big diaphragm (it's not unusual to see electrostatics that are $2m^2$).

- Starting prices for these things are in the thousands of dollars.

Advantages

- The diaphragm can be *very* thin, making it practically massless – or at least approaching the mass of the air it's moving. This in turn, gives electrostatic loudspeakers a very good transient response due to an extremely high high-frequency cutoff.

- You can see through them (this impresses visitors)

- They have symmetrical distortion – we'll talk a little more about this in the following section on enclosures.

Electrical Impedance

Hmmmmmm... loudspeaker impedance... We're going to only look at moving coil loudspeakers. (If you want to know more about this – or anything about the other kinds of drivers, get a copy of the Borwick book I mentioned earlier.)

We'll begin by looking at the impedance of a resistor:

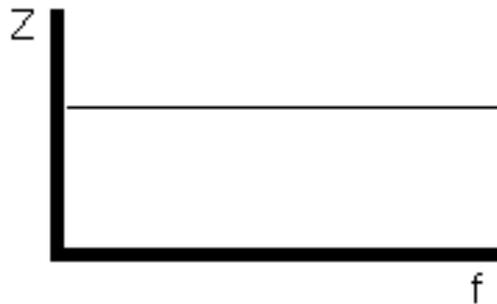


Figure 5.112: The impedance is the same at all frequencies.

Next, we learned the impedance characteristic of a capacitor

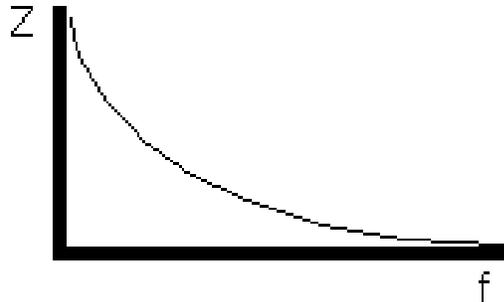


Figure 5.113: The higher the frequency, the lower the impedance.

Finally, we learned that the impedance characteristic of an inductor (a coil of wire) is the opposite to that of a capacitor.

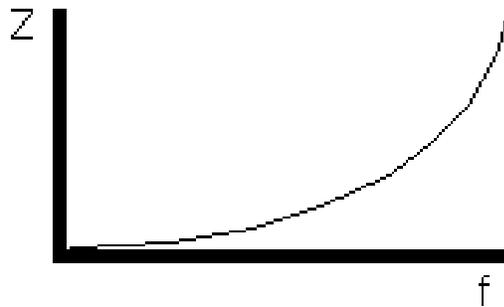


Figure 5.114: The higher the frequency, the higher the impedance.

Therefore, even before we attach a diaphragm or a magnet, a loudspeaker coil has the above impedance. The problem occurs when you add the diaphragm, magnet and enclosure which, together create a resonant frequency for the whole system. At that frequency, a little energy will create a lot of movement because the system is naturally resonating. This causes something interesting to happen: the loudspeaker starts acting more like a microphone of sorts and produces its own current in the opposite direction to that which you are sending in. This is called *back EMF* (*Electro-Motive Force*) and its end result is that the impedance of the speaker is increased at its resonant frequency. The resulting curve looks something like this (notice the “bump” in the impedance at the resonant frequency of the system):

Note the following things:

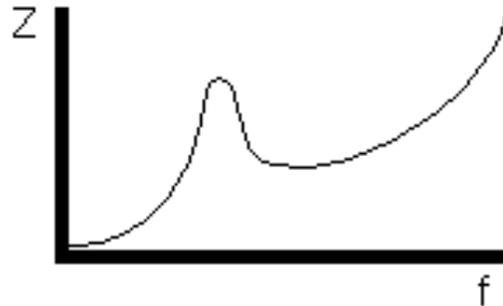


Figure 5.115: INSERT CAPTION

- The curve is for a single driver in an enclosure
- For a driver rated at an impedance of 8Ω , the curve varies from about 7Ω to 80Ω (at the resonance point).

As you add more drivers to the system, the impedance curve gets more and more complicated. The important thing to remember is that an 8Ω speaker is almost never 8Ω .

Suggested Reading List

5.9 Loudspeaker acoustics

Note that this is just an outline at this point. Bear with me while I think about this before I start writing it.

5.9.1 Enclosures

- *Enclosure Design* (i.e. “what does the box look like?”) of which there are a number of possibilities including the following:

- *Dipole Radiator* (no enclosure) - *Infinite Baffle* - *Finite Baffle* - *Folded Baffle* - *Sealed Cabinet* (aka Acoustic Suspension) - *Ported Cabinet* (aka Bass Reflex)

- *Horns* (aka Compression Driver)

Dipole Radiator

(no enclosure)

aka a Doublet Radiator

Since both sides of a dipole radiator’s diaphragm are exposed to the outside world (better known as your listening room) there are opposite polarity pressures being generated in the same space (albeit at different locations) simultaneously. As the diaphragm moves “outwards” (i.e. towards the listener), the resulting air pressure at the “front” of the diaphragm is positive while the pressure at the back of the diaphragm is of equal magnitude but opposite polarity.

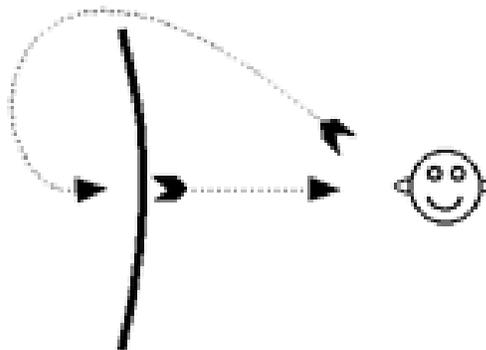


Figure 5.116: INSERT CAPTION

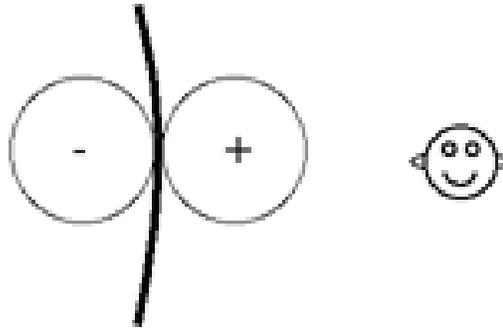


Figure 5.117: INSERT CAPTION

At high frequencies, the diaphragm *beams* (meaning that it's directional – it only sends the energy out in a narrow beam rather than dispersing it in all directions) as if the rear side of the diaphragm were sealed. The result of this is that the negative pressure generated by the rear of the diaphragm does not reach the listener. For more information on beaming see section 1.6 of the *Loudspeaker and Headphone Handbook* edited by John Borwick and published by Butterworth.

At low frequencies, the delay time required for energy to travel from the back of the diaphragm to the front and towards the listener is a negligible part of the phase of the wave (i.e. the distance around the diaphragm is very small compared to the wavelength of the frequency). Since the energy at the rear of the diaphragm is opposite in polarity to the front, the two cancel each other out at the listening position.

The result of all this is that a dipole radiator doesn't have a very good low frequency response – in fact, it acts much like a high pass filter with a slope of 6 dB/octave and a cutoff frequency which is dependent on the physical dimensions of the diaphragm.

How do we fix this?

Infinite Baffle

The biggest problem with a dipole radiator is caused by the fact that the energy from the rear of the diaphragm reaches the listener at the front of the diaphragm. The simplest solution to this issue is to seal the back of the diaphragm as shown.

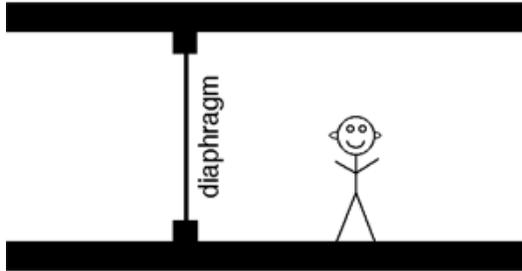


Figure 5.118: INSERT CAPTION

This is called an “infinite baffle because the diaphragm is essentially mounted on a wall of infinite dimensions. (Another way to draw this would have been to put the diaphragm mounted in a small hole in a wall that goes out forever. The problem with that kind of drawing is that my stick man wouldn’t have had anything to stand on.)

Disadvantages

- You’re automatically throwing away half of your energy (the half that’s radiated by the rear of the diaphragm)
- Costs... How do you build a wall of infinite dimensions?

Advantages

- Lower low-frequency cutoff than a dipole radiator.

Now the low cutoff frequency is dependent on the diameter of the diaphragm according to the following equation:

$$f_c = \frac{c}{\pi D} \quad (5.32)$$

Where f_c is the low frequency cutoff in Hz (-3 dB point) c is the speed of sound in air in m/s D is the diameter of the loudspeaker in m

Unlike the dipole radiator, frequencies below the cutoff have a pressure response of 12 dB/octave (instead of 6 dB/octave). Therefore, if we’re going down in frequency comparing a dipole radiator and an infinite baffle with identical diaphragms, the infinite baffle will go lower, but drops off more quickly past that frequency.

Finite Baffle

Instead of trying to build a baffle with infinite dimensions, what would happen if we mounted the diaphragm on a circular baffle of finite dimensions as

shown below?

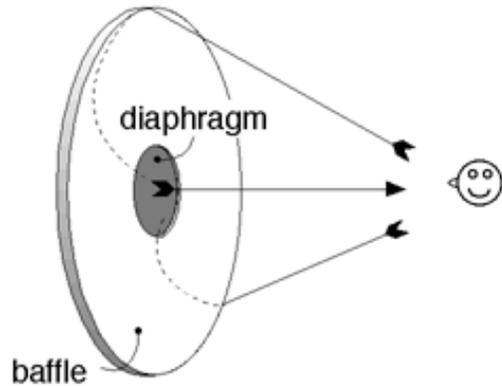


Figure 5.119: INSERT CAPTION

Now, the circular baffle causes the energy from the rear of the driver to reach the listener with a given delay time determined by the dimensions of the baffle. This causes a comb-filtering effect with the first notch happening where the path length from the rear of the driver to the front of the baffle equals one wavelength as shown below:

One solution to this problem is to create multiple path lengths by using an irregularly-shaped baffle. This causes multiple delay times for the pressure from the rear of the diaphragm to reach the front of the baffle. The technique will eliminate the comb filtering effect (no notches above the transition frequency) but we still have a 6 dB/octave slope in the low end (below the transition frequency). If the baffle is very big (approaching infinite relative to the wavelength of the lowest frequencies) then the slope in the low end approaches 12 dB/octave – essentially, we have the same effect as if it were an infinite baffle.

If the resonant frequency of the driver, below which the roll-off occurs at a slope of 6 dB/octave, is the same as (or close to) the transition frequency of the baffle, the slope becomes 12 or 18 dB/octave (dependent on the size of the baffle – the slopes add). The resonant frequency of the driver is the rate at which the driver would oscillate if you thumped it with your finger – though we'll talk about that a little more in the section on impedance.

What do you do if you don't have enough room in your home for big baffles? You fold them up!

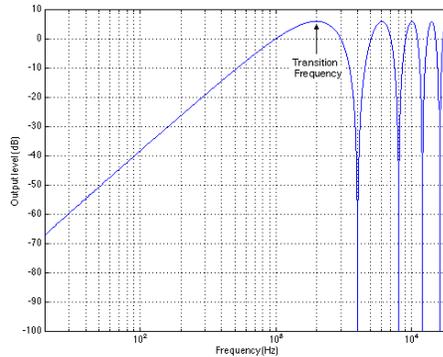


Figure 5.120: The natural frequency response of a radiator mounted in the centre of a circular finite baffle. The transition frequency is where the extra path length from the rear of the diaphragm to the listener is one half of a wavelength. Note that the plotted transition frequency is unnaturally high for a real loudspeaker. Also, compare this frequency response to the natural frequency response of a bidirectional microphone shown in Figure XXX in Section 5.7.

Folded Baffle

(aka Open-back Cabinet)

A folded baffle is essentially a large flat baffle that has been “folded” into a tube which is open on one end (the back) and sealed by the driver at the other end as is shown below.

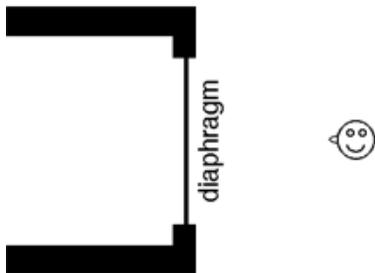


Figure 5.121: INSERT CAPTION

The fact that there’s a half-open tube in the picture causes a nasty resonance (see the section on resonance in the Acoustics section of this textbook for more info on why this happens) at a frequency determined by the length of the tube.

wavelength of the resonant frequency = 4 * length of the tube

The low frequency response of this system is the same as in the case of finite baffles.

Sealed Cabinet

(aka Acoustic Suspension)

We can eliminate the resonance of an open-back cabinet by sealing it up, thus turning the tube into a box.

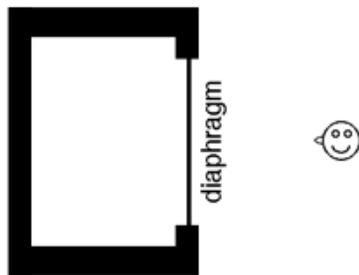


Figure 5.122: INSERT CAPTION

Now the air sealed inside the enclosure acts as a spring which pushes back against the rear of the diaphragm. This has a number of subsequent effects:

- increased resonant frequency of the driver
- reduced efficiency of the driver (because it has to push and pull the “spring” in addition to the air in the room)
- non-symmetrical distortion (because pushing in on the “spring” has a different response than pulling out on it)

We’re throwing away a lot of energy as in the case of the infinite baffle because the back of the driver is sealed off to the world.

Usually the enclosure has an acoustic resonant frequency (sort of like little room modes) which is lower than the (now raised...) resonant frequency of the driver. This effectively lowers the low cutoff frequency of the entire system, below which the slope is typically 12 dB/octave.

Ported Cabinet

(aka Bass Reflex)

You can achieve a lower low cutoff frequency in a sealed cabinet if you cut a hole in it – the location is up to you (although, obviously, different locations will have different effects...)

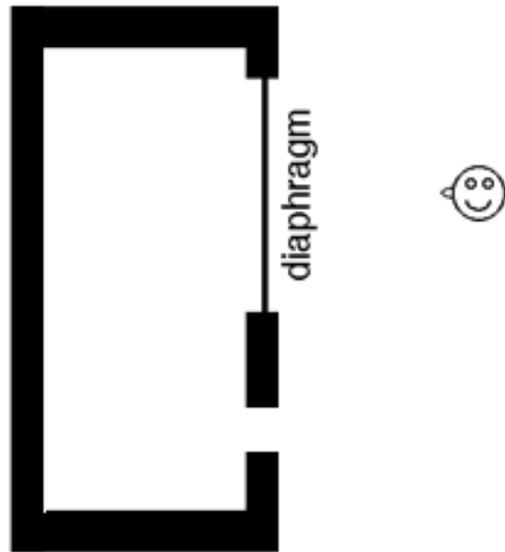


Figure 5.123: INSERT CAPTION

If you match the cabinet resonance, which is determined by the volume of the cabinet and the diameter and length of the hole (look up Helmholtz resonators in the Acoustics section), with the speaker driver resonance properly (NOTE: Could someone explain to me what a “properly matched” cabinet and driver resonance means?), you can get a lower low frequency cutoff than a sealed cabinet, but the low frequencies now roll off at a slope of 24 dB/octave.

Horns

(aka Compression Drivers)

The efficiency of a system in transferring power is determined by how the power delivered to the system relates to the power received in the room. For example, if we were to send a signal to a resistor from a function generator with an output impedance of 50Ω , I would get the most power dissipation in the resistor if it had a value of 50Ω . Any other value would mean that less power would be dissipated by the device. This is because its impedance

would be matched to the internal impedance of the function generator.

Consider loudspeakers to have the same problem: the transfer motion (*kinetic energy*, if you're a geek) or power from a moving diaphragm to the air surrounding it. The problem is that the air has much lower acoustic impedance than the driver (meaning, it's easier to move back and forth), therefore our system isn't very efficient.

Just as we could use a transformer to match electrical impedances in a circuit (providing a R_{eq} rather than the actual R) we can use something called a *horn* to match the acoustic impedances between a driver and the air.

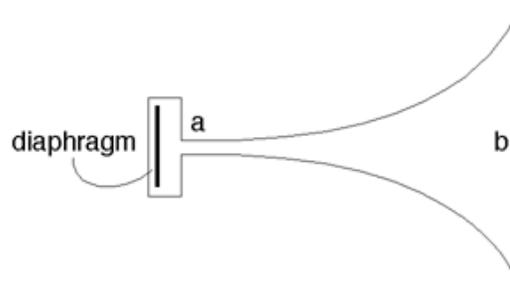


Figure 5.124: INSERT CAPTION

There are a couple of things to point out regarding this approach:

- The diaphragm is mounted in a small chamber which opens to the horn
- At point "a" (the small area of the horn) there is a small area of air undergoing a large excursion. The purpose of the horn is to slowly change this into the energy at point "b" where we have a large area with a small excursion. In other words, the horn acts as an impedance matching device, ensuring that we have an optimal power transfer from the diaphragm to the air in the room.

- Notice that the diaphragm is large relative to the opening to the horn. This means that, if the diaphragm moves as a single unit (which it effectively does) there is a phase difference (caused by propagation delay differences to the horn) between the pressure caused by the centre of the diaphragm and the pressure caused by the edge of the diaphragm. Luckily, this problem can be avoided with the insertion of a plug which ensures that the path lengths from different parts of the diaphragm to the centre of the throat of the horn are all equal.

In the above diagram, the large black area is the cross section of the

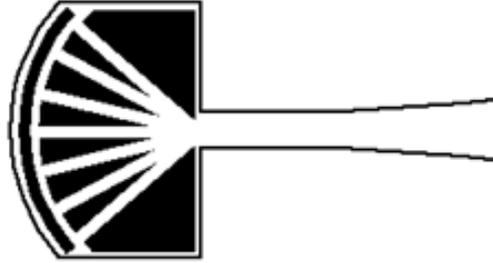


Figure 5.125: INSERT CAPTION

plug with hole drilled in it (the white lines). Each hole is the same length, ensuring that there is no interference, either constructive or destructive, caused by multiple path lengths from the diaphragm to the horn.

5.9.2 Other Issues

Beaming

As a general rule of thumb, a loudspeaker is almost completely omnidirectional (it radiates energy spherically – in all directions) when the driver is equal to or less than $1/4$ the wavelength of the frequency being produced. The higher the frequency, the more directional the driver.

Why is this a problem?

High frequencies are very directional, therefore if your ear is not in the direct “line of fire” of the speaker (that is to say, “on axis”) then you are getting some high-frequency roll-off.

In the case of low frequencies, the loudspeaker is omnidirectional, therefore you are getting energy radiating behind and to the sides of the speaker. If there’s a nearby wall, floor or ceiling, the pressure which bounces off it will add to that which is radiating forward and you’ll get a boost in the low end.

- If you’re near 1 surface, you get 3 dB more power in the low end
- If you’re near 2 surfaces, you get 6 dB more power in the low end
- If you’re near 3 surfaces, you get 9 dB more power in the low end

The moral of this story? Unless you want more low end, don’t put your loudspeaker in the corner of the room.

Multiple drivers

We said much earlier that drivers are usually optimized for specific frequency bands. Therefore, an easy way to get a wide-band loudspeaker is to use multiple drivers to cover the entire audio range. Most commonly seen these days are 2-way loudspeakers (meaning 2 drivers, usually in 1 enclosure), although there are many variations on this.

If you're going to use multiple drivers, then you have some issues to contend with:

1 – Crossovers

- You don't want to send high frequencies to a low frequency driver (aka *woofer* or, more destructively, low frequencies to a high-frequency driver (aka *tweeter*). In order to avoid this, you filter the signals getting sent to each driver – the woofer's signal is routed through a low-pass filter, while the tweeter's is filtered using a high-pass. If you are using mid-range drivers as well, then you use a band-pass filter for the signal. This combination of filters is known as a crossover.

Most filters in crossovers have slopes of 12 or 18 dB/octave (although it's not uncommon to see steeper slopes) and have very specific designs to minimize phase distortion around the *crossover frequencies* (the frequencies where the two filters overlap and therefore the two drivers are producing roughly equal energy) This is particularly important because crossover frequencies are frequently (sorry... I couldn't resist at least one pun) around the 1 – 3 kHz range – the most sensitive band of our hearing.

There are two basic types of crossovers, *active* and *passive*.

Passive crossovers

These are probably what you have in your home. Your power amplifier sends a full-bandwidth signal to the crossover in the loudspeaker enclosure which, in turn, sends filtered signals to the various drivers. This system has a drawback in that it wastes power from your amplifier – anything that is filtered out is lost power. “Audiophiles” (translation: “people with too much disposable income who spend more time reading reviews of their audio equipment than they do actually listening to it”) also complain about issues like back EMF and cabinet vibrations which may or may not affect the crossover. (I'll believe this when I see some data on it...). The advantage of passive crossovers is that they're idiot-proof. You plug in one amplifier to one speaker input and the system works and works well. Also – they're inexpensive.

Active crossovers

These are filters that precede the power amplifier in the audio chain (the amplifier is then directly connected to the individual driver). They are more efficient, since you only amplify the required frequency band for each driver – then again, they’re more expensive because you have to buy the crossover plus extra amplifiers.

2 – *Crossover Distortion*

There is a band of frequencies around the cutoffs of the crossover filters where the drivers overlap. At this point you have roughly equal energy being emitted by at least two diaphragms. There is an acoustic interaction between these two which must be considered, particularly because it is in the middle of the audio range usually.

In order to minimize problems in this band, the drivers must have matched distances to the listener, otherwise, you’ll get comb filtering due to propagation delay differences. This so-called *time aligning* can be done in one of two ways. You can either build the enclosure such that the tweeter is set back into the face of the cabinet, so its voice coil is vertically aligned with the voice coil of the deeper woofer. Or, alternately you can use an electronic delay to retard the arrival of the signal at the closer driver by an appropriate amount.

3 – *Interaction between drivers*

Remember that the air inside the enclosure acts like a spring which pushes and pulls the drivers contrary to the direction you want them to move in. Imagine a woofer moving into a cabinet. This increases the air pressure inside the enclosure which pushes out against both the woofer AND the tweeter. This is a bad thing. Some manufacturers get around this problem by putting divided sections inside the cabinet – others simply build separate cabinets – one for each driver (as in the B and W 801).

Electrical Impedance

Hmmmmmm... loudspeaker impedance... We’re going to only look at moving coil loudspeakers. (If you want to know more about this – or anything about the other kinds of drivers, get a copy of the Borwick book I mentioned earlier.)

We’ll begin by looking at the impedance of a resistor:

Next, we learned the impedance characteristic of a capacitor

Finally, we learned that the impedance characteristic of an inductor (a coil of wire) is the opposite to that of a capacitor.

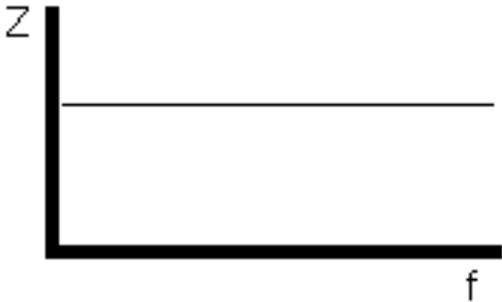


Figure 5.126: The impedance is the same at all frequencies.

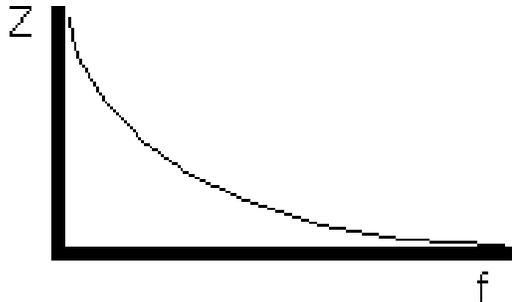


Figure 5.127: The higher the frequency, the lower the impedance.

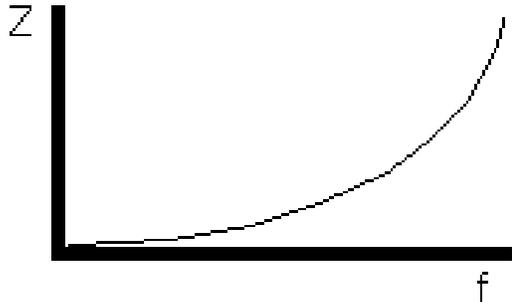


Figure 5.128: The higher the frequency, the higher the impedance.

Therefore, even before we attach a diaphragm or a magnet, a loudspeaker coil has the above impedance. The problem occurs when you add the diaphragm, magnet and enclosure which, together create a resonant frequency for the whole system. At that frequency, a little energy will create a lot of movement because the system is naturally resonating. This causes something interesting to happen: the loudspeaker starts acting more like a microphone of sorts and produces its own current in the opposite direction to that which you are sending in. This is called *back EMF* (*Electro-Motive Force*) and its end result is that the impedance of the speaker is increased at its resonant frequency. The resulting curve looks something like this (notice the “bump” in the impedance at the resonant frequency of the system):

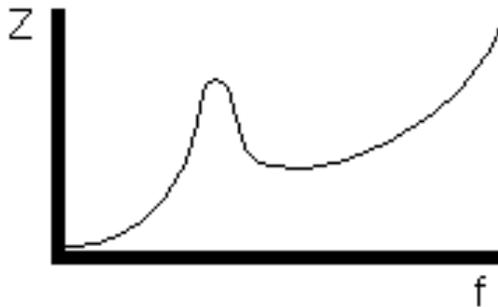


Figure 5.129: INSERT CAPTION

Note the following things:

- The curve is for a single driver in an enclosure
- For a driver rated at an impedance of 8Ω , the curve varies from about 7Ω to 80Ω (at the resonance point).

As you add more drivers to the system, the impedance curve gets more and more complicated. The important thing to remember is that an 8Ω speaker is almost never 8Ω .

Suggested Reading List

5.10 Power Amplifiers

NOT YET WRITTEN

5.10.1 Introduction

Chapter 6

Electroacoustic Measurements

6.1 Tools of the Trade

I could describe a piece of paper using a list of its measurements. If I said, “one sheet of white paper, 8.5 inches by 11 inches” you’d have a pretty clear idea of what I was talking about. The same can be done for audio gear – we can use various measurements to describe the characteristics of a given piece of equipment. Although it would be impossible to perform a set of measurements that describe the characteristics exactly, we can get a good idea of what’s going on.

Before we look at the various electroacoustic measurements you can perform on a piece of audio equipment, we’ll look at the various tools that we use.

6.1.1 Digital Multimeter

A *digital multimeter* (also known as a *DMM* for short) is a handy device that replaces a number of different analog tools such as a *voltmeter* and a *galvanometer*. The particular tools incorporated into one DMM depends on the manufacturer and the price range.

At the very least, a DMM will have meters for measuring voltage and current – both AC and DC. These are typically chosen in advance, so you set the switch on the meter to “AC Voltage” to do that particular measurement, for example. In addition, the DMM will invariably include a meter for measuring the resistance of a device (such as a resistor, for example...)

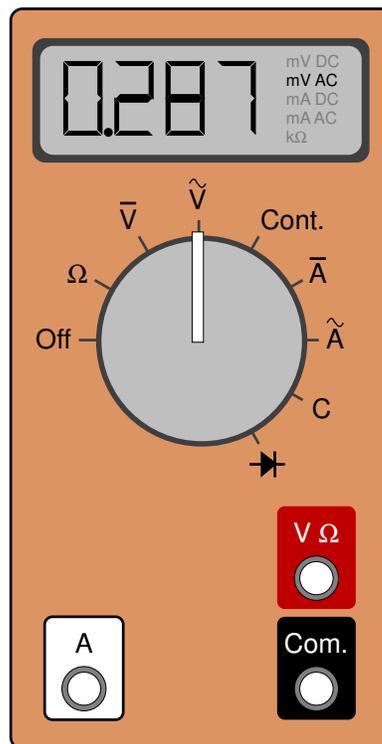


Figure 6.1: A cheap digital multimeter. This thing is about the size of a credit card, not much thicker and cost around \$20Cdn.

Usually, the last meter you'll be guaranteed is a great little tool called a *continuity tester*. This is essentially a simplified version of a resistance meter where the only output of the meter is a "beep" when there is a low-resistance connection between the two leads.

Some fancier (and therefore more expensive) units will also include a diode tester and a meter that can measure capacitance.

There are a couple of things to be concerned about when using one of these devices.

First, we'll just look at the basic operation. Let's say that you wanted to measure the DC voltage between two points on your circuit. You set the DMM to a DC Voltage measurement and touch the two leads on the two points. Be careful not to touch more than one point on the circuit with each lead of the DMM. This could do some damage to your circuit... Note that one of the leads is labelled "ground" which does not necessarily mean that it's actually connected to ground in your circuit. All it really means is that, for the DMM, that particular lead is the reference. The voltage displayed is the voltage of the other lead relative to the "ground" lead.

When you're measuring voltage, the two leads of the DMM theoretically have an infinite impedance between them. This is necessary because, if they didn't, you'd change the voltage that you were measuring by measuring it. For example, let's pretend that you have two identical resistors in series across a voltage source. If your DMM had a small impedance across the two leads, and you used it to measure the voltage across one of the resistors, then the small impedance of the DMM would be in parallel with the resistor being measured. Consequently, the voltage division will change and you won't get an accurate measurement. Since the impedance between the leads is infinity Ω (or at least, pretty darned close...), when it's in parallel with a resistor, it's as if it's not there at all, and it consequently has no effect on the circuit being measured.

If you're measuring current, then you have to use the DMM a little differently. Whereas a voltage measurement is performed by putting the DMM in parallel with the resistor being measured, a current measurement is performed by putting the DMM in series with the circuit. This ensures that the current goes through the DMM on its way through the circuit. Also, as a result, it's good to remember that the impedance across the leads for a current measurement is 0Ω – otherwise, it would change the current in the circuit and you'd get an inaccurate measurement.

Now, there's an important thing to remember here – when the DMM is in voltage measuring mode, the impedance is infinity. When it's in current measuring mode, it's 0Ω . Therefore, if you are measuring current, and

you accidentally forget to set the meter back to voltage measuring mode to measure the voltage across a power supply or a battery, you'll probably melt something. Remember that power supplies and batteries don't like being shorted out with a 0Ω impedance – it means that you're asking for infinite current...

Another thing: a DMM measures resistance (and continuity) by putting a small voltage difference across its leads and measuring the resulting current through it. The smaller the current, the bigger the resistance. This also means, however, that your DMM in this case, is its own little power supply – so you don't want to go poking at a circuit to measure resistors without lifting the resistor out – one leg of the resistor will do.

One last thing to remember: the AC measurements in a typical DMM are supposed to be RMS measurements. In fact, they're pseudo-RMS, so you can't usually trust them. Remember from way back that the RMS of a sine wave is about 0.707 of its peak value? Well, a DMM assumes that this is always the case – so it measures the peak value, multiplies by 0.707 and displays that number. This is fine if all you ever do is measure sine waves (like most electricians do...) but we audio types deal with waveforms other than sinusoids, so we need to use something a little more accurate. Just remember, if you're using a DMM and it's not a sine wave, then you're looking at the wrong answer.

A small thing that you can usually ignore: a DMM is only accurate (such as it is...) within a limited frequency range. Don't expect a \$10 DMM to be able to give you a reliable reading on a 30 kHz sine tone.

6.1.2 True RMS Meter

A *true RMS meter* is just a fancy DMM that actually does a real RMS voltage measurement instead of some quick and dirty short cut to the wrong answer. Unfortunately, because there's not much demand for these things, they're a LOT more expensive than a run-of-the-mill DMM. Bet on spending tens to hundreds of times the cost of a DMM to get a true RMS measurement. They look exactly the same as a normal DMM, but they have the words "True RMS" written on it somewhere – that and the price tag will be higher (on the order of about \$ 300 Cdn).

Note that a true RMS meter usually suffers from the same limited frequency range problem as a cheap DMM. The good thing is that a true RMS meter will have a much wider range, and the manual will probably tell you what that range is, so you know when to stop trusting it.

You may also see a DMM that has a switchable low-pass filter on it.

This may seem a little silly at first, but it's actually really useful for doing noise measurements. For example, if you want to measure the noise output of a microphone preamp, you'll have to do a true RMS measurement, but you're not really worried about noise above, say, 100 kHz (then again, maybe you are... I don't know...). If you just did the measurement without a low pass filter, you'd get a much bigger value than you'll like, and one that isn't representative of the audible noise level. So, in this case, the low pass filter is your friend. If your true RMS meter doesn't have one, you might want to build one out of a simple RC circuit. Just remember, though, that if you're stating a noise measurement that you did, include the LPF magnitude characteristic with your values.

6.1.3 Oscilloscope

Digital multimeters and true RMS meters are great for doing a quick measurement of a signal and getting a single number to describe it – the problem is that you really don't know all that much about the signal's shape. For example, if I handed you two wires and asked you to give me a measurement of their voltage difference with a true RMS meter, you'd tell me one, maybe two numbers (the AC and DC components can be measured separately, remember...). But you don't know if it's a sine wave or a square wave or a Britney Spears tune.

in order to see the shape of the voltage waveform, we use an *oscilloscope*. This is a device which can be seen in the background of every B science fiction film ever made. All evil scientists and aliens seem to have oscilloscopes in their labs measuring sine waves all the time.

An oscilloscope displays a graph showing the voltage being measured as it changes in time. It shows this by sweeping a bright little dot across the screen from left to right and moving it up when the voltage is positive and down when it's negative.

For the remainder of this discussion, keep referring back to the drawing of an oscilloscope in Figure 6.2. The first thing to notice is that the display screen is divided into squares – 10 across and 8 down. These squares are called *divisions* and each is further divided into five sections using little tick marks on the centre lines – the X- and Y-axes. As we'll see, these divisions are our only reference to the particular characteristics of the waveform.

You basically have control over two parameters on a 'scope – the speed of the dot as it goes across the screen and the vertical magnification of the voltage. There are a couple of extra controls that we'll look at one by one. We'll do this section by section on the panel in Figure 6.2.

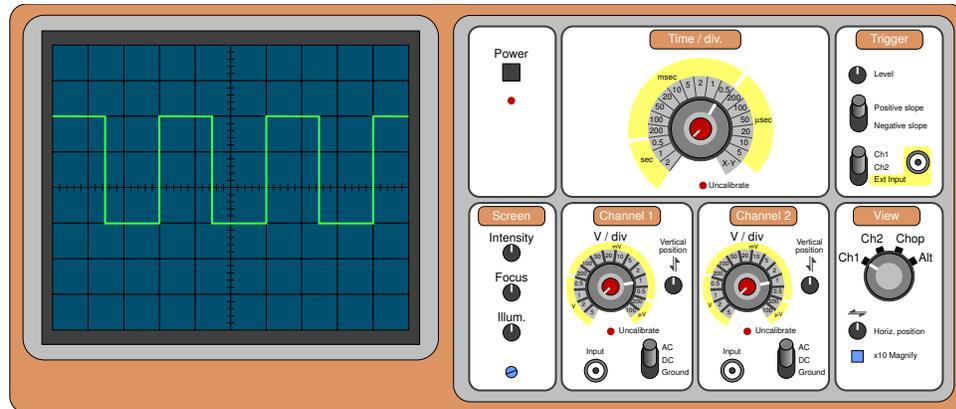


Figure 6.2: A simple dual-channel oscilloscope.

Power

This section is simple – it contains the Power switch – the on/off button.

Screen

Typically, there is a section on the oscilloscope that lets you control some basic aspects of the display like the *Intensity* of the beam – a measure of its brightness, the *focus* and the screen *Illumination*. In addition, there is also probably a small hole showing the top of a screwdriver potentiometer. If this is labeled, it is called the *trace rotation*. This controls the horizontal angle of the dot on the screen and shouldn't need to be adjusted very often – once a month maybe.

Time / div.

The horizontal speed of the dot is controlled by the knob marked *Time/div* which stands for “Time per Division.” For example, when this is set to 1 second, it means that the dot will move from left to right, taking 10 seconds to move across the screen (because there are 10 divisions and the dot is traveling at a rate of 1 division per second). As the knob is turned to smaller and smaller divisions of a second, the dot moves faster and faster until it is moving so fast that it appears that there is a line on the screen rather than a single dot.

You'll notice on the drawing that there are actually two knobs – one

gray and one red. The gray one is the usual one. The red one is rarely used so it's normally in the off position. If it's turned on, it uncalibrates the larger knob, meaning that if the time-per-division number can no longer be trusted. This is why, when the red knob is turned on, a red light comes on on the front panel next to the word "Uncalibrate," warning the user that any time readings will be incorrect.

Channel 1

Most oscilloscopes that you'll see have two input channels which can be used to make completely independent measurements of two different signals simultaneously. We'll just look at one of the channel's controls, since the second one has identical parameters.

To begin with, the input is a BNC connector, shown on the lower left of the panel. This is normally connected to either of two things. The first is a probe, which has an alligator clip for the ground connection (almost always black) and a pointy part for the probe that you'll use to make the measurement. The second is somewhat simpler – it's just two alligator clips, one black one for the ground and one red one for the probe.

The large knob on this panel is marked "V/div" which stands for "Volts per Division." This is essentially a vertical magnification control, telling you how many divisions the green dot on the screen will move vertically for a given voltage. For example, when this knob is set to 100 mV per division and the dot moves upwards by 2 divisions, then the incoming voltage is 200 mV. A downwards movement of the dot indicates a negative voltage.

This knob can also be uncalibrated using the smaller red knob within it which is normally turned off. Again, there is a warning light to tell you not to trust the display when this knob is turned on.

Moving to the lower right side of the panel, you'll see a toggle switch that can be used to select three different possibilities – AC, DC and Ground. These don't mean what your intuition would think, so we'll go through them. When the toggle is set to Ground, the dot displays the vertical location of 0 V. Since this vertical position is adjustable using the knob just above the toggle switch, it can be set to any location on the screen. Normally, we would set the toggle to Ground which results in a flat horizontal line on the display, then we turn the *Vertical position* so that the line is perfectly aligned with the centre line on the screen. This means that you have the same positive and negative voltage ranges. If you're looking at a DC voltage, or an AC voltage with a DC component, then you may want to move the Ground position to a more appropriate vertical location.

When the toggle is set to DC, the dot displays the actual instantaneous voltage. This is true whether the signal is a DC signal or an AC signal. Note that this does not mean that you're looking at just the DC component of an AC signal. If you have a $1 V_{RMS}$ sine wave with a -1 V DC offset coming into the oscilloscope and if it's set to DC mode, then what you see is what you get – both the AC and the DC components of the signal.

When the toggle is set to AC, the dot displays the AC component of the waveform, omitting the DC offset component. So, if you try to measure the voltage of a battery using an oscilloscope in AC mode, then you'll see a value of 0 V. However, if you're measuring the ripple of a power supply rail without seeing the DC component, this mode is very useful. For example, let's say that you've got an 18 V rail with a 10 mV of ripple. If the oscilloscope is in DC mode and you set the V/div knob to have a high magnification so that you can accurately measure the 10 mV AC, then you can't see the signal because it's somewhere up in the ceiling. If you decrease the magnification so that you can see the signal, it looks like a perfectly flat line because 10 mV AC is so much smaller than 18 V DC. So, you set the mode to AC – this removes the 18 V DC component and lets you look at only the 10 mV AC ripple component at a high magnification.

Trigger

Let's say that you've got a 1 kHz sine wave that you're measuring with the 'scope. The dot on the screen just travels from left to right at a speed determined by the user. When it gets to the right side of the screen, it just starts on the left side again. Chances are that this rotation doesn't correspond in any way to the period of the signal. Consequently, the sine wave appears to move on the screen because every time the trace starts, it starts at a new point in the wave. In order to prevent ourselves from getting seasick or at least a headache, we have to tell the oscilloscope to start the trace at the same point in the waveform every time. This is controlled by the *trigger*. If we set the trigger to a level of 0 V, then the trace will wait until the signal crosses the level of 0 V before moving across the screen. This works great, but we need one more control – the *slope*. If we just set the trigger level to 0 V and hoped for the best, then sometimes the sine wave would start off on the left side at 0 V heading in the positive direction, sometimes at 0 V in the negative direction. This would result in a shape made of two sine waves, each a vertical mirror image of the other. Therefore, we set the trigger to a particular slope – for example, 0 V heading positive.

We can also select which signal is sent to the triggering circuit. This is

done using the toggle switch at the bottom of the panel. The user can select either the signal coming into Channel 1, Channel 2 or a separate External input connected to the BNC connector right on the trigger panel.

View

Finally, there is a panel that contains the controls to determine what you're actually looking at on the screen. There is a knob that permits you to select between looking at either Channel 1, Channel 2 or both simultaneously. There are two options for seeing both signals – *Chop* and *Alternate*. There is a slight difference between these two.

In actual fact, there is only one dot tracing across the screen on the 'scope, so it cannot display both signals simultaneously. It can, however, do a pretty good job of fooling your eyes so that it looks like you're seeing both signals at once. When the 'scope is in Alternate mode, the dot traces across the screen showing Channel 1, then it goes back and traces Channel 2, alternating back and forth. If the horizontal speed of the trace is fast enough, then two lines appear on the screen. This works really well for fast traces used for high frequency signals, but if the trace is moving very slowly, then it takes too long to wait for the other channel to re-appear.

Consequently, when you're using slower trace speeds for lower frequencies, you put the scope in Chop mode. In this case, the dot moves across the screen, hopping back and forth between the Channel 1 and 2 inputs, essentially showing half of each. This doesn't work well for high trace speeds because the display gets too dim because you're only showing half of the trace at any given time.

In addition, the View panel also has a couple of other controls. One is a knob which allows you to change the *Horizontal position* of the traces. In fact, if you turn this knob enough, you should be able to move the trace right off the screen.

The other control is a magnification switch which is rarely used. This is usually a *x 10 Magnification* which multiplies the speed of the trace by a factor of 10. Typically, this magnified trace will be shown in addition to the regular trace.

A couple of extra things...

There are a couple of things that I've left out here. The first is an option on the Time control knob called *X-Y mode*. This changes the function of the oscilloscope somewhat, turning off the time control of the dot. When

this mode is selected, the vertical movement of the dot is determined by the Channel 1 input in a normal fashion. The difference is that the horizontal movement of the dot is determined by the Channel 2 input, positive voltage moving the dot to the right, and negative to the left. This permits you to measure the phase relationship of two signals using what is known as a *Lissajous pattern*. This is discussed elsewhere in this book.

The last thing to talk about with oscilloscopes is a minor issue related to grounding. Although the signal input to the 'scope has an infinite impedance to ground, its ground input is directly connected to the third pin in the AC connector for the device. This means that, if you're using it to measure typical pro audio gear, the ground on the 'scope is likely already connected to the ground on the gear through your AC power connections (unless your gear has a floating ground). Consequently, you can't just put the ground connection anywhere in your circuit as you can with a free-floating DMM.

6.1.4 Function Generators

More often than not, you'll be trying to measure a piece of audio equipment that doesn't have an output without a signal being sent to its input. For example, a mixing console or an equalizer doesn't (or at least shouldn't) have an output when there's no input. Consequently, you need something to simulate a signal that would normally be fed into the input portion of the *Device Under Test* (also known as a *DUT*). This device is called a *Signal Generator* or a *Function Generator*.

A function generator produces an analog voltage output with a waveform, amplitude and frequency determined by the user. In addition, there are also probably controls for adding a DC voltage component to the signal as well as a VCF input for controlling the frequency using an external voltage.

The great thing about a good function generator is that you can set it to a chosen output amplitude and then sweep its frequency without any effect on the output level. You'll find that this is a very useful characteristic.

Figure 6.3 shows a simple function generator with a dial-type control. Newer models will have a digital readout which allows the user to precisely set the frequency.

Frequency and Range

On the far left of the panel is a dial that is used to control the frequency of the output. This knob usually has a limited range on its own, so it's used in

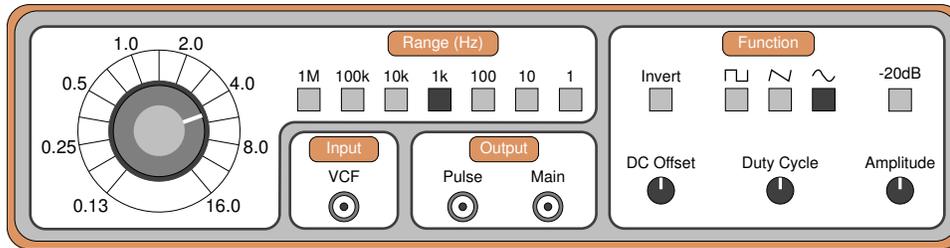


Figure 6.3: A simple function generator.

conjunction with a number of selector switches to determine the range. For example, in this case, if you wanted the output at 12 kHz, then you'd set the dial to 1.2 and press the range button marked 10 k. This would mean $1.2 * 10 \text{ k} = 12 \text{ kHz}$.

Function

On the far right of the panel is a section marked “Function” which contains the controls for various components of the waveform. In the top middle are three buttons for selecting between a square, triangle or sine wave. To the left of this is a button which allows the user to invert the polarity of the signal. On the far right is a button which provides a -20 dB attenuation. This button is used in conjunction with the Amplitude control knob just below it. On the bottom left is the knob which controls the DC Offset. Typically, this is turned off by pushing the knob in and turned on by pulling away from the panel.

The final control here is for the *Duty Cycle* of the waveform. This is a control which determined the relative time spent in the positive and negative portions of the waveform. For example, a square wave has a duty cycle of 50% because it is high for 50% of the cycle and low for the remaining 50%. If this relationship is changed, the wave approaches a pulse wave. Similarly, a triangle wave has a 50% duty cycle. If this is reduced to 0% then you have a sawtooth wave.

Output

The average function generator will have two outputs. One, which is usually marked “Main” is the normal output of the device. The second is a pulse wave with the same frequency as the main output. This can be used for

externally triggering an oscilloscope rather than triggering from the signal itself – a technique which is useful when measuring delay times, for example.

Input

Finally, some function generators have a VCF input which permits the user to control the frequency of the generator with an external voltage source. This may be useful in doing swept sine tone measurements, where the frequency of the sine wave is swept from a very low to very high frequency to see the entire range of the DUT.

6.1.5 Automated Testing Equipment

Unfortunately, measuring a DUT using these devices is typically a tedious affair. For example, if you wanted to do a frequency response measurement of an equalizer, you'd have to connect a function generator to its input and its output to an oscilloscope. Then you'd have to go through the frequency range of the device, sweeping from a low to a high frequency, checking for changes in the output level on the oscilloscope.

Fortunately, there are a number of various automated testing devices which do all the hard work for you. These are usually computer-controlled systems that use different techniques to arrive at a reliable measurement of the device. Two of these are the *Swept Sine Tone* and the *Maximum Length Sequence* techniques.

6.1.6 Swept Sine Tone

NOT YET WRITTEN

6.1.7 MLS-Based Measurements

Back in Section 3.3.2 we looked at a mathematical concept called a *maximum length sequence* or *MLS*. An MLS is a series of values (1's and -1's) that have a very particular order. We're not going to look at how these are created – too complicated. We'll just look at some of their properties. We've already seen that they can be used to make diffusion panels for your studio walls. It turns out that they can also be used to make electroacoustic measurements.

Let's take a small MLS of 15 values (in other words, $N = 15$). This sequence is shown below.

+ + + - - - + - - + - +

If we were to graph this sequence, it would look like Figure 6.4.

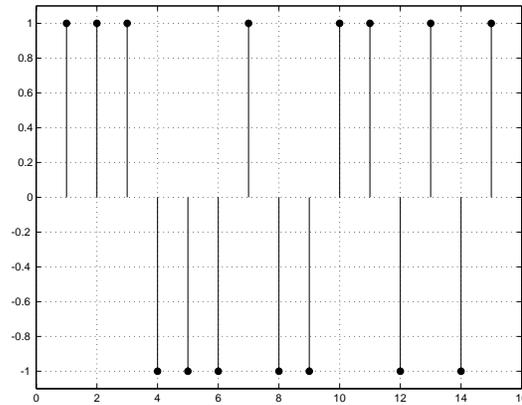


Figure 6.4: A graphic representation of the MLS shown above ($N=15$)

Later on in the section on DSP we're going to learn about a special mathematical process called a *cross-correlation function*. For the purposes of this discussion, we're not going to worry about what that means – as far as we're concerned for now, it's math that's done to a bunch of numbers that result in another bunch of numbers. Basically, a cross-correlation function tells you how much two signals are alike if you compare them to each other at different times.

There's a wonderful property of an MLS that makes it so special. If you take an MLS and do a cross-correlation function on it with itself, magic occurs. Now, at first it might seem strange to do math to figure out how much a signal is like itself – of course it's identical to itself... but remember that the cross-correlation function also checks to see how alike the signals are if they are not time-aligned.

Let's begin by using the MLS with $N = 15$ shown in Figure 6.4 as a digital audio signal. If we do a cross-correlation function on this with itself, we get the result shown in Figure 6.5.

What does this tell us? Reading from left to right on the graph, we can see that if the MLS is compared to itself with a time offset of 14 samples, then they have a correlation of about 0.1 which means that they're not very similar (a correlation of 0 means that they are completely different). As we make the time offset smaller and smaller, the correlation stays very low (less than 0.2) until we get to an offset of 0. At this point, the two signals are identical, so they have a correlation of 1.

If we used a longer MLS sequence, the correlation values at time offsets

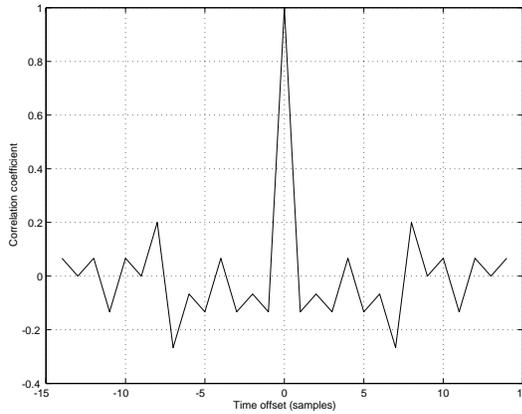


Figure 6.5: The cross correlation function of the MLS sequence shown in Figure 6.4 with itself.

other than 0 would become very small. For example, Figure 6.6 shows the cross-correlation function of an 8191-point MLS.

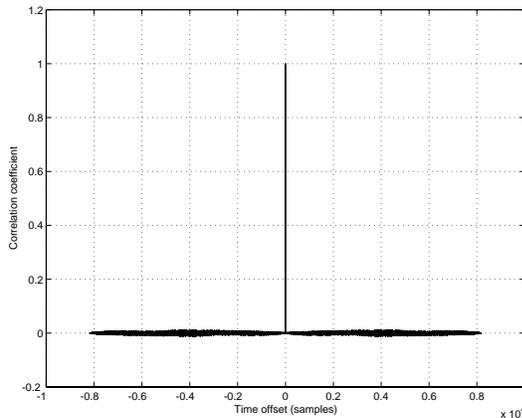


Figure 6.6: The cross correlation function of an 8191-point MLS sequence with itself.

If we idealize this graph, we could say that, at a time offset of 0 we have a correlation coefficient of 1, and at all other time offsets, we have a correlation coefficient of 0. If we think of this as a time signal instead of a cross-correlation function, then it's an impulse. This has some pretty significant implications on our lives.

Let's say that you wanted to make an impulse response measurement of

a room. You'd have to make an impulse, record it and look at the recording. So, you snap your fingers and magically, you can see direct sound, reflections and reverberation in your recording. Unfortunately, this won't work. You see, in the real world, there's noise, and if you snap your fingers, you won't be able to distinguish the finger snap from the noise because your impulse isn't strong enough – your signal-to-noise ratio is far too small. “No problem!” I hear you cry. “We make a bigger impulse. I'll get my gun.”

In fact, this would work pretty well. For many years, acousticians carried around pistols and buckets of sand to fire bullets into in order to make impulse response measurements. There are some drawbacks to this procedure, however. You have to own a gun, you need a microphone that can handle a VERY big signal, you need to buy ear protection... All sorts of reasons that work against this option.

There is a better way. Play an MLS out of a loudspeaker. (It will sound a bit like white noise – in fact the signal is sometimes called a *pseudo-random signal* because it sounds random but it isn't.) Record the signal in the room with a microphone. Then go home.

When you get home, do a cross-correlation function on the MLS signal that you sent to the loudspeaker with the signal that came into the microphone. When you do, the resulting output will be an impulse response of the room with a very good signal-to-noise ratio.

Let's look at a brief example. I made a 1023-point MLS. Then I multiplied by 0.5 delayed it by 500 samples and added the result back to the original. So, we have two overlapping MLS's with a time offset of 500 samples, and the second one is lowered in level by 6 dB. I then did a cross correlation function of the two signals using the *xcorr* function in MATLAB. The result is shown in Figure 6.7. As you can see here, the result is a noisy representation of the impulse response of the resulting signal. If I wanted to improve my signal-to-noise ratio, all I have to do is to increase the length of the MLS.

All of this was a lot of work that you don't have to do because you can buy measurement systems that do all the math for you. The most popular one is the MLSSA system (pronounced “Melissa”) from DRA Laboratories (www.mlssa.com), but other companies make them. For example, the AudioPrecision system has it as an option on their devices (www.ap.com). Of course, if you have MATLAB and the option to playback and record digital audio from your computer, you could just do it yourself.

6.1.8 Suggested Reading List

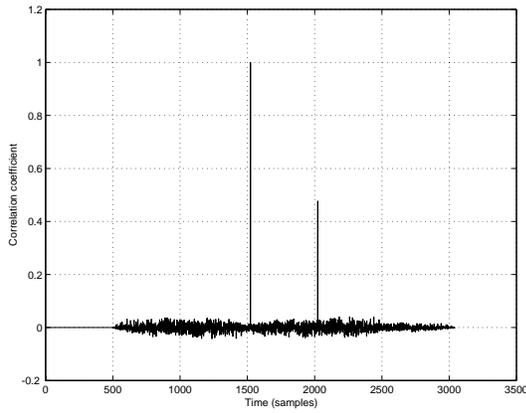


Figure 6.7: The cross correlation function of an 1023-point MLS sequence with a modified version of itself (see text for details).

6.2 Electrical Measurements

6.2.1 Output impedance

The output impedance of a device under test (DUT) is measured through the comparison of a given signal being output by the device with and without a load of a known impedance. For example, if we send a 1 kHz sine tone through a DUT and measure its output to be 0 dBV (1 V_{RMS}) with an oscilloscope and no other load, then we put a 600 Ω resistor across the output and measure its new output level (we've changed nothing but the load on the output) to be -6.02 dBV (0.5 V_{RMS}), then we know that the DUT has a 600 Ω output impedance at 1 kHz. Effectively, we have built a voltage divider using the output of the DUT and a 600 Ω resistor. The less the output drops when we add the resistor, the lower the output impedance (expect values around 50 Ω nowadays...).

The equation for calculating this could be something like

$$Z_{OUT} = 600 * \frac{V_1}{V_2} - 1 \quad (6.1)$$

Where Z_{OUT} is the output impedance of the DUT (in Ω)

V_1 is the level of the unloaded output signal of the DUT (in volts)

and

V_2 is the level of the output signal of the DUT loaded with a 600 Ω resistor (in volts)

6.2.2 Input Impedance

The input impedance of a device is measured in a similar fashion to the output impedance, except that we are now using the DUT as the load on a function generator with a known impedance, usually 600 Ω . For example, if we produce a 1 kHz, 0 dBV sine tone with a function generator with a 600 Ω output impedance and the output of the generator drops by 6.02 dB, then we know that the input impedance of the device is 600 Ω .

The equation for calculating this could be something like

$$Z_{IN} = \frac{600}{\frac{V_1}{V_2} - 1} \quad (6.2)$$

Where Z_{IN} is the input impedance of the DUT (in Ω)

V_1 is the level of the unloaded output signal of the function generator (with a 600 Ω output impedance) (in volts)

and

V_2 is the level of the output signal of the same function generator loaded with the input of the DUT (in volts)

6.2.3 Gain

The gain of a DUT is measured by comparing its input to its output. This gain can either be expressed as a multiplier or in decibels.

For example : if the input of a DUT is a 1 kHz sine tone with an amplitude of $100mV_{RMS}$ and its output is a 1 kHz sine tone with an amplitude of $1V_{RMS}$, then the gain is

$$\frac{1V_{RMS}}{100mV_{RMS}} \quad (6.3)$$

$$\frac{1}{0.1} \quad (6.4)$$

$$Gain = 10 \quad (6.5)$$

This could also be expressed in decibels as follows

$$20 \log_{10} \left(\frac{1V_{RMS}}{100mV_{RMS}} \right) \quad (6.6)$$

$$20 \log_{10}(10) \quad (6.7)$$

$$Gain = 20dB \quad (6.8)$$

6.2.4 Gain Linearity

In theory, a DUT should have the same gain on a signal irrespective of the level of the input (assuming that we're not measuring a DUT that is supposed to make dynamic gain changes, like a compressor, limiter, expander or gate, for example...). Therefore, the gain, once set to a value of 2 using a 0 dBV 1kHz sine tone should be 2 for all other amplitudes of a 1 kHz sine tone (it may change at other frequencies, as in the case of a filter, for example, but we'll come to that...) If this is indeed the case, then the DUT is said to have a linear gain response. This will likely not be the case – so we measure the gain of the DUT using various amplitudes at its input. The resulting measurement is most easily graphed on an X-Y plot, with the

input amplitude on the x-axis and the deviation from the nominal gain on the y-axis.

This measurement is useful for showing the distortion of low-level signals caused by digital converters in a DUT, for example.

6.2.5 Frequency Response

A frequency response measurement is much like a gain linearity measurement with the exception that we are now sweeping through various frequencies, keeping the amplitude of the input signal constant rather than vice versa. This is probably the most common and most commonly abused measurement for audio equipment (for example, when a manufacturer states that a device has a frequency response from 20 Hz to 20 kHz... this means nothing without some indication of a deviation in gain)

We make this measurement by calculating the gain of a DUT at a number of frequencies and plotting them on an X-Y graph, showing frequency (usually plotted logarithmically) on the X-axis and the gain on the Y-axis. The more frequencies we use for the measurement, the more accurate the graph. If we are using a computer based measurement device such as the Audio-Precision or the MLSSA, keep in mind that the frequencies plotted are the result of an FFT-analysis of a time measurement, therefore the frequencies displayed are equally spaced linearly.

6.2.6 Bandwidth

All DUT's will exhibit a roll off in their frequency response at low and high frequencies. The bandwidth of the DUT is the width in Hz between the low and high points where we first reach a gain of -3 dB relative to the nominal gain measured at a mid frequency (usually 1 kHz).

6.2.7 Phase Response

The phase response of a DUT is similar to the frequency response, with the exception that we are now looking for a change in phase of a signal at various frequencies rather than a change in gain. It is measured by again, comparing the input of a DUT to its output at various frequencies. There are frequently cases when the *group delay* of the DUT must also be considered in making a phase response measurement. This will be discussed below.

Phase response measurements are usually graphed similarly to frequency response measurements, on a semi-log graph, showing frequency on the X-axis and phase on the Y-axis. The phase will usually be plotted in degrees,

either *wrapped* or *unwrapped*. If the phase is wrapped, then the graph is limited from -180° to 180° . any phase measured to be outside of that window is then wrapped around to the other side of the window – for example, -190° is plotted as 170° , since 190° behind and 170° ahead will have the same result. If the plot is unwrapped, the phase will be plotted as measured.

6.2.8 Group Delay

NOT YET WRITTEN

6.2.9 Slew Rate

The slew rate of a device, whether it's a big unit like a mixing console or a little unit like an op amp, is the rate at which it can change from one voltage level to another voltage level. In theory (and in a perfect world), we'd probably like all of our devices to be able to change from any voltage to any other voltage instantly. This, however, is not the case for a number of reasons.

In order to measure the slew rate of a DUT, connect a function generator generating a pulse wave to its input. The greater the amplitude of the signal (assuming that we're not clipping anything) the easier to make this measurement. Look at the transition between the two voltage levels in the pulse wave both at the input and output of the DUT. Set the scope to a trace time fast enough that the transition no longer looks vertical – this will likely be somewhere in the microsecond range.

If the output of the function generator has instantaneous transitions between voltages, then, looking at the slope of the transition of the output signal, measure the change in voltage in 1 microsecond. This is your slew rate – usually expressed in Volts per microsecond.

6.2.10 Maximum Output

Almost every device has some maximum positive voltage (and minimum negative voltage) level which cannot be exceeded. This level is determined by a number of factors including (but not limited to) the voltage levels of the power supply rails, the operational amplifiers being used, and the gain structure of the device. The principal question in discussing the term “maximum output” is how we define “maximum” – we can say that it's the absolute maximum voltage which the device is able to output, although this is not usually the case. More often, the maximum output is defined to be

the level at which some pre-determined value of distortion of the signal is incurred.

If we send a sine wave into our DUT and start turning things up (either the gain of the DUT or the level of the input signal) while monitoring the output, eventually we'll reach a level where the DUT is unable to produce a high enough voltage to accurately reproduce the sine waveform (make sure that the fault lies in the DUT and not the function generator). This appears visually as a flattening of the peaks and troughs in the sine wave – called *clipping* because the wave is “clipped.” If we continue to increase the level, the flat sections will occupy more and more of the period of the signal, approaching a square wave. The more we approach the square wave, the more we are distorting the original shape of the wave. Obviously, the now-flat section of the wave is the maximum level. This is generally not the method of making this measurement.

If we go back to a level just before the wave started clipping and measure the amount of distortion (we'll talk about how to do this a little later) we'll find that the distortion of the wave occurs before we can see it on an oscilloscope display. In fact, at the level at which start to see the clipping, you've already reached approximately 10% distortion. Usually, the maximum output level is the point when we have either 0.1% or 1% distortion, although this will change from manufacturer to manufacturer. Usually, when a maximum output voltage is given, the percentage of distortion will be included to tell you what the company in question means by the word “maximum.”

Remember that the maximum output of the device is independent of its gain (because we won't know the level of the input signal). It's simply a maximum level of the output signal.

6.2.11 Noise measurements

Every device produces noise. Even resistors cause some noise (caused by the molecules moving around due to heat – hence the term *thermal noise*) If there's no input (see next paragraph) to a DUT and we measure its output using a true RMS meter, we'll see a single number expressing its total output noise. This value, however, isn't all that useful to us, since much of the noise the meter is measuring is outside of the audio range. Consequently, it's common (and recommended) practice to limit the bandwidth of the noise measurement. Unfortunately, there is no standard bandwidth for this, so you have to rely on a manufacturer to tell you what the bandwidth was. If you're doing the measurement, pick your frequency band, but tell be sure to include that information when you're specifying the noise level of the device

How do ensure that there's no signal coming into a device? This is not as simple as simply disconnecting the previous device. If we just leave the input of our DUT *unterminated* (a fancy word for “not connected to anything”) then we could be incurring some noise in the device that we normally wouldn't get if we were plugged into another unit. (Case in point: listen to the hum of a guitar amp when you plug in an input cable that's not connected to an electric guitar.) Most people will look after this problem by grounding the input – that is to say they just connect the input wire(s) to the ground wire. This is done most frequently, largely due to convenience. If you want to get a more accurate measurement, then you'll have to terminate the input with an impedance that matches the output impedance of the device that's normally connected to it.

Remember, when specifying noise measurements, include the bandwidth and the input termination impedance.

6.2.12 Dynamic Range

The dynamic range of a device is it the difference in level (usually expressed in dB) between its maximum output level and its noise floor (below which, your audio signal is effectively unusable). This gives you an idea of the absolute limits of the dynamics of the signal you'll be sending through the DUT. Since a DNR measurement relays in part on the noise floor measurement, you'll have to specify the bandwidth and input termination impedance.

Also, in spite of everything you see, don't confuse the dynamic range of a device with its signal to noise ratio.

6.2.13 Signal to Noise Ratio (SNR or S/N Ratio)

The signal to noise ratio is similar to the dynamic range measurement in that it expresses a difference in dB between the noise floor of a DUT and another level. Unlike the dynamic range measurement, however, the signal to noise expresses the difference between the nominal recommended operating level and the noise floor. Usually, this operating level will be shown as 0 dB VU on the meters of the device in question and will correspond at the output to either +4 dBu or -10 dBV, depending on the device.

6.2.14 Equivalent Input Noise (EIN)

There is one special case for noise measurements which occurs when you're trying to measure the input noise for a microphone preamplifier. In this case, the input noise levels will be so low that they're difficult to measure

accurately (if the input noise levels aren't too low to measure accurately, then you need a new mic pre...). Luckily, however, we have lots of gain available to us in the preamp itself which we can use to boost the noise in order to measure it. The general idea is that we set the preamp to some known amount of gain, measure the output noise, and then subtract the gain from our noise measurement to calculate what the level is – this is the reason for the “equivalent” in the title, because it's not actually measured – it's calculated.

There are a couple of things to worry about when making an EIN measurement. The first thing is the gain of the microphone preamplifier. Using a function generator with a very small output level, set the gain of the preamp to +60 dB. This is done by simultaneously measuring the input and the output on an oscilloscope and changing the gain until the output is 1000 times greater in amplitude than the input. Then you disconnect the function generator and the oscilloscope (but don't change the gain of the preamp!).

Since it's a noise measurement, we're concerned with the input impedance. In this case, it's common practice to terminate the input of the microphone preamp with a 150 Ω resistor to match the output impedance of a typical microphone. The easiest way to do this is to solder a 150 Ω resistor between pins 2 and 3 of a spare male XLR connector you have lying around the house. Then you can just plug that directly into the mic preamp input with a minimum of hassle.

With the 150 Ω termination connected, and the gain set to +60 dB, measure the level of the noise at the output of the mic preamp using a true RMS meter, with a band-limited input. Take the measurement of the output noise (expressed in dBu) and subtract 60 dB (for the gain of the preamp). That's the EIN – usually somewhere around -120 dBu.

6.2.15 Common Mode Rejection Ratio (CMRR)

Any piece of “professional” audio gear has what we commonly term “balanced” inputs (be careful – you may not necessarily know what a balanced input is – if you can define it without using the word “impedance” you'll be wrong). One of the components that makes a balanced input balanced is what is called a differential input. This means that there are two input legs for a single audio signal, one labelled “positive” or “non-inverting” and the other “negative” or “inverting.” If the differential input is working properly, it will subtract the voltage at the inverting input from the voltage at the non-inverting input and send the resulting difference into the DUT. This

works well for audio signals because we usually send a balanced input a balanced signal which means that we're sending the audio signal on two wires, with the voltages on each of these wires being the negative of the other – so when one wire is at +1 V, the other is at -1 V. The +1 V is sent to the non-inverting input, and the -1 V is sent to the inverting input and the differential input does the math $+1 - -1 = 2$ and gets the signal out with a gain of 6.02 dB.

The neat thing is that, since any noise picked up on the balanced transmission cable before the balanced input will theoretically be the same on both wires, the differential input subtracts the noise from itself and therefore eliminates it. For more info on this (and the correct definition of “balanced”) see the chapter on grounding and shielding in the electronics section

Since the noise on the cable is *common* to both inputs (the inverting and the non-inverting) we call it a *common mode* and since this common mode cancels itself, we say that it has been *rejected*. Therefore, the ability for a differential input to eliminate a signal sent to both input legs is called the common mode rejection ratio. It's expressed in dB, and here's how you measure it.

Take a function generator emitting a sine wave. Send the output of the generator to both input legs of the balanced input of the DUT (pins 2 and 3 on an XLR). Measure the input level and the output level of the DUT. Hopefully, the level of the output will be considerably lower (at least 60 dB lower) than the level of the input. The difference between them, in dB is the CMRR.

This method is the common way to find the CMRR but there is a slight problem with it. Since a balanced input is designed to work best when connected to a balanced output (meaning that the impedances of the two legs to ground are matched – not necessarily that the voltages are symmetrical) the CMRR of our DUT's input may change if we connected it to unmatched impedances. Two people I have talked to regarding this subject have both commented on the necessity to implement a method of testing the CMRR of a device under unmatched impedances. This may become a recommended standard...

6.2.16 Total Harmonic Distortion (THD)

If you go out to the waveform store and throw out a lot of money on the perfect sine wave, take it home and analyze it, you'll find that it has very few interesting characteristics beyond its simplicity. The thing that makes it useful to us is that it is an irreducible component – the quarks (NOTE:

are quarks irreducible?) of the audio world. What this means is that a sine wave has no energy in any of its harmonics above the fundamental. It is the fundamental and that's that.

If you then take your perfect sine wave and send it to the input of a DUT and look at the output, you'll find that the shape of your wave has been modified slightly. Just enough to put some energy in the harmonics above the fundamental. The distortion of the shape of the wave results in energy in the upper harmonics. Therefore, if we measure the total amount of energy in the harmonics relative to the level of the fundamental, we have a quantifiable method of determining how much the signal has been distorted. The greater the amplitude of the harmonics, the greater the distortion.

This measurement is not a common one, but its counterpart, explained below is seen everywhere. It's made by sending a perfect sine wave into the input of the DUT, and looking at an FFT of the output (which shows the amplitudes of individual frequency "bins"). You add the squares of the amplitudes of the harmonics other than the fundamental, and take the square root of this sum and divide it by the amplitude of the fundamental at the output. If you multiply this value by 100, you get the THD in percent. If you take the log of the number and multiply by 20, you'll get it in dB.

A couple of things about THD measurements – usually the person doing the measurement will only measure the amplitudes of the harmonics up to a given harmonic, so you'll see specifications like "THD through to the 7th harmonic" meaning that there is more energy above that, but it's probably too low or close to the noise floor to measure.

Secondly, unless you have a smart computer which is able to do this measurement, you probably will notice that's relative time-consuming and not really worth the effort. This is especially true when you compare the procedure to getting a THD+N measurement as is explained below. That's why you rarely see it.

6.2.17 Total Harmonic Distortion + Noise (THD+N)

The distortion of a signal isn't the only reason why a DUT will make things sound worse than when they came in. The noise (meaning wide-band noise as well as hums and buzzes and anything else that you don't expect) incurred by the signal going through the device is also a contributing factor to signal degradation. As a result, manufacturers use a measurement to indicate the extra energy coming out of the DUT other than the original signal. This is measured using a similar method to the THD measurement, however, it takes considerably less time.

Since we're looking for all of the energy other than the original, we'll send a perfect sine wave into the input of the DUT. We then acquire a notch-filter tuned to the frequency of the input sine, which theoretically eliminates the original inputted signal, leaving all other signal (distortion produces and noise which is produced by the DUT itself) to get through. This filtered output is measured using a true RMS meter (we usually band-limit the signal as well, because it's a noise measurement – see above). This measurement is divided by the RMS level of the fundamental which gives a ratio which can either be converted to a percentage or a dB equivalent as explained above.

6.2.18 Crosstalk

6.2.19 Intermodulation Distortion (IMD)

NOT YET WRITTEN

6.2.20 DC Offset

NOT YET WRITTEN

6.2.21 Filter Q

See section [5.1.2](#).

6.2.22 Impulse Response

NOT YET WRITTEN

6.2.23 Cross-correlation

NOT YET WRITTEN

6.2.24 Interaural Cross-correlation

NOT YET WRITTEN

6.2.25 Suggested Reading List

6.3 Loudspeaker Measurements

Measuring the characteristics of a loudspeaker is not an simple task. There are two basic problems that cause us to be unsure that we are getting a clear picture of what the loudspeaker is actually doing.

The first of these is the microphone that you're using to make the measurement. The output of the loudspeaker is converted to a measurable electrical signal by a microphone which cannot be a perfect transducer. So, for example, when you make a frequency response measurement, you are not only seeing the frequency response of the loudspeaker, but the response of the microphone as well. The answer to this problem is to spend a lot of money on a microphone that includes extensive documentation of its particular characteristics. In many cases today, this documentation is included in the form of measurement data that can be incorporated into your measurement system and subtracted from your measurements to make the microphone effectively "invisible."

The second problem is that of the room that you're using for the measurements. Always remember that a loudspeaker is, more often than not, in a room. That room has two effects. Firstly, different rooms make different loudspeakers behave differently. (Take the extreme case of a large loudspeaker in a small room such as a closet or a car. In this case, the "room" is more like an extra loudspeaker enclosure than a room. This will cause the driver units to behave differently than in a large room because the loudspeaker "sees" the load imposed on the driver by the space itself.) Secondly, the room itself imposes resonance, reflections and reverberation that will be recorded at the microphone position. These effects may be indistinguishable from the loudspeaker (for example, a ringing room mode compared to a ringing loudspeaker driver, or a ringing EQ stage in front of the power amplifier).

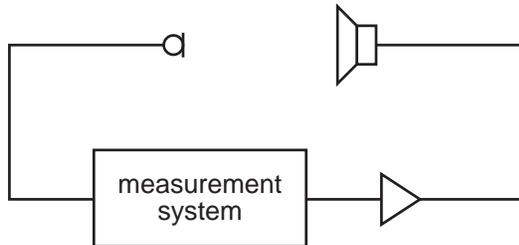


Figure 6.8: A simple block diagram showing the basic setup used for typical loudspeaker measurements. The measurement device may be anything from a signal generator and oscilloscope to a sophisticated self-contained automated computer system.

6.3.1 Frequency Response

6.3.2 Phase Response

6.3.3 Distortion

6.3.4 Off-Axis Response

6.3.5 Power Response

6.3.6 Impedance Curve

6.3.7 Sensitivity

6.3.8 Suggested Reading List

6.4 Microphone Measurements

NOT YET WRITTEN

6.4.1 Polar Pattern

6.4.2 Frequency Response

6.4.3 Phase Response

6.4.4 Off-Axis Response

6.4.5 Sensitivity

6.4.6 Equivalent Noise Level

6.4.7 Impedance

6.4.8 Suggested Reading List

6.5 Acoustical Measurements

NOT YET WRITTEN

6.5.1 Room Modes

6.5.2 Reflections

6.5.3 Reflection and Absorption Coefficients

6.5.4 Reverberation Time (RT60)

6.5.5 Noise Level

6.5.6 Sound Transmission

6.5.7 Intelligibility

6.5.8 Suggested Reading List

Chapter 7

Digital Audio

7.1 How to make a (PCM) digital signal

7.1.1 The basics of analog to digital conversion

Once upon a time, the only way to store or transmit an audio signal was to use a change in voltage (or magnetism) that had a waveform that was analogous to the pressure waveform that was the sound itself. This *analog* signal (analog because the voltage wave is an analog of the pressure wave) worked well, but suffered from a number of issues, particularly the unwanted introduction of noise. Then someone came up with the great idea that those issues could be overcome if the analog wave was converted into a different way of representing it.

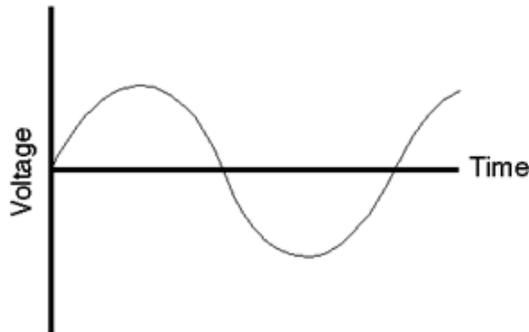


Figure 7.1: The analog waveform that we're going to convert to digital representation.

The first step in digitizing an analog waveform is to do basically the same thing film does to motion. When you sit and watch a movie in a cinema, it

appears that you are watching a moving picture. In fact, you are watching 24 still pictures every second – but your eyes are too slow in responding to the multiple photos and therefore you get fooled into thinking that smooth motion is happening. In technical jargon, we are changing an event that happens in *continuous time* into one that is chopped into slices of *discrete time*.

Unlike a film, where we just take successive photographs of the event to be replayed in succession later, audio uses a slightly different procedure. Here, we use a device to *sample* the voltage of the signal at regular intervals in time as is shown below in Figure 7.2.

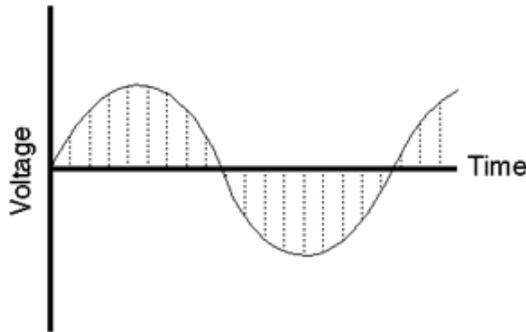


Figure 7.2: The audio waveform being sliced into moments in time. A sample of the signal is taken at each vertical dotted line.

Each sample is then temporarily stored and all the information regarding what happened to the signal between samples is thrown away. The system that performs this task is what is known as a *sample and hold* circuit because it samples the original waveform at a given moment, and holds that level until the next time the signal is sampled as can be seen in Figure 7.3.

Our eventual goal is to represent the original signal with a string of numbers representing measurements of each sample. Consequently, the next step in the process is to actually do the measurement of each sample. Unfortunately, the “ruler” that’s used to make this measurement isn’t infinitely precise – it’s like a measuring tape marked in millimeters. Although you can make a pretty good measurement with that tape, you can’t make an accurate measurement of something that’s 4.23839 mm long. The same problem exists with our measuring system. As can be seen in Figure 7.4, it is a very rare occurrence when the level of each sample from the sample and hold circuit lands exactly on one of the levels in the measuring system.

If we go back to the example of the ruler marked in millimeters being

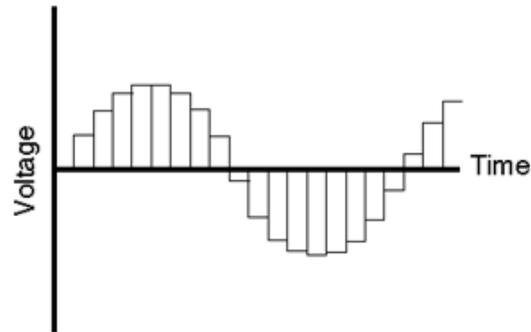


Figure 7.3: The output of the sample and hold circuit. Notice that, although we still have the basic shape of the original waveform, the smooth details have been lost.

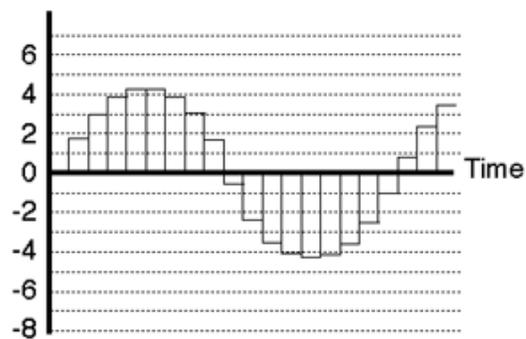


Figure 7.4: The output of the sample and hold circuit shown against the allowable levels plotted as horizontal dotted lines.

used to measure something 4.23839 mm long, the obvious response would be to round off the measurement to the nearest millimeter. That's really the best you could do... and you wouldn't worry too much because the worst error that you could make is about a half a millimeter. The same is true in our signal measuring circuit – it rounds the level of the sample to the nearest value it knows. This procedure of rounding the signal level is called *quantization* because it is changing the signal (which used to have infinite resolution) to match quanta, or discrete values. (Actually, a “quantum” according to my dictionary is “the minimum amount by which certain properties ... of a system that can change. Such properties do not, therefore, vary continuously, but in integral multiples of the relevant quantum.” (A Concise Dictionary of Physics – Oxford))

Of course, we have to keep in mind that we're creating error by just rounding off these values arbitrarily to the nearest value that fits in our system. That error is called *quantization error* and is perceivable in the output of the system as noise whose characteristics are dependent on the signal itself. This noise is commonly called *quantization noise* and we'll come back to that later.

In a perfect world, we wouldn't have to quantize the signal levels, but unfortunately, the world isn't perfect... The next best thing is to put as many possible gradations in the system so that we have to round off as little as possible. That way we minimize the quantization error and therefore reduce the quantization noise. We'll talk later about what this implies, but just to get a general idea to start, a CD has 65,536 possible levels that it can use when measuring the level of the signal (as compared to the system shown in Figure 7.5 where we only have 16 possible levels...)

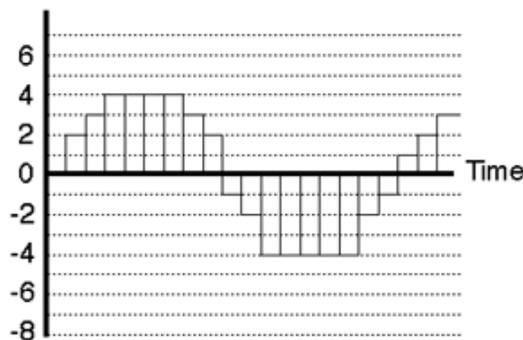


Figure 7.5: The output of the quantizing circuit. Notice that almost all the samples have been rounded off to the nearest dotted line.

At this point, we finally have our digital signal. Looking back at Figure 7.5 as an example, we can see that the values are

0 2 3 4 4 4 4 3 2 -1 -2 -4 -4 -4 -4 -4 -2 -1 1 2 3

These values are then stored in (or transmitted by) the system as a digital representation of the original analog signal.

7.1.2 The basics of digital to analog conversion

Now that we have all of those digits stored in our digital system, how do we turn them back into an audio signal? We start by doing the reverse of the sample and hold circuit. We feed the new circuit the string of numbers which it converts into voltages, resulting in a signal that looks just like the output of the quantization circuit (see Figure 7.5).

Now we need a small extra piece of knowledge. Compare the waveform in Figure 7.1 to the waveform in Figure 7.4. One of the biggest differences between them is that there are instantaneous changes in the *slope* of the wave – that is to say, the wave in Figure 7.4 has sharper corners in it, while the one in Figure 7.1 is nice and smooth. The presence of those sharp corners indicates that there are high frequency components in the signal. No high frequencies, no sharp corners.

Therefore, if we take the signal shown in Figure 7.5 and remove the high frequencies, we remove the sharp corners. This is done using a filter that blocks the high frequency information, but allows the low frequencies to pass. Generally speaking, the filter is called a *low pass filter* but in this specific use in digital audio it's called a *reconstruction filter* (although some people call it a *smoothing filter*) because it helps to reconstruct (or smooth) the audio signal from the ugly staircase representation as shown in Figure 7.6.

The result of the output of the reconstruction filter, shown by itself in Figure 7.7 is the output of the system. As you can see, the result is a continuous waveform (no sharp edges...). Also, you'll note that it's exactly the same as the analog waveform we sent into the system in the first place – well... not exactly... but keep in mind that we used an example with VERY bad quantization error. You'd hopefully never see anything this bad in the real world.

7.1.3 Aliasing

I remember when I was a kid, I'd watch the television show M*A*S*H every week, and every week, during the opening credits, they'd show a shot of the

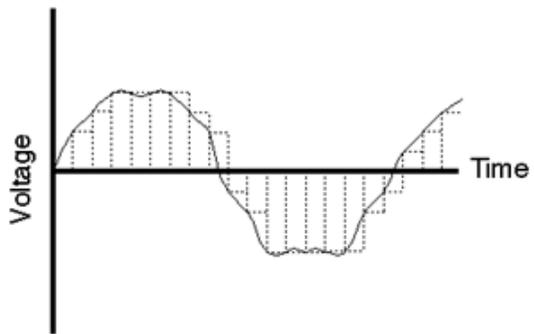


Figure 7.6: The results of the reconstruction filter showing the original staircase waveform from which it was derived as a dotted line.

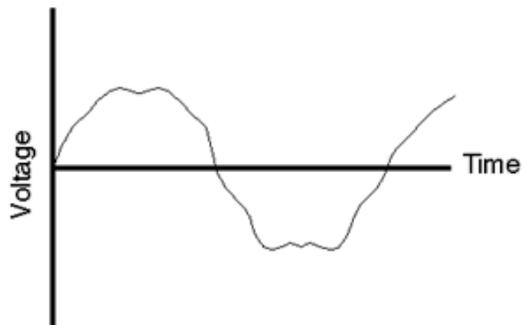


Figure 7.7: The output of the system.

jeep accelerating away out of the camp. Oddly, as the jeep got going faster and faster forwards, the wheels would appear to speed up, then stop, then start going backwards... What didn't make sense was that the jeep was still going forwards. What causes this phenomenon, and why don't you see it in real day-to-day life?

Let's look at this by considering a wheel with only one spoke as is shown in the top left of Figure 8. Each column of Figure 8 represents a different rotational speed for the wheel, each numbered row represents a frame of the movie. In the leftmost column, the wheel makes one sixth of a rotation per frame. As can be seen in the animation in Figure 9, this results in a wheel that appears to be rotating clockwise as expected. In the second column, the wheel is making one third of a rotation per frame and the resulting animation is a faster turning wheel, but still in the clockwise rotation. In the third column, the wheel is turning slightly faster, making one half of a rotation per frame. As can be seen in the corresponding animation, this results in the appearance of a 2-spoked wheel that is stopped.

If the wheel is turning faster than one rotation every two frames, an odd thing happens. The wheel, making more than one half of a rotation per frame, results in the appearance of the wheel turning backwards and more slowly than the actual rotation... This is a problem caused by the fact that we are slicing continuous time into discrete time, thus distorting the actual event. This result which appears to be something other than what happened is known as an *alias*.

The same problem exists in digital audio. If you take a look at the waveform in Figure 7.9, you can see that we have less than two samples per period of the wave. Therefore the frequency of the wave is greater than one half the sampling rate.

Figure 7.10 demonstrates that there is a second waveform with the same amplitude as the one in Figure 7.9 which could be represented by the same samples. As can be seen, this frequency is lower than the one that was recorded

The whole problem of aliasing causes two results. Firstly, we have to make sure that no frequencies above half of the sampling rate (typically called the *Nyquist frequency*) get into the sample and hold circuit. Secondly, we have to set the sampling rate high enough to be able to capture all the frequencies we want to hear. The second of these issues is a pretty easy one to solve: textbooks say that we can only hear frequencies up to about 20 kHz, therefore all we need to do is to make sure that our sampling rate is at least twice this value – therefore at least 40,000 samples per second.

The only problem left is to ensure that no frequencies above the Nyquist

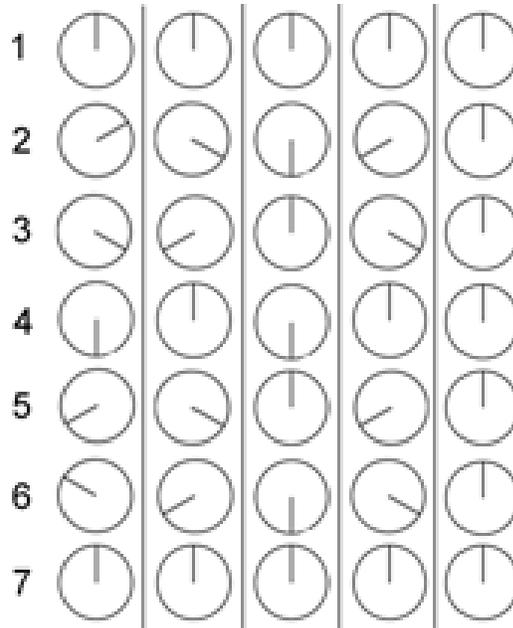


Figure 7.8: Frame-by-frame shots of a 1-spoked wheel turning at different speeds and captured by the same frame rate.

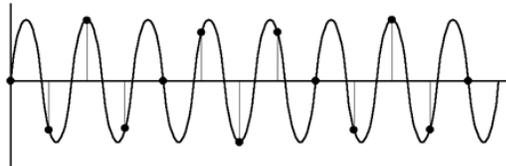


Figure 7.9: Waveform with a frequency that is greater than one-half the sampling rate.

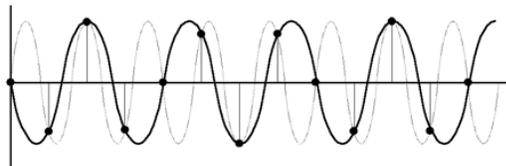


Figure 7.10: The resulting alias frequency caused by sampling the waveform as shown in Figure 7.9.

frequency get into the sample and hold circuit to begin with. This is a fairly easy task. Just before the sample and hold circuit, a low-pass filter is used to eliminate high frequency components in the audio signal. This low-pass filter, usually called an *anti-aliasing filter* because it prevents aliasing, cuts out all energy above the Nyquist, thus solving the problem. Of course, some people think that this creates a huge problem because it leaves out a lot of information that no one can really prove isn't important.

There is a more detailed discussion of the issue of aliasing and antialiasing filters in Section 7.3.

7.1.4 Binary numbers and bits

If you don't understand how to count in binary, please read Section 1.7.

As we'll talk about a little later, we need to convert the numbers that describe the level of each sample into a *binary* number before storing or transmitting it. This just makes the number easier for a computer to recognize.

The reason for doing this conversion from decimal to binary is that computers – and electrical equipment in general – are happier when they only have to think about two digits. Let's say, for example that you had to invent a system of sending numbers to someone using a flashlight. You could put a dimmer switch on the flashlight and say that, the bigger the number, the brighter the light. This would give the receiver an intuitive idea of the size of the number, but it would be extremely difficult to represent numbers accurately. On the other hand, if we used binary notation, we could say “if the light is on, that's a 1 – if it's off, that's a 0” then you can just switch the light on and off for 1's and 0's and you send the number. (of course, you run into problems with consecutive 1's or 0's – but we'll deal with that later...)

Similarly, computers use voltages to send signals around – so, if the voltage is high, we call that a 1, if it's low, we call it 0. That way we don't have to use 10 different voltage levels to represent the 10 digits. Therefore, in the computer world, binary is better than decimal.

7.1.5 Two's complement

Let's back up a bit (no pun intended...) to the discussion on binary numbers. Remember that we're using those binary numbers to describe the signal level. This would not really be a problem except for the fact that the signal is what is called *bipolar* meaning that it has positive and negative components. we

could use positive and negative binary numbers to represent this but we don't. We typically use a system called "two's complement." There are really two issues here. One is that, if there's no signal, we'd probably like the digital representation of it to go to 0 – therefore zero level in the analog signal corresponds to zeros in the digital signal. The second is, how do we represent negative numbers? One way to consider this is to use a circular plotting of the binary numbers. If we count from 0 to 7 using a 3-bit "word" we have the following:

000
001
010
011
100
101
110
111

Now if we write these down in a circle starting at 12 o'clock and going clockwise as is shown in Figure 7.11, we'll see that the value 111 winds up being adjacent to the value 000. Then, we kind of ignore what the actual numbers are and starting at 000 turn clockwise for positive values and counterclockwise for negative values. Now, we have a way of representing positive and negative values for the signal where one step above 000 is 001 and one step below 000 is 111. This seems a little odd because the numbers don't really line up the way we'd like them as can be seen in Figure 7.12 – but does have some advantages. Particularly, digital zero corresponds to analog zero – and if there's a 1 at the beginning of the binary word, then the signal is negative.

One issue that you may want to concern yourself here is the fact that there is one more quantization level in the negative area than there is in the positive. This is because there are an even number of quantization levels (because that number is a power of two) but one of them is dedicated to the zero level. Therefore, the system is slightly asymmetrical – so it is, in fact possible to distort the signal in the positive before you start distorting in the negative. But keep in mind that, in a typical 16-bit system we're talking about a VERY small difference.

7.1.6 Suggested Reading List

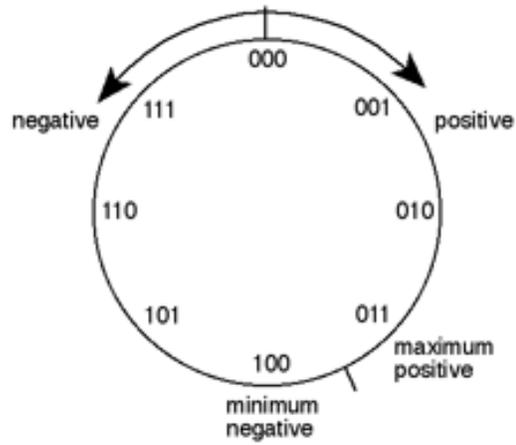


Figure 7.11: Counting from 0 to 7 using a 3-bit word around a circle.

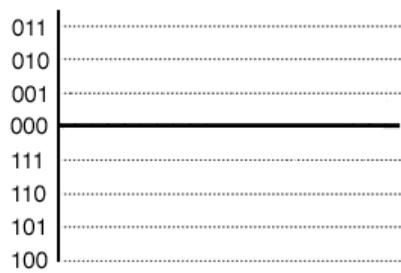


Figure 7.12: Binary words corresponding to quantization levels in a two's complement system.

7.2 Quantization Error and Dither

The fundamental difference between digital audio and analog audio is one of resolution. Analog representations of analog signals have a theoretically infinite resolution in both level and time. Digital representations of an analog sound wave are *discretized* into quantifiable levels in slices of time. We've already talked about discrete time and sampling rates a little in the previous section and we'll elaborate more on it later, but for now, let's concentrate on quantization of the signal level.

As we've already seen, a PCM-based digital audio system has a finite number of levels that can be used to specify the signal level for a particular sample on a given channel. For example, a compact disc uses a 16-bit binary word for each sample, therefore there are a total of 65,536 (2^{16}) quantization levels available. However, we have to always keep in mind that we only use all of these levels if the signal has an amplitude equal to the maximum possible level in the system. If we reduce the level by a factor of 2 (in other words, a gain of -6.02 dB) we are using one fewer bits worth of quantization levels to measure the signal. The lower the amplitude of the signal, the fewer quantization levels that we can use until, if we keep attenuating the signal, we arrive at a situation where the amplitude of the signal is the level of 1 *Least Significant Bit* (or *LSB*).

Let's look at an example. Figure 9.1 shows a single cycle of a sine wave plotted with a pretty high degree of resolution (well... high enough for the purposes of this discussion).

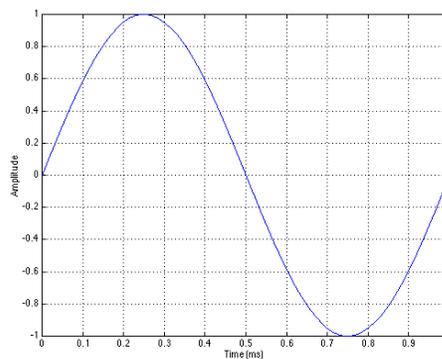


Figure 7.13: A single cycle of a sine wave. We'll consider this to be the analog input signal to our digital converter.

Let's say that this signal is converted into a PCM digital representation using a converter that has 3 bits of resolution – therefore there are a total of 8 different levels that can be used to describe the level of the signal. In a two's complement system, this means we have the zero line with 3 levels above it and 4 below. If the signal in Figure 9.1 is aligned in level so that its positive peak is the same as the maximum possible level in the PCM digital representation, then the resulting digital signal will look like the one shown in Figure 7.14.

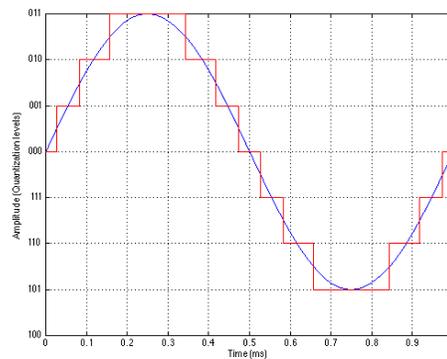


Figure 7.14: A single cycle of a sine wave after conversion to digital using 4-bit, PCM, two's complement where the signal level is rounded to the nearest quantization level at each sample. The blue plot is the original waveform, the red is the digital representation.

Not surprisingly, the digital representation isn't exactly the same as the original sine wave. As we've already seen in the previous section, the cost of quantization is the introduction of errors in the measurement. However, let's look at exactly how much error is introduced and what it looks like.

This error is the difference between what we put into the system and what comes out of it, so we can see this difference by subtracting the red waveform in Figure 7.14 from the blue waveform.

There are a couple of characteristics of this error that we should discuss. Firstly, because the sine wave repeats itself, the error signal will be periodic. Also, the period of this complex waveform made up of the will be identical to the original sine wave – therefore it will be comprised of harmonics of the original signal. Secondly, notice that the maximum quantization error that we introduce is one half of 1 LSB. The significant thing to note about this is its relationship to the signal amplitude. The quantization error will never be greater than one half of an LSB, so the more quantization levels

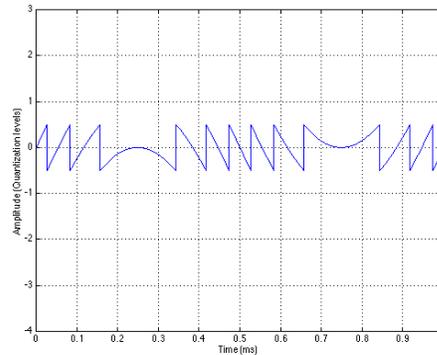


Figure 7.15: A plot of the quantization error generated by the conversion shown in Figure 7.14.

we have, the louder we can make the signal we want to hear relative to the error that we don't want to hear. See Figures 7.16 through 7.18 for a graphic illustration of this concept.

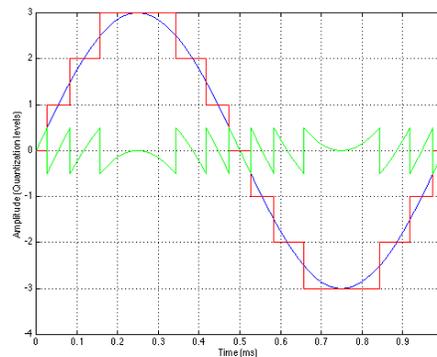


Figure 7.16: A combined plot of the original signal, the quantized signal and the resulting quantization error in a 3-bit system.

As is evident from Figures 7.16, 7.17 and 7.18, the greater the number of bits that we have available to describe the instantaneous signal level, the lower the apparent level of the quantization error. I use the apparent here in a strange way – no matter how many bits you have, the quantization error will be a signal that has a peak amplitude of one half of an LSB in the worst case. So, if we're thinking in terms of LSB's – then the amplitude of the

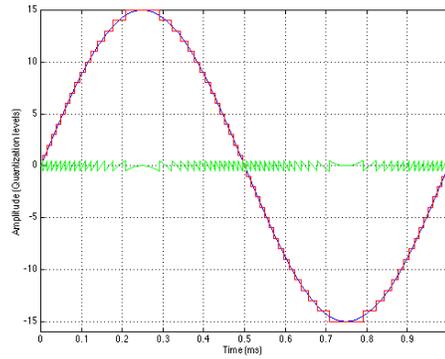


Figure 7.17: A combined plot of the original signal, the quantized signal and the resulting quantization error in a 5-bit system.

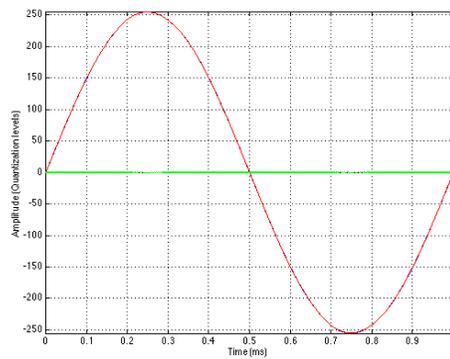


Figure 7.18: A combined plot of the original signal, the quantized signal and the resulting quantization error in a 9-bit system.

quantization error is the same no matter what your resolution. However, that's not the way we normally think – typically we think in terms of our signal level, so, relative to that signal, the higher the number of available quantization levels, the lower the amplitude of the quantization error.

Given that a CD has 65,536 quantization levels available to us, do we really care about this error? The answer is “yes” – for two reasons:

1. We have to always remember that the only time all of the bits in a digital system are being used is when the signal is at its maximum possible level. If you go lower than this – and we usually do – then you're using a subset of the number of quantization levels. Since the quantization error stays constant at ± 0.5 LSB and since the signal level is lower, then the relative level of the quantization error to the signal is higher. The lower the signal, the more audible the error. This is particularly true at the end of the decay of a note on an instrument or the reverberation in a large room. As the sound decays from maximum to nothing, it uses fewer and fewer quantization levels and the perceived quality drops because the error becomes more and more evident because it is less and less masked.
2. Since the quantization error is periodic, it is a distortion of the signal and is therefore directly related to the signal itself. Our brains are quite good at ignoring unimportant things. For example, you walk into someone's house and you smell a new smell – the way that house smells. After 5 minutes you don't smell it anymore. The smell hasn't gone away – your brain just ignores it when it realizes that it's a constant. The same is true of analog tape noise. If you're like most people you pay attention to the music, you stop hearing the noise after a couple of minutes. Your brain is able to do this all by itself because the noise is unrelated to the signal. It's a constant, unrelated sound that never changes with the music and is therefore unrelated – the brain decides that it doesn't change so it's not worth tracking. Distortion is something different. Distortion, like noise, is typically comprised entirely of unwanted material (I'm not talking about guitar distortion effects or the distortion of a vintage microphone here...). Unlike noise, however, distortion products modulate with the signal. Consequently the brain thinks that this is important material because it's trackable, and therefore you're always paying attention. This is why it's much more difficult to ignore distortion than noise. Unfortunately, quantization error produces distortion – not noise.

7.2.1 Dither

Luckily, however, we can make a cute little trade. It turns out that we can effectively eliminate quantization error simply by adding noise called *dither* to our signal. It seems counterproductive to fix a problem by adding noise – but we have to consider that what we’re essentially doing is to make a trade – distortion for noise. By adding dither to the audio signal with a level that is approximately one half the level of the LSB, we generate an audible, but very low-level constant noise that effectively eliminates the program-dependent noise (distortion) that results from low-level signals.

Notice that I used the word “effectively” at the beginning of the last paragraph. In fact, we are not eliminating the quantization error. By adding dither to the signal before quantizing it, we are randomizing the error, therefore changing it from a program-dependent distortion into a constant noise floor. The advantage of doing this is that, although we have added noise to our final signal, it is constant, and therefore not trackable by our brains. Therefore, we ignore it,

So far, all I have said is that we add “noise” to the signal, but I have not said what kind of noise - is it white, pink, or some other colour? People who deal with dither typically don’t use these types of terms to describe the noise – they talk about *probability density functions* or *PDF* instead. When we add dither to a signal before quantizing it, we are adding a random number that has a value that will be within a predictable range. The range has to be controlled, otherwise the level of the noise would be unnecessarily high and therefore too audible, or too low, and therefore ineffective.

Probability density functions

Flip a coin. You’ll get a heads or a tails. Flip it again, and again and again, each time writing down the result. If you flipped the coin 1000 times, chances are that you’ll see that you got a heads about 500 times and a tails about 500 times. This is because each side of the coin has an equal chance of landing up, therefore there is a 50% probability of getting a heads, and a 50% probability of getting a tails. If we were to draw a graph of this relationship, with “heads” and “tails” being on the x-axis and the probability on the y-axis, we would have two points, both at 50%.

Let’s do basically the same thing by rolling a die. If we roll it 600 times, we’ll probably see around 100 1’s, 100 2’s, 100 3’s, 100 4’s, 100 5’s and 100 6’s. Like the coin, this is because each number has an equal probability of being rolled. I tried this, and kept track of each number that was rolled an

got the following numbers.

| Result | 1 | 2 | 3 | 4 | 5 | 6 |
|------------------------|-----|----|-----|----|----|-----|
| Number of times rolled | 111 | 98 | 101 | 94 | 92 | 104 |

Table 7.1: The results of rolling a die 600 times.

If we were to graph this information, it would look like Figure 7.19.

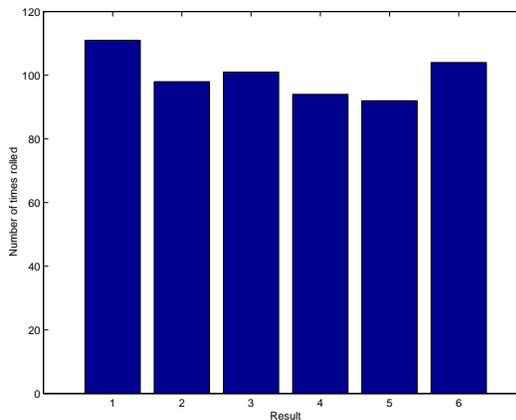


Figure 7.19: The results of 600 rolls of a die.

Let's say that we didn't know that there was an equal probability of rolling each number on the die. How could we find this out experimentally? All we have to do is to take the numbers in Table 7.1 and divide by the number of times we rolled the die. This then tells us the probability (or the chances) of rolling each number. If the probability of rolling a number is 1, then it will be rolled every time. If the probability is 0, then it will never be rolled. If it is 0.5, then the number will be rolled half of the time.

Notice that the numbers didn't work out perfectly in this example, but they did come close. I was expecting to get each number 100 times, but there was a small deviation from this. The more times I roll the dice, the more reality will approach the theoretical expectation. To check this out, I did a second experiment where I rolled the die 60,000 times.

This graph tells us a number of things. Firstly, we can see that there is a 0 probability of rolling a 7 (this is obvious because there is no "7" on a die, so we can never roll and get that result). Secondly, we can see that

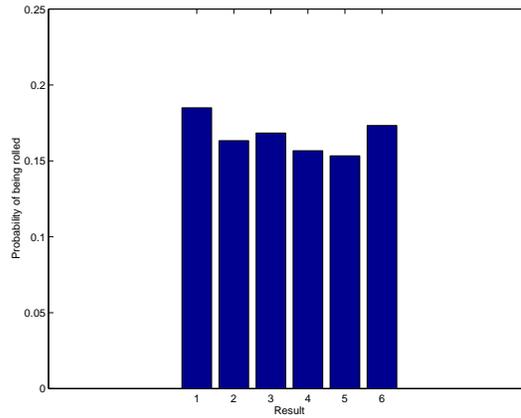


Figure 7.20: The calculated probability of rolling each number on the die using the results shown in Table 7.1.

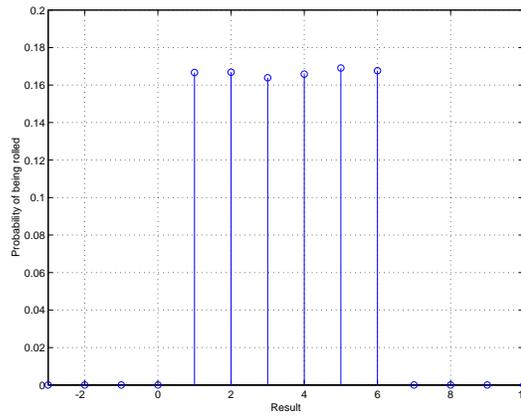


Figure 7.21: The calculated probability of rolling each number on the die using the results after 60,000 rolls. Notice that the graph has a rectangular shape.

there is an almost exactly equal probability of rolling the numbers from 1 to 6 inclusive. Finally, if we look at the shape of this graph, we can see that it makes a rectangle. So, we can say that rolling a die results in a *rectangular probability density function* or *RPDF*.

It is possible to have other probability density functions. For example, let's look at the ages of children in Grade 5. If we were to take all the Grade 5 students in Canada, ask them their age, and make a PDF out of the results, it might look like Figure 7.22.

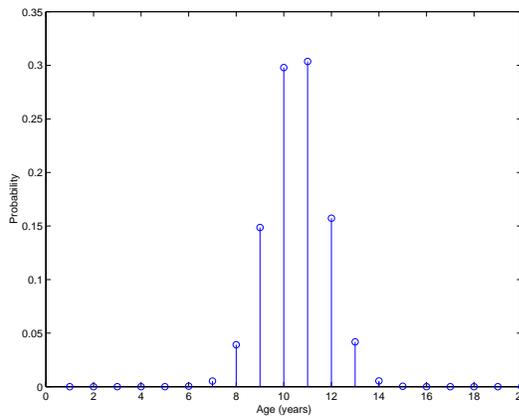


Figure 7.22: The calculated probability of the age of students in Grade 5 in Canada.

This is obviously not an RPDF because the result doesn't look like a rectangle. In fact, it is what statisticians call a *normal distribution*, better known as a *bell curve*. What this tells us is that the probability a Canadian Grade 5 student of being either 10 or 11 years old is higher than for being any other age. It is possible, but less likely that the student will be 8, 9, 12 or 13 years old. It is extremely unlikely, but also possible for the student to be 7 or 14 years old.

Dither examples

There are three typical dither PDF's used in PCM digital audio, RPDF, *TPDF* (*triangular PDF*) and *Gaussian PDF*. We'll look at the first two.

For this section, I used MATLAB to make a sine wave with a sampling rate of 32.768 kHz. I realize that this is a strange sampling rate, but it made the graphs cleaner for the FFT analyses. The total length of the sine wave was 32768 samples (therefore 1 second of audio.) MATLAB typically

calculates in 64-bit floating point, so we have lots of resolution for analyzing an 8-bit signal, as I'm doing here.

To make the dither, I used the MATLAB function for generating random numbers called `RAND`. The result of this is a number between 0 and 1 inclusive. The PDF of the function is rectangular as we'll see below.

RPDF dither should have a rectangular probability density function with extremes of -0.5 and 0.5 LSB's. Therefore, a value of more than half of an LSB is not possible in either the positive or negative directions. To make RPDF dither, I made a vector of 32768 numbers using the command `RAND(1, n) - 0.5` where n is the length of the dither signal, in samples. The result is equivalent to white noise.

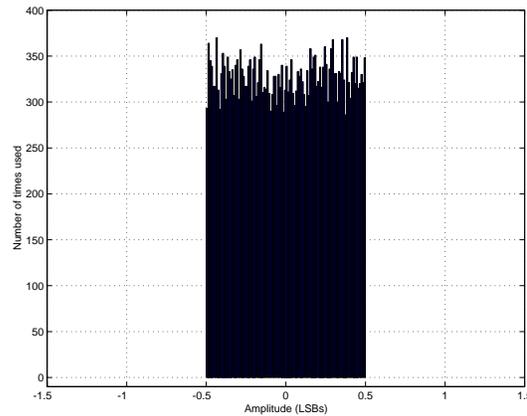


Figure 7.23: Histogram of RPDF dither for 32k block. MATLAB equation is `RAND(1, 32768) - 0.5`

TPDF dither has the highest probability of being 0, and a 0 probability of being either less than -1 LSB or more than 1 LSB. This can be made by adding two random numbers, each with an RPDF together. Using MATLAB, this is most easily done using the the command `RAND(1, n) - RAND(1, n)` where n is the length of the dither signal, in samples. The reason they're subtracted here is to produce a TPDF that ranges from -1 to 1 LSB.

Let's look at the results of three options: no dither, RPDF dither and TPDF dither. Figure 7.27 shows a frequency analysis of 4 signals (from top to bottom): (1) a 64-bit 1 kHz sine wave, (2) an 8-bit quantized version of the sine wave without dither added, (3) an 8-bit quantized version with RPDF added and (4) an 8-bit quantized version with TPDF added.

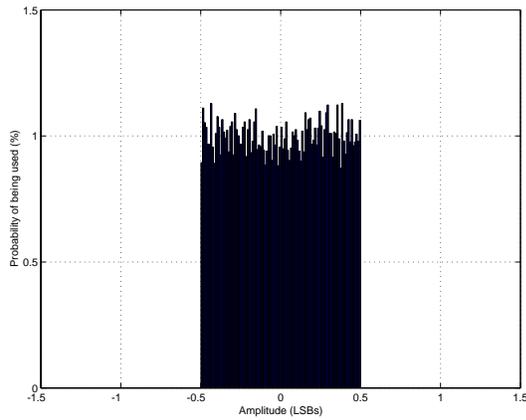


Figure 7.24: Probability distribution function of RPDF dither for 32k block.

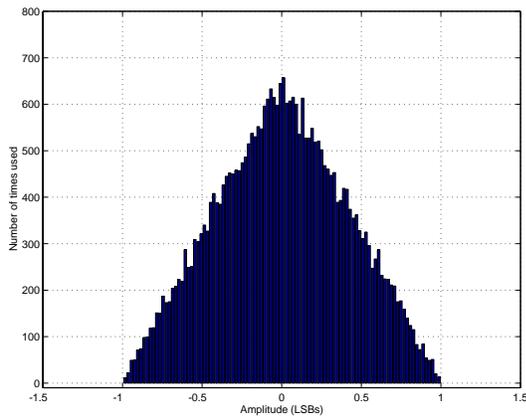


Figure 7.25: Histogram of TPDF dither for 32k block. MATLAB equation is $\text{RAND}(1, 32768) - \text{RAND}(1, 32768)$

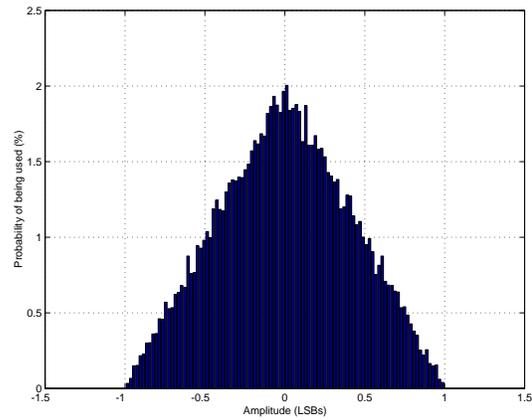


Figure 7.26: Probability distribution function of TPDF dither for 32k block.

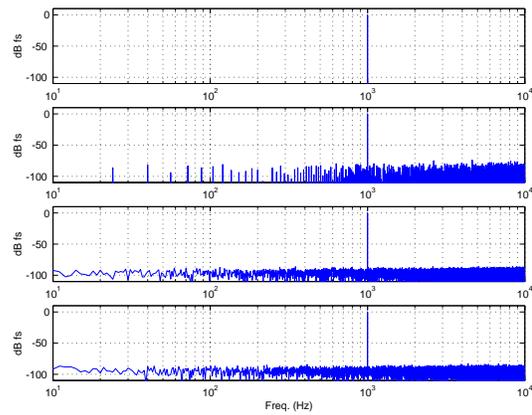


Figure 7.27: From top to bottom, a 64-bit sine 1 kHz wave in MATLAB, 8-bit no dither, 8-bit RPDF dither, 8-bit TPDF dither. $F_s = 32768$ Hz, FFT window is rectangular, FFT length is 32768 point.

One of the important things to notice here is that, although the dithers raised the overall noise floor of the signal, the resulting artifacts are wide-band noise, rather than spikes showing up at harmonic intervals as can be seen in the no-dither plot. If we were to look at the artifacts without the original 1 kHz sine wave, we get a plot as shown in Figure 7.28.

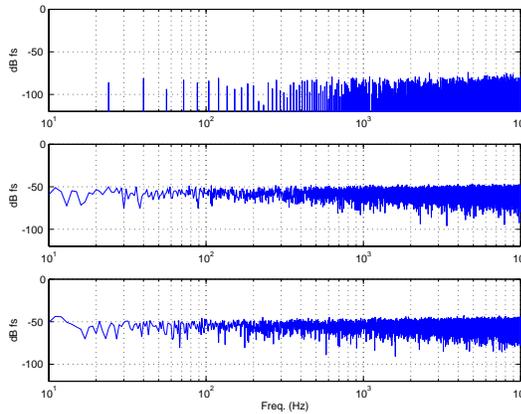


Figure 7.28: Artifacts omitting the 1 kHz sine wave. From top to bottom, 8-bit no dither, 8-bit RPDF dither, 8-bit TPDF dither. $F_s = 32768$ Hz, FFT window is rectangular, FFT length is 32768 point.

7.2.2 When to use dither

I once attended a seminar on digital audio measurements given by Julian Dunn. I think that, in a room of about 60 or 70 people, I was the only one in the room who was not an engineer working for a company that made gear (I was working on my Ph.D. in music at the time...) At lunch, I sat at a table of about 12 people, and someone asked me the simple question “what do you think about dither?” I responded that I thought it was a good idea. Then the question was re-phrased – “yes, but when do you use it?” The answer is actually pretty simple – you use dither whenever you have to re-quantize a signal. Typically, we do DSP on our audio signals using word lengths much greater than the original sample, or the resulting output. For example, we typically record at 16 or 24 bits (with dither built into the ADC’s), and the output is usually at one of these two bit depths as well. However, most DSP (like equalization, compression, mixing and reverberation) happens with an accuracy of 32 bits (although there are some devices such as those from Eventide that run at 64-bit internally). So, a 16-bit signal comes into your

mixer, it does math with an accuracy of 32 bits, and then you have to get out to a 16-bit output. The process of converting the 32-bit signal into a 16-bit signal must include dither.

Remember, if you are quantizing, you dither.

7.2.3 Suggested Reading List

7.3 Aliasing

Back in Section 7.1.3, we looked very briefly at the problem of aliasing, but now we have to dig a little deeper to see more specifically what it does and how to avoid it.

As we have already seen, aliasing is an artifact of chopping continuous events into discrete time. It can be seen in film when cars go forwards but their wheels go backwards. It happens under fluorescent or strobe lights that prevent us from seeing intermediate motion. Also, if we're not careful, it happens in digital audio.

Let's begin by looking at a simple example: we'll look at a single sampling rate and what it does to various sine waves from analog input to analog output from the digital system.

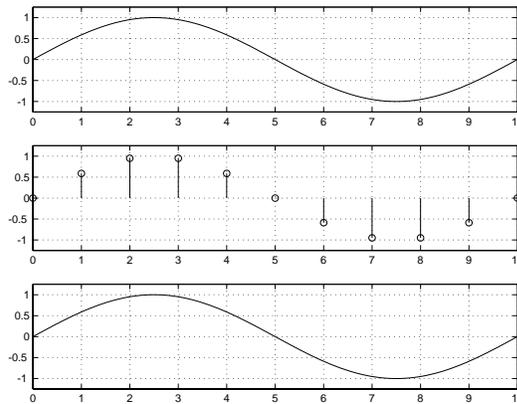


Figure 7.29: INSERT CAPTION

7.3.1 Antialiasing filters

NOT YET WRITTEN

7.3.2 Suggested Reading List

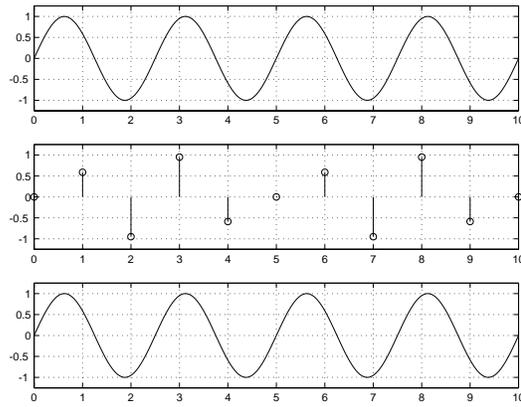


Figure 7.30: INSERT CAPTION

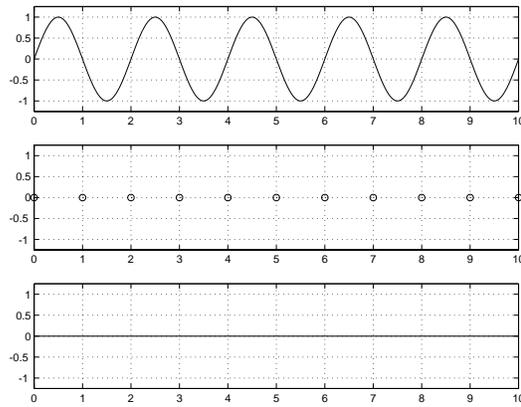


Figure 7.31: INSERT CAPTION

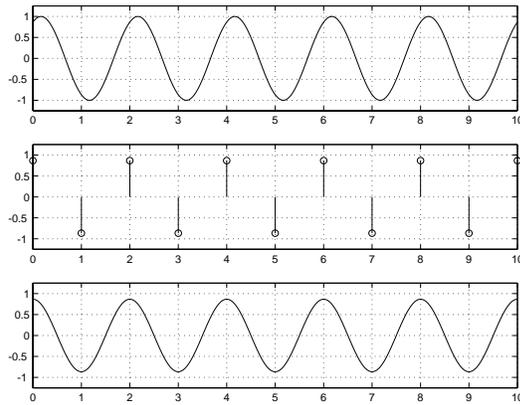


Figure 7.32: INSERT CAPTION

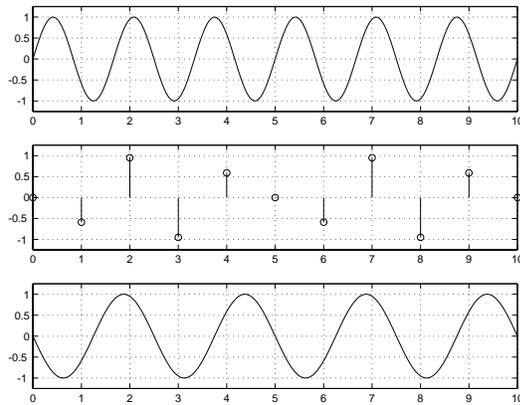


Figure 7.33: INSERT CAPTION

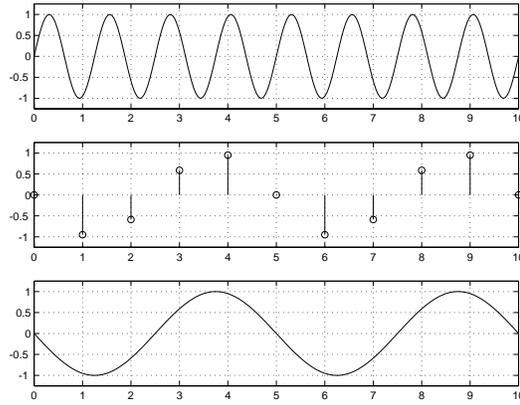


Figure 7.34: INSERT CAPTION

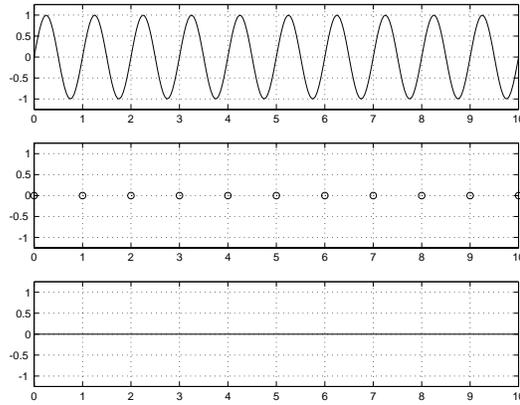


Figure 7.35: INSERT CAPTION

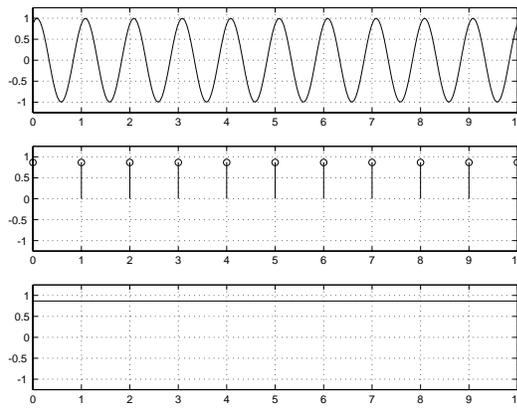


Figure 7.36: INSERT CAPTION

7.4 Advanced A/D and D/A conversion

NOT YET WRITTEN

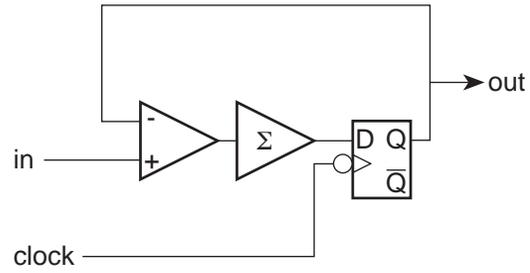


Figure 7.37: A simple 1-bit $\Delta - \Sigma$ (Delta-Sigma) analog to digital converter [Watkinson, 1988].

7.4.1 Suggested Reading List

7.5 Digital Signal Transmission Protocols

7.5.1 Introduction

If we want to send a digital (we'll assume that this, in turn means "binary") signal across a wire, the simplest way to do it is to have a ground reference and some other DC voltage (5 V is good...) and we'll call one "0" (being denoted by ground) and the other "1" (5 V). If we then know WHEN to look at the electrical signal on the wire (this timing can be determined by an internal clock...), we can know whether we're receiving a 0 or 1. This system would work – in fact it does, but it has some inherent problems which would prevent us from using it as a method of transmitting digital audio signals around the real world.

Back in the early 1980's a committee was formed by the Audio Engineering Society to decide on a standard protocol for transmitting digital audio. And, for the most part, they got it right... They decided at the outset that they wanted a system that had the following characteristics and the resulting implications :

The protocol should use existing cables, connectors and jackfields. In addition, it should withstand transmission through existing equipment. The implications of this are :

- There can be no DC content which would be wiped out by transformers or capacitors in the transmission chain. (Think of a string of 1's in the 0 V / 5 V system explained above... it would be 5 V DC which wouldn't get through your average system)
- The signal cannot be phase-critical. That is to say, if we take the signal and flip its polarity, it cannot change 1's to 0's and vice versa. This is just in case someone plugs the wrong cable in.
- The signal must be self-clocking
- The signal must have 2 channels of audio on a single cable.

The result was the AES/EBU protocol (also known as IEC-958 Type 1). It's a *bi-phase mark coding protocol* which fulfills all of the above requirements. "What's a bi-phase mark coding protocol?" I hear you cry... Well what that means is that, rather than using two discrete voltages to denote 1 and 0, the distinction is made by voltage transitions.

In order to transmit a single bit down a wire, the AES/EBU system carves it into two "cells." If the cells are the same voltage, then the bit is a

0 : if the cells are different voltages, then the bit is a 1. In other words, if there is a transition between the cells, the bit is a 1. If there is no transition, the bit is a 0.

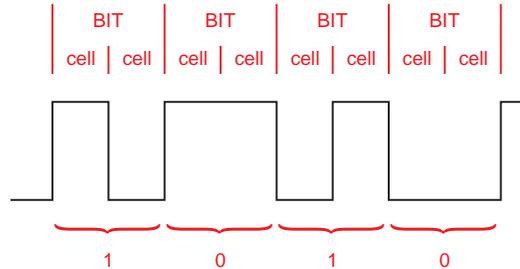


Figure 7.38: The relationship between cell transitions and the value of the bit in a bi-phase mark. Note that if both cells in one bit are the same, the represented value is 0. If the two cells have a different value, the bit value is 1. This is independent of the actual high or low value in the signal.

The peak-peak level of this signal is between 2 V and 7 V. The source impedance is 150Ω (this is the impedance between pins 2 and 3 on the XLR connector).

Note that there is no need for a 0 V reference in this system. The AES/EBU receiver only looks at whether there is a transition between the cells – it doesn’t care what the voltage is – only whether it’s changed.

The only thing we’ve left out is the “self-clocking” part... This is accomplished by a circuit known as a Phase-Locked Loop (or PLL to its friends...). This circuit creates a clock using an oscillator which derives its frequency from the transition rate of the voltage it receives. The AES/EBU signal is sent into the PLL which begins in a “capture” mode. This is a wide-angle view where the circuit tries to get a basic idea of what the frequency of the incoming signal is. Once that’s accomplished, it moves into “pull-in” mode where it locks on the frequency and stays there. This PLL then becomes the clock which is used by the receiving device’s internals (like buffers and ADC’s).

7.5.2 What’s being sent?

AES/EBU data is sent in *Blocks* which are comprised of 192 *Frames*. Each frame contains 2 *Sub-Frames*, each of which contains 32 *bits* of information. The layout goes like this :

The information in the sub-frame can be broken into two categories, the *channel code* and the *source code*, each of which comprises various pieces of

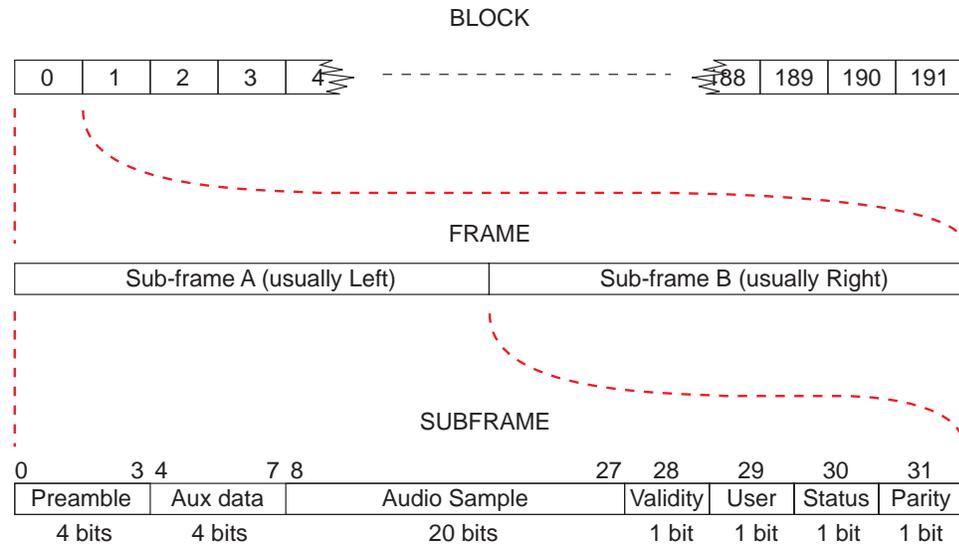


Figure 7.39: The relationship of the structures of a Block, Frame and Sub-Frame.

| Code | Channel Code | Source Code |
|-----------------|------------------------|--|
| Contents | Preamble
Parity Bit | Audio Sample
Auxiliary data
Validity Bit
User Bit
Status Bit |

Table 7.2: The contents of the Channel Code and the Source Code

information.

Channel Code

This is information regarding the transmission itself – data that keeps the machines talking to each other. It consists of 5 bits making up the Preamble (or Sync Code) and the Parity Bit.

Preamble (also known as the Sync Code)

These are 4 bits which tell the receiving device that the transmission is at the beginning of a block or a subframe (and which subframe it is...) Different specific codes tell the receiver what’s going on as follows...

Note that these codes violate the bi-phase mark protocol (because there

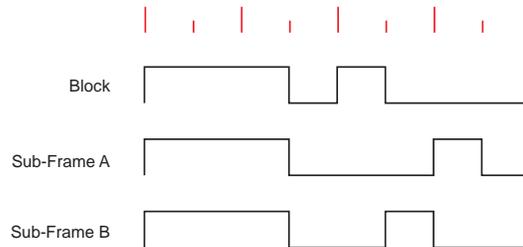


Figure 7.40: The structure of the preamble at the start of each Sub-Frame. Note that each of these breaks the bi-phase mark rule that there must be a transition on every bit since all preambles start with 3 consecutive cells of the same value.

is no transition at the beginning of the second bit.) but they do not violate the no-DC rule.

Note as well, that these are sometimes called the *X*, *Y*, and *Z* preambles. An *X* Preamble indicates that the Sub-Frame is an audio sample for the Left. A *Y* Preamble indicates that the Sub-Frame is an audio sample for the Right. A *Z* Preamble indicates the start of a Block.

Parity Bit

This is a single bit which ensures that all of the preambles are in phase. It doesn't matter to the receiving device whether the preambles start by going up in voltage or down (I drew the above examples as if they are all going up...) but all of the preambles must go the same way. The parity bit is chosen to be a 1 or 0 to ensure that the next preamble will be headed in the right direction.

Source Code

This is the information that we're trying to transmit. It uses the other 27 bits of the sub-frame comprising the Audio Sample (20 bits), the Auxiliary Data (4 bits), the Validity Bit (1 bit), the User Bit (1 bit) and the Status Bit (1 bit).

Audio sample

This is the sample itself. It has a maximum of 20 bits, with the Least Significant Bit sent first.

Auxiliary Data

This is 4 bits which can be used for anything. These days it's usually used for 4 extra bits to be attached to the audio sample – bringing the resolution up to 24 bits.

Validity Bit

This is simply a flag which tells the receiving device whether the data is valid or not. If the bit is a 1, then the data is non-valid. A 0 indicates that the data is valid.

User Bit

This is a single bit which can be used for anything the user or manufacturer wants (such as time code, for example).

For example, a number of user bits from successive sub-frames are strung together to make a single word. Usually this is done by collecting all 192 user bits (one from each sub frame) for each channel in a block. If you then put these together, you get 24 bytes of information in each channel.

Typically, the end user in a recording studio doesn't have direct access to how these bits should be used. However, if you have a DAT machine, for example, that is able to send time code information on its digital output, then you're using your user bits.

Status Bit

This is a single-bit flag which can be used for a number of things such as :

- Emphasis on / off
- Sampling rate
- Stereo / Mono signal
- Word length of audio sample
- ascii (8 bits) for channel origin and destination

This information is arranged in a similar method to that described for the User Bits. 192 status bits are collected per channel per block. Therefore, you have 192 bits for the A channel (left) and 192 for the B channel (right). If you string these all together, then you have 24 bytes of information in each channel. The AES/EBU standard dictates what information goes where in this list of bytes. This is shown in the diagram in Figure ??

For specific information regarding exactly what messages are given by what arrangement of bits, see [Sanchez and Taylor, 1998] available as Application Note AN22 from www.crystal.com.

7.5.3 Some more info about AES/EBU

The AES/EBU standard (IEC 958 Type 1) was set in 1985.

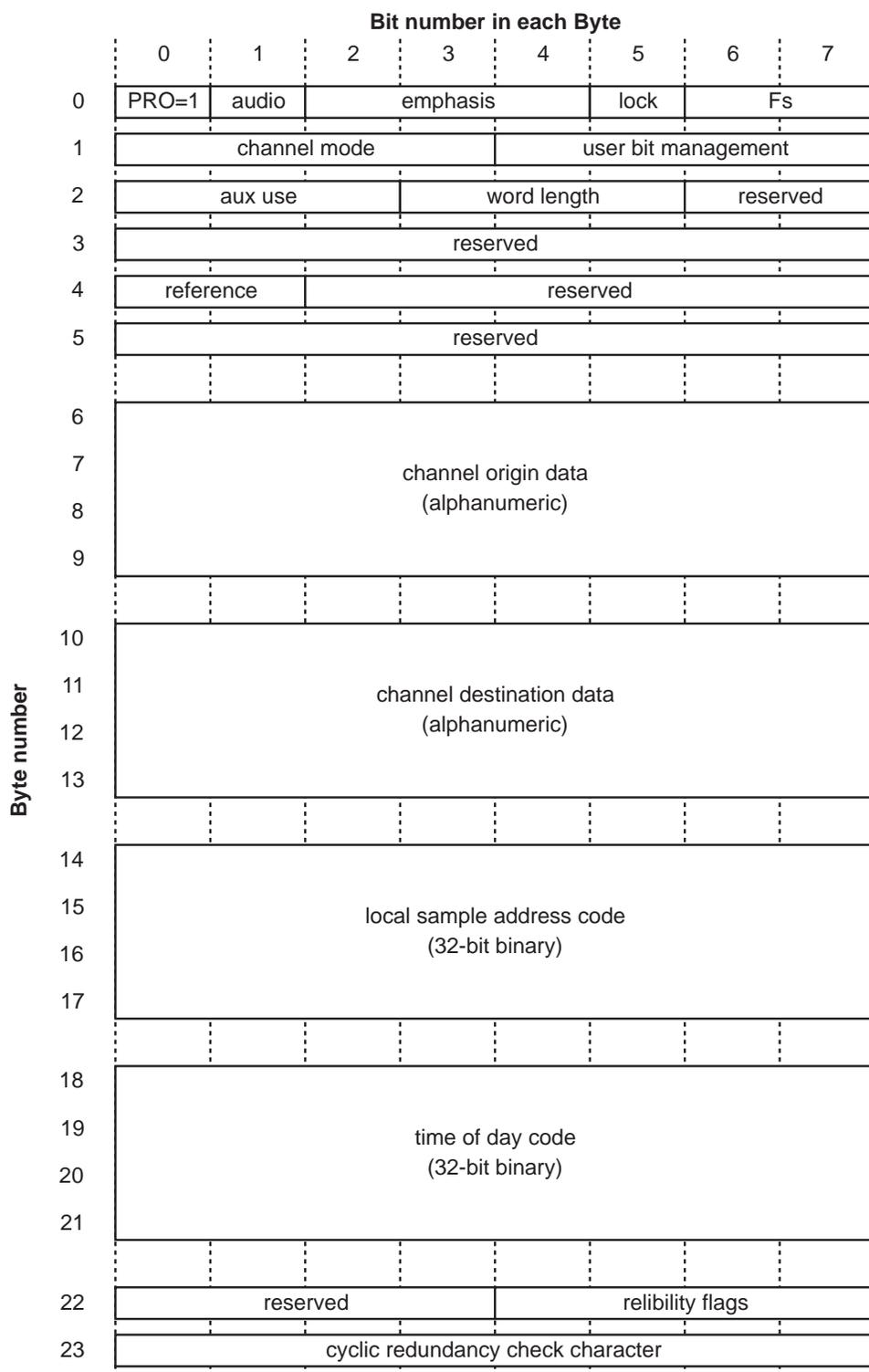


Figure 7.41: The structure of the bytes made out of the status bits in the channel code information in a single Block. This is sometimes called the Channel Status Block Structure [Sanchez and Taylor, 1998].

The maximum cable run is about 300 m balanced using XLR connectors. If the signal is unbalanced (using a transformer, for example) and sent using a coaxial cable, the maximum cable run becomes about 1 km.

Fundamental Frame Rates

If the Sampling Rate is 44.1 kHz, 1 frame takes 22.7 microsec. to transmit (the same as the time between samples)

If the Sampling Rate is 48 kHz, 1 frame takes 20.8 microsec. to transmit

At 44.1 kHz, the bit rate is 2.822 Mbit/s

At 48 kHz, the bit rate is 3.072 Mbit/s

Just for reference (or possibly just for interest), this means that 1/4 wavelength of the cell in AES/EBU is about 19 m on a wire.

7.5.4 S/PDIF

S/PDIF was developed by Sony and Philips (hence the S/P) before AES/EBU. It uses a single unbalanced coaxial wire to transmit 2 channels of digital audio and is specified in IEC 958 Type 2. The Source Code is identical to AES/EBU with the exception of the channel status bit which is used as a copy prohibit flag.

Some points :

The connectors used are RCA with a coaxial cable

The voltage alternates between 0V and 1V 20% (note that this is not independent of the ground as in AES/EBU)

The source impedance is 75Ω

S/PDIF has better RF noise immunity than AES/EBU because of the coax cable (please don't ask me to explain why... the answer will be "dunno... too dumb...")

It can be sent as a "video" signal through existing video equipment

Signal loss will be about -0.75 dB / 35 m in video cable

7.5.5 Some Terms That You Should Know...

Synchronous

Two synchronous devices have a single clock source and there is no delay between them. For example, the left windshield wiper on your car is synchronous with the right windshield wiper.

Asynchronous

Two asynchronous devices have absolutely no relation to each other. They are free-running with separate clocks. For example, your windshield wipers are asynchronous with the snare drum playing on your car radio.

Isochronous

Two isochronous devices have the same clock but are separated by a fixed propagation delay. They have a phase difference but that difference remains constant.

7.5.6 Jitter

Jitter is a modulation in the frequency of the digital signal being transmitted. As the bit rate changes (and assuming that the receiving PLL can't correct variations in the frequency), the frequency of the output will modulate and therefore cause distortion or noise.

Jitter can be caused by a number of things, depending on where it occurs :

Intrinsic jitter within a device

- Parasitic capacitance within a cable
- Oscillations within the device
- Variable timing stability

Transmission Line Jitter

Reflections off stubs

- Impedance mismatches

Jitter amounts are usually specified as a time value (for example, $X ns_p - p$).

The maximum allowable jitter in the AES/EBU standard is 20 ns (10 ns on either side of the expected time of the transition).

See Bob Katz's 'Everything you wanted to know about jitter but were afraid to ask' (www.digido.com/jitteressay.html) for more information.

7.5.7 Suggested Reading List

Sanchez, C. and Taylor, R. (1998) Overview of Digital Audio Interface Data Structures. Application Note AN22REV2, Cirrus Logic Inc. (available at

<http://www.crystal.com>)

7.6 Jitter

Go and make a movie using a movie camera that runs at 24 frames per second. Then, play back the movie at 30 fps. Things in the movie will move faster than they did in real life because the frame rate has speeded up. This might be a neat effect, but it doesn't reflect reality. The point so far is that, in order to get out what you put in, a film must be played back at the same frame rate at which it was recorded.

Similarly, when an audio signal is recorded on a digital recording system, it must be played back at the same sampling rate in order to ensure that you don't result in a frequency shift. For example, if you increase the sampling rate by 6% on playback, you will produce a shift in pitch of a semitone.

There is another assumption that is made in digital audio (and in film, but it's less critical). This is that the sampling rate does not change over time – neither when you're recording nor on playback.

Let's think of the simple case of a sine tone. If we record a sine wave with a perfectly stable sampling rate, and play it back with a perfectly stable sampling rate with the same frequency as the recording sampling rate, then we get out what we put in (ignoring any quantization or aliasing effects...). We know that if we change the sampling rate of the playback, we'll shift the frequency of the sine tone. Therefore, if we modulate the sampling rate with a regular signal, shifting it up and down over time, then we are subjecting our sine tone to *frequency modulation* or *FM*.

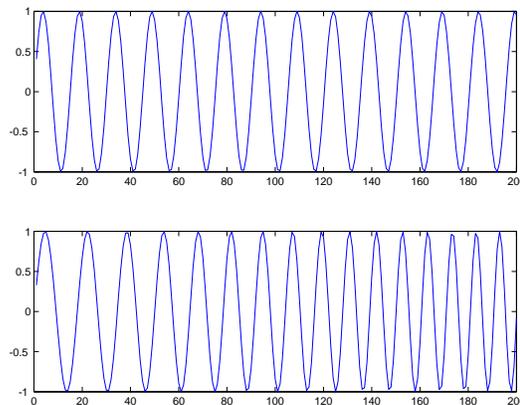


Figure 7.42: INSERT CAPTION

NOT YET WRITTEN

7.6.1 When to worry about jitter

There are many cases where the presence of jitter makes absolutely no difference to your audio signal whatsoever.

NOT YET WRITTEN

7.6.2 What causes jitter, and how to reduce it

NOT WRITTEN YET

7.6.3 Suggested Reading List

7.7 Fixed- vs. Floating Point

7.7.1 Suggested Reading List

MATLAB paper

 Jamie's textbook

7.8 Noise Shaping

NOT YET WRITTEN

7.8.1 Suggested Reading List

7.9 “High-Resolution” Audio

Back in the days when digital audio was first developed, it was considered that a couple of limits on the audio quality were acceptable. Firstly, it was decided that the sampling rate should be at least high enough to provide the capability of recording and reproducing a 20 kHz signal. This means a sampling rate of at least 40 kHz. Secondly, a dynamic range of around 100 dB was considered to be adequate. Since a word length of 16 bits (a convenient power of 2) gives a dynamic range of about 93 dB after proper dithering, that was decided to be the magic number.

However, as we all know, the business of audio is all about providing *more* above all else, therefore the resolution of commercially-available digital audio, both in the temporal and level domains, had to be increased. There are some less-cynical arguments for increasing these resolutions.

One argument is that the rather poor (sic) quality of 44.1/16 (kHz and bits respectively) audio is inadequate for any professional recording. This might be supportable were it not for the proliferation of mp3 files throughout the world. Consumption indicates that the general consumer is more and more satisfied with the quality of data-compressed audio which is considerably worse than CD-quality audio. Therefore one might conclude that it makes no sense for manufacturers to be developing systems that have a higher audio quality than a CD.

Another argument is that it is desirable that we return to the old hierarchy where the professionals had access to a much higher quality of recording and reproduction device than the consumers. Back in the old days, we had 1/2-inch 30-ips reel-reel analog tape in the studios and 33-1/3 RPM vinyl LP's in the consumers' homes. This meant that the audio quality of the studio equipment was much, much higher than that in the homes, and that we in the studios could get away with murder. You see, if the resolution of your playback system is so bad that you can't hear the errors caused by the recording, editing and mastering system then the recording, editing and mastering engineers can hear things that the consumer can't – an excellent situation for quality control. The introduction of digital media to the professional and consumer markets meant that you could suddenly hear at home *exactly* what the professionals heard. And, if you spent a little more than they did on a DAC, amplifier and loudspeaker, you could hear things that the pro's couldn't... a very bad situation for quality control...

To quote a friend of mine, I stand very firmly on both sides of the fence – I can see that there might be some very good reasons for going to high-resolution digital audio, but I also think that the biggest reason is probably

marketing. I'll try to write this section in an unbiased fashion, but you'd better keep my personal cynicism in mind as you read...

7.9.1 “What is “High-Resolution Audio?”

As it became possible, in terms of processing power, storage and transmission bandwidth, people started demanding more from their digital audio. As we've seen, there are two simple ways to describe the resolution (if not the quality level) of digital audio – the sampling rate and the word length. In order to make digital audio systems that were better (or at least appeared on paper to be better) these two numbers had to get bigger. As ADC and DAC design improved, it became easier to increase the word length to 24 bits. Twenty-four was chosen to be a good number since the AES/EBU protocol was designed to support a 24-bit word length and since 141 dB was decided to be an adequate signal to noise ratio. In order to increase the sampling rate, we just started multiplying by 2, starting with the two standard sampling rates. We now have 88.1, 96, 176.4 and 192 kHz, among others.

When the folks that came up with DVD-Audio sat down to decide on the specifications for the media, they made it possible for the disc to support 24 bits and the various sampling rates listed above. (However, people were using higher sampling rates before DVD-Audio came around – it was just harder to get it into people's homes.) The term “high resolution” audio was already being thrown around by the time of the introduction of DVD-Audio to describe sampling rates higher than 44.1 or 48 kHz and word lengths longer than 16 bits.

One of the problems with DVD-Audio is a marketing problem – it is not backwards compatible with the CD format. Although a DVD-Audio player can play a CD, a CD player cannot play a DVD-Audio. So, the folks at Sony and Philips saw a marketing niche. Instead of supporting the DVD-Audio format, they decided to come up with a competing format called Super Audio Compact Disc or SACD. As is described below, this format is based on Delta-Sigma conversion at a very high sampling rate of approximately 2.83 MHz. Through some smart usage of the two layers on the disc, the format was designed with a CD-compatible layer so that a hybrid disc could be manufactured – one that could play on an old-fashioned CD player.

We'll look at the specifics of the formats in later sections.

7.9.2 Wither High-Resolution audio?

It didn't take long for the fascination with compact discs and Digital Audio Tapes to wear off and for people to want more. There were many complaints in the early 80's about all of the problems with digital audio. Words like "harsh" and "brittle" were being thrown around a lot. There are a couple of reasonable explanations for this assessment. The early ones suggested that digital audio inherently produced symmetrical distortion (producing artifacts in odd harmonics) whereas analog produced asymmetrical distortion (therefore producing even harmonics). People say that even harmonic distortion sounds "better" than odd harmonic distortion, therefore analog is better than digital. This argument may not necessarily be the case, but it gave people something to talk about. (In fact, this argument is a throwback to the reasons for buying tubes (or valves, depending on which side of the Atlantic Ocean you're from) versus transistors.)

Some other explanations hinged on the general idea of distortion. Namely, that analog reproduction systems, particularly vinyl, distorted the recorded signal. Things like wow and flutter or analog tape compression were commonplace in the days of analog recording. When CD's hit the scene, these things disappeared (except maybe for tape compression from the original analog recordings...). It is possible that people developed a preference for such types of distortion on their audio, so when the distortion was removed, they didn't like it as much.

Another reason that might be attributable to the problems in early digital audio was the quality of the converters. Twenty years ago, it wasn't so easy to build a decent analog-to-digital converter with an accuracy equivalent to 16 bits. In fact, if you do a measurement of some of the DAT machines from the early 80's you'll find that the signal-to-noise ratio was equivalent to about 12 or 13 bits. On top of this, people really didn't know how to implement dither properly, so the noise floor was primarily distortion. This automatically puts you in the areas of harsh and brittle...

Yet another possible explanation for the problems in early digital audio lies in the anti-aliasing filters used in the DAC's. Nowadays, we use analog filters with gentle slopes feeding oversampled DAC's followed by digital filters before the digital signal is downsampled to the required resolution for storage or transmission. Early converters were slightly different because we didn't have enough computational power to run either the oversampled DAC or the digital filters, so the filters were implemented in analog circuitry. These analog filters had extremely high slopes, and therefore big phase shifts around the cutoff frequency, resulting in ringing. This ringing is still audible

in the recordings that were made in that era.

Today, many people still aren't happy with the "standard" sampling rates of 44.1 and 48 kHz, nor are they satisfied with a 16-bit word length, although to the best of my knowledge, there haven't been any trustworthy listening tests whose results have conclusively proven that going to higher resolutions produces audible results.

There are a number of suggestions that people have put forward regarding why higher resolutions are necessary – or at least desirable in digital audio. I've listed a couple of these reasons below with some explanations as to why they may or may not be worth listening to.

Extended frequency range

This is one of the first arguments you'll hear for higher sampling rates. Many people claim that there are multiple benefits to increasing the upper cutoff frequency of our recording systems, requiring higher sampling rates. Remember that the original sampling rates were chosen to be adequate to record audio content up to and including 20 kHz. This number is ubiquitous in our literature regarding the upper limits of human hearing – every student that learns anything about psychoacoustics starts off on day one learning that audible sound ranges from 20 Hz to 20 kHz, as if we had two brick wall filters in our ears.

One group makes its claims on frequency response measurements of the outputs of instruments. (REFERENCE TO BOYK ARTICLE) The logic is that since the instruments (such as a trumpet with a harmon mute as a notable example) produce relatively high-level content above 20 kHz, our recording systems must capture this content. Therefore, we need higher sampling rates. One can show that humans are unable to hear sine tones of a reasonable level with frequencies above 20 kHz. However, it is possible that complex tones with harmonic content above 20 kHz produce resultant tones that may or may not be audible. It could be argued that these resultant tones are not audible with real sources in real life, but would be audible but undesirable in loudspeakers resulting from intermodulation distortion (described in Section ??). (If you would like to test yourself in this area, do an experiment where you're asked to tell the difference between a square wave and a sine wave, both with a fundamental frequency of 7 kHz. If you can hear the difference between these two (make sure that they're the same level!) then these people have a point. You see, the only difference between a sine wave a square wave is the energy in the odd harmonics above the fundamental in the square wave. Since the first odd harmonic above the

fundamental is 3 times the fundamental, then all of the differences between the two tones at 7 kHz will be content at 21 kHz and higher. In case you're interested, I tried this experiment and none of my subjects (including me) were able to tell the difference with any reliability.)

Another group relies on the idea that our common understanding of human limits of hearing is incorrect. For many years, many people have argued that our hearing does not stop at 20 kHz, regardless of what the tests and textbooks tell us. These folks say that we are actually able to perceive spectral content above 20 kHz in one way or another. When it proved to be impossible to get people to identify such things in standard listening tests (i.e. can you hear the difference between two sounds, one band-limited and one not) people resorted to looking at EKG's to see if high-frequency content changed Alpha-waves (CHECK DETAILS – REFERENCE TO YAMAMOTO ARTICLE).

Passband ripple

Once upon a time, there was a voice of sanity in the midst of the noise. Listserves and newgroups on Usenet would be filled with people spouting opinions on why bigger was better when it came to numbers describing digital audio. All sorts of strange ideas were (are) put forward by people who don't know the quote by George Eliot – “Blessed is the man who, having nothing to say, abstains from giving in words evidence to that fact” or a similar piece of advice from Abraham Lincoln – “Better to be thought a fool than to open your mouth and remove all doubt.” That lonely voice belonged to a man named Julian Dunn. Mr. Dunn wrote a paper that suggested that there was a very good reason why higher sampling rate *may* result in better sounding audio even if you can't hear above 20 kHz. He showed that the antialiasing filters used within ADC's do not have a flat frequency response in their passband. And, not only was their frequency response not flat, but they typically have a periodic ripple in the frequency domain. Of course, there's a catch – the ripple that we're talking about is on the order of 0.1 dB peak-to-peak, so we're not talking about a big problem here...

The interesting thing is that this frequency response irregularity can be reduced by increasing your sampling rate and reducing the slope of the antialiasing filters. Therefore, it's possible that higher sampling rates sound better because of reduced artifacts caused by the filters.

Dunn also noted that, if you're smart, you can design your reconstruction filter in your DAC to have the same ripple with the opposite phase (in the

frequency domain), thus canceling the effects of both filters and producing a perfectly flat response of the total system. Of course, this would mean that all manufacturers of ADC's and DAC's would have to use the same filters and that would, in turn mean that no converter would sound better than another which would screw up the pricing structure of that market... So most people that make converters (especially expensive ones) probably think that this is a bad idea.

You can download a copy of this paper from the web at www.nanophon.com.

The myth of temporal resolution

When students are learning PCM digital audio for the very first time, an analogy is made to film. Film takes 24 snapshots per second of the action and plays them back at the same speed to resemble motion. Similarly, PCM digital audio takes a much larger number of "snapshots" of the voltage level of the signal each second and plays them back later in the same order to reconstruct the signal.

This is a pretty good analogy (which is why I used it as well in Section ??). However, it causes a misconception later on. If we stick with thinking about film for a moment, we have a limited temporal resolution. For example, if an event happens, lasting for a very short period of time and occurring between two frames of the film, then the event will never be recorded on the film. Let's say that you're making a movie of a wedding and somebody snaps a photograph with a flash. Let's also pretend that the flash lasts only for 1/60th of a second (faster than your 1/24th of a second frame rate) and that flash happens between frames of your movie. When you play back the movie, you'll never see the flash because it happened at a moment in time that your film didn't record.

There are a large number of people in the audio world who are under the misconception that this also holds true in digital audio. The belief is that an event that happens between two samples will not be recorded. Consequently, things like precise time of arrival at a microphone or multiple reflections arriving at very closely spaced intervals will not be recorded because their arrival doesn't correspond to a sample time. Or, another belief is that the time of arrival is quantized in time to the nearest sample, so differences in times of arrival of a wavefront at two microphones will be slightly altered, rounding off to the nearest sample.

I want to express this as explicitly as possible... **These beliefs are just plain wrong.** In fact, if these people were right, then there would be no such thing as an interpolated delay, and we know that those exist.

The sampling period has no relationship whatsoever with the ability of a digital recording system (either PCM or Delta-Sigma) to resolve the time of a signal. This is because the signal is filtered by the anti-aliasing filters. A discrete time audio system (i.e. digital) has an infinite resolution in the temporal domain if the signal is properly band-limited (REFERENCES TO GO HERE).

| Sampling Rate | Sampling Period |
|---------------|-----------------|
| 44.1 kHz | 22.7 μ sec |
| 48 kHz | 20.8 μ sec |
| 88.2 kHz | 11.3 μ sec |
| 96 kHz | 10.4 μ sec |
| 192 kHz | 5.3 μ sec |
| 384 kHz | 2.6 μ sec |
| 2.83 MHz | 0.35 μ sec |

Table 7.3: Sampling periods for various standard sampling rates. You will see these listed as the temporal resolution of the various sampling rates. This is incorrect. The temporal resolutions of all of these sampling rates and systems are the same – infinite.

7.9.3 Is it worth it?

Do not take everything I've said in this chapter as the definitive reference on the subject of high-resolution audio. Go read the papers by Julian Dunn and Malcom Hawkesford and a bunch of other people before you make up your mind on whether it's worth the disc space to do recordings at 96 kHz or higher. However, don't base your decisions on one demo from a marketing rep of a company who's trying to sell you a product. And don't get sucked in by the simple marketing ploy that "more" and "better" are equivalent terms.

7.9.4 Suggested Reading List

Malcolm Hawkesford
Julian Dunn

7.10 Data Compression and Reduction

NOT YET WRITTEN

7.10.1 History

NOT YET WRITTEN

7.10.2 ATRAC

NOT YET WRITTEN

7.10.3 PASC

NOT YET WRITTEN

7.10.4 Dolby

NOT YET WRITTEN

AC-1 & AC-2

NOT YET WRITTEN

AC-3 (aka Dolby Digital)

NOT YET WRITTEN

7.10.5 MPEG

NOT YET WRITTEN

Layer II

NOT YET WRITTEN

AAC

NOT YET WRITTEN

7.10.6 MLP

NOT YET WRITTEN

7.10.7 Suggested Reading List

7.11 Transmission Techniques for Internet Distribution

NOT YET WRITTEN

7.11.1 Suggested Reading List

7.12 CD – Compact Disc

Back in 1983, a new format for playing audio was introduced to the consumer market that represented a radical shift in the relationship between professional and consumer quality. The advertisements read “Perfect sound... forever...” Of course, we now know that most early digital recordings were far from “perfect” (whatever that word entails) and that compact discs don’t last forever.

The CD format was developed as a cooperative effort between Sony and Philips. It was intended from the very beginning to be a replacement format for vinyl LP’s – a position that it eventually secured in all but two markets (being the DJ and the hard-core audiophile markets).

In initially developing the CD format, Sony and Phillips had to make many decisions regarding the standard. Three of the most basic requirements that they set were the sampling rate, the bit depth and the total maximum playing time.

7.12.1 Sampling rate

Nowadays, when we want to do a digital recording, we either show up with a DAT machine, or a laptop with an audio input, writing data straight to our hard drives. Back in the 1980’s however, DAT hadn’t been invented yet, and hard discs just weren’t fast enough to cope with the data rate required by digital audio. The only widely-available format that could be used to record the necessary bandwidth was a video recorder. Consequently, machines were built that took a two-channel audio input, converted that to digital and sent a video output designed to be recorded on either a Beta tape if you didn’t have a lot of money (yes... Beta... if you’re too young, you may not have even heard of this format. If you’re at least as old as me, you probably have some old recordings still lying around on Beta tapes...) or U-Matic if you did. (U-Matic is an old analog professional-format video tape that uses 3/4” tape.) The process of converting from digital audio to video was actually pretty simple, a 1 was a white spot and a 0 was a black spot. So, if you looked at your digital audio recording on a television, you’d see a bunch of black and white strips, looking mostly like noise (or snow as it’s usually called when you see noise on a TV).

Since the recording format (video tape) was already on the market, the conversion process was, in part, locked to that format. In addition, the manufacturers had to ensure that tapes could be shipped across the Atlantic Ocean and still be useable. This means that the sampling rate had to be

derived from, among other things, the frame rates of NTSC and PAL video.

To begin with, it was decided that the minimum sampling rate was 40 kHz to allow for a 20 kHz minimum Nyquist frequency. Remember that the audio samples were stored as black and white stripes in the video signal, so a number above 40 kHz had to be found that fit both formats nicely. NTSC video has 525 lines per frame (of which 490 are usable lines for recording signals) at a frame rate of 29.97 Hz. This can be further divided into 245 usable lines per field (there are 2 fields per frame) at a field rate of 59.95 Hz. If we put 3 audio samples on each line of video, then we arrive at the following equation [Watkinson, 1988]:

$$59.94 \text{ Hz} * 245 \text{ lines per field} * 3 \text{ samples per line} = 44.0559 \text{ Hz}$$

PAL is slightly different. Each frame has 625 lines (with 588 usable lines) at 25 Hz. This corresponds to 294 usable lines per field at a field rate of 50 Hz. Again, with 3 audio samples per line of video, we have the equation [Watkinson, 1988]:

$$50.00 * 294 \text{ lines per field} * 3 \text{ samples per line} = 44.1000 \text{ Hz}$$

These two resulting sampling rates were deemed to be close enough (only a 0.1% difference in sampling rate) to be compatible (this difference in sampling rate corresponds to a pitch shift of about 0.0175 of a semitone).

This is perfect, but we're forgetting one small thing... most people record in stereo. Therefore, the EIAJ format was developed from these equations, resulting in 6 samples per video line (3 for each channel).

There is one odd addition to the story. Technically speaking, the compact disc format really had no ties with video (back in 1983, you couldn't play video off a CD yet) but the equipment that was used for recording and mastering was video-based. Interestingly, NTSC professional video gear (the U-Matic format) can run at frame rate of 30 fps, and is not locked to the 29.97 of your television at home. Consequently, if you re-do the math with this frame rate, you'll find that the resulting sampling rate is exactly 44.1 kHz. Therefore, to ensure maximum compatibility and still keep a technically achievable sampling rate, 44.1 kHz was chosen to be the standard.

7.12.2 Word length

The next question was that of word length. How many bits are enough to adequately reproduce an audio signal? And, possibly more importantly, what is technically feasible to implement, both in terms of storing the data as well as converting it from and back to the analog domain. We have already seen in Section ?? that there is a direct connection between the number of bits used to describe the signal level and the signal-to-noise ratio incurred

in the conversion process.

NOT YET WRITTEN

7.12.3 Storage capacity (recording time)

NOT YET WRITTEN

7.12.4 Physical construction

A compact disc is a disc made primarily of a polycarbonate plastic made by Bayer and called Makrolon [Watkinson, 1988]. This is supplied to the CD manufacturer as small beads shipped in large bags. The disc has a total outer diameter of 120 mm with a 15 mm hole, and the thickness is 1.2 mm.

The data on the disc plays from the inside to the outside and has a constant bit rate. As a result, as the laser tracks closer and closer to the edge of the disc, the rotational speed must be reduced to ensure that the disc-to-laser speed remains constant at between 1.2 and 1.4 m/s [Watkinson, 1988]. At the very start of the CD, the disc spins at about 230 rpm and at the very end of a long disc, it's spinning at about 100 rpm. On a full CD, the total length of the spiral path read by the laser is about 5.7 km [Watkinson, 1988].

ADD MORE HERE ABOUT PITS AND LANDS, OR BUMPS...

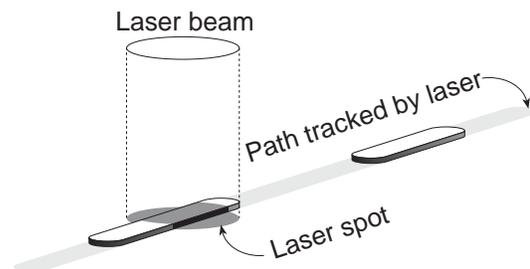


Figure 7.43: A single 14-bit channel word represented as “bumps” on the CD. Notice that the spot formed by the laser beam is more than twice as wide as the width of the bump. This is intentional. The laser spot diameter is approximately $1.2 \mu\text{m}$. The bump width is $0.5 \mu\text{m}$, and the bump height is $0.13 \mu\text{m}$. The track pitch (the distance between this row of bumps and an adjacent row) is $1.6 \mu\text{m}$ [Watkinson, 1988]. Also remember that, relative to most CD players, this drawing is upside down – typically the laser hits the CD from below.

The wavelength λ of the laser light is $0.5 \mu\text{m}$. The bump height is $0.13 \mu\text{m}$, corresponding to approximately $\frac{\lambda}{4}$ for the laser. As a result, when the laser spot is hitting a bump on the disc, the reflections from both the bump and the adjacent lands (remember that the laser spot is wider than the

bump) results in destructive interference and therefore cancellation of the reflection. Therefore, seen from the point of view of the pickup, there is no reflection from a bump.

7.12.5 Eight-to-Fourteen Modulation (EFM)

There is a small problem in this system of representing the audio data with bumps on the CD. Let's think about a worst-case situation where, for some reason, the audio data is nothing but a string of alternating 1's and 0's. If we were representing each 0 with a bump and each 1 with a not-bump, then, even if we didn't have additional information to put on the disc (and we do...), we'd be looking at 1,411,200 bump-to-not-bump transitions per second or a frequency of about 1.4 MHz (44.1 kHz x 16 bits per sample x 2 channels). Unfortunately, it's not possible for the optical system used in CD's to respond that quickly – there's a cutoff frequency for the data rate imposed by a couple of things (such as “the numerical aperture of the optics and the track velocity” [Watkinson, 1988]). In addition, we want to keep our data rate below this cutoff to gain some immunity from problems caused by “disc warps and focus errors” [Watkinson, 1988].

So, one goal is to somehow magically reduce the data rate from the disc as low as possible. However, this will cause us another problem. The data rate of the bits going to the DAC is still just over 705 kHz (44.1 kHz x 16 bits per sample). This clock rate has to be locked to, or derived from the data coming off the disc itself, which we have already established, can't go that fast... If the data coming off the disc is too slow, then we'll have problems locking our slow data rate from the disc with the fast data rate of information getting to the DAC. If this doesn't make sense at the moment, stick with me for a little while and things might clear up.

So, we know that we need to reduce the amount of data written to (and read from) the disc without losing any audio information. However, we also know that we can't reduce the data rate too much, or we'll introduce jitter at best and locking problems at worst. The solution to this problem lies in a little piece of magic called *eight to fourteen modulation* or *EFM*.

To begin with, eight to fourteen modulation is based on a lookup table. A small portion of this table is shown in Table 7.4.

We begin by taking each 16-bit sample that we want to put on the disc and slice it into two 8-bit bytes. We then go to the table and look up the 8-bit value in the middle column of Table 7.4. The right-hand column shows a different number containing 14 bits, so we write this down.

For example, let's say that we want to put the number 01101001 on the

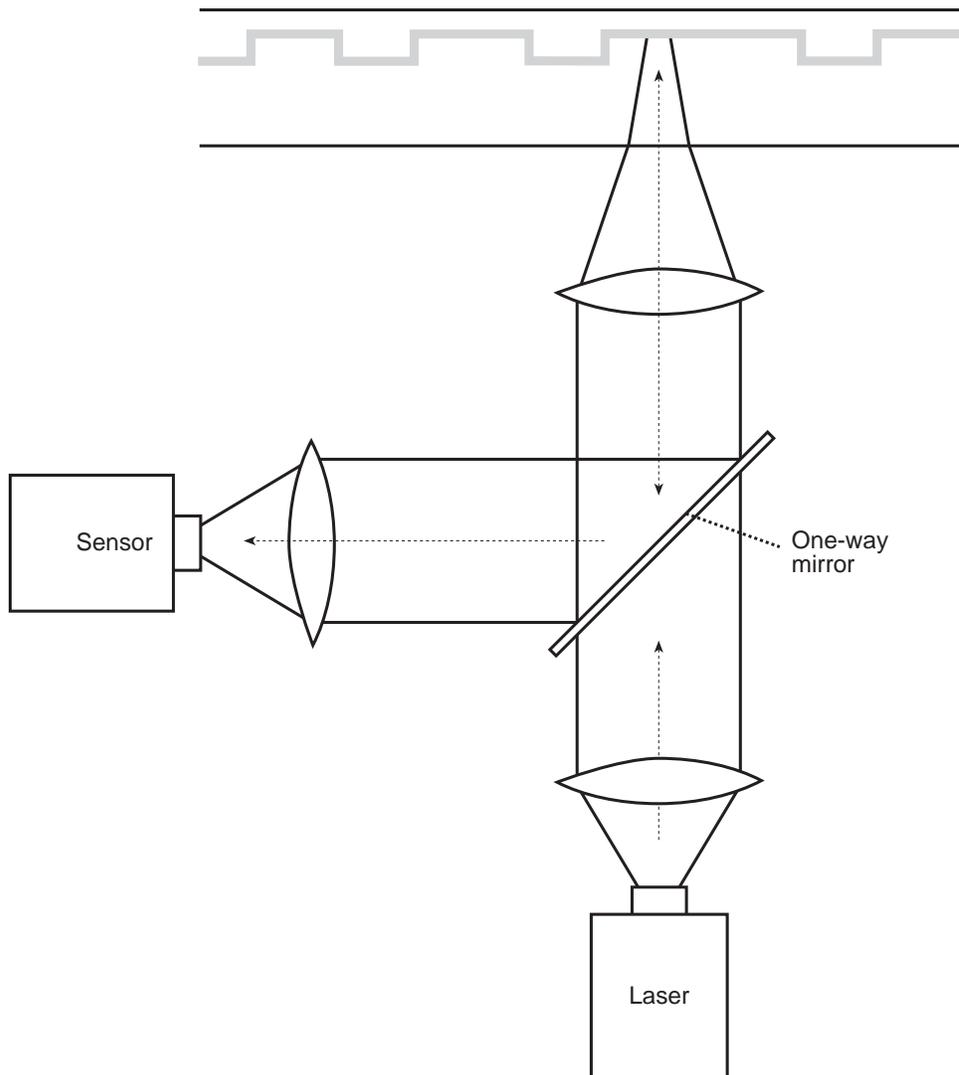


Figure 7.44: Simplified diagram showing how the laser is reflected off the CD. The laser shines through a semi-reflective mirror, bounces off the CD, reflects off the mirror and arrives at the sensor [Watkinson, 1988].

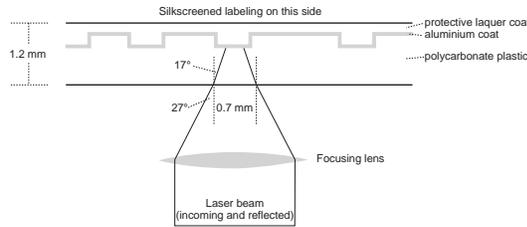


Figure 7.45: Cross section of a CD showing the polycarbonate base, the reflective aluminum coating as well as the protective lacquer coating. The laser has a diameter of 0.7 mm when it hits the surface of the disc, therefore giving it a reasonable immunity to dirt and scratches [Watkinson, 1988].

disc. We go to the table, look up that number and we get the corresponding number 10000001000010.

| Data value
(decimal) | Data bits
(binary) | Channel bits
(binary) |
|-------------------------|-----------------------|--------------------------|
| 101 | 01100101 | 00000000100010 |
| 102 | 01100110 | 01000000100100 |
| 103 | 01100111 | 00100100100010 |
| 104 | 01101000 | 01001001000010 |
| 105 | 01101001 | 10000001000010 |
| 106 | 01101010 | 10010001000010 |
| 107 | 01101011 | 10001001000010 |
| 108 | 01101100 | 01000001000010 |
| 109 | 01101101 | 00000001000010 |
| 110 | 01101110 | 00010001000010 |

Table 7.4: A small portion of the table of equivalents in EFM. The value that we are trying to put on the disc is the 8-bit word in the middle column. The actual word printed on the disc is the 14-bit word in the right column [Watkinson, 1988].

Okay, so right about now, you should be saying “I thought that we wanted to *reduce* the amount of data... not increase it from 8 up to 14 bits...” We’re getting there.

What we now do is to take our 14-bit word 10000001000010 and draw an irregular pulse wave where we have a transition (from high to low or low to high) for every “1” in the word. This is illustrated in Figure 7.46. Compare the examples in this figure with the corresponding 14-bit values in Table 7.4.

Okay, we’re not out of the woods yet. We can still have a problem. What

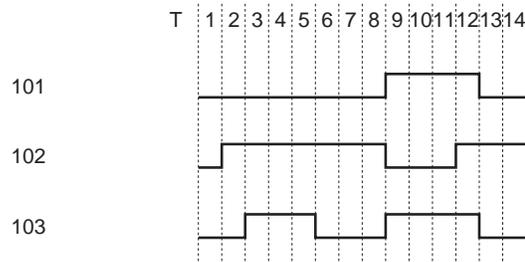


Figure 7.46: Three examples of the representation of the data word from Table 7.4 being represented as pits and lands using EFM.

if the 14-bit word has a string of 1's in it? Aren't we still stuck with the original problem, only worse? Well, yes. But, the clever people that came up with this idea were very careful about choosing their 14-bit representative words. They made sure that there are no 14-bit values with 1's separated be less than two 0's. Huh? For example, non of the 14-bit words in the lookup table contain the codes 11 or 101 anywhere. Take a look at the small example in Table 7.4. You won't find any 1's that close together - minimum separation of two 0's at any time. In real textbooks they talk about a minimum period between transitions of $3T$ where T is the period of 1 bit in the 14-bit word. (This period T is 231.4 ns, corresponding to a data rate of 4.3218 MHz [Watkinson, 1988] – but remember, that's the data rate of the 14-bit word, not the signal stamped on the disc.) This guarantees that the transition rate on the disc cannot exceed 720 kHz, which is acceptably high.

So, that looks after the highest frequency, but what about the lowest possible frequency of bump transitions? This is looked after by setting a maximum period between transitions of $11T$, therefore there are no 14-bit words with more than ten 0's between 1's. This sets our minimum transition frequency to 196 kHz which is acceptably low.

Let's talk a little more about why we have this low-frequency limitation on the data rate. Remember that when we talk about a “period between transitions of $11T$ ” we're directly talking about the length of the bump (or not-bump) on the disc surface. We're already seen that the rotational speed of the disc is constantly changing as the laser gets further and further away from the centre. This speed change is done in order to keep the data rate constant – the physical length of a bump of $9T$ at the beginning of the disc is the same as that of a bump of $9T$ at the end of the disc. The problem is, if you're the sensor responsible for converting bump length into a number,

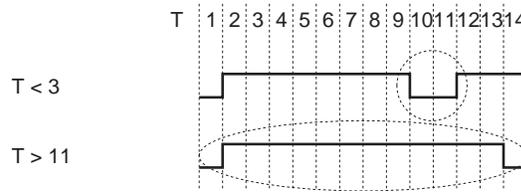


Figure 7.47: Two examples of invalid codes. The errors are circled. The top code cannot be used since one of the pits is shorter than $3T$. The bottom code is invalid because the land is longer than $11T$.

you *really* need to know how to measure the bump length. The longer the bump, the more difficult it is to determine the length, because it's a longer time since the last transition.

To get an idea of what this would be like, stand next to a train track and watch a slowing train as it goes by. Count the cars, and get used to the speed at which they're going by, and how much they're slowing down. Then close your eyes and keep counting the cars. If you had to count for 3 cars, you'd probably be pretty close to being right in synch with the train. If you had to count 9 cars, you'd probably be wrong, or at least losing synchronization with the train. This is exactly the same problem that the laser sensor has in estimating pit lengths. The longer the pit, the more likely the error, so we keep a maximum of $11T$ to minimize the likelihood of errors.

7.12.6 Suggested Reading List

7.13 DVD-Audio

NOT YET WRITTEN

7.13.1 Suggested Reading List

7.14 SACD and DSD

Super Audio Compact Disc and Direct Stream Digital
NOT YET WRITTEN

7.14.1 Suggested Reading List

7.15 Hard Disk Recording

7.15.1 Bandwidth and disk space

One problem that digital audio introduces is the issue of bandwidth and disk space. If you want to record something to tape and you know that it's going to last about an hour and a half, then you go and buy a two-hour tape. If you're recording to a hard drive, and you have 700 MB available, how much time is that?

In order to calculate this, you'll need to consider the bandwidth of the signal you're going to send. Consider that, for each channel of audio, you're going to some number of samples, depending on the sampling rate, and each of those samples is going to have some number of bits, depending on the word length.

Let's use the example of CD to make a calculation. CD uses a sampling rate of 44.1 kHz (44100 samples per second) and 16-bit word lengths for two channels of audio. Therefore, each second, in each of two channels, 44100, 16-bit numbers come through the system. So:

2 channels
* 44100 samples per second
* 16 bits per sample
= 1,411,200 bits per second

What does this mean in terms of disc space? In order to calculate this, we just have to convert the number of bits into the typical storage unit for computers – a byte (eight bits).

1,411,200 bits per second
/ 8 bits per byte
= 176,400 bytes per second

divide that by 1024 (bytes per kilobyte) and we get the value in kilobytes per second, resulting in a value of 172.27 kB per second.

From there we can calculate into what are probably more meaningful terms:

172.27 kB per second
* 60 seconds per minute
= 10,335.94 kilobyte per minute
/ 1024 kilobyte per megabyte
= 10.1 MB per minute.

So, when you're storing uncompressed, CD-quality audio on your hard drive, it's occupying a little more than 10 MB per minute of space, so 700 MB of free space is about 70 minutes of music.

7.15.2 Suggested Reading List

7.16 Digital Audio File Formats

NOT YET WRITTEN

7.16.1 AIFF

NOT YET WRITTEN

7.16.2 WAV

NOT YET WRITTEN

7.16.3 SDII

NOT YET WRITTEN

7.16.4 mu-law

NOT YET WRITTEN

7.16.5 Suggested Reading List

Chapter 8

Digital Signal Processing

8.1 Introduction to DSP

One of my first professors in Digital Signal Processing (better known as DSP) summarized the entire field using the simple flowchart shown in Figure 8.1.

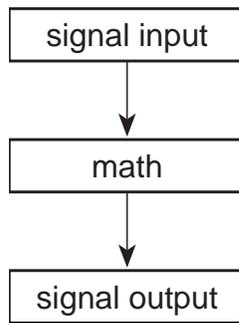


Figure 8.1: DSP in a very small nutshell. The “math” that is done is the processing done to the digital signal. This could be a delay, an EQ or a reverb unit – either way, it’s just math.

Essentially, that’s just about it. The idea is that you have some signal that you have converted from analog to a discrete representation using the procedure we saw in Section 7.1. You want to change this signal somehow – this could mean just about anything... you might want to delay it, filter it, mix it with another signal, compress it, make reverberation out of it – anything... Everything that you want to do to that signal means that you are going to take a long string of numbers and turn them into a long string of different numbers. This is processing of the digital signal or, *Digital Signal Processing* – it’s the math that is applied to the signal to turn it into the

other signal that you're looking for.

8.1.1 Block Diagrams and Notation

Usually, a block diagram of a DSP algorithm won't run from top to bottom as is shown in Figure 8.1. Just like analog circuit diagrams, DSP diagrams typically run from left to right, with other directions being used either when necessary (to make a feedback loop, for example) or for different signals like control signals as in the case of analog VCA's (See section 5.2.2).

Let's start by looking at a block diagram of a simple DSP algorithm.

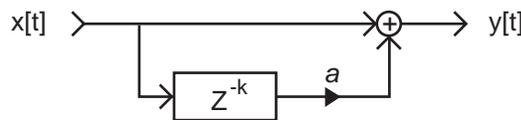


Figure 8.2: Basic block diagram of a comb filter implemented in the digital domain.

Let's look at all of the components of Figure 8.2 to see what they mean. Looking from left to right you can see the following:

- x_t This actually has a couple of things in it that we have to talk about. Firstly, there's the x which tells us that this is an input signal with a value of x (x is usually used to indicate that it's an input, y for output). The subscript t is the sample number of the signal – in other words, it's the time that the signal's value is x .
- z^{-k} This is an indication of a delay of k samples. We know that t is a sample number, and we're just subtracting some number of samples to that (in fact, we're subtracting k samples). Subtracting k from the sample number means that we're getting earlier in time, so $[t-k]$ means that we're delaying the signal at time $[t]$ by k samples. We use a delay to hear something now that actually came in earlier. So, this block is just a delay unit with a delay time of k samples. Since the delay time is expressed in a number of samples, we call it an *integer delay* to distinguish it from other types that we'll see later. We'll also talk later (in Section 8.7) about why it's written using a z .
- a There is a little triangle with an a over it in the diagram. This is a gain function where a is the multiplier. Everything that passes

through that little triangle (which means, in this case, everything that comes out of the delay box) gets multiplied by a .

- The circle with the $+$ sign in it indicates that the two signals coming in from the left and below are added together and sent out the right. Sometimes you will also see this as a circle with a Σ in it instead. This is just another way of saying the same thing.
- y_t As you can probably guess, this is the value of the sample at time t at the output (hence the y) of the system.

There is another way of expressing this block diagram using math. Equation 8.1 shows exactly the same information without doing any drawing.

$$y_t = x_t + ax_{t-k} \quad (8.1)$$

8.1.2 Normalized Frequency

If you start reading books about DSP, you'll notice that they use a strange way of labeling the frequency of a signal. Instead of seeing graphs with frequency ranges of 20 Hz to 20 kHz, you'll usually see something called a *normalized frequency* ranging from 0 to 0.5 and no unit attached (it's not 0 to 0.5 Hz).

What does this mean? Well, think about a digital audio signal. If we record a signal with a 48 kHz sample rate and play it back with a 48 kHz sampling rate, then the frequency of the signal that went in is the same as the frequency of the signal that comes out. However, if we play back the signal with a 24 kHz sampling rate, then the frequency of the signal that comes out will be one half that of the recorded signal. The ratio of the input frequency to the output frequency of the signal is the same as the ratio of the recording sampling rate to the playback sampling rate. This probably doesn't come as a surprise.

What happens if you really don't know the sampling rate? You just have a bunch of samples that represent a time-varying signal. You know that the sampling rate is constant, you just don't know its frequency. This is the what a DSP processor "knows." So, all you know is what the frequency of the signal is *relative to* the sampling rate. If the samples all have the same value, then the signal must have a frequency of 0 (the DSP assumes that you've done your anti-aliasing properly...). If the signal bounces back between positive and negative on every sample, then its frequency must be the Nyquist Frequency – one half of the sampling rate.

So, according to the DSP processor, where the sampling rate has a frequency of “1”, the signal will have a frequency that can range from 0 to 0.5. This is called the normalized frequency of the signal.

There’s an important thing to remember here. Usually people use the word “normalize” to mean that something is changed (you normalize a mixing console by returning all the knobs to a default setting, for example). With normalized frequency, nothing is changed – it’s just a way of describing the frequency of the signal.

PUT A SHORT DISCUSSION HERE REGARDING THE USE OF ωt

8.1.3 Suggested Reading List

8.2 FFT's, DFT's and the Relationship Between the Time and Frequency Domains

NOTE TO SELF: CHECK ALL THE COMPLEX NUMBERS IN THIS SECTION

Note:

If you're unhappy with the concepts of real and imaginary components in a signal, and how they're represented using complex numbers, you'd better go back and read Chapter 1.5.

8.2.1 Fourier in a Nutshell

We saw briefly in Section 3.1.16 that there is a direct relationship between the time and frequency domains for a given signal. In fact, if we know everything there is to know about the frequency domain, we already know its shape in the time domain and vice versa. Now let's look at that relationship a little more closely.

Take a sine wave like the top plot shown in Figure 8.3 and add it to another sine wave with one third the amplitude and three times the frequency (the middle plot, also in Figure 8.3). The result will be shaped like the bottom plot in Figure 8.3.

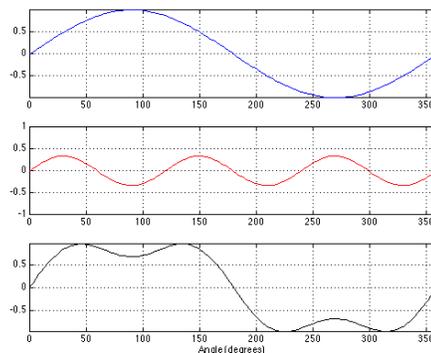


Figure 8.3: The bottom plot is the sum of the top two plots.

If we continue with this series, adding a sine wave at 5 times the frequency and 1/5th the amplitude, 7 times the frequency and 1/7th the am-

plitude and so on up to 31 times the frequency (and $1/31$ the amplitude) you get a waveform that looks like Figure 8.4.

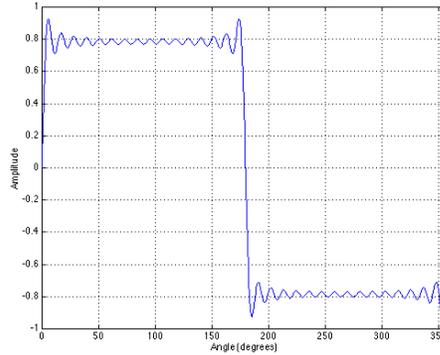


Figure 8.4: The sum of odd harmonics of a sine wave where the amplitude of each is $1/n$ where ‘ n ’ is the harmonic number up to the 31st harmonic.

As is beginning to become apparent, the result starts to approach a square wave. In fact, if we kept going with the series up to ∞ Hz, the sum would be a perfect square wave.

The other issue to consider is the relative phase of the harmonics. For example, if we take the same sinusoids as are shown in Figures 8.3 and 8.4 and offset each by 90° before summing, we get a very different result as can be seen in Figures 8.5 and 8.6.

This isn’t a new idea – in fact, it was originally suggested by a guy named Jean Baptiste Fourier that any waveform (such as a square wave) can be considered as a sum of a number of sinusoidal waves with different frequencies and phase relationships. This means that we can create a waveform out of sinusoidal waves like we did in Figures 8.3 and 8.4, but it also means that we can take any waveform and look at its individual components. This concept will come as no surprise to musicians, who call these components “harmonics” – the timbres of a violin and a viola playing the same note are different because of the different relationships of the harmonics in their *spectra* (one *spectrum*, two *spectra*)

Using a lot of math – and calculus, it is possible to calculate what is known as a *Fourier Transform* of a signal to find out what its sinusoidal components are. We won’t do that. There is also a way to do this quickly – called a *Fast Fourier Transform* or *FFT*. We won’t do that either. The FFT is used for signals that are continuous in time. As we already know, this is

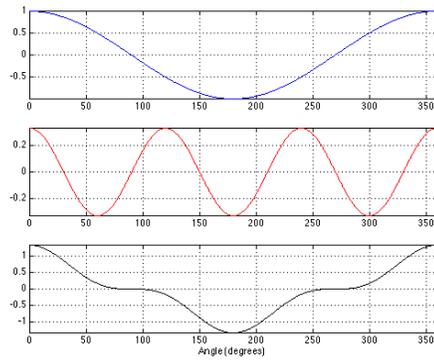


Figure 8.5: The bottom plot is the sum of the top two plots. Note that the frequencies and amplitudes of the two components are identical to those shown in Figure 1, however, the result of adding the two produces a very different waveform.

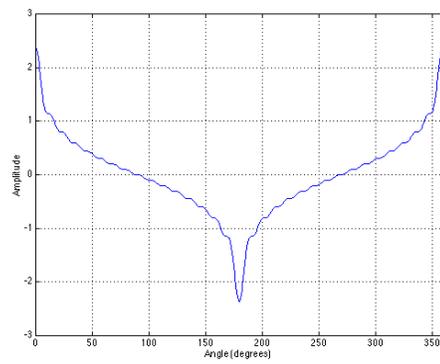


Figure 8.6: The sum of odd harmonics of a sine wave where the amplitude of each is $1/n$ where 'n' is the harmonic number up to the 31st harmonic.

not the case for digital audio signals, where time and amplitude are divided into discrete divisions. Consequently, in digital audio we use a variation on the FFT called a *Discrete Fourier Transform* or *DFT*. This is what we'll look at. One thing to note is that most people in the digital world use the term FFT when they really mean DFT – in fact, you'll rarely hear someone talk about DFT's – even though that's what they're doing. Just remember, if you're doing what you think is an FFT to a digital signal, you're really doing a DFT.

8.2.2 Discrete Fourier Transforms

Let's take a digital audio signal and view just a portion of it, shown in Figure 8.7. We'll call that portion the *window* because the audio signal extends on both sides of it and it's as if we're looking at a portion of it through a window. For the purposes of this discussion, the window length, usually expressed in *points* is only 1024 samples long (therefore, in this case we have a 1024-point window).

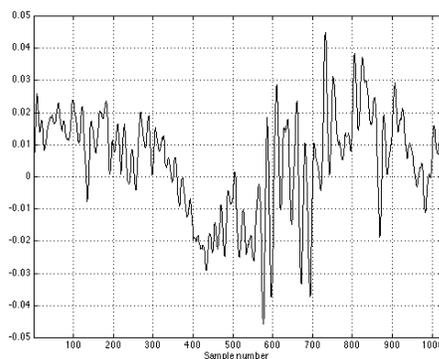


Figure 8.7: A plot of a digital audio signal 1024 samples long.

As we already know, this signal is actually a string of numbers, one for each sample. To find the amount of 0 Hz in this signal, all we need to do is to add the values of all the individual samples. Technically speaking, we really should make it an average and divide the result by the number of samples, but we're in a rush, so don't bother. If the wave is perfectly symmetrical around the 0 line (like a sinusoidal wave, for instance...) then the total will be 0 because all of the positive values will cancel out all of the negative values. If there's a DC component in the signal, then adding the

values of all the samples will show us this total level.

So, if we add up all the values of each of the 1024 samples in the signal shown, we get the number 4.4308. This is a measure of the level of the 0 Hz component (in electrical engineering jargon, the *DC offset*) in the signal. Therefore, we can say that the *bin* representing the 0 Hz component has a value of 4.4308. Note that the particular value is going to depend on your signal, but I'm giving numbers like "4.4308" just as an example.

We know that the window length of the audio signal is, in our case, 1024 samples long. Let's create a cosine wave that's the same length, counting from 0 to 2π (or, in other words, 0 to 360°). (Technically speaking, its period is 1025 samples, and we're cutting off the last one...) Now, take the signal and, sample by sample, multiply it by its corresponding sample in the cosine wave as shown in Figure 8.8.

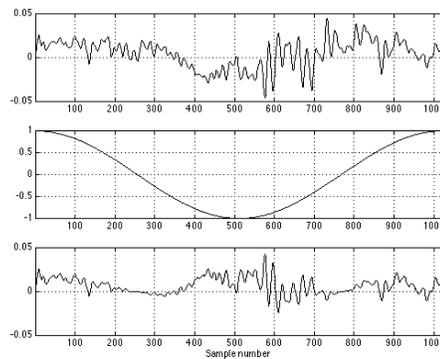


Figure 8.8: The top plot is the original signal. The middle plot is one period of a cosine wave (minus the last sample). The bottom plot is the result when we multiply the top two, sample by sample.

Now, take the list of numbers that you've just created and add them all together (for this particular example, the result happens to be 6.8949). This is the "real" component at a frequency whose period is the length of the audio window. (In our case, the window length is 1024 samples, so the period for this component is $\frac{f_s}{1024}$ where f_s is the sampling rate.)

Repeat the process, but use a sine wave instead of a cosine and you get the imaginary component for the same frequency, shown in Figure 8.9.

Take that list of numbers and add them and the result is the "imaginary" component at a frequency whose period is the length of the sample (for this particular example, the result happens to be 0.9981).

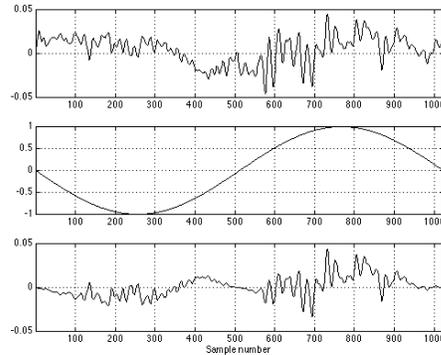


Figure 8.9: The top plot is the original signal. The middle plot is one period of an inverted sine wave (minus the last sample). The bottom plot is the result when we multiply the top two, sample by sample.

Let's assume that the sampling rate is 44.1 kHz, this means that our bin representing the frequency of 43.0664 Hz (remember, $44100/1024$) contains the complex value $6.8949 + 0.9981i$. We'll see what we can do with this in a moment.

Now, repeat the same procedure using the next harmonic of the cosine wave, shown in Figure 8.10.

Take that list of numbers and add them and the result is the “real” component at a frequency whose period is the one half the length of the audio window (for this particular example, the result happens to be -4.1572).

And again, we repeat the procedure with the next harmonic of the sine wave, shown in Figure 8.11.

Take that list of numbers and add them and the result is the “imaginary” component at a frequency whose period is the length of the sample (for this particular example, the result happens to be -1.0118).

If you want to calculate the frequency of this bin, it's 2 times the frequency of the last bin (because the frequency of the cosine and sine waves are two times the fundamental). Therefore it's $2 \left(\frac{f_s}{1024} \right)$, or, in this example, $2 * (44100/1024) = 86.1328$ Hz.

Now we have the 86.1328 Hz bin containing the complex number $-4.1572 - 1.0118i$.

This procedure is repeated, using each harmonic of the cosine and sine until you get up to a frequency where you have 1024 periods of the cosine and sine in the window. (Actually, you just go up to the frequency where

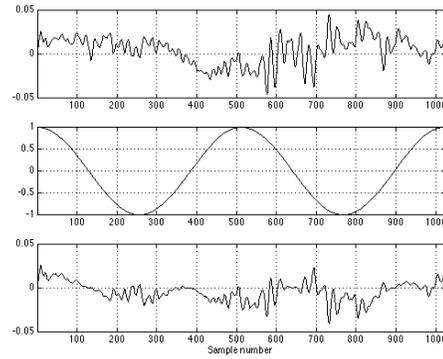


Figure 8.10: The top plot is the original signal. The middle plot is two periods of a cosine wave (minus the last sample). The bottom plot is the result when we multiply the top two, sample by sample.

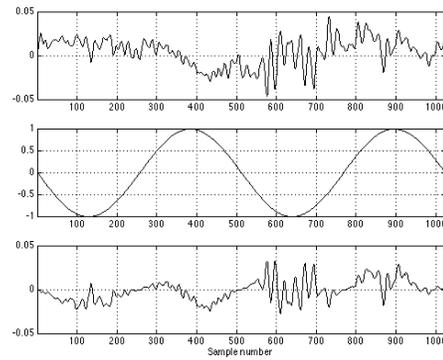


Figure 8.11: The top plot is the original signal. The middle plot is two periods of an inverted sine wave (minus the last sample). The bottom plot is the result when we multiply the top two, sample by sample.

the number of periods in the cosine or the sine is equal to the length of the window in samples.)

Using these numbers, we can create Table 8.1.

| Bin Number | Frequency (Hz) | Real component | Imaginary component |
|------------|----------------|----------------|---------------------|
| 0 | 0 Hz | 4.4308 | N/A |
| 1 | 43.0664 Hz | 6.8949 | 0.9981i |
| 2 | 86.1328 Hz | -4.1572 | -1.0118i |

Table 8.1: The results of our multiplication and averaging described above for the first three bins. These are just the first three bins of 1024 (bin numbers 0 to 1023).

How can we use this information? Well, remember from the chapter on complex numbers that the magnitude of a signal – essentially, the amplitude of the signal that we see – is calculated from the real and imaginary components using the Pythagorean theorem. Therefore, in the example above, the *magnitude response* can be calculated by taking the square root of the sum of the squares of the real and imaginary results of the DFT. Huh? Check out Table 8.2.

| Bin Number | Frequency (Hz) | $\sqrt{real^2 + imag^2}$ | Magnitude |
|------------|----------------|--------------------------------|-----------|
| 0 | 0 Hz | $\sqrt{4.4308^2}$ | 4.4308 |
| 1 | 43.0664 Hz | $\sqrt{6.8949^2 + 0.9981^2}$ | 6.9668 |
| 2 | 86.1328 Hz | $\sqrt{-4.1572^2 + -1.0118^2}$ | 4.2786 |

Table 8.2: The magnitude of each bin, calculated using the data in Table 8.1

If we keep filling out this table up to the 1024th bin, and graphed the results of Magnitude vs. Bin Frequency we'd have what everyone calls the *Frequency Response* of the signal. This would tell us, frequency by frequency the amplitude relationship of the various harmonics in the signal. The one thing that it wouldn't tell us is what the phase relationship of the various harmonics are. How can we calculate that? Well, remember from the chapter on trigonometry that the relative levels of the real and imaginary components can be calculated using the phase and amplitude of a signal. Also, remember from the chapter on complex numbers that the phase of the signal can be calculated using the relative levels of the real and imaginary components using the equation:

$$\phi = \arctan\left(\frac{\textit{imaginary}}{\textit{real}}\right) \quad (8.2)$$

So, now we can create a table of phase relationships, bin by bin as shown in Table 8.3:

| Bin Number | Frequency (Hz) | $\arctan\left(\frac{\textit{imaginary}}{\textit{real}}\right)$ | Phase (degrees) |
|------------|----------------|--|-----------------|
| 0 | 0 Hz | $\arctan \frac{0}{4.4308}$ | 0° |
| 1 | 43.0664 Hz | $\arctan \frac{0.9981}{6.8949}$ | 8.2231° |
| 2 | 86.1328 Hz | $\arctan \frac{-1.0118}{-4.1572}$ | 13.6561° |

Table 8.3: The phase of each bin, calculated using the data in Table 8.1

So, for example, the signal shown in Figure 3 has a component at 86.1328 Hz with a magnitude of 4.2786 and a phase offset of 13.6561°.

Note that you'll sometimes hear people saying something along the lines of the real component being the signal and the imaginary component containing the phase information. If you hear this, ignore it – it's wrong. You need both the real and the imaginary components to determine both the magnitude and the phase content of your signal. If you have one or the other, you'll get an idea of what's going on, but not a very good one.

8.2.3 A couple of more details on detail

When you do a DFT, there's a limit on the number of bins you can calculate. This total number is dependent on the number of samples in the audio signal that you're using, where the number of bins equals the number of samples. Just to make computers happier (and therefore faster) we tend to do DFT's using window lengths which are powers of 2, so you'll see lengths like 256 points, or 1024 points. So, a 256-point DFT will take in a digital audio signal that is 256 samples long and give you back 256 bins, each containing a complex number representing the real and imaginary components of the signal.

Now, remember that the bins are evenly spaced from 0 Hz up to the sampling rate, so if you have 1024 bins and the sampling rate is 44.1 kHz then you get a bin every 44100/1024 Hz or 43.0664 Hz. The longer the DFT window, the better the resolution you get because you're dividing the sampling rate by a bigger number and the spacing in frequency gets smaller. Intuitively, this makes sense. For example, let's say that we used a DFT length of 4 points. At a very low frequency, there isn't enough change in

4 samples (which is all the DFT knows about...) to see a difference, so the DFT can't calculate a low frequency. The longer the window length, the lower the frequency we can look at. If you wanted to see what was happening at 1 Hz, then you're going to have to wait for at least 1 full period of the waveform to understand what's going on. This means that your DFT window length has to be at least the size of all the samples in one second (because the period of a 1 Hz wave is 1 second). Therefore, the better low-frequency resolution you want, the longer a DFT window you're going to have to use.

This is great, but if you're trying to do this in real time, meaning that you want a DFT of a signal that you're listening to while you're listening to it, then you have to remember that the DFT can't be calculated until all of the samples in the window are known. Therefore, if you want good low-frequency resolution, you'll have to wait a little while for it. For example, if your window size is 8192 samples long, then the first DFT result won't come out of the system until 8192 samples after the music starts. Essentially, you're always looking at what has just happened – not what is happening. Therefore, if you want a faster response, you need to use a smaller window length.

The moral of the story here is that you can choose between good low-frequency resolution or fast response – but you really can't have both (but there are a couple of ways of cheating...).

8.2.4 Redundancy in the DFT Result

If you go through the motions and do a DFT bin by bin, you'll start to notice that the results start mirroring themselves. For example, if you do an 8-point DFT, then bins 0 up to 4 will all be different, then bin 5 will be identical to bin 3, bins 6 and 2 are the same, and bins 7 and 1 are the same. This is because the frequencies of the DFT bins go from 0 Hz to the sampling rate, but the audio signal only goes to half of the sampling rate, normally called the Nyquist frequency. Above the Nyquist, aliasing occurs and we get redundant information. The odd thing here is the fact that we actually eliminated information above the Nyquist frequency on the conversion from analog to digital, but there is still stuff there – it's just a mirror image of the signal we kept.

Consequently, when we do a DFT, since we get this mirror effect, we typically throw away the redundant data and keep a little more than half the number of bins – in fact, it's one more than half the number. So, if you do a 256-point DFT, then you are really given 256 frequency bins, but only

129 of those are usable ($256/2 + 1$). In real books on DSP, they'll tell you that, for an N-point DFT, you get $N/2+1$ bins. These bins go from 0 Hz up to the Nyquist frequency or $\frac{f_s}{2}$.

Also, you'll notice that at the first and last bins (at 0 Hz and the Nyquist frequency) only contain real values – no imaginary components. This is because, in both cases, we can't calculate the phase. There is no phase information at 0 Hz, and since, at the Nyquist frequency, the samples are always hitting the same point on the sinusoid, we don't see its phase.

8.2.5 What's the use?

Good question. Well, what we've done is to look at a signal represented in the *Time domain* (in other words, what does the signal do if we look at it over a period of time) and convert that into a representation in the *Frequency domain* (in other words, what are the represented frequencies in this signal). These two domains are completely inter-related. That is to say that if you take any signal in the time domain, it has only one representation in the frequency domain. If you take that signal's representation in the frequency domain, you can convert it back to the time domain. Essentially, you can calculate one from the other because they are just two different ways of expressing the same signal.

For example, I can use the word “two” or the symbol “2” to express a quantity. Both mean exactly the same thing – there is no difference between “2 chairs” and “two chairs.” The same is true when you're moving back and forth between the frequency domain and time domain representations of the same signal. The signal stays the same – you just have two different ways of writing it down.

8.2.6 Suggested Reading List

[[Strawn, 1985](#)]

8.3 Windowing Functions

Now that we know how to convert a time domain signal into a frequency domain representation, let's try it out. We'll start by creating a simple sine wave that lasts for 1024 samples and is comprised of 4 cycles as is shown in Figure 8.12.

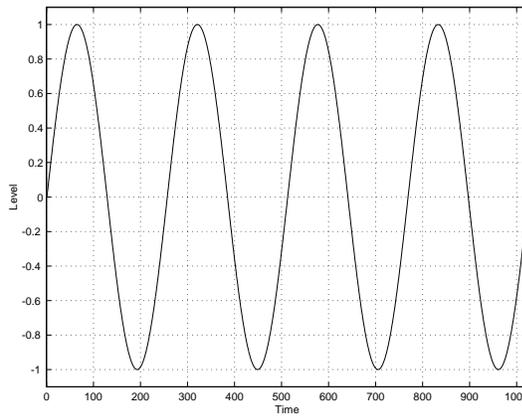


Figure 8.12:

If we do a 1024-point DFT of this signal and show the modulus of each bin (not including the redundancy in the mirror image), it will look like Figure 8.13.

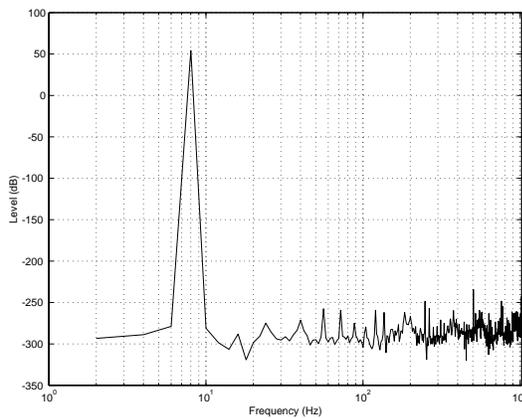


Figure 8.13:

You'll notice that there's a big spike at one frequency and a little noise (very little... -350 dB is a VERY small number) in all of the other bins. In case you're wondering, the noise is caused by the limited resolution of MATLAB which I used for creating these graphs. MATLAB calculates numbers with a resolution of 64 bits. That gives us a dynamic range of about 385 dB or so to work with – more than enough for this textbook... More than enough for most things, actually...

Now, what would happen if we had a different frequency? For example, Figure 8.14 shows a sine wave with a frequency of 0.875 times the one in Figure 8.12.

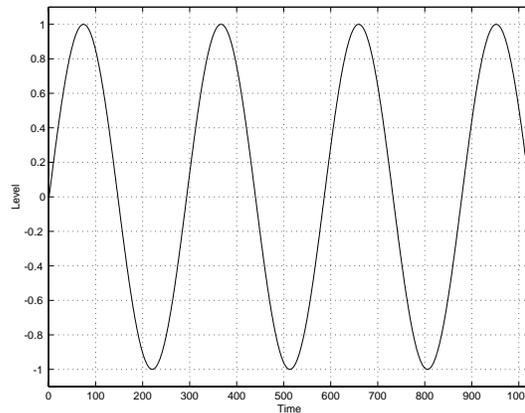


Figure 8.14:

If we do a 1024-point DFT on this signal and just look at the modulus of the result, we get the plot shown in Figure 8.15.

You'll probably notice that Figure 8.13 is very different from Figure 8.15. We can tell from the former almost exactly what the signal is, and what it isn't. In the latter, however, we get a big much of information. We can see that there's more information in the low end than in the high end, but we really can't tell that it's a sine wave. Why does this happen?

The problem is that we've only taken a slice of time. When you do a DFT, it assumes that the time signal you're feeding it is periodic. Let's make the same assumption on our two sine waves shown above. If we take the first one and repeat it, it will look like Figure 8.16. You can see that the repetition joins smoothly with the first presentation of the signal, so the signal continues to be a sine wave. If we kept repeating the signal, you'd get smooth connections between the signals and you'd just make a longer and

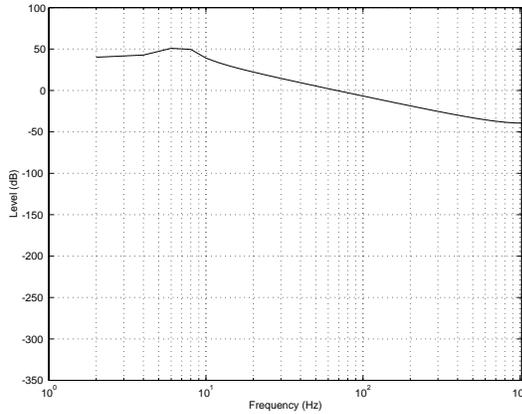


Figure 8.15:

longer sine wave.

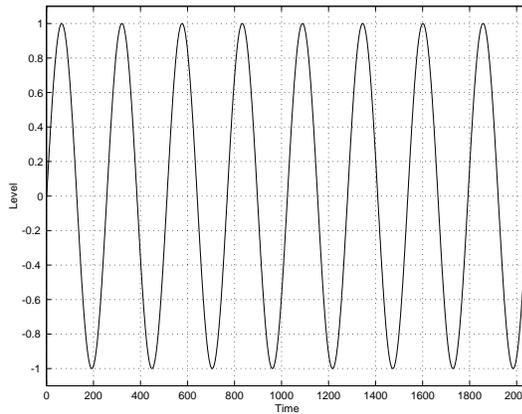


Figure 8.16:

If we take the second sine wave and repeat it, we get Figure 8.17. Now, we can see that things aren't so pretty. Because the length of the signal is not an integer number of cycles of the sine wave, when we repeat it, we get a nasty-look change in the sine wave. In fact, if you look at Figure 8.17, you can see that it can't be called a sine wave any more. It has some parts that look like a sine wave, but there's a spike in the middle. If we keep repeating the signal over and over, we've got a spike for every repetition.

That spike (also called a *discontinuity* in the time signal) contains energy

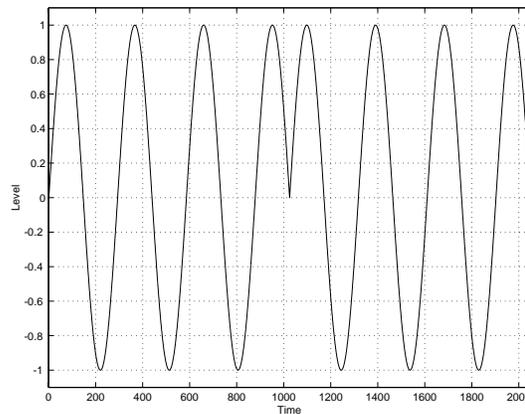


Figure 8.17:

in frequency bins other than where the sine wave is. In fact, this energy can be seen in the DFT that we did in Figure 8.15.

The moral of the story thus far is that if your signal's period is not the same length as 1 more sample than the window of time you're doing the DFT on, then you're going to get a strange result. Why does it have to be 1 more sample? This is because if the period was equal to the window length, then when you repeated it, you'd get a repetition of the signal because the first sample in the window is the same as the last. (In fact, if you look carefully at the end of the signal in Figure 8.12, you'll see that it doesn't quite get back to 0 for exactly this reason.)

How do we solve this problem? Well, we have to do something to the signal to make sure that the nasty spike goes away. The concept is basically the same as doing a crossfade between two sounds – we're just going to make sure that the signal in the window starts at 0, fades in, and then fades away to 0 before we do the DFT. We do this by introducing something called a *windowing function*. This is a list of gains that are multiplied by our signal as is shown in Figure 8.18.

Let's take the signal in Figure 8.14 – the one that caused us all the problems. If we multiply each of the samples in that signal with its corresponding gain shown in Figure 8.18, then we get a signal that looks like the one shown in Figure 8.19.

You'll notice that the signal still has the sine wave from Figure 8.14, but we've changed its level over time so that it starts and ends with a 0 value. This way, when we repeat it, the ends join together seamlessly. Now, if we

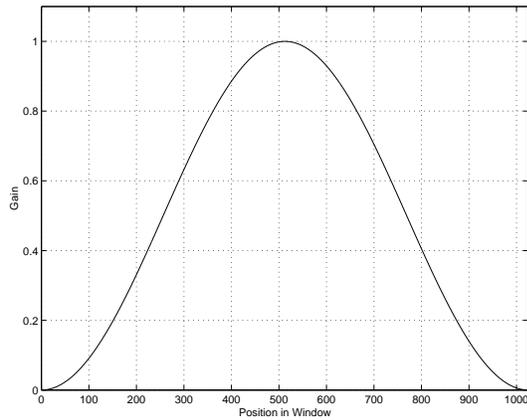


Figure 8.18: An example of a windowing function. The X-value of the signal corresponds to a position within the window. The Y-value is a gain multiplier used to change the level of the signal we're measuring.

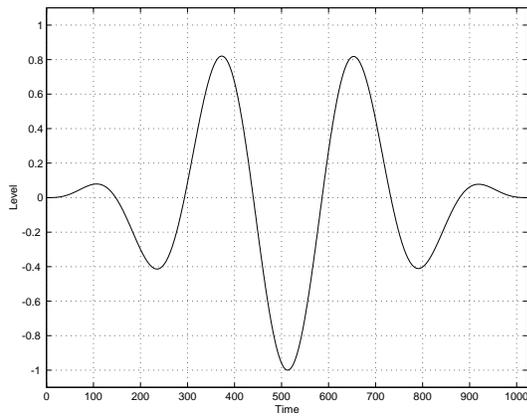


Figure 8.19: The result of the signal in Figure 8.14 multiplied by the gain function shown in Figure 8.18.

do a 1024-point DFT on this signal we get Figure 8.20.

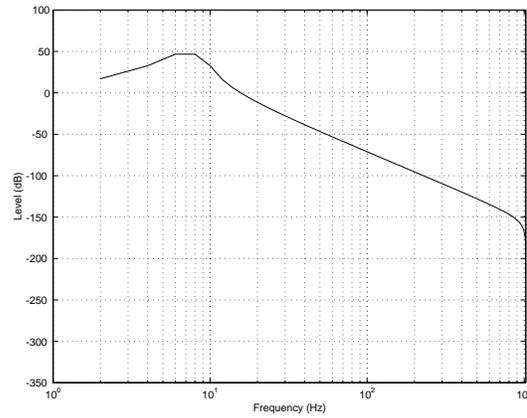


Figure 8.20:

Okay, so the end result isn't perfect, but we've attenuated the junk information by as much as 100 dB, which, if you ask me is pretty darned good.

Of course, like everything in life, this comes at a cost. What happens if we apply the same windowing function to the well-behaved signal in Figure 8.12 and do a DFT? The result will look like Figure 8.21.

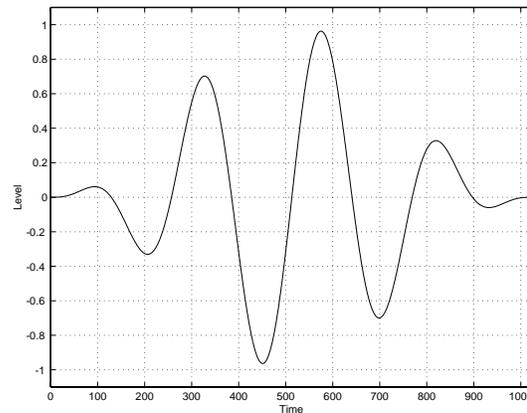


Figure 8.21:

So, you can see that applying the windowing function made bad things better but good things worse. The moral here is that you need to know that

using a windowing function will have an effect on the output of your DFT calculation. Sometimes you should use it, sometimes you shouldn't. If you don't know whether you should or not, you should try it with and without and decide which worked best.

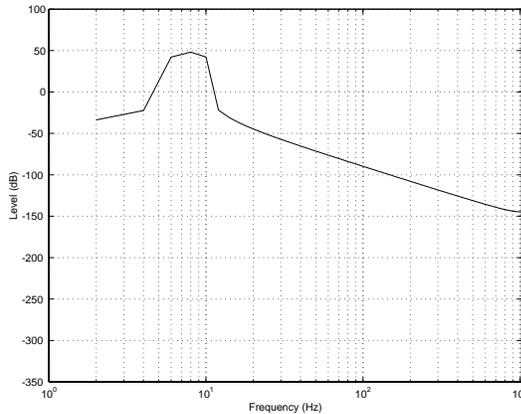


Figure 8.22:

So, we've seen that applying a windowing function will change the resulting frequency response. The good thing is that this change is predictable, and different for different functions, so you can not only choose whether or not to window your signal, but you can also choose what kind of window to use according to your requirements.

There are essentially an infinite number of different windowing functions available for you, but there are a number of standard ones that everyone uses for different reasons. We'll look at only three – the rectangular, Hanning and Hamming functions.

8.3.1 Rectangular

The rectangular window is the simplest of all the windowing functions because you don't have to do anything. If you multiply each value in your time signal by 1 (or do nothing to your signal) then your gain function will look like Figure 8.23. This graph looks like a rectangle, so we call doing nothing a rectangular window.

What does this do to our frequency response? Take a look at Figure 8.24 which is an 8192-point DFT of the signal in Figure ??.

This shows the general frequency response curve that is applied to the

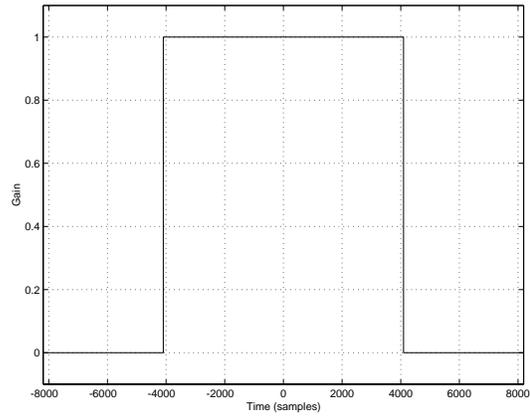


Figure 8.23: Time vs. gain response of rectangular windowing function 8192 samples long.

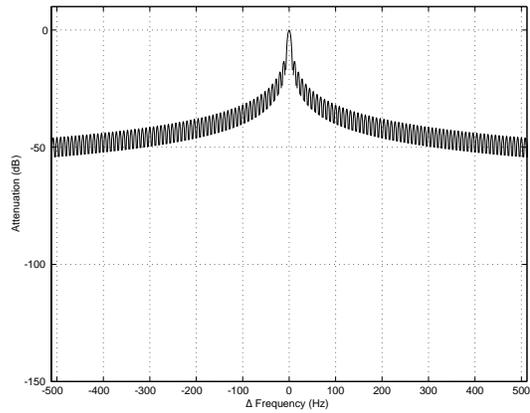


Figure 8.24: Frequency response of rectangular windowing function. The holes in the response are where the magnitude drops to $-\infty$.

DFT of your signal. In essence, you can sort of consider this to be an EQ curve that is added to the frequency response measurement of your signal. In fact, if you take a look at Figure 8.15, you'll see a remarkable similarity to the general shape of Figure 8.24.

DISCUSSION GOES HERE WHY IT'S DELTA FREQUENCY IN THESE PLOTS.

Now, let's zoom in on the middle of Figure ?? as is shown in Figure 8.25. Here, you can see that the frequency response of the rectangular window consists of multiple little lobes, with the biggest and widest in the middle. In between each of these lobes, the level drops to 0 (or $-\infty$ dB).

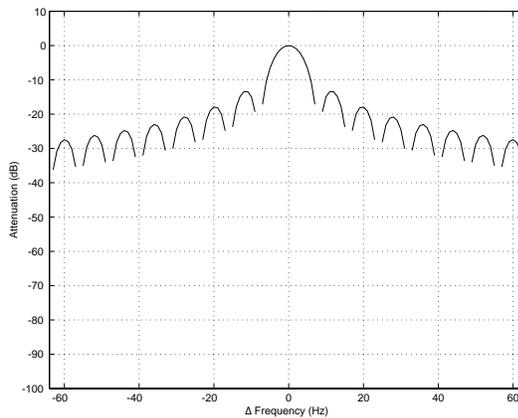


Figure 8.25: Frequency response of rectangular windowing function.

8.3.2 Hanning

You have already seen one standard windowing function in Figure 8.18. This is known as a Hanning window and is defined using the equation below [Morfe, 2001].

$$w(t) = \begin{cases} \frac{1}{2} \left(1 + \cos \frac{2\pi t}{T} \right) & \text{for } -\frac{1}{2}T \leq t \leq \frac{1}{2}T \\ 0 & \text{otherwise} \end{cases}$$

This looks a little complicated, but if you spend some time with it, you'll see that it is exactly the same math as a cardioid microphone. It says that, within the window, you have the same response as a cardioid microphone, and outside the window, the gain is 0. This is shown in Figure 8.26.

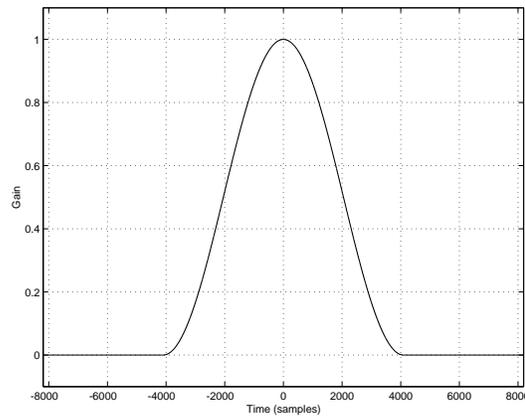


Figure 8.26: Time vs. gain response of Hanning windowing function 8192 samples long.

If we do a DFT of the signal in Figure 8.26, we get the plot shown in Figure 8.27. You can see here that, compared to the rectangular window, we get much more attenuation of unwanted signals far away from the frequencies we're interested in. But this comes at a cost...

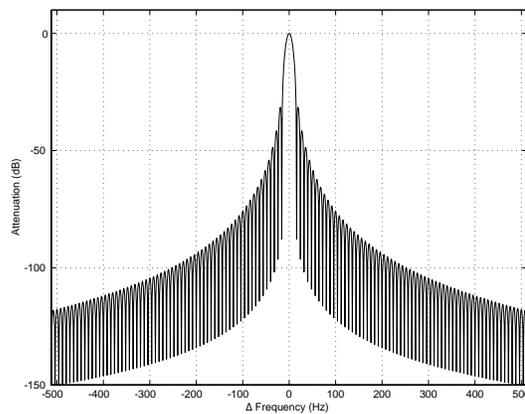


Figure 8.27: Frequency response of Hanning windowing function.

Take a look at Figure 8.28 which shows a close-up of the frequency response of a Hanning window. You can see here that the centre lobe is much wider than the one we saw for a rectangular window. So, this means that, although you'll get better rejection of signals far away from the frequencies that you're interested in, you'll also get more garbage leaking in very close

to those frequencies. So, if you're interested in a broad perspective, this window might be useful, if you're zooming in on a specific frequency, you might be better off using another.

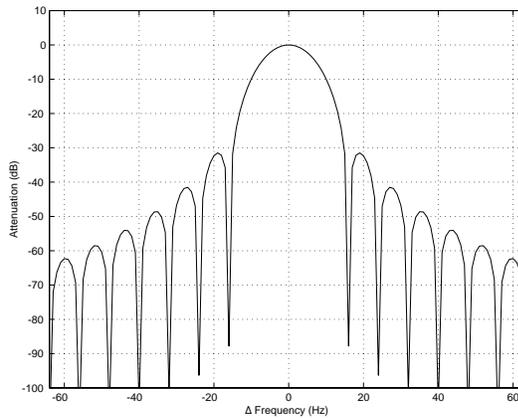


Figure 8.28: Frequency response of Hanning windowing function.

8.3.3 Hamming

Our third windowing function is known as the Hamming window. The equation for this is

EQUATION FOR HAMMING TO GO HERE

This gain response can be seen in Figure 8.29. Notice that this one is slightly weird in that it never actually reaches 0 at the ends of the window, so you don't get a completely smooth transition.

The frequency response of the Hamming window is shown in Figure 8.30. Notice that the rejection of frequencies far away from the centre is better than with the rectangular window, but worse than with the Hanning function.

So, why do we use the Hamming window instead of the Hanning if its rejection is worse away from the 0 Hz line? The centre lobe is still quite wide, so that doesn't give us an advantage. However, take a look at the lobes adjacent to the centre in Figure 8.31. Notice that these are quite low, and very narrow, particularly when they're compared to the other two functions. We'll look at them side-by-side in one graph a little later, so no need to flip pages back and forth at this point.

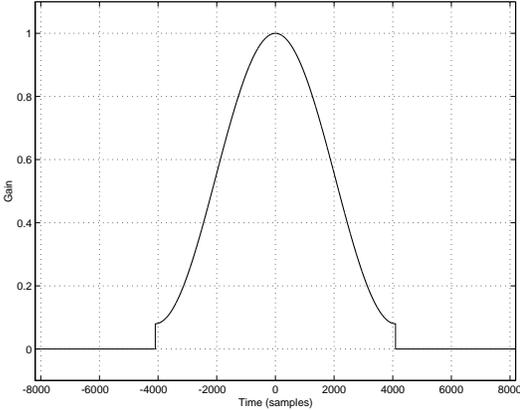


Figure 8.29: Time vs. gain response of Hamming windowing function 8192 samples long.

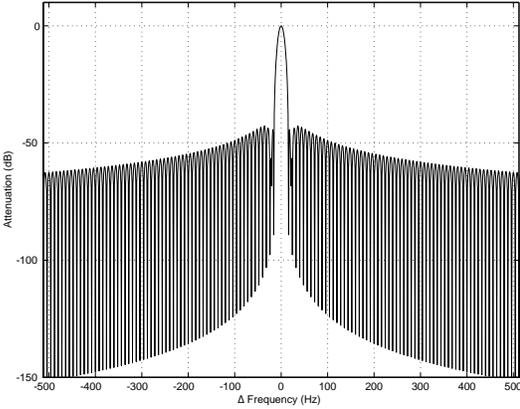


Figure 8.30: Frequency response of Hamming windowing function.

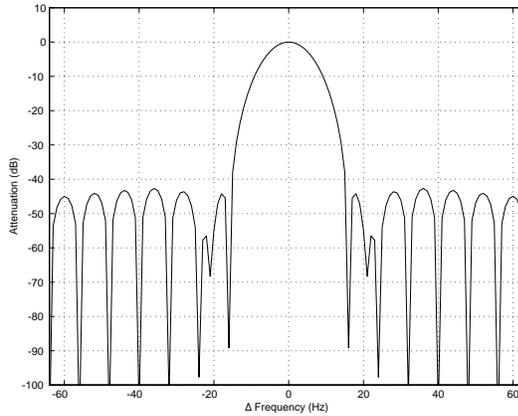


Figure 8.31: Frequency response of Hamming windowing function.

8.3.4 Comparisons

Figure 8.32 shows the three standard windows compared on one graph. As you can see, the rectangular window in blue has the narrowest centre lobe of the three, but the least attenuation of its other lobes. The Hanning window in black has a wider centre lobe but good rejection of its other lobes, getting better and better as we get further away in frequency. Finally, the Hamming window in red has a wide centre lobe, but much better rejection in its adjacent lobes.

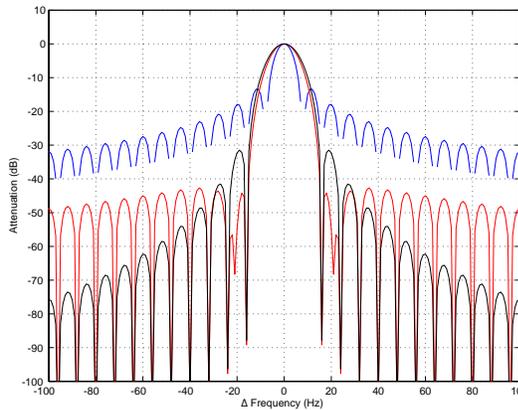


Figure 8.32: Frequency response of three common windowing functions. Rectangular (blue), Hamming (red) and Hanning (black).

8.3.5 Suggested Reading List

8.4 FIR Filters

We've now seen in Sections 8.2 that there is a direct link between the time domain and the frequency domain. If we make a change in a signal in the time domain, then we must incur a change in the frequency content of the signal. Way back in the section on analog filters we looked at how we can use the slow time response of a capacitor or an inductor to change the frequency response of a signal passed through it.

Now we are going to start thinking about how to intentionally change the frequency response of a signal by making changes to it in the time domain digitally.

8.4.1 Comb Filters

We've already seen back in Section 3.2.4 that a comb filter is an effect that is caused when a signal is mixed with a delayed version of itself. This happens in real life all the time when a direct sound meets a reflection of the same sound at your eardrum. These two signals are mixed acoustically and the result is a change in the timbre of the direct sound.

So, let's implement this digitally. This can be done pretty easily using the block diagram shown in Figure 8.33 which corresponds to Equation 8.3.

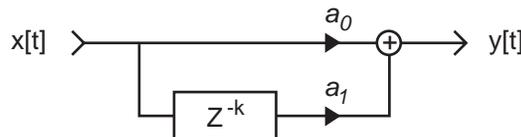


Figure 8.33: Basic block diagram of a comb filter implemented in the digital domain.

$$y_t = a_0x_t + a_1x_{t-k} \quad (8.3)$$

This implementation is called a *Finite Impulse Response comb filter* or *FIR comb filter* because, as we'll see in the coming sections, its impulse response is finite (meaning it ends at some predictable time) and that its frequency response looks a little like a hair comb.

As we can see in the diagram, the output consists of the addition of two signals, the original input and a gain-modified delayed signal (the delay is the block with the Z^{-k} in it. We'll talk later about why that notation is used,

but for now, you'll need to know that the delay time is k samples.). Let's assume for the remainder of this section that the gain value a is between -1 and 1, and is not 0. (If it was 0, then we wouldn't hear the output of the delay and we wouldn't have a comb filter, we'd just have a "through-put" where the output is identical to the input.)

If we're thinking in terms of acoustics, the direct sound is simulated by the non-delayed signal (the through-put) and the reflection is simulated by the output of the delay.

Delay with Positive Gain

Let's take an FIR comb filter as is described in Figure 8.46 and Equation 8.3 and make the delay time equal to 1 sample, and $a = 1$. What will this algorithm do to an audio signal?

We'll start by thinking about a sine wave with a very low frequency – in this case the phase difference between the input and the output of the delay is very small because it's only 1 sample. The lower the frequency, the smaller the phase difference until, at 0 Hz (DC) there is no phase difference (because there is no phase...). Since the output is the addition of the values of the two samples (now, and 1 sample ago), we get more at the output than the input. At 0 Hz, the output is equal to exactly two times the input. As the frequency goes higher, the phase difference caused by the delay gets bigger and the output gets smaller.

Now, let's think of a sine wave at the Nyquist frequency (see Section 7.1.3 if you need a definition). At this frequency, a sample and the previous sample are separated by 180° , therefore, they are identical but opposite in polarity. Therefore, the output of this comb filter will be 0 at the Nyquist Frequency because the samples are cancelling themselves. At frequencies below the Nyquist, we get more and more output from the filter.

If we were to plot the resulting frequency response of the output of the filter it would look like Figure 8.34. (Note that the plot uses a normalized frequency, explained in Section 8.1.2.)

What would happen if the delay in the FIR comb filter were 2 samples long? We can think of the result in the same way as we did in the previous description. At DC, the filter will behave in the same way as the FIR comb with a 1-sample delay. At the Nyquist Frequency, however, things will be a little different... At the Nyquist frequency (a normalized frequency of 0.5), every second sample has an identical value because they're 360° apart. Therefore, the output of the filter at the Nyquist Frequency will be two times the input value, just as in the case of DC.

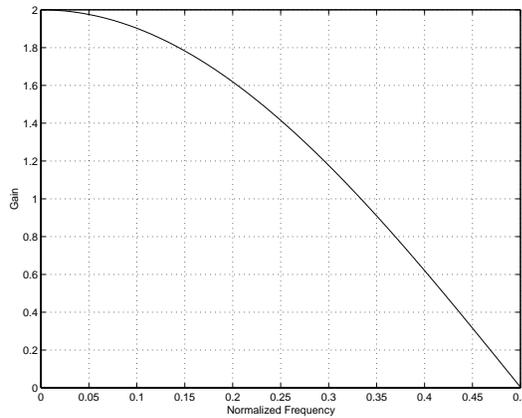


Figure 8.34: Frequency response of an FIR comb filter with a delay of 1 sample, $a_0 = 1$, and $a_1 = 1$

At one-half the Nyquist Frequency (a normalized frequency of 0.25) there is a 90° phase difference between samples, therefore there is a 180° phase difference between the input and the output of the delay. Therefore, at this frequency, our FIR comb filter will have no output.

The final frequency response is shown in Figure 8.35.

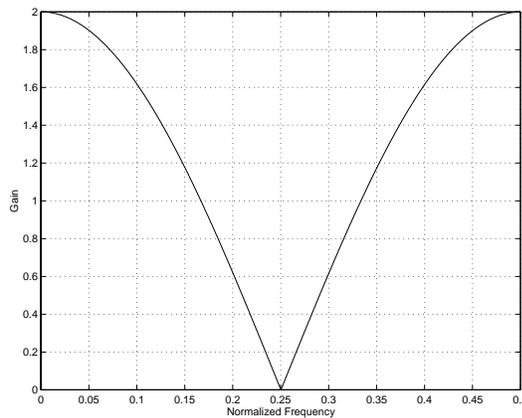


Figure 8.35: Frequency response of an FIR comb filter with a delay of 2 samples, $a_0 = 1$, and $a_1 = 1$

As we increase the delay time in the FIR comb filter, the first notch in the frequency response drops lower and lower in frequency as can be seen in Figures 8.36 and 8.37.

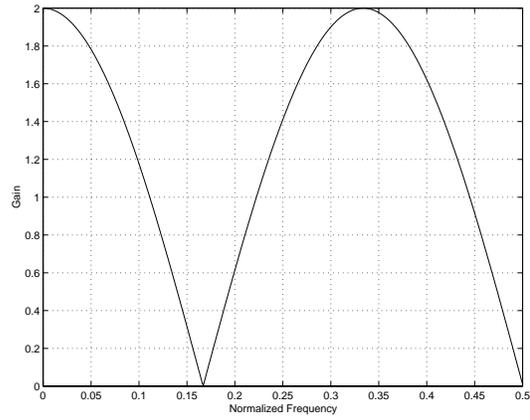


Figure 8.36: Frequency response of an FIR comb filter with a delay of 3 samples, $a_0 = 1$, and $a_1 = 1$

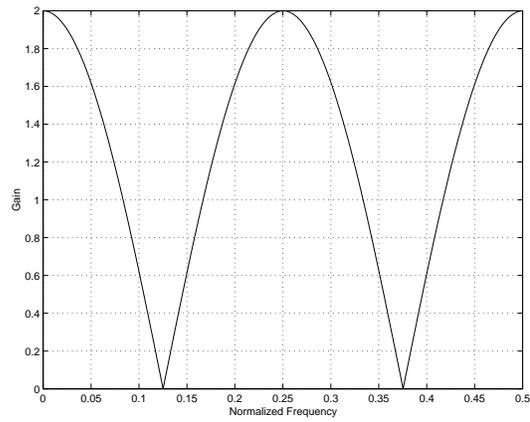


Figure 8.37: Frequency response of an FIR comb filter with a delay of 4 samples, $a_0 = 1$, and $a_1 = 1$

Up to now, we have been looking at the frequency response graphs on a linear scale. This has been to give you an idea of the behaviour of the FIR comb filter in a mathematical sense, but it really doesn't provide an intuitive feel for how it will sound. In order to get this, we have to plot the frequency response on a semi-logarithmic plot (where the X-axis is on a logarithmic scale and the Y-axis is on a linear scale). This is shown in Figure 8.38.

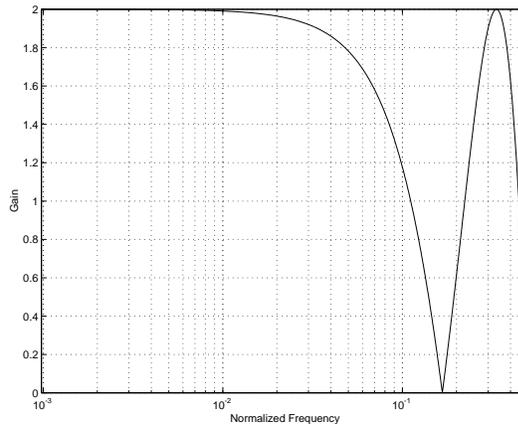


Figure 8.38: Frequency response of an FIR comb filter with a delay of 3 samples, $a_0 = 1$, and $a_1 = 1$. Note that this is the same as the graph in Figure 8.36

So far, we have kept the gain on the output of the delay at 1 to make things simple. What happens if this is set to a smaller (but still positive) number? The bumps and notches in the frequency response will still be in the same places (in other words, they won't change in frequency) but they won't be as drastic. The bumps won't be as big and the notches won't be as deep.

Delay with Negative Gain

In the previous section we limited the value of the gain applied to the delay component to positive values only. However, we also have to consider what happens when this gain is set to a negative value. In essence, the behaviour is the same, but we have a reversal between the constructive and destructive interferences. In other words, what were bumps before become notches, and the notches become bumps.

For example, let's use an FIR comb filter with a delay of 1 sample and

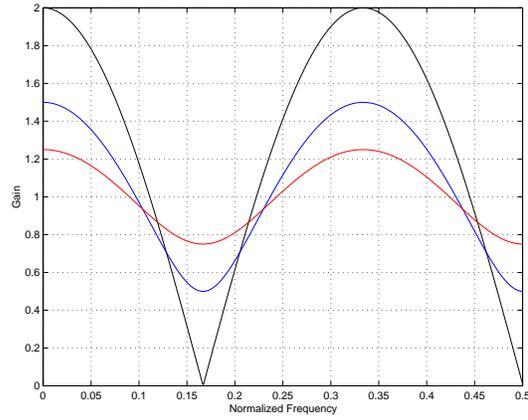


Figure 8.39: Frequency response of an FIR comb filter with a delay of 3 samples, $a_0 = 1$. Black $a_1 = 1$, blue $a_1 = 0.5$, red $a_1 = 0.25$.

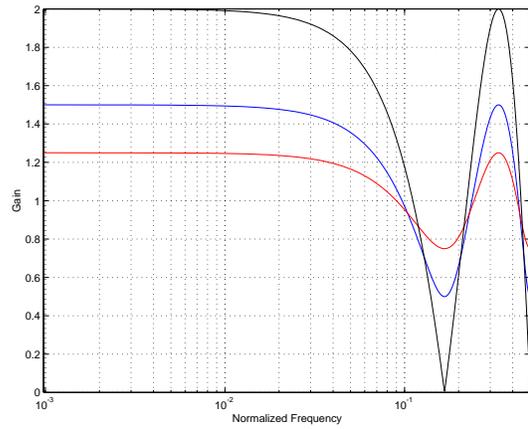


Figure 8.40: Frequency responses of an FIR comb filter with a delay of 3 samples, $a_0 = 1$. Black $a_1 = 1$, blue $a_1 = 0.5$, red $a_1 = 0.25$.

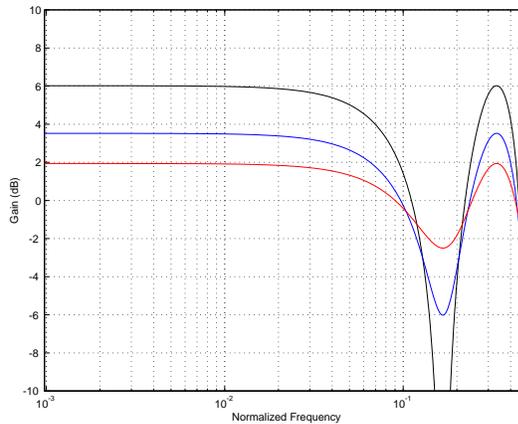


Figure 8.41: Frequency responses of an FIR comb filter with a delay of 3 samples, $a_0 = 1$. Black $a_1 = 1$, blue $a_1 = 0.5$, red $a_1 = 0.25$.

where $a_0 = 1$, and $a_1 = -1$. At DC, the output of the delay component will be identical to but opposite in polarity with the non-delayed component. This means that they will cancel each other and we get no output from the filter. At a normalized frequency of 0.5 (the Nyquist Frequency) the two components will be 180° out of phase, but since we're multiplying one by -1 , they add to make twice the input value.

The end result is a frequency response as is shown in Figures 8.42.

If we have a longer delay time, then we get a similar behaviour as is shown in Figure ??.

If the gain applied to the output of the delay is set to a value greater than -1 but less than 0 , we see a similar reduction in the deviation from a gain of 1 as we saw in the examples with FIR comb filters with a positive gain delay component.

8.4.2 Frequency Response Deviation

If you want to get an idea of the effect of an FIR comb filter on a frequency response, we can calculate the levels of the maxima and minima in its frequency response. For example, the maximum value of the frequency response shown in Figure ?? is 6 dB. The minimum value is $-\infty$ dB. Therefore a total peak-to-peak variation in the frequency response is ∞ dB.

We can calculate these directly using Equations 8.4 and 8.5 [Martin, 2002a].

$$mag_{max} = 20 \log_{10} (a_0 + a_1) \quad (8.4)$$

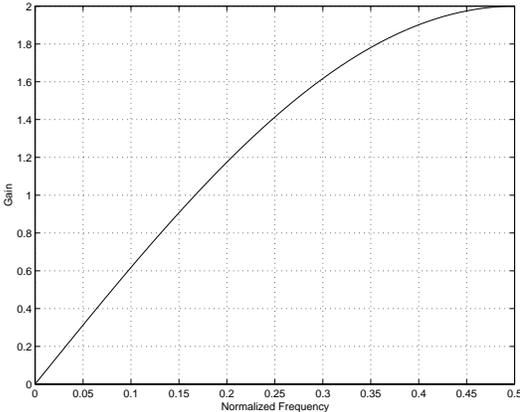


Figure 8.42: Frequency response of an FIR comb filter with a delay of 1 sample, $a_0 = 1$, and $a_1 = -1$

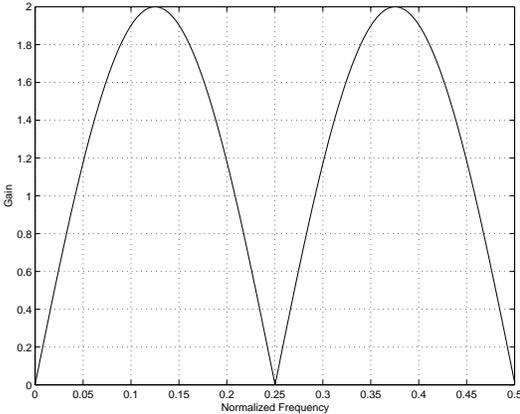


Figure 8.43: Frequency response of an FIR comb filter with a delay of 4 samples, $a_0 = 1$, and $a_1 = -1$

$$mag_{min} = 20 \log_{10} (a_0 - a_1) \quad (8.5)$$

In order to find the total peak-to-peak variation in the frequency response of the FIR comb filter, we can use Equation 8.6.

$$mag_{p-p} = 20 \log_{10} \left| \frac{a_0 + a_1}{a_0 - a_1} \right| \quad (8.6)$$

8.4.3 Impulse Response vs. Frequency Response

We have seen in earlier sections of this book that the time domain and the frequency domain are just two different ways of expressing the same thing. This rule holds true for digital filters as well as signals.

Back in Section 3.5.3 that we used an impulse response to find the behaviour of a string on a musical instrument. We can do the same for a filter implemented in DSP, with a little fudging here and there... First we have to make the digital equivalent of an impulse. This is fairly simple, we just have to make an infinite string of 0's with a single 1 somewhere. Usually when this signal is described in a DSP book, we think of the "1" as happening now, therefore we see a description like Equation 8.4.3.

$$\delta = \begin{cases} 0 & n < 0 \\ 1 & n = 0 \\ 0 & n > 0 \end{cases}$$

This digital equivalent to an impulse is called a *Dirac impulse* (named after Paul Dirac, a French mathematician. FIND OUT MORE ABOUT DIRAC). Although in the digital domain it looks a lot like an impulse, it really isn't because it isn't infinitely short of infinitely loud. On the other hand, it behaved in a very similar way to a real impulse since, in the digital domain, it has a flat frequency response from DC to the Nyquist Frequency.

What happens at the output when we send the Dirac impulse through the FIR comb filter with a 3-sample delay? First, we see the Dirac come out the output at the same time it arrives at the input, but multiplied by the gain value a_0 (up to now, we have used 1 for this value). Then, three samples later, we see the impulse again at the output, this time multiplied by the gain value a_1 .

Since nothing happens after this, the impulse response has ended (we could keep measuring it, but we would just get 0's forever...) which is why we call these FIR (for *finite impulse response*) filters. If an impulse goes in, the filter will stop giving you an output at a predictable time.

Figure 8.44 shows three examples of different FIR comb filter impulse responses and their corresponding frequency responses. Note that the delay values for all three filters are the same, therefore the notches and peaks in the frequency responses are all matched. Only the value of a_1 was changed, therefore modifying the amount of modulation in the frequency responses.

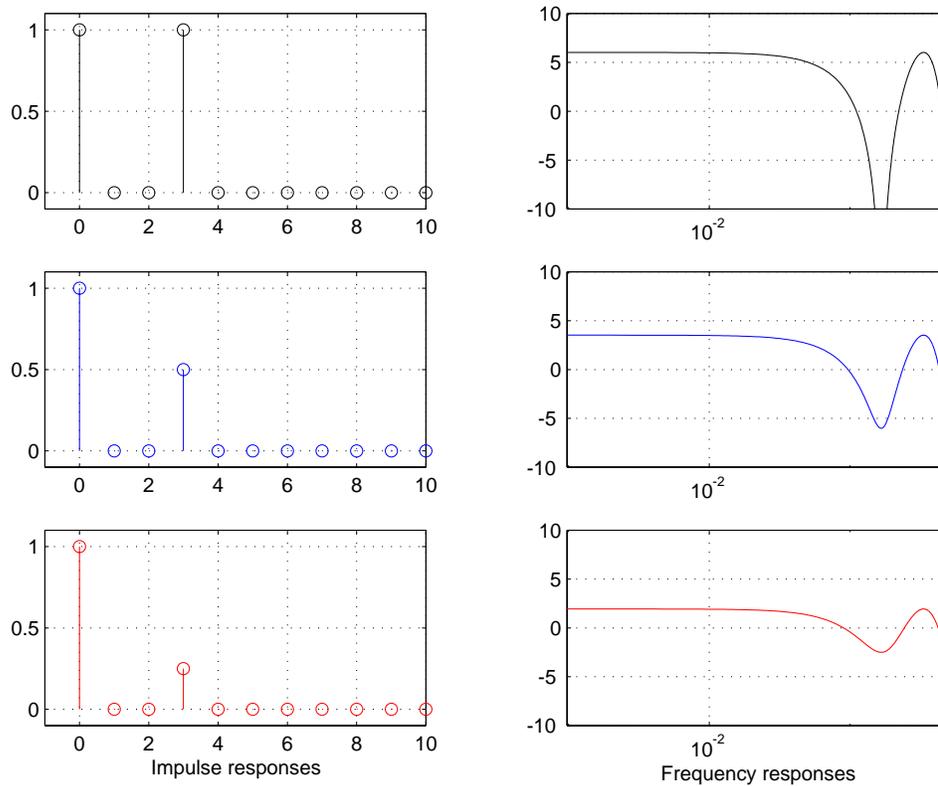


Figure 8.44: Impulse and corresponding frequency responses of FIR comb filters with a delay of 3 samples. Black $a_1 = 1$, blue $a_1 = 0.5$, red $a_1 = 0.25$.

8.4.4 More complicated filters

So far, we have only looked at one simple type of FIR filter – the FIR comb. In fact, there are an infinite number of other types of FIR filters but most of them don't have specific names (good thing too, since there is an infinite number of them... we would spend forever naming things...).

Figure 8.45 shows the general block diagram for an FIR filter. Notice

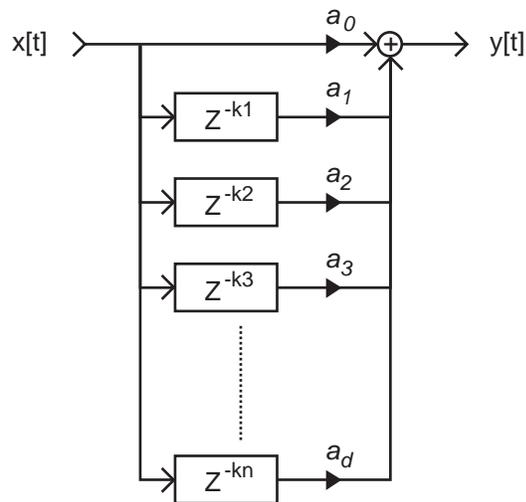


Figure 8.45: General block diagram for an FIR filter.

that there can be any number of delays, each with its own delay time and gain.

8.4.5 Suggested Reading List

[Steiglitz, 1996]

8.5 IIR Filters

Since we have a class of filters that are specifically called “Finite Impulse Response” filters, then it stands to reason that they’re called that to distinguish them from another type of filter with an infinite impulse response. If you already guess this, then you’d be right. That other class is called *Infinite Impulse Response* (or *IIR*) filters for reasons that will be explained below.

8.5.1 Comb Filters

We have already seen in Section 8.4.1 that a comb filter is one with a frequency response that looks like a hair comb. In an FIR comb filter, this is accomplished by adding a signal to a delayed copy of itself.

This can also be accomplished as an IIR filter, however, both the implementation and the results are slightly different.

Figure 8.46 shows a simple IIR comb filter

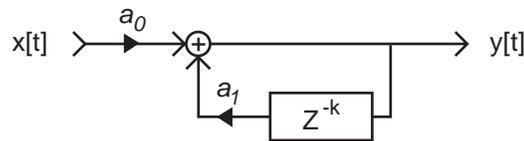


Figure 8.46: Basic block diagram of an IIR comb filter implemented in the digital domain.

As can be seen in the block diagram, we are now using feedback as a component in the filter’s design. The output of the filter is fed backwards through a delay and a gain multiplier and the result is added back to the input, which is fed back through the same delay and so on...

Since the output of the delay is connected (through the gain and the summing) back to the input of the delay, it keeps the signal in a loop forever. That means that if we send a Dirac impulse into the filter, the output of the filter will be busy until we turn off the filter, therefore it has an infinite impulse response.

The feedback loop can also be seen in Equation 8.8 which shows the general equation for a simple IIR comb filter. Notice that there is a y on both sides of the equation – the output at the moment y_t is formed of two components, the input at the same time (now) x_t and the output from k

samples ago indicated by the y_{t-k} . Of course, both of these components are multiplied by their respective gain factors.

$$y_t = a_0x_t + a_1y_{t-k} \quad (8.7)$$

Positive Feedback

Let's use the block diagram shown in Figure 8.33 and make an IIR comb. We'll make $a_0 = 1$, $a_1 = 0.5$ and $k = 3$. The result of this is that a Dirac impulse comes in the filter and immediately appears at the output (because $a_0 = 1$). Three samples later, it also comes out the output at one-half the level (because $a_1 = 0.5$ and $k = 3$).

The resulting impulse response will look like Figure 8.47.

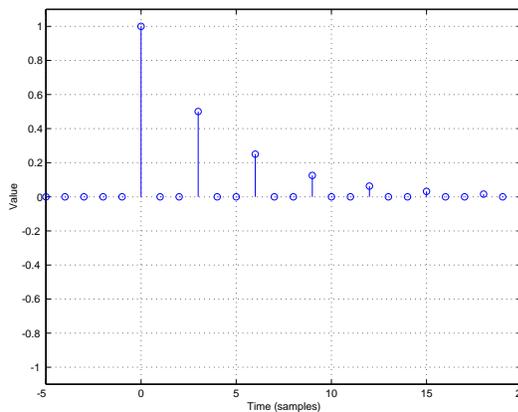


Figure 8.47: The impulse response of the IIR comb filter shown in Figure 8.33 and Equation 8.8 with $a_0 = 1$, $a_1 = 0.5$ and $k = 3$.

Note that, in Figure 8.47, I only plotted the first 20 samples of the impulse response, however, it in fact extends to infinity.

What will the frequency response of this filter look like? This is shown in Figure 8.48.

Compare this with the FIR comb filter shown in Figure 8.39. There are a couple of things to notice about the similarity and difference between the two graphs.

The similarity is that the peaks and dips in the two graphs are at the same frequencies. They aren't the same basic shape, but they appear at the same place. This is due to the matching 3-sample delays and the fact that the gain applied to the delays are both positive.

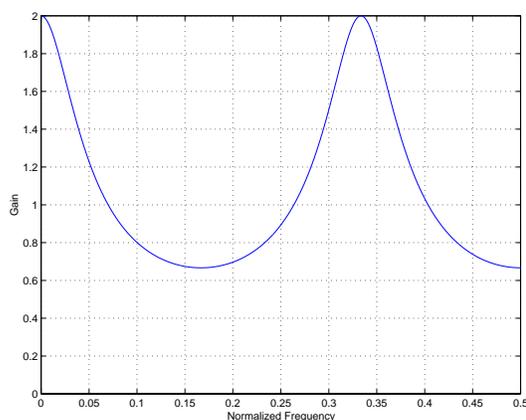


Figure 8.48: The frequency response of the IIR comb filter shown in Figure 8.33 and Equation 8.8 with $a_0 = 1$, $a_1 = 0.5$ and $m = 3$.

The difference between the graphs is obviously the shape of the curve itself. Where the FIR filter had broad peaks and narrow dips, the IIR filter has narrow peaks and broad dips. This is going to cause the two filters to sound very different. Generally speaking, it is much easier for us to hear a boost than a cut. The narrow peaks in the frequency response of an IIR filter are immediately obvious as boosts in the signal. This is not only caused by the fact that the narrow frequency bands are boosted, but that there is a smearing of energy in time at those frequencies known as *ringing*. In fact, if you take an IIR filter with a fairly high value of a_1 – say between 0.5 and 0.999 of a_0 , and put in an impulse, you’ll hear a tone ringing in the filter. The higher the gain of a_1 , the longer the ringing and the more obvious the tone. This frequency response change can be seen in Figure 8.49.

Negative Feedback

Just like the FIR counterpart, an IIR comb filter can have a negative gain at the delay output. As can be seen in Figure 8.48, positive feedback causes a large boost in the low frequencies with a peak at DC. This can be avoided by using a negative feedback value.

The interesting thing here is that the result of the negative feedback through a delay causes the impulse response to flip back and forth in polarity as can be seen in Figure 8.50.

The resulting frequency response for this filter is shown in Figure 8.51.

IIR comb filters with negative feedback suffer from the same ringing

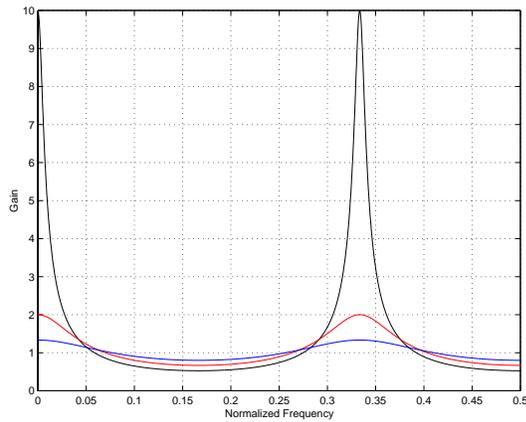


Figure 8.49: The frequency response of the IIR comb filter shown in Figure 8.33 and Equation 8.8 with $a_0 = 1$ and $m = 3$. Black $a_0 = 0.999$, Red $a_0 = 0.5$ and blue $a_0 = 0.25$.

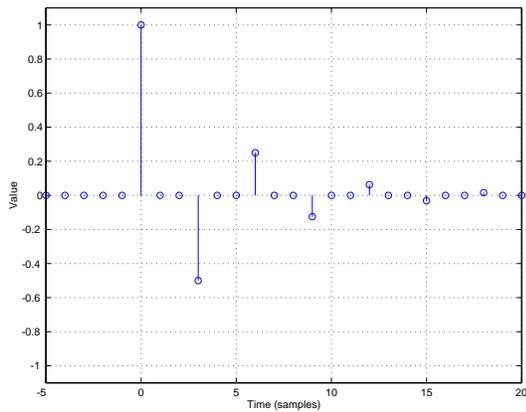


Figure 8.50: The frequency response of the IIR comb filter shown in Figure 8.33 and Equation 8.8 with $a_0 = 1$, $a_1 = -0.5$ and $m = 3$.

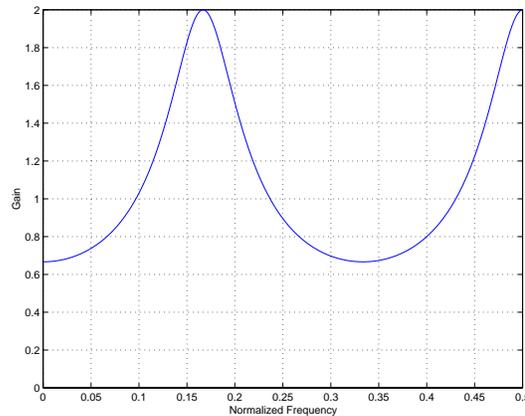


Figure 8.51: The frequency response of the IIR comb filter shown in Figure 8.33 and Equation 8.8 with $a_0 = 1$, $a_1 = -0.5$ and $m = 3$.

problems as those with positive feedback as can be seen in the frequency response graph in Figure 8.52.

8.5.2 Danger!

There is one important thing to beware of when using IIR filters. Always remember that feedback is an angry animal that can lash out and attack you if you're not careful. If the value of the feedback gain goes higher than 1, then things get ugly very quickly. The signal comes out of the delay louder than it went in, and circulates back to the input of the delay where it comes out even louder and so on and so on. Depending on the delay time, it will take a small fraction of a second for the filter to overload. And, since it has an infinite impulse response, even if you pull the feedback gain back to less than 1, the distortion that you caused will always be circulating in the filter. The only way to get rid of it is to drop a_1 to 0 until the delay clears out and then start again. (Although some IIR filters allow you to send a *clear* command, telling them to forget everything that's happened before now, and to continue on as if you were normal. Take a look at some of the filters in [Max/MSP](#), for example.)

8.5.3 Biquadratic

While it's fun to make comb filters to get started (every budding guitarist has a flanger or a phaser in their kit of pedal effects. The rich kids even have

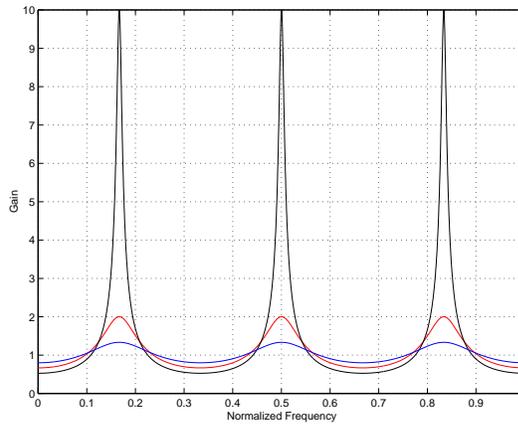


Figure 8.52: The frequency response of the IIR comb filter shown in Figure 8.33 and Equation 8.8 with $a_0 = 1$ and $k = 3$. Black $a_0 = 0.999$, Red $a_0 = 0.5$ and blue $a_0 = 0.25$.

both and claim that they know the difference!) IIR filters can be a little more useful. Now, don't get me wrong, FIR filters are also useful, but as you'll see if you ever have to start working with them, they're pretty expensive in terms of processing power. Basically, if you want to do something interesting, you have to do it for a long time (sort of Zen, no?). If you have an FIR filter with a long, impulse response, then that means a lot of delays (or long ones) and lots of math to do. We'll see a worst-case scenario in Section 8.6.

A simple IIR filter has the advantage of having only three operations (two multiplies and one add) and one delay, and still having an infinite impulse response, so you can do interesting things with a minimum of processing.

One of the most common building blocks in any DSP algorithm is a little package called a *biquadratic filter* or *biquad* for short. This is a sort of mini-algorithm that contains a just small number of delays, gains and additions as is shown in Figure 8.53, but it turns out to be a pretty powerful little tool – sort of the op amp of the DSP world in terms of usefulness. CHECK THIS DIAGRAM. I DON'T THINK THAT IT'S CORRECT

Okay, let's look at what a biquad does, going through it, gain by gain.

- Take a look at a_0 . The signal x_t comes in, gets multiplied by a_0 and gets added the output at the same time. This is just a gain applied to the input.
- Now let's look at a_1 . The signal x_t comes in, gets delayed by 1 sample, multiplied by a_1 and added to the output. By itself, this is just the

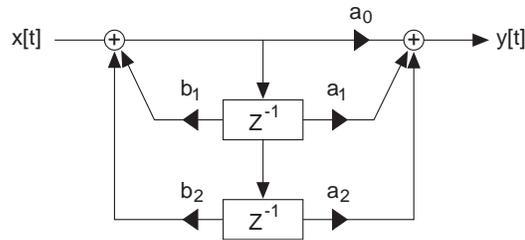


Figure 8.53: Block diagram of one implementation of a biquad filter. There are other ways to achieve the same algorithm, but we won't discuss that in this book. Check this reference for more information [1].

feed-forward component of an FIR comb filter with a delay of 1 sample.

- a_2 . The signal x_t comes in, gets delayed by 2 samples, multiplied by a_2 and added to the output. By itself, this is just the feed-forward component of an FIR comb filter with a delay of 2 samples.
- b_1 . The signal x_t comes in, gets delayed by 1 sample, multiplied by b_1 and added to the *input*. By itself, this is just the feedback component of an IIR comb filter with a delay of 1 sample.
- b_2 . The signal x_t comes in, gets delayed by 2 samples, multiplied by b_1 and added to the input. By itself, this is just the feedback component of an IIR comb filter with a delay of 2 samples.

The end result is an algorithm described by Equation ??.

$$y_t = a_0x_t + a_1x_{t-1} + a_2x_{t-2} + b_1y_{t-1} + b_1x_{t-2} \quad (8.8)$$

CHECK THIS EQUATION. I DON'T THINK THAT IT'S CORRECT

So, what we've seen is that a biquad is an assembly of two FIR combs and two IIR combs, all in one neat little package. The interesting thing is that this algorithm is extremely powerful. So, I'll bet you're sitting there wondering what it's used for? Let's look at just a couple of examples.

INTRODUCTION TO GO HERE ABOUT HOW YOU'LL NEED THE FOLLOWING EQUATIONS TO CONTINUE

$$A = \sqrt{10^{\frac{gain}{20}}} \quad (8.9)$$

where *gain* is in dB.

$$\omega = 2\pi f_c \quad (8.10)$$

where f_c is the *cutoff frequency* (or *centre frequency* in the case of peaking filters or the *shelf midpoint frequency* for shelving filters). Note that the frequency f_c here is given as a normalized frequency between 0 and 0.5. If you prefer to think in terms of the more usual way of describing frequency, in Hertz, then you'll need to do an extra little bit of math as is shown in Equation 8.11.

$$\omega = \frac{2\pi f_c}{f_s} \quad (8.11)$$

where f_c is the *cutoff frequency* (or *centre frequency* in the case of peaking filters or the *shelf midpoint frequency* for shelving filters) in Hz and f_s is the sampling rate in Hz. **Important note:** You should use *either* Equation 8.10 or Equation 8.11, depending on which way you prefer to state the frequency. However, I would recommend that you get used to thinking in terms of normalized frequency, so you should be happy with using Equation 8.10.

$$cs = \cos \omega \quad (8.12)$$

$$sn = \sin \omega \quad (8.13)$$

$$Q = \frac{sn}{\ln(2)bw\omega} \quad (8.14)$$

Use the previous equation to convert *bandwidth* or *bw* into a Q value.

$$\alpha = sn \sinh\left(\frac{1}{2Q}\right) \quad (8.15)$$

$$\beta = \sqrt{\frac{A^2 + 1}{S} - (A - 1)^2} \quad (8.16)$$

where S is a *shelf slope* parameter. “When $S = 1$, the shelf slope is as steep as you can get it and remain monotonically increasing or decreasing gain with frequency (for lowShelf or highShelf).” [?]

Low-pass filter

We already know from Section ?? what a low-pass filter is. One way to implement this in a DSP chain is to insert a biquad and calculate its coefficients using the equations below.

$$b_0 = \frac{1 - cs}{2} \quad (8.17)$$

$$b_1 = 1 - cs \quad (8.18)$$

$$b_2 = \frac{1 - cs}{2} \quad (8.19)$$

$$a_0 = 1 + \alpha \quad (8.20)$$

$$a_1 = -2cs \quad (8.21)$$

$$a_2 = 1 - \alpha \quad (8.22)$$

Low-shelving filter

Likewise, we could, instead, create a low-shelving filter using coefficients calculated in the equations below.

$$b_0 = A((A + 1) - (A - 1)cs + \beta sn) \quad (8.23)$$

$$b_1 = 2A((A - 1) - (A + 1)cs) \quad (8.24)$$

$$b_2 = A((A + 1) - (A - 1)cs - \beta sn) \quad (8.25)$$

$$a_0 = (A + 1) + (A - 1)cs + \beta sn \quad (8.26)$$

$$a_1 = -2((A - 1) + (A + 1)cs) \quad (8.27)$$

$$a_2 = (A + 1) + (A - 1)cs - \beta sn \quad (8.28)$$

Peaking filter

The following equations will result in a reciprocal peak-dip filter configuration.

$$b_0 = 1 + \alpha A \quad (8.29)$$

$$b_1 = -2cs \quad (8.30)$$

$$b_2 = 1 - \alpha A \quad (8.31)$$

$$a_0 = 1 + \frac{\alpha}{A} \quad (8.32)$$

$$a_1 = -2cs \quad (8.33)$$

$$a_2 = 1 - \frac{\alpha}{A} \quad (8.34)$$

High-shelving filter

The following equations will produce a high-shelving filter.

$$b_0 = A((A + 1) - (A - 1)cs + \beta sn) \quad (8.35)$$

$$b_1 = -2A((A - 1) - (A + 1)cs) \quad (8.36)$$

$$b_2 = A((A + 1) - (A - 1)cs - \beta sn) \quad (8.37)$$

$$a_0 = (A + 1) + (A - 1)cs + \beta sn \quad (8.38)$$

$$a_1 = 2((A - 1) + (A + 1)cs) \quad (8.39)$$

$$a_2 = (A + 1) + (A - 1)cs - \beta sn \quad (8.40)$$

High-pass filter

Finally, the following equations will produce a high-pass filter.

$$b0 = \frac{1 + cs}{2} \quad (8.41)$$

$$b1 = -1 - cs \quad (8.42)$$

$$b2 = \frac{1 + cs}{2} \quad (8.43)$$

$$a0 = 1 + \alpha \quad (8.44)$$

$$a1 = -2cs \quad (8.45)$$

$$a2 = 1 - \alpha \quad (8.46)$$

8.5.4 Allpass filters

There is an interesting type of filter that is frequently used in DSP that we didn't talk about very much in the chapter on analog electronics. This is the *allpass filter*.

As its name implies, an allpass filter allows all frequencies to pass through it without a change in magnitude. Therefore the frequency response of an allpass filter is flat. This may sound a little strange – why build a filter that doesn't do anything to the frequency response? Well, the nice thing about an allpass is that, although it doesn't change the frequency response, it does change the phase response of your signal. This can be useful for various situations as we'll see below.

FINISH THIS OFF

Decorrelation

FINISH THIS OFF

Fractional delays

FINISH THIS OFF

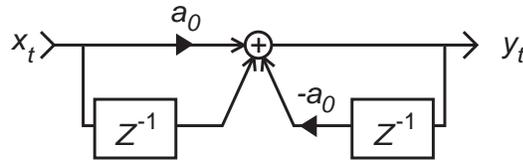


Figure 8.54: Block diagram of a first-order allpass filter [Steiglitz, 1996].

Reverberation simulation

FINISH THIS OFF

8.5.5 Suggested Reading List

[Steiglitz, 1996]

[?]

[?]

[?] LIST THIS PAPER BY KNUD BANK

8.6 Convolution

Let's think of the most inefficient way to build an FIR filter. Back in Section ?? we saw a general diagram for an FIR that showed a stack of delays, each with its own independent gain. We'll build a similar device, but we'll have an independent delay and gain for every possible *integer delay time* (meaning a delay value of an integer number of samples like 4 or 13, but not 2.4 or 15.2). When we need a particular delay time, we'll turn on the corresponding delay's gain, and all the rest we'll leave at 0 all the time.

For example, a smart way to do an FIR comb filter with a delay of 6 samples is shown in Figure 8.55 and Equation 8.47.

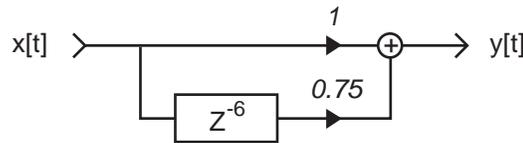


Figure 8.55: A smart way to implement an FIR comb filter with a delay of 6 samples and where $a_0 = 1$ and $a_1 = 0.75$.

$$y_t = 1x_t + 0.75x_{t-6} \quad (8.47)$$

There are stupid ways to do this as well. For example take a look at Figure 8.56 and Equation 8.48. In this case, we have a lot of delays that are implemented (therefore taking up lots of memory) but their output gains are set to 0, therefore they're not being used.

$$y_t = 1x_t + 0x_{t-1} + 0x_{t-2} + 0x_{t-3} + 0x_{t-4} + 0x_{t-5} + 0.75x_{t-6} + 0x_{t-7} + 0x_{t-8} \quad (8.48)$$

We could save a little memory and still be stupid by implementing the algorithm shown in Figure 8.57. In this case, each delay is 1 sample long, arranged in a type of bucket-brigade, but we still have a bunch of unnecessary computation being done.

Let's still be stupid and think of this in a different way. Take a look at the impulse response of our FIR filter, shown in Figure 8.58.

Now, consider each of the values in our impulse response to be the gain value for its corresponding delay time as is shown in Figure 8.56.

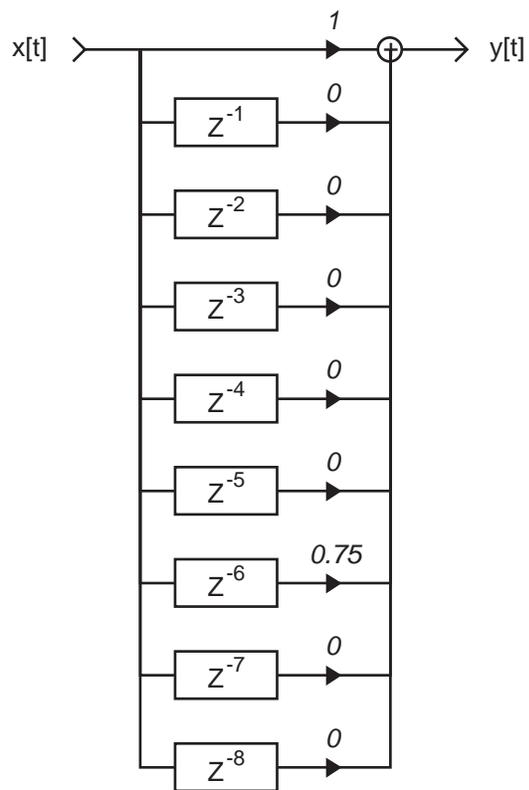


Figure 8.56: A stupid way to implement the same FIR comb filter shown in Figure 8.55.

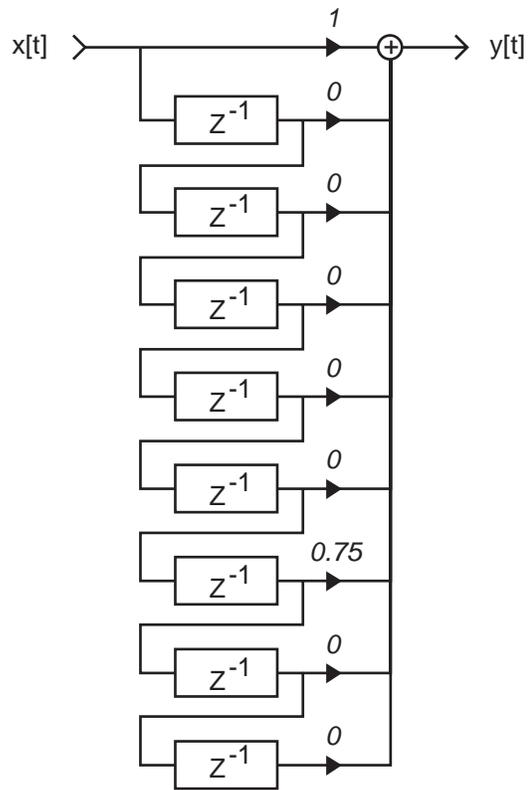


Figure 8.57: A slightly less stupid way to implement the FIR comb filter.

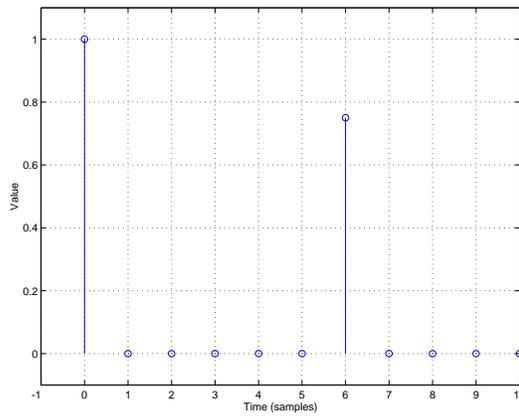


Figure 8.58: The impulse response of the FIR comb filter shown in Figure 8.55.

At time 0, the first signal sample gets multiplied by the first value in the impulse response. That's the value of the output.

At time 1 (1 sample later) the first signal sample gets "moved" to the second value in the impulse response and is multiplied by it. At the same time, the second signal sample is multiplied by the first value in the impulse response. The results of the two multiplications are added together and that's the output value.

At time 2 (1 sample later again...) the first signal sample gets "moved" to the third value in the impulse response and is multiplied by it. The second signal sample gets "moved" to the second value in the impulse response and is multiplied by it. The third signal sample is multiplied by the first value in the impulse response. The results of the three multiplications are added together and that's the output value.

As time goes on, this process is repeated over and over. For each sample period, each value in the impulse response is multiplied by its corresponding sample in the signal. The results of all of these multiplications are added together and that is the output of the procedure for that sample period.

In essence, we're using the values of the samples in the impulse response as individual gain values in a multi-tap delay. Each sample is its own tap, with an integer delay value corresponding to the index number of the sample in the impulse response.

This whole process is called *convolution*. What we're doing is *convolving* the incoming signal with an impulse response.

Of course, if your impulse response is as simple as the one shown above, then it's still really stupid to do your filtering this way because we're essentially doing the same thing as what's shown in Figure 8.57. However, if you have a really complicated impulse response, then this is the way to go (although we're be looking at a smart way to do the math later...).

One reason convolution is attractive is that it gives you an *identical* result as using the original filter that's described by the impulse response (assuming that your impulse response was measured correctly). So, if you can go down to your local FIR filter rental store, rent a good FIR comb filter for the weekend, measure its impulse response and return the filter to the store on Monday. After that, if you convolve your signals with the measured impulse response, it's the same as using the filter. Cool huh?

The reason we like to avoid doing convolution for filtering is that it's just so expensive in terms of computational power. For every sample that comes out of the convolver, its brain had to do as many multiplications as there are samples in the impulse response, and only one fewer additions. For example, if your impulse response was 8 samples long, then the convolver

does 8 multiplications (one for every sample in the impulse response) and 7 additions (adding the results of the 8 multiplications) for every sample that comes out. That's not so bad if your impulse response is only 8 samples long, but what if it's something like 100,000 samples long? That's a lot of math to do on every sample period!

So, now you're probably sitting there thinking, "Why would I have an impulse response of a filter that's 100,000 samples long?" Well, think back to Section 3.16 and you'll remember that we can make an impulse response measurement of a room. If you do this, and store the impulse response, you can convolve a signal with the impulse response and you get your signal in that room. Well, technically, you get your signal played out of the loudspeaker you used to do the IR measurement at that particular location in the room, picked up by the measurement microphone at its particular placement... If you do this in a big concert hall or a church, you could easily get up to a 4 or 5 second-long impulse response, corresponding to a 220,500-sample long FIR filter at 44.1 kHz. This then means 440,999 mathematical operations (multiplications and additions) for every output sample, which in turn means 19,448,066,900 operations per second per channel of audio... That's a lot – far more than a normal computer can perform these days.

So, here's the dilemma, we want to use convolution, but we don't have the computational power to do it the way I just described it. That method, with all the multiplications and additions of every single sample is called *real convolution*.

So, let's think of a better way to do this. We have a signal that has a particular frequency content (or response), and we're sending it through a filter that has a particular frequency response. The resulting output has a frequency content equivalent to the multiplication of these two frequency responses as is shown in Figure 8.59

So, we now know two interesting things:

- putting a signal through filter produces the same result as convolving the signal with the impulse response of the filter.
- putting a signal through a filter produces the same frequency content as multiplying the frequency content of the signal and the frequency response of the filter.

Luckily, some smart people have figured out some clever ways to do a DFT that don't take much computational power. (If you want to learn about this, go get a good DSP textbook and look up the word *butterfly*.)

So, what we can do is the following:

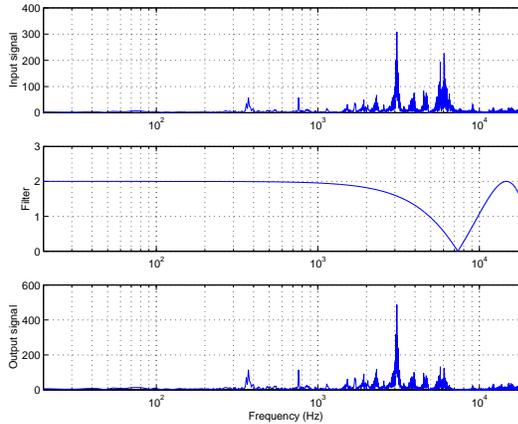


Figure 8.59: The top graph is the frequency content of an arbitrary signal (actually it's a recording of Handel). The middle plot is the frequency response of an FIR comb filter with a 3-sample delay. The bottom graph is the frequency content of the signal filtered through the comb filter. Notice that the result is the same as if we had multiplied the top plot by the middle plot, bin by bin.

1. take a slice of our signal and do a DFT on it. This gives us the frequency content of the signal.
2. take the impulse response of the filter and do a DFT on it.
3. multiply the results of the DFT's keeping the real and imaginary components separate. In other words, you multiply the real components together, and multiply the imaginary components together, bin by bin.
4. take the resulting real and imaginary components and do an IDFT (inverse discrete fourier transform), converting from the frequency domain to the time domain.
5. send the time domain out.

This procedure, called *fast convolution* will give you exactly the same results as if you did real convolution, however you use a lot less computational power.

There are a couple of things to worry about when you're doing fast convolution.

- The window lengths of the two DFT's must be the same.
- When you convolve two signals of length n , then the resulting output has a length of $2n - 1$. This isn't immediately obvious, but it can

be if we think of a situation where one of the signals is the impulse response of reverberation and the other is music. If you play 1024 samples of music in a room with a reverb time of 1024 samples, then the resulting total sound will be 2047 samples because we have to wait for the last sample of music to get through the whole reverberation tail. Therefore, if you're doing fast convolution, then the IDFT has to be twice the length of the two individual DFT's to make the output window twice the length of the input windows. This means that your output is half-way through playing when the new input windows are starting to get through. Therefore you have to mix the new output with the hold. This is what's known as the *overlap-add process*.

8.6.1 Correlation

NOT YET WRITTEN

8.6.2 Autocorrelation

NOT YET WRITTEN

8.6.3 Suggested Reading List

8.7 The z-domain

8.7.1 Introduction

So far, we have seen that digital audio means that we are taking a continuous signal and slicing it into discrete time for transmission, storage, or processing. We have also seen that we can do some math on this discrete time signal to determine its frequency content. We have also seen that we can create digital filters that basically consist of three simple building blocks, addition, multiplication and delay, which can be used to modify the signal.

What we haven't yet seen, however, is how to figure out what a filter does, just by looking at its structure. Nor have we talked about how to design a digital filter. This is where this chapter will come in handy.

One note of warning... This whole chapter is basically a paraphrase of a much, much better introduction to digital signal processing by Ken Steiglitz [Steiglitz, 1996]. If you're really interested in learning DSP, go out and buy this book.

Okay, to get started, let's do a little review. The next couple of equations should be burnt into your brain for the remainder of this chapter, although I'll bring them back now and again to remind you.

The first equation is the one for radian frequency, discussed in Section ??.

$$\omega = 2\pi f \tag{8.49}$$

where f is the normalized frequency (where DC = 0, the Nyquist frequency is 0.5 and the sampling rate is 1) and ω is the radian frequency in radians / samples. Notice that we're measuring time in samples, not in seconds. Unfortunately, our DSP computer can't tell time in seconds, it can only tell time in samples, so we have to accommodate. Therefore, whenever you see a time value listed in this section, it's in samples.

Another equation that you probably learned in high school, and should have read back in Section ?? states that

$$\frac{1}{a} = a^{-1} \tag{8.50}$$

The next equation, which we saw in Section 1.6 is called Euler's Identity. This states that

$$e^{j\omega t} = \cos(\omega t) + j \sin(\omega t) \tag{8.51}$$

You might notice that I made a slight change in Equation 8.51 compared with Equation 1.60 in that I replaced the θ with an ωt to make things a little easier to deal with in the time domain. Remember that ω is just another way of thinking of frequency (see section ??) and t is the time in samples.

We have to make a variation on this equation by adding an extra delay of k samples:

$$e^{j\omega(t-k)} = \cos(\omega(t-k)) + j \sin(\omega(t-k)) \quad (8.52)$$

You should also remember that we can play with exponents as follows:

$$e^{j\omega(t-k)} = e^{(j\omega t)(-j\omega k)} \quad (8.53)$$

$$= e^{(j\omega t)} e^{(-j\omega k)} \quad (8.54)$$

Another basic equation to remember is the following

$$\cos^2(\theta) + \sin^2(\theta) = 1 \quad (8.55)$$

for any value of θ

Therefore, if we make a plot where the x-axis is the value of $\cos(\theta)$ and the y-axis is the value of $\sin(\theta)$ and we plot all the angles from 0 to 2π (or 0° to 360°), we get a circle. (Go back to the discussion on Pythagoras in Section ?? if this isn't obvious.)

If we put a j in front of the $\sin^2(\theta)$ in the equation (so it looks like $\cos^2(\theta) + j \sin^2(\theta)$) a similar thing happens. This means that the $\sin^2(\theta)$ component is now imaginary, and can't be mixed with the $\cos^2(\theta)$. We can still use this equation to draw a perfect circle, but now we're using a real and an imaginary axes as we saw way back in Figure ??.

Okay... that's about it. Now we're ready.

8.7.2 Filter frequency response

Let's take a very simple filter, shown in Figure 8.60 where we add the input signal to a delayed version of the input signal that has been delayed by k samples and multiplied by a gain of a_1 .

We have already seen that this filter can be written as Equation 8.56:

$$y_t = x_t + a_1 x_{t-k} \quad (8.56)$$

Let's put the output of a simple phasor (see Section ??) into the input of this filter. Remember that a phasor is just a sinusoidal wave generator

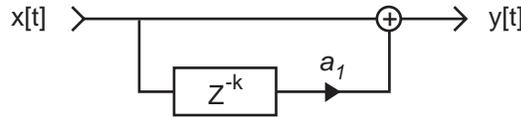


Figure 8.60: A simple filter using one delay of k samples multiplied by a gain and added to the input.

that is “thinking” in terms of a rotating wheel, therefore it has a real and imaginary component. Consequently, we can describe the phasor’s output using $\cos(\omega t) + j \sin(\omega t)$, which, as we saw above can also be written as $e^{j\omega t}$.

Since the output of that phasor is connected to the input of the filter then

$$x_t = e^{j\omega t} \quad (8.57)$$

Therefore, we can rewrite Equation 8.56 for the filter as follows:

$$y_t = e^{j\omega t} + a_1 e^{j\omega(t-k)} \quad (8.58)$$

$$= e^{j\omega t} + a_1 e^{j\omega t} e^{-j\omega k} \quad (8.59)$$

$$= e^{j\omega t} (1 + a_1 e^{-j\omega k}) \quad (8.60)$$

This last equation is actually pretty interesting. Remember that the input, x_t is equal to $e^{j\omega t}$. So what? Well, this means that we can think of the equation as is shown in Figure 8.61.

$$y[t] = e^{j\omega t} (1 + a_1 e^{-j\omega k})$$

input what the input
 is multiplied
 by to get the
 output

Figure 8.61: An intuitive way of thinking of Equation 8.60.

Notice that the end of Equation 8.60 has a lot of ω ’s in it. This means that it is frequency-dependent. We already know that if you multiply a signal by something, that “something” is gain. We also know that if that

gain is dependent on frequency (meaning that it is different at different frequencies) then it will change the frequency response. So, since we know all that, we should also see that the “what the input is multiplied by to get the output” part of Equation 8.60 is a frequency-dependent gain, and therefore is a mathematical description of the filter’s *frequency response*.

Please make sure that this last paragraph made sense before you move on...

Let’s be lazy and re-write Equation 8.60 a little simpler. We’ll invent a symbol that means “the frequency response of the filter.” We know that this will be a frequency-dependent function, so we’ll keep an ω in there, so let’s use the symbol $H(\omega)$. Although, if we need to describe another filter, we can use a different letter, like $G(\omega)$ for example...

We know that the frequency response of our filter above is $1 + a_1 e^{-j\omega k}$ and we’ve already decided that we’re going to use the general symbol $H(\omega)$ to replace that messy math, but just to make things clear, let’s write it down...

$$H(\omega) = 1 + a_1 e^{-j\omega k} \quad (8.61)$$

One important thing to notice here is that the above equation doesn’t have any t ’s in it. This means that it is not dependent on what time it is, so it never changes. This is what is commonly known as a *Linear Time Invariant* or *LTI* system. This means two things. The “linear” part means that everything that comes in is treated the same way, no matter what. The “time invariant” part means that the characteristics of the filter never change (or vary) over time.

Things are still a little complicated because we’re still working with complex numbers, so let’s take care of that. We have seen that the reason for using complex notation is to put magnitude and phase information into one neat little package (see Section ?? for this discussion.) We have also seen that we can take a signal represented in complex notation and pull the magnitude and phase information back out if we need to. This can also be done with $H(\omega)$ (the frequency response) of our filter. Equation ?? is a general form showing how to do this.

$$H(\omega) = |H(\omega)| e^{j\theta(\omega)} \quad (8.62)$$

This divides the frequency response $H(\omega)$ into the filter’s magnitude response $|H(\omega)|$ and the phase shift $\theta(\omega)$ at frequency ω . Be careful here not to just to conclusions. The phase shift is not θ , it’s $\theta(\omega)$ but we can calculate to make a little more sense intuitively.

$$\theta(\omega) = \arctan \frac{\Im H(\omega)}{\Re H(\omega)} \quad (8.63)$$

Where the symbols \Im means “The imaginary component of...” and \Re means “The real component of...”

Don’t panic – it’s not so difficult to see what the real and imaginary components of the $H(\omega)$ are because the imaginary components have a j in them. For example, if we look at the filter that we started with at the beginning of this chapter, its frequency response was:

$$H(\omega) = 1 + a_1 e^{-j\omega k} \quad (8.64)$$

$$= 1 + a_1 (\cos(\omega k) - j \sin(\omega k)) \quad (8.65)$$

$$= 1 + a_1 \cos(\omega k) - a_1 j \sin(\omega k) \quad (8.66)$$

So this means that

$$\Re H(\omega) = 1 + a_1 \cos(\omega k) \quad (8.67)$$

and

$$\Im H(\omega) = -a_1 \sin(\omega k) \quad (8.68)$$

Notice that I didn’t put the j in there because I don’t need to say that it’s imaginary... I already know that it’s the imaginary component that I’m looking for.

So, to calculate the phase of our example filter, we just put those two results into Equation 8.63 as follows:

$$\theta(\omega) = \arctan \frac{\Im H(\omega)}{\Re H(\omega)} \quad (8.69)$$

$$= \arctan \frac{-a_1 \sin(\omega k)}{1 + a_1 \cos(\omega k)} \quad (8.70)$$

8.7.3 The z-plane

Let’s look at an even simpler example before we go too much further. Figure 8.62 shows a simple delay line of k samples with the result multiplied by a gain factor of a_1 .

This delay can be expressed as the equation:

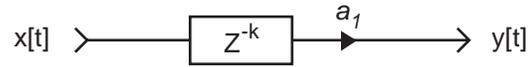


Figure 8.62: A delay line of k samples multiplied by a gain.

$$y_t = a_1 x_{t-k} \quad (8.71)$$

$$= x \left(a_1 e^{-j\omega k} \right) \quad (8.72)$$

As you've probably noticed by now, we tend to write $e^{j\omega}$ a lot, so let's be lazy again and use another symbol instead – we'll use z , therefore

$$z = e^{j\omega} \quad (8.73)$$

What if you want to write a delay, as in $e^{-j\omega k}$? Well, we'll just write (you guessed it...) z^{-k} . Therefore z^2 is a 2-sample delay because $e^{-j\omega 2}$ is a 2-sample delay as well.

Always remember that z is a complex variable. In other words,

$$z = e^{j\omega} \quad (8.74)$$

$$= \cos(\omega) + j \sin \omega \quad (8.75)$$

Therefore, we can plot the value of z on a Cartesian graph with a real and an imaginary axis, just like we're always been doing.

8.7.4 Transfer function

Let's take a signal and send it through two gain functions connected in series as is shown in Figure 8.63.

We can think of this as two separate modules as is shown in Figure 8.64.

So, let's calculate how y_t relates to x_t . Using Figure 8.64, we can write the following equations:

$$w[t] = a_0 x_t \quad (8.76)$$

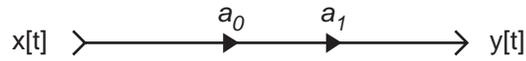


Figure 8.63: Two gain stages in series.

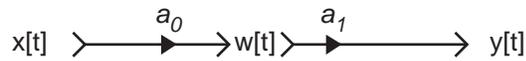


Figure 8.64: A simplified way of thinking of Figure 8.63.

and

$$y_t = a_1 w[t] \quad (8.77)$$

Therefore

$$y_t = a_1 w[t] \quad (8.78)$$

$$= a_1 (a_0 x_t) \quad (8.79)$$

$$= a_0 a_1 x_t \quad (8.80)$$

What I took a long while to say here is that, if you have two gain stages (or, more generally, two filters) connected in series, you can multiply the effects that they have on the input to find out the output.

So far, we have been thinking in terms of the instantaneous value of each sample, sample by sample. That's why we have been using terms like x_t and y_t – we've been quite specific about which sample we're talking about. However, since we're dealing with LTI systems, we can say that the same thing happens to every sample in the signal. Consequently, we don't have to think in terms of individual samples going through the filter, we can think of the whole signal (say, a Johnny Cash tune, for example). When we speak of the whole signal, we use things like X and Y instead of x_t and y_t . That whole signal is what is called an *operator*.

In the case of the last filter we looked at (the one with the two gains in series shown in Figure 8.63), we can write the equation in terms of operators.

$$Y = [a_0 a_1] X \quad (8.81)$$

In this case, we can see that the operator X is multiplied by $a_0 a_1$ in order to become Y . This is a simple way to think of DSP if you're not working in real time. You take the Johnny Cash tune on your hard disc. Multiply its contents by $a_0 a_1$ and you get an output, which is a Johnny Cash tune at a different level (assuming that $a_0 a_1 \neq 1$).

That thing that the input operator is multiplied by is called the *transfer function* of the filter and is represented by the symbol $H(z)$. So, the transfer function of our previous filter is $a_0 a_1$.

Notice that the transfer function in the equation is held in square brackets.

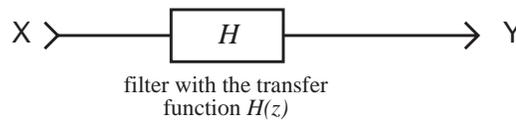


Figure 8.65: Transfer function of a filter with operators as inputs and outputs.

Let's look at another simple example using two 1-sample delays connected in series as is shown in Figure 8.66



Figure 8.66: Two delays connected in series.

The equation for this filter would be

$$y_t = x_t z^{-1} z^{-1} \quad (8.82)$$

$$= x_t z^{-2} \quad (8.83)$$

and therefore

$$Y = X [z^{-2}] \quad (8.84)$$

Notice that I was able to simply multiply the two delays z^{-1} and z^{-1} together to get z^{-2} . This is one of the slick things about using z^{-k} to indicate a delay. If you multiply them together, you just add the exponents, therefore adding the delay times.

Let's do one more filter before moving on...

Figure 8.67 shows a fairly simple filter which adds a gain-modified version of the input with a delayed, gain-modified version of the input.

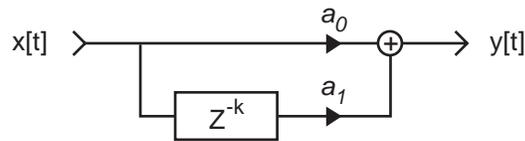


Figure 8.67: A filter which adds a gain-modified version of the input with a delayed, gain-modified version of the input.

The equation for this filter is

$$y_t = a_0 x_t + a_1 x_{t-k} \quad (8.85)$$

$$= a_0 x_t + a_1 z^{-k} x_t \quad (8.86)$$

$$= (a_0 + a_1 z^{-k}) x_t \quad (8.87)$$

Therefore

$$Y = [a_0 + a_1 z^{-k}] X \quad (8.88)$$

and so $a_0 + a_1 z^{-k}$ is the transfer function of the filter.

8.7.5 Calculating filters as polynomials

Let's take this a step further. We'll build a big filter out of two smaller filters as is shown in Figure 8.68. The transfer function $G(z) = a_0 + a_1 z^{-1}$ while the transfer function $H(z) = b_0 + b_1 z^{-1}$.

$$Y = H(z)W \quad (8.89)$$

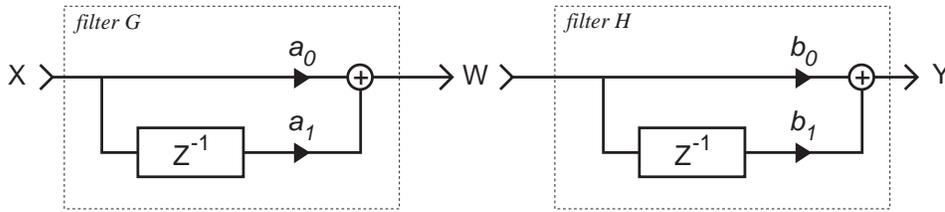


Figure 8.68: Two filters in series.

$$W = G(z)X \quad (8.90)$$

Therefore

$$Y = G(z)H(z)X \quad (8.91)$$

$$= (a_0 + a_1z^{-1})(b_0 + b_1z^{-1})X \quad (8.92)$$

$$= a_0b_0 + a_0b_1z^{-1} + b_0a_1z^{-1} + a_1z^{-1}b_1z^{-1} \quad (8.93)$$

$$= (a_0b_0 + (a_0b_1 + b_0a_1)z^{-1} + a_1b_1z^{-1})X \quad (8.94)$$

Therefore

$$y_t = a_0b_0x_t + (a_0b_1 + b_0a_1)x_{t-1} + a_1b_1x_{t-2} \quad (8.95)$$

In other words, we could implement exactly the same filter as is shown in Figure 8.69. Then again, this might not be the smartest way to implement it. Just keep in mind that there are other ways to get the same result. We know that these will give the same result because their mathematical expressions are identical.

The thing to notice about this section was the way I manipulated the equation for the cascaded filters. I just treated the notation as if it was a polynomial straight out of a high-school algebra textbook. The cool thing is that, as long as I don't screw up, every step of the way describes the same filter in an understandable way (as long as you remember what z^{-k} means...).

8.7.6 Zeros

Let's go back to a simpler filter as is shown in Figure 8.70.

This gives us the equation

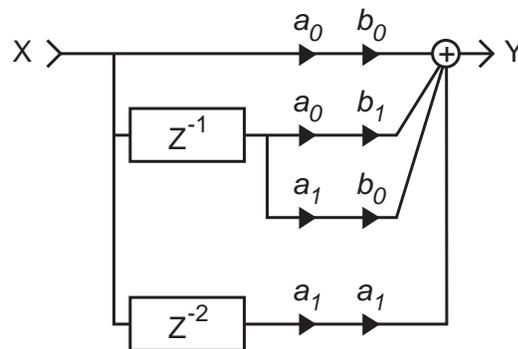


Figure 8.69: An equivalent result to the cascaded filters shown in Figure 8.68

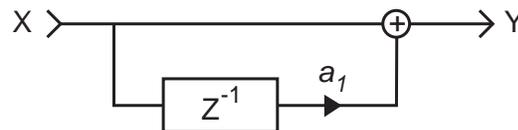


Figure 8.70: A simple filter.

$$y_t = x_t + a_1 x_{t-1} \quad (8.96)$$

If we write this out the “old way” – that is, before we knew about the letter z , the equivalent would look like this:

$$y_t = e^{j\omega t} + a_1 e^{j\omega t} e^{-j\omega} \quad (8.97)$$

$$= [1 + a_1 e^{-j\omega}] e^{j\omega t} \quad (8.98)$$

So $1 + a_1 e^{-j\omega}$ is “what the filter does” to the input $e^{j\omega t}$. Think of $1 + a_1 e^{-j\omega}$ as a function of frequency. It is complex (therefore it causes a phase shift in the signal). Consequently, we say that it is a complex function of ω . In the z -domain, the equivalent is $1 - a_1 z^{-1}$.

So, since we know that

$$1 - a_1 z^{-1} = 1 + a_1 e^{-j\omega} \quad (8.99)$$

then we can say that

$$H(z) = H(\omega) \quad (8.100)$$

In other words, *the transfer function is the same as the frequency response of the filter.*

We know that the transfer function of this filter is $1 - a_1 z^{-1}$. You should also remember that $\frac{1}{z} = z^{-1}$. Therefore,

$$H(z) = 1 + a_1 z^{-1} \quad (8.101)$$

$$= \frac{z}{z} + \frac{a_1}{z} \quad (8.102)$$

$$= \frac{z + a_1}{z} \quad (8.103)$$

Okay, now I have a question. What values in the above equation make the numerator (the top part of the fraction) equal to 0. The answer, in this particular case is $z = -a_1$. If the filter had been different, we would have gotten a different answer. (Maybe we would have even gotten more than one answer... but that comes later.) Notice that if the numerator is 0, then $H(z) = 0$. This will be important later.

Another question: What values will make the denominator (the bottom part of the equation) equal to 0? The answer, in this particular case is $z = 0$.

Again, it might have been possible to get more than one answer. Notice in this case, that if the denominator is 0, then $H(z) = \infty$.

Let's graph z on a cartesian plot. We'll make the x-axis the real axis where we plot the real component of z , and the y-axis is the imaginary axis.

We know that

$$z = e^{j\omega} \quad (8.104)$$

$$= \cos(\omega) + j \sin(\omega) \quad (8.105)$$

We also know that

$$\cos^2(\omega) + \sin^2(\omega) = 1 \quad (8.106)$$

for any value of ω .

Therefore, if we make a Cartesian plot of z then we get a circle with a radius of 1 as is shown in Figure 8.71. Notice that the amount of rotation corresponds to the radian frequency, starting at DC where $z = 1$ (notice that there is no imaginary component) and going to the Nyquist frequency at $z = -1$. When $z = j1$, we are at a normalized frequency of 0.25, or one-quarter of the sampling rate. This might not make sense yet, but stick with me for a while.

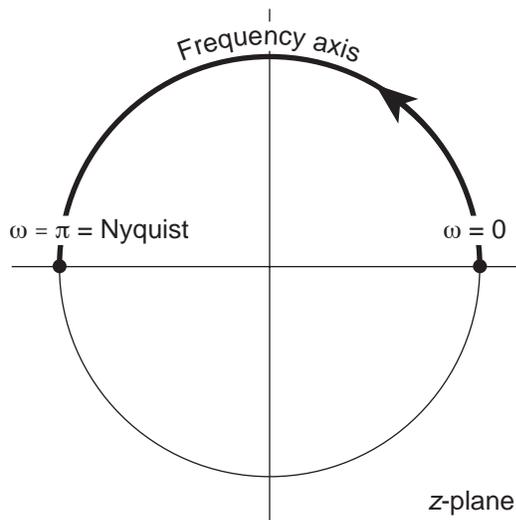


Figure 8.71: The frequency axis in the z -plane. The top half of the circle corresponds to positive frequencies, the bottom half corresponds to negative frequencies [?].

If it doesn't make sense, it might be because the problem is that it appears that we are confusing frequency and phase. However, in one sense, they are directly connected, because we are describing frequency in terms of radians per sample instead of the more usual complete cycles of the waveform per second. Normally we describe the frequency of a waveform by how many complete cycles go by each second. For example, we say "a 1 kHz tone" meaning that we get 1000 complete cycles each second. Alternatively, we could say "360000 degrees per second" which says the same thing. If we counted faster than a second, then we would say something like "360 degrees per millisecond" or "0.36 degrees per microsecond." These all say the same thing. Notice that we're using the rate of change of the phase to describe the frequency, just like we do when we use ω .

If we back up a bit to the last filter we were working on, we said that its transfer function was

$$H(z) = \frac{z + a_1}{z} \quad (8.107)$$

We can therefore easily calculate the magnitude of $H(z)$ using the equation:

$$|H(z)| = \frac{|z + a_1|}{|z|} \quad (8.108)$$

An important thing to remember here is that, when you're dealing with complex numbers, as we are at the moment, the bars on the sides of the values do not mean "the absolute value of..." as they normally do. They mean "the magnitude of..." instead. Also remember that you calculate the magnitude by finding the hypotenuse of the triangle where the real and imaginary components are the other two sides (see Figure ??).

Since

$$z = \cos(\omega) + j \sin(\omega) \quad (8.109)$$

and

$$\cos^2(\omega) + j \sin^2(\omega) = 1 \quad (8.110)$$

and, because of Pythagoras

$$|z| = \cos^2(\omega) + j \sin^2(\omega) \quad (8.111)$$

then

$$|z| = 1 \quad (8.112)$$

So, this means that

$$|H(z)| = \frac{|z + a_1|}{|z|} \quad (8.113)$$

$$= \frac{|z + a_1|}{1} \quad (8.114)$$

$$= |z + a_1| \quad (8.115)$$

We found out from our questions and answers at the end of Section 8.7.6 that (for our filter that we're still working on) $H(z) = 0$ when $z = -1$. Let's then mark that point with a "0" (a *zero*) on the graph of the z -plane. Note that a_1 has only a real component, therefore $-a_1$ has only a real component as well. Consequently, the "0" that we put on the graph sits right on the real axis.

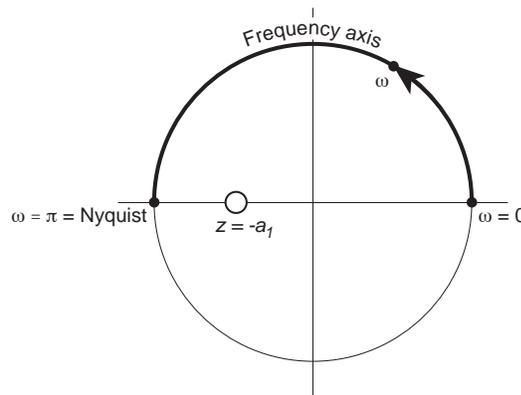


Figure 8.72: A zero plotted on the z -plane

Here's where things get interesting (finally!). The magnitude response (what most normal people call the frequency response) of this filter can be found by measuring the distance from the "0" that we plotted to the circle on the graph. See Figure 8.73

You might be sitting there wondering "what happens if I have more than one zero on the plot?" This is actually pretty simple, you just pick a value for ω , then measure the distance from each zero on the plot to the

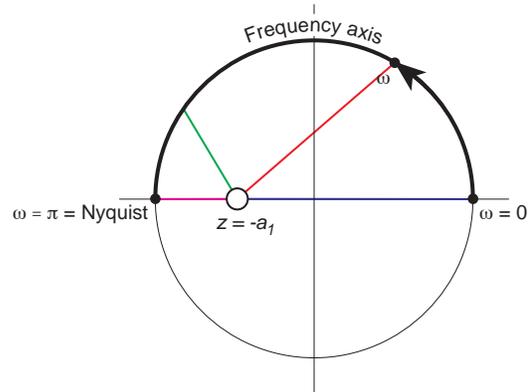


Figure 8.73: Finding the magnitude response of the filter using the zero plotted on the z-plane.

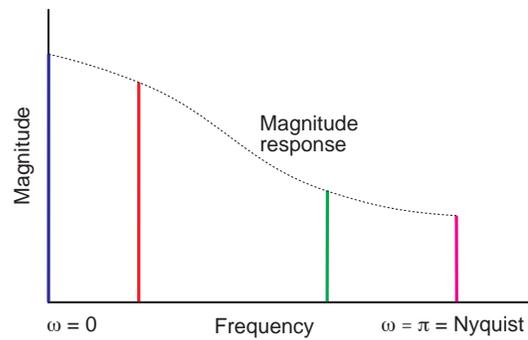


Figure 8.74: An equivalent plot to 8.73, if we “unwrapped” the frequency axis from a circle to a straight line. Note that this plot is only roughly to scale – it is not terribly accurate.

frequency ω . Then you multiply all your distances together, and the result is the magnitude response of the filter at that frequency.

There is another, possibly more intuitive way of thinking about this. Cut a circle out of a sheet of heavy rubber, and magically suspend it in space. The circle directly corresponds to the z -plane. If you have a zero on the z -plane, you push the rubber down with a pointy stick as far as possible, preferably infinitely. This will put a dent in the rubber that looks a bit like a funnel. This will pull down the edge of the circle. The closer the zero is to the edge, the more the edge will get pulled down.

If you were able to unwrap the edge of the circle, keeping its vertical shape, you would have a picture of the frequency response.

MAKE A 3-D PLOT OF THIS IN MATHEMATICA TO ILLUSTRATE THIS POINT.

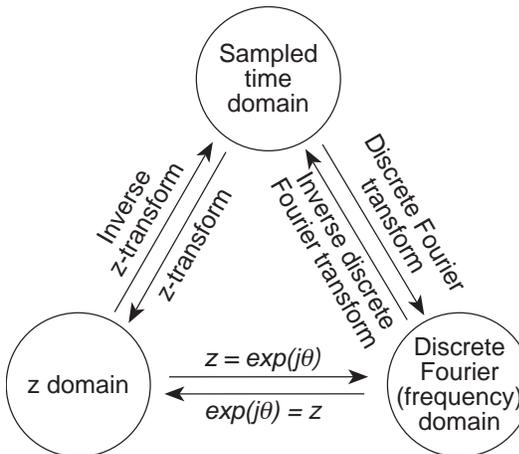


Figure 8.75: The relationship between the sampled time, discrete frequency and z domains [Watkinson, 1988].

8.7.7 Poles

So far we have looked at how to calculate the frequency response (therefore the magnitude and phase responses) of a digital filter using zeros in the z -plane. However, you may have noticed that we have only discussed FIR filters in this section. We have not looked at what happens when you introduce feedback, therefore creating an IIR filter.

So, let's make a simple IIR filter with a single delayed feedback with gain as is shown in Figure 8.76.

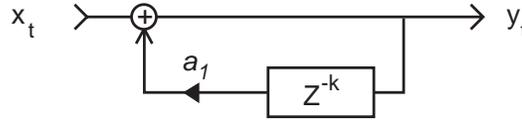


Figure 8.76: A simple IIR filter with one delay and one gain.

The equation for this filter is

$$y_t = x_t + a_1 y_{t-k} \quad (8.116)$$

Notice now that the output is the sum of the input (the x_t) and a gain-modified version (using the gain a_1) of the output y_t with a delay of k samples.

We already know from the previous section how to write this equation symbolically using operators (think back to the Johnny Cash example) as is shown in the equation below.

$$Y = X + a_1 z^{-k} Y \quad (8.117)$$

Let's do a little standard high-school algebra on this equation...

$$Y = X + a_1 z^{-k} Y \quad (8.118)$$

$$Y - a_1 z^{-k} Y = X \quad (8.119)$$

$$X = Y - a_1 z^{-k} Y \quad (8.120)$$

$$X = [1 - a_1 z^{-k}] Y \quad (8.121)$$

So, in a weird way, we can think of this IIR filter that relies on feedback as a FIR filter that only contains feedforward, except that we have to think of the filter backwards. The input, X is simply the output of the filter, Y , minus a delayed, gain-varied, version of the output. This can be seen intuitively if you compare Figure 8.60 with Figure 8.76. You'll notice that, if you look at one of the diagrams backwards (following the signal path from right to left), they're identical.

Let's continue with the algebra, picking up where we left off...

$$X = [1 - a_1 z^{-k}] Y \quad (8.122)$$

$$\frac{X}{1 - a_1 z^{-k}} = Y \quad (8.123)$$

$$Y = \frac{X}{1 - a_1 z^{-k}} \quad (8.124)$$

$$Y = X \frac{1}{1 - a_1 z^{-k}} \quad (8.125)$$

Therefore, we can see right away that the *transfer function* of this IIR filter is

$$H(z) = \frac{1}{1 - a_1 z^{-k}} \quad (8.126)$$

because that's what gets multiplied by the input to get the output.

This, in turn, means that the *magnitude response* of the filter is

$$|H(z)| = \left| \frac{1}{1 - a_1 z^{-k}} \right| \quad (8.127)$$

$$= \frac{1}{|1 - a_1 e^{-i\omega k}|} \quad (8.128)$$

This raises an interesting problem. What happens if we make the denominator equal to 0? First of all, let's find out what values make this happen.

$$0 = 1 - a_1 z^{-k} \quad (8.129)$$

$$1 = a_1 z^{-k} \quad (8.130)$$

$$\frac{1}{a_1} = z^{-k} \quad (8.131)$$

$$\frac{1}{a_1} = e^{-j\omega k} \quad (8.132)$$

Remember Euler's identity which states that

$$e^{-j\omega k} = \cos(\omega k) + j \sin(\omega k) \quad (8.133)$$

We therefore know that, for any value of ω and k , $-1 \leq e^{-j\omega k} \leq 1$. In other words, a sinusoidal waveform just swings back and forth over time between -1 and 1.

We saw that the denominator of the equation for the magnitude response of our filter will be 0 when

$$\frac{1}{a_1} = e^{-j\omega k} \quad (8.134)$$

Therefore, the magnitude response of the filter will be 0 (at some values for ω and k) when $a_1 \geq 1$.

So what? Well, if the denominator in the equation describing the magnitude response of the filter goes to 0, then the magnitude response goes to ∞ . Therefore, no matter what the input of the filter (other than an input of 0), the output will be ∞ . This is bad because an infinite output signal is very, very loud... An intuitive way to think of this is that the filter goes nuts if the feedback causes it to get louder and louder over time. In the case of a single feedback delay, this is easy to think of, since it's just a question of whether the gain applied to that delay outputs a bigger result than the input, which is fed into the delay again and made bigger and bigger and so on. In the case of a complicated filter lots of feedback and feedforward paths (and therefore lots of poles and zeros) you'll have to do some math to figure out what will make the filter get out of control.

We saw that, if the numerator of the magnitude response equation of the filter goes to 0, then we put a *zero* in the z -plane plot of the response. Similarly, if the denominator of the equation goes to 0, then we put a *pole* on the z -plane plot of the response. Whereas a zero can be thought of as a deep “dent” in the 3D plot of the z -plane, a pole is a very high pointy shape coming up off the z -plane. This is shown in Figures 8.77 and ??.

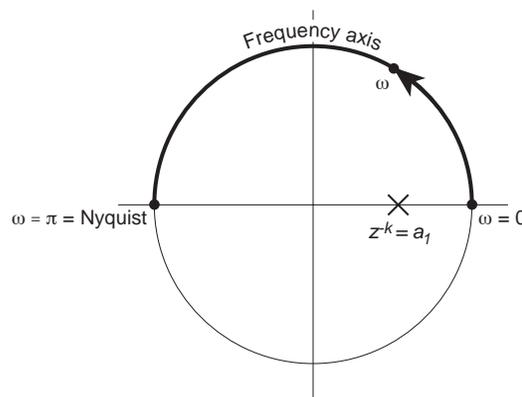


Figure 8.77: A pole plotted on the z -plane

MAKE A 3-D PLOT OF THIS IN MATHEMATICA TO ILLUSTRATE THIS POINT.

Notice in Figure 8.77 that I have marked the pole using an “X” and placed it where $z^{-k} = a_1$. This may seem odd, since we said a couple of equations ago that the magic number is where $\frac{1}{a_1} = z^{-k}$. I agree – it is a little odd, but just trust me for now. We put the pole on the plot where $z^{-k} =$ the bottom half of the equation (in this case, a_1) and we’ll see in a couple of paragraphs how it ties back to $\frac{1}{a_1}$.

As we saw before, one way to intuitively think of the result of a pole on the magnitude response of the filter is to unwrap the edge of the circle and see how the pole pulls up the edge. This will give you an idea of where the total gain of the filter is higher than 1 relative to frequency.

Alternately, we can think of the magnitude response, again using the distance from the pole’s location to the edge of the circle in the z -plane. However, there is a difference in the way we think of this when compared with the zeros. In the case of the zeros, the magnitude response at a given frequency was simply the distance between the zero and the edge of the circle as we saw in Figures 8.73 and 8.74. We also saw that, if you have more than one zero, you multiply their distances (to the location of a single frequency on the circle) together to get the total gain of the filter for that frequency. The same is almost true for poles, however you divide the distance instead of multiplying. So, in the case of a filter with only one pole, you divide 1 by the distance from the pole’s location to a frequency on the circle to get the gain of the filter at that frequency. This is shown in Figures 8.78 and 8.79. (See, I *told* you that I’d link it back to $\frac{1}{a_1}$...)

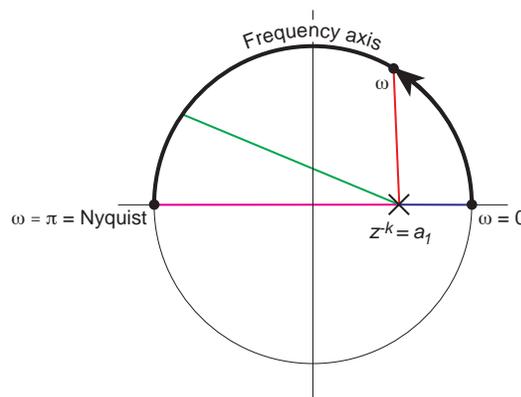


Figure 8.78: Finding the magnitude response of the filter using the pole plotted on the z -plane.

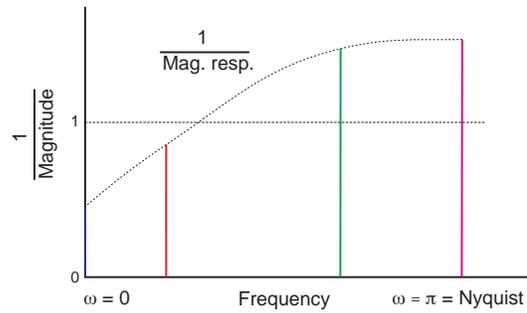


Figure 8.79: An equivalent plot to 8.78, if we “unwrapped” the frequency axis from a circle to a straight line. Notice that the plot shows $\frac{1}{|H(z)|}$. Note that this plot is only roughly to scale – it is not terribly accurate.

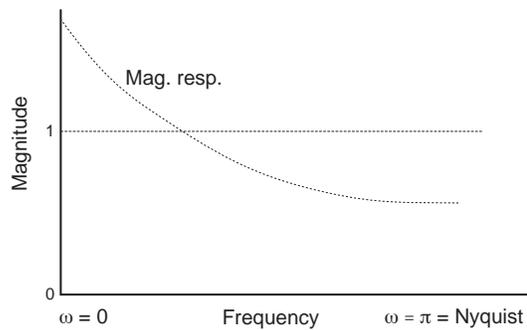


Figure 8.80: The magnitude response calculated from the plot in Figure 8.79. Note that this plot is only roughly to scale – it is not terribly accurate.

8.7.8 Suggested Reading List

[Steiglitz, 1996]

Chapter 9

Audio Recording

9.1 Levels and Metering

Thanks to Claudia Haase and Thomas Lischker at RTW Radio–Technische (www.rtw.de) for their kind permission to use graphics from their product line for this page.

9.1.1 Introduction

When you sit down to do a recording – any recording, you have two basic objectives:

- 1) make the recording sound nice aesthetically
- 2) make sure that the technical quality of the recording is high.

Different people and record labels will place their priorities differently (I’m not going to mention any names here, but you know who you are...)

One of the easiest ways to guarantee a high technical quality is to pay particular attention to your gain and levels at various points in the recording chain. This sentence is true not only for the signal as it passes out and into various pieces of equipment (i.e. from a mixer output to a tape recorder input), but also as it passes through various stages within one piece of equipment (in particular, the signal level as it passes through a mixer). The question is: “what’s the best level for the signal at *this* point in the recording chain?”

There are two beasts hidden in your equipment that you are constantly trying to avoid and conceal as you do your recording. On a very general level, these are *noise* and *distortion*.

Noise

Noise can be generally defined as any audio in the signal that you don't want there. If we restrict ourselves to electrical noise in recording equipment, then we're talking about hiss and hum. The reasons for this noise and how to reduce it are discussed in a different chapter, however, the one inescapable fact is that noise cannot be avoided. It can be reduced, but never eliminated. If you turn on any piece of audio equipment, or any component within any piece of equipment, you get noise. Normally, because the noise stays at a relatively constant level over a long period of time and because we don't bother recording signals lower in level than the noise, we call it a *noise floor*.

How do we deal with this problem? The answer is actually quite simple: we turn up the level of the signal so that it's much louder than the noise. We then rely on *psychoacoustic masking* (and, if we're really lucky, the *threshold of hearing*) to cover up the fact that the noise is there. We don't eliminate the noise, we just hide it – and the louder we can make the signal, the better it's hidden. This works great, except that we can't keep increasing the level of the signal because at some point, we start to distort it.

Distortion

If the recording system was absolutely perfect, then the signal at its output would be identical to the signal at the input of the microphone. Of course, this isn't possible. Even if we ignore the noise floor, the signals at the two ends of the system are not identical – the system itself modifies or *distorts* the signal a little bit. The less the modification, the lower the distortion of the signal and the better it sounds.

Keep in mind that the term “distortion” is extremely general – different pieces of equipment and different systems will have different detrimental effects on different signals. There are different ways of measuring this – these are discussed in the section on electroacoustic measurements – but we typically look at the amount of distortion in percent. This is a measurement of how much extra power is included in the signal that shouldn't be there. The higher the percentage, the more distortion and the worse the signal. (See the chapter on distortion measurements in the Electroacoustic Measurements section.)

There are two basic causes of distortion in any given piece of equipment. The first is the normal day-to-day error of the equipment in transmitting or recording the signal. No piece of gear is perfect, and the error that's added to the signal at the output is basically always there. The second, however,

is a distortion of the signal caused by the fact that the level of the signal is too high. The output of every piece of equipment has a maximum voltage level that cannot be exceeded. If the level of the signal is set so high that it should be greater than the maximum output, then the signal is *clipped* at the maximum voltage as is shown in Figure 9.2.

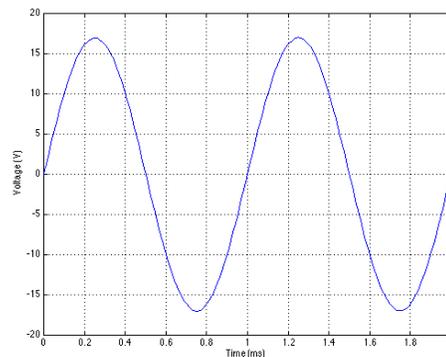


Figure 9.1: A 1 kHz sine wave without distortion worth talking about.

For our purposes at this point in the discussion, I’m going to oversimplify the situation a bit and jump to a hasty conclusion. Distortion can be classified as a process that generates unwanted signals that are added to our program material. In fact, this is exactly what happens – but the unwanted signals are almost always harmonically related to the signal whereas your run-of-the-mill noise floor is completely unrelated harmonically to the signal. Therefore, we can group distortion with noise under the heading “stuff we don’t want to hear” and look at the level of that material as compared to the level of the program material we’re recording – in other words the “stuff we do want to hear.” This is a small part of the reason that you’ll usually see a measurement called “THD+N” which stands for “Total Harmonic Distortion plus Noise” – the stuff we don’t want to hear.

Maximizing your quality

So, we need to make the signal loud enough to mask the noise floor, but quiet enough so that it doesn’t distort, thus maximizing the level of the signal compared to the level of the distortion and noise components. How do we do that? And, more importantly, how do we keep the signal at an optimal level so that we have the highest level of technical quality? In

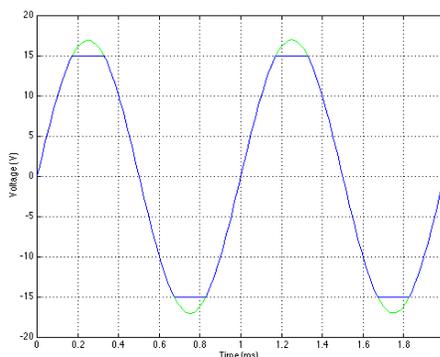


Figure 9.2: The same 1 kHz sine wave in a piece of equipment that has a maximum voltage of 15 V (and a minimum voltage of -15 V). Note that the top and bottom of the sine wave are clipped at the voltage rails of the equipment. This clipping causes a high distortion level because the signal is significantly changed or distorted. The green waveform is the original undistorted sine wave and the blue is the clipped output.

order to answer this question, we have to know the exact behaviour of the particular piece of gear that we're using – but we can make some general rules that apply for groups of gear. These three groups are 1) digital gear, 2) analog electronics and 3) analog tape.

9.1.2 Digital Gear in the PCM World

As we've seen in previous chapters, digital gear has relatively easily defined extremes for the audio signal. The noise floor is set by the level of the dither, typically with a level of one half of an LSB. The signal to noise ratio of the digital system is dependent on the number of bits that are used for the signal – increasing by 6.02 dB per bit used. Since the level of the dither is typically half a bit in amplitude, we subtract 3 dB from our signal to noise ratio calculated from the number of bits. For example, if we are recording a sine wave that is using 12 of the 16 bits on a CD and we make the usual assumptions about the dither level, then the signal to noise ratio for that particular sine wave is:

$$\begin{aligned} & (12 \text{ bits} * 6 \text{ dB per bit}) - 3 \text{ dB} \\ & = 69 \text{ dB} \end{aligned}$$

Therefore, in the above example, we can say that the noise floor is 69 dB below the signal level. The more bits we use for the signal (and therefore the higher its peak level) the greater the signal to noise ratio and therefore

the better the technical quality of the recording. (Do not confuse the signal to noise ratio with the dynamic range of the system. The former is the ratio between the signal and the noise floor. The latter is the ratio between the maximum possible signal and the noise floor – as we’ll see, this raises the question of how to define the maximum possible level...)

We also know from previous chapters that digital systems have a very unforgiving maximum level. If you have a 16 bit system, then the peak level of the signal can only go to the maximum level of the system defined by those 16 bits. There is some debate regarding what you can get away with when you hit that wall – some people say that 2 consecutive samples at the maximum level constitutes a clipped signal. Others are more lenient and accept one or two more consecutively clipped samples. Ignoring this debate, we can all agree that, once the peak of a sine wave has reached the maximum allowable level in a digital system, any increase in level results in a very rapid increase in distortion. If the system is perfectly aligned, then the sine wave starts to approach a square wave very quickly (ignoring a very small asymmetry caused by the fact that there is one extra LSB for the negative-going portion of the wave than there is for the positive side in a PCM system). See Figure 9.2 to see a sample input and output waveform. The “consecutively clipped samples” that we’re talking about is a measurement of how long the flattened part of the waveform stays flat.

If we were to draw a graph of this behaviour, we would result in the plot shown in Figure 9.3. Notice that we’re looking at the Signal to THD+N ratio vs. the level of the signal.

The interesting thing about this graph is that it’s essentially a graph of the peak signal level vs. audio quality (at least technically speaking... we’re not talking about the quality of your mix or the ability of your performers...). We can consider that the X-axis is the peak signal level in dB FS and the Y-axis is a measurement of the quality of the signal. Consequently, we can see that the closer we can get the peak of the signal to 0 dB FS the better the quality, but if we try to increase the level beyond that, we get very bad very quickly.

Therefore, the general moral of the story here is that you should set your levels so that the highest peak in the signal for the recording will hit as close to 0 dB FS as you can get without going over it. In fact, there are some problems with this – you may actually wind up with a signal that’s greater than 0 dB FS by recording a signal that’s less than 0 dB FS in some situations... but we’ll look at that later... this is still the introduction.

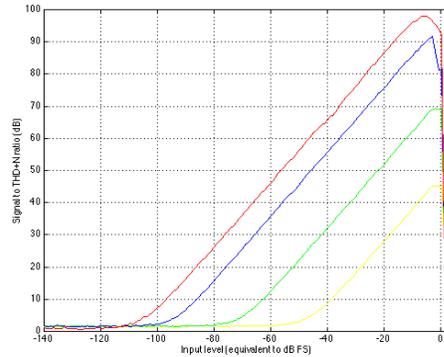


Figure 9.3: A plot of a measurement of the signal to THD+N (caused by noise and distortion byproducts) ratio vs. the signal level in a typical digital converter with a dither level of one half an LSB measured with a 997 Hz sine tone. The curves are 8-bit (yellow), 12-bit (green), 16-bit (blue) and 24-bit (red). The resolution on the input level is 1 dB. The positive slope on the left is the result of the increase in the signal level over the static noise floor. The nasty drop on the right is caused by the sudden increase in distortion when you try to make the sine tone go beyond 0 dB FS.

9.1.3 Analog electronics

Analog electronics (well, operational amplifiers really... but pretty well everything these days is built with op amps, so we'll stick with the generalized assumption for now...) have pretty much the same distortion characteristics as digital, but with a lower noise floor (unless you have a very high resolution digital system or a really crappy analog system). As can be seen in Figure 9.4, the general curve for an analog microphone preamplifier, mixing console, or equalizer (note that we're not talking about dynamic range controllers like compressors, limiters, expanders and gates) looks the same as the curve for digital gear shown in Figure 9.3. If you've got a decent piece of analog gear (even something as cheap as a Mackie mixer these days) then you should be able to hit a maximum signal to noise ratio of about 125 dB or so when the signal is at some maximum level where the peak level is bordering on the clipping level (somewhere around +/- 13.5 V or +/- 16.5 V, depending on the power supply rails and the op amps used). Any signal that goes beyond that peak level causes the op amps to start clipping and the distortion goes up rapidly (and bringing the quality level down quickly).

So, the moral of the story here is the same as in the digital world. As a general rule, it's good for analog electronics to keep your signal as high as possible without hitting the maximum output level and therefore clipping

your signal.

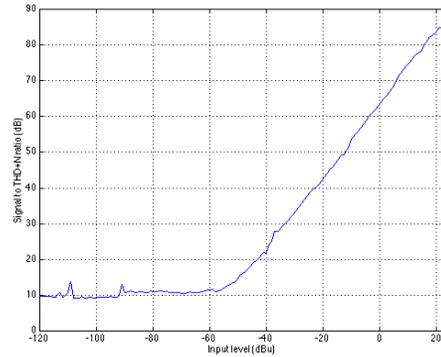


Figure 9.4: A plot of the signal to THD+N ratio vs. the signal level for a simple analog equalizer set to bypass mode and measured with a 1 kHz sine tone. The resolution on the input level is 1 dB. Note the similarity to the curve for PCM digital systems shown in Figure 3.

One minor problem in the analog electronics world is knowing exactly what level causes your gear to distort. Typically, you can't trust your meters as we'll see later, so you'll either have to come up with an alternate metering method (either using an oscilloscope, an external meter, or one of your other pieces of gear as the meter) or just keep your levels slightly lower than optimal to ensure that you don't hit any brick walls.

One nice trick that you can use is in the specific case where you're coming from an analog microphone preamplifier or analog console into a digital converter (depending on its meters). In this case, you can pre-set the gain at the input stage of the mic pre such that the level that causes the output stage of the mixer to clip is also the level that causes the input stage of the ADC to clip. In this case, the meters on your converter can be used instead of the output meters on your microphone preamplifier or console. If all the gear clips at the same level and you stay just below that level at the recording's peak, then you've done a good job. The nice thing about this setup is that you only need to worry about one meter for the whole system.

9.1.4 Analog tape

Analog tape is a different kettle of fish. The noise floor in this case is the same as in the analog and digital worlds. There is some absolute noise floor that is inherent on the tape (the reasons for which are discussed in the

chapter on analog tape, oddly enough...) but the distortion characteristics are different.

When the signal level recorded on an analog tape is gradually increased from a low level, we see an increase in the signal to noise ratio because the noise floor stays put and the signal comes up above it. At the same time however, the level of distortion gradually increases. This is substantially different from the situation with digital signals or op amps because the clipping isn't immediate – it's a far more gradual process as can be seen in Figure 9.5.

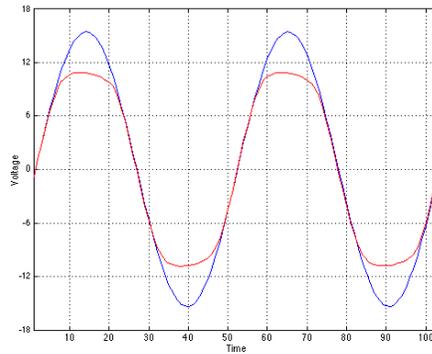


Figure 9.5: A measurement of a 1 kHz sine tone that is “clipped” by analog tape. Notice that, although the peaks and troughs are distorted and limited to the boundaries, the clipping process is much more gradual than was seen in Figure 9.3 with the digital gear and op amps. The blue waveform is the original undistorted sine wave and the red is the output from the analog tape.

The result of this softer, more gradual clipping of the waveform is twofold. Firstly, as was mentioned above, the increase in distortion is more gradual as the level is increase. In addition, because the change in the slope of the waveform is less abrupt, there are fewer very high frequency components resulting from the distortion. Consequently, there are a large number of people who actually use this distortion as an integral part of their processing. This *tape compression* as it is commonly known, is most frequently used for tracking drums.

Assuming that we are trying to maintain the highest possible technical quality and assuming that this does not include tape compression, then we are trying to keep the signal level at the high point on the graph in Figure 5. This level of 0 dB VU is a so-called *nominal* level at which it has been decided (by the tape recorder manufacturer, the analog tape supplier and the technician that works in your studio) that the signal quality is best. Your

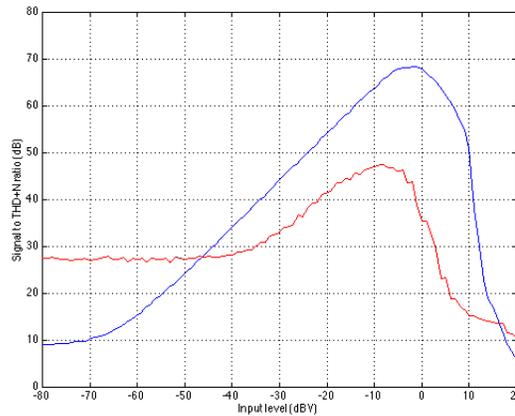


Figure 9.6: A plot of the signal to noise (caused by noise and distortion byproducts) ratio vs. the signal level in a typical analog tape recording. The blue signal is the response of the electronics in the tape recorder (measured using the “Input” monitor). The red signal is the response of the tape. (This is an old Revox A77 that needs a little maintenance, recording on some spare Ampex 456 tape that I had lying around, in case you’re wondering...)

goal in this case is to keep the average level of the signal for the recording hovering around the 0 dB VU mark. You may go above or below this on peaks and dips – but most of the time, the signal will be at an optimal level.

Notice that there are two fundamentally different ways of thinking presented above. In the case of digital gear or analog electronics, you’re determining your recording level based on the absolute maximum peak for the entire recording. So, if you’re recording an entire symphony, you find out what the loudest part will be and make that point in the recording as close to maximum as possible. Look after the peak and the rest will look after itself. In contrast, in the case of analog tape, we’re not thinking of the peak of the signal, we’re concentrating on the average level of the signal – the peaks will look after themselves.

9.1.5 Meters

So, now that we’ve got a very basic idea of the objective, how do we make sure that the levels in our recording system are optimized? We use the meters on the gear to give us a visual indication of the levels. The only problem with this statement is that it assumes that the meter is either telling you what you want to know, or that you know how to read the meter. This isn’t necessarily as dumb as it sounds.

A discussion of meters can be divided into two subtopics. The first is

the issue of *scale* – what actual signal level corresponds to what indication on the meter. The second is the issue of *ballistics* – how the meter responds in time to changes in level.

Before we begin, we'll take a quick review of the difference between the peak and the RMS value of a signal. Figure 9.7 shows a portion of a recorded sound wave. In fact, it's an excerpt of a recording of male speech.

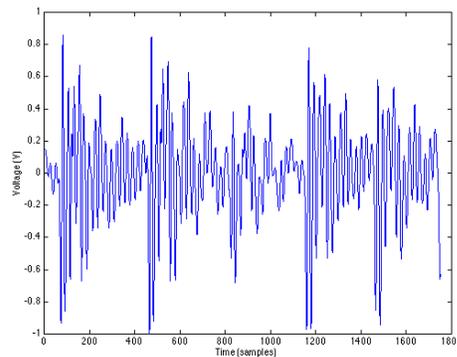


Figure 9.7: The instantaneous voltage level of a recording of male speech.

One simple measurement of the signal level is to continuously look at its absolute value. This is simply done by taking the absolute value of the signal shown in Figure 9.8.

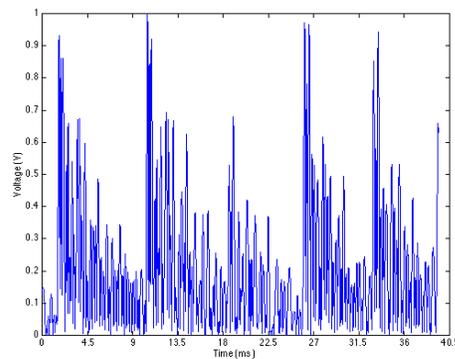


Figure 9.8: The absolute value of the signal shown in Figure 9.7

A second, more complex method is to use the running RMS of the signal.

As we've already discussed in an earlier chapter, the relationship between the RMS and the instantaneous voltage is dependent on the *time constant* of the RMS detection circuit. Notice in Figure 9.9 that not only do the highest levels in the RMS signals differ (the longer the time constant, the higher the level) but their attack and decay slopes differ as well.

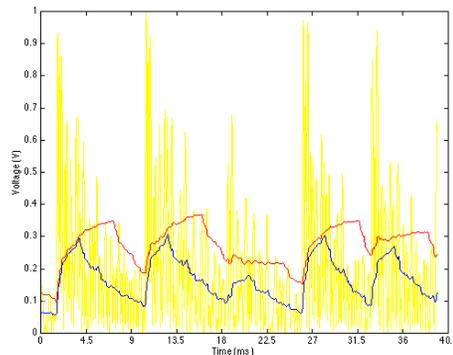


Figure 9.9: Two running measurements of the RMS value of the displayed signal. The blue signal is an RMS using a time constant of 2.27 ms. The red signal uses a time constant of 5.67 ms.

A level meter tells you the level of the signal – either the peak or the RMS value of the level depending on the meter – on a relative scale. We'll look at these one at a time, and deal with the respective scale and ballistics for each.

Peak Light (also known as an Overload indicator)

Scale

If you look at a microphone preamplifier or the input module of a mixing console, you'll probably see a red LED. This peak light is designed to light up as a warning signal when the peak level (the instantaneous voltage – or more likely the absolute value of the instantaneous voltage) approaches the voltage where the components inside the equipment will start to clip. More often than not, this level is approximately 3 dB below the clipping level. Therefore, if the device clips at ± 16.5 V then the peak light will come on if the signal hits 11.6673 V or -11.6673 V (3 dB less than 16.5 V or $16.5 \text{ V} / \sqrt{2}$). Remember that the level at which the peak indicator lights is dependent on the clip level of the device in question – unlike many other meters, it does not indicate an absolute signal strength. So, without

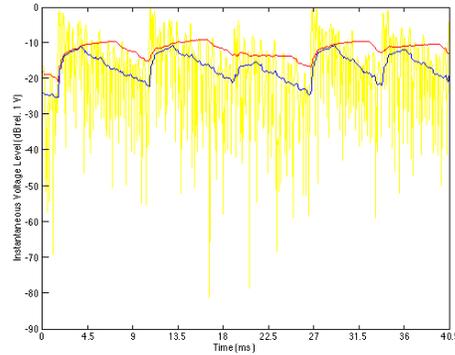


Figure 9.10: The same plot as Figure 8 with the Y-axis changed to a decibel scale. There are a couple of things to note here. Firstly, the decibel scale is relative to 1 V, similar to a dBV scale – the difference is that this plot uses an instantaneous measurement of the voltage compared to 1 V rather than an RMS value relative to $1V_{RMS}$ as in the dBV scale. Secondly, note that the decay curves of both RMS measurements (the red and blue plots) are more linear on this dB scale when compared to a linear absolute voltage scale. Also, note that the red plot (with a time constant of 5.67 ms) reads a signal level that gets closer to the maxima in level whereas the blue plot (with a time constant of 2.27 ms) gives a result that is closer to the minima. In both cases, however, there is approximately a 10 dB error in the RMS values relative to the instantaneous peak voltages.

knowing the exact characteristics of the equipment, we cannot know what the exact level of the signal is when the LED lights. Of course, the moral of that issue is “know your equipment.”



Figure 9.11: Two typical peak indicators. On the left is an Overload indicator light on a GML microphone preamplifier. On the right is a Peak light on a Sony/MCI recording console input strip. The red LED is the peak indicator – the green LED is a signal indicator which lights at a much lower level.

For example, take a look at Figure 9.12. Let’s assume that the signal is passing through a piece of equipment that clips at a maximum voltage of 10 V. The peak indicator will more than likely light up when the signal is 3 dB below this level. Therefore any signal greater than 7.07 V or less than -7.07 V will cause the LED to light up.

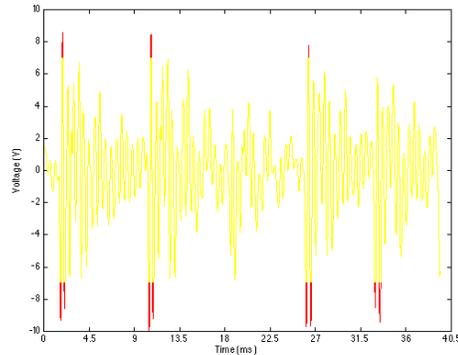


Figure 9.12: The same male speech shown earlier passing through a hypothetical device that clips at ± 10 V and has a peak level indicator that is calibrated to light at 3 dB below clipping (at ± 7.07 V). All of the signal components drawn in red are the signals that will cause the indicator to light.

Ballistics

Note that a peak indicator is an instantaneous measurement. If all is working properly, then any signal of any duration (no matter how short) will cause the indicator to light if the signal strength is high enough.

Also note that the peak indicator lights when the signal level is slightly lower than the level where clipping starts, so just because the light lights doesn't mean that you've clipped your signal... but you're really close.

Volume Unit (VU) Meter

Scale

The Volume Unit Meter (better known as a VU Meter) shows what is theoretically an RMS level reading of the signal passing through it. Its display is calibrated in decibels that range from -20 dB VU up to $+3$ dB VU (the range of 0 dB VU to $+3$ dB VU are marked in red). Because the VU meter was used primarily for recording to analog tape, the goal was to maintain the RMS of the signal at the “optimal” level on the tape. As a result, VU meters are centered around 0 dB VU – a nominal level that is calibrated by the manufacturer and the studio technician to match the optimal level on the tape.

In the case of an analog tape recorder, we can monitor the signal that is being recorded to the tape or the signal coming off the tape. Either way, the



Figure 9.13: A typical Type A VU Meter on a Joe Meek compressor.

meter should be showing us an indication of the amount of magnetic flux on the medium. Depending on the program material being recorded, the policy of the studio, and the tape being used, the level corresponding to 0 dB VU will be something like 250 nWb/m or so. So, assuming that the recorder is calibrated for 250 nWb/m, then when the VU Meter reads 0 dB VU, the signal strength on the tape is 250 nWb per meter. (If the term “nWb/m” is unfamiliar, or if you’re unsure how to decide what your optimal level should be, check out the chapter on analog tape.)

In the case of other equipment with a VU Meter (a mixing console, for example), the indicated level on the meter corresponds to an electrical signal level, not a magnetic flux level. In this case, in almost all professional recording equipment, 0 dB VU corresponds to +4 dBu or, in the case of a sine tone that’s been on for a while, $1.228V_{RMS}$. So, if all is calibrated correctly, if a 1 kHz sine tone is passed through a mixing console and the output VU meters on the console read 0 dB VU, then the sine tone should have a level of $1.228V_{RMS}$ between pins 2 and 3 on the XLR output. Either pin 2 or 3 to ground (pin 1) will be half of that value.

In addition to the decibel scale on a VU Meter, it is standard to have a second scale indicated in percentage of 0 dB VU where 0 dB VU = 100%. VU Meters are subdivided into two types – the *Type A* scale has the decibel scale on the top and the 0% to 100% in smaller type on the bottom as is shown in Figure 9.13. The *Type B* scale has the 0% to 100% scale on the top with the decibel equivalents in smaller type on the bottom.

If you want to get really technical, the official definition of the VU Meter specifies that it reads 0 dB VU when it is bridging a 600 Ω line and the signal level is +4 dBm.

Ballistics

Since VU Meters are essentially RMS meters, we have to remember that they do not respond to instantaneous changes in the signal level. The ballistics for VU Meters have a carefully defined rise and decay time – meaning that we know how fast they respond to a sudden attack or a sudden decay in the sound – slowly. These ballistics are defined using a sine tone that is suddenly switched on and off. If there is no signal in the system and a sine tone is suddenly applied to the VU Meter, then the indicator (either a needle or a light) will reach 99% of the actual RMS level of the signal in 300 ms. In technical terms, the indicator will reach 99% of full-scale deflection in 300 ms. Similarly, when the sine tone is turned off and the signal drops to 0 V instantaneously, the VU meter should take 300 ms to drop back 99% of the way (because the meter only sees the lack of signal as a new signal level, therefore it gets 99% of the way there in 300 ms – no matter where it's going).

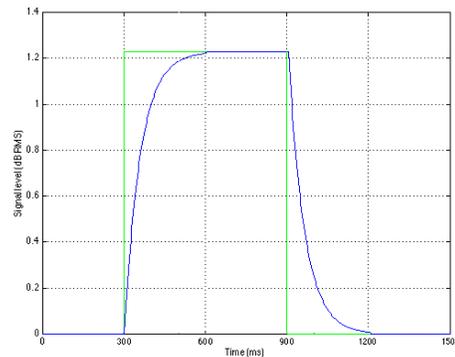


Figure 9.14: A simplified example of the ballistics of a VU meter. Notice that the signal (plotted in green) changes from a $0 V_{RMS}$ to a $1.228 V_{RMS}$ sine wave instantaneously (I know, I know... you can't have an instantaneous change to an RMS value – but I warned you that it was a simplified description!) The level displayed by the VU Meter takes 300 ms to get to 99% of the signal level. Similarly, when the signal is turned off instantaneously, it takes 300 ms for the VU Meter to drop to 0. Notice that the attack and decay curves are reciprocals.

Also, there is a provision in the definition of a VU Meter's ballistics for something called *overshoot*. When the signal is suddenly applied to the meter, the indicator jumps up to the level it's trying to display, but it typically goes slightly over that level and then drops back to the correct level. That amount of overshoot is supposed to be no more than 1.5% of the actual signal level. (If you're picky, you'll notice that there is no overshoot

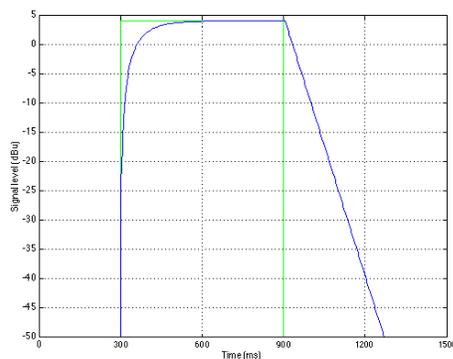


Figure 9.15: The same graph as is shown in Figure 9.14 plotted in a decibel scale. Note that the logarithmic decay of the VU Meter appears as a linear drop in the decibel scale, whereas the attack curve is not linear.

plotted in Figures 9.14 and 9.15.)

Peak Program Meter (PPM)

Scale

The good thing about VU Meters is that they show you the average level of the signal – so they’re great for recording to analog tape or for mastering purposes where you want to know the overall general level of the signal. However, they’re very bad at telling you the peak level of the signal – in fact, the higher the crest factor, the worse they are at telling you what’s going on. As we’ve already seen, there are many applications where we need to know exactly what the peak level of the signal is. Once upon a time, the only place where this was necessary was in broadcasting – because if you overload a transmitter, bad things happen. So, the people in the broadcasting world didn’t have much use for the VU Meter – they needed to see the peak of the program material, so the Peak Program Meter or PPM was developed in Europe around the same time as the VU Meter was in development in the US.

A PPM is substantially different from a VU Meter in many respects. These days it has many different incarnations – particularly in its scale, but the traditional one that most people think of is the UK PPM (also known as the BBC PPM). We’ll start there.

The UK PPM looks very different from a VU Meter – it has no decibel

markings on it – just numbered indications from “Mark 0” up to “Mark 7.” In fact, the PPM is divided in decibels, they just aren’t marked there – generally, there are 4 decibels between adjacent marks – so from Mark 2 to Mark 3 is an increase of 4 dB. There are two exceptions to this rule – there are 6 decibels between Marks 0 and 1 (but note that Mark 0 is not marked). In addition, there are 2 decibels between Mark 7 and Mark 8 (which is also not marked).

Because we’re thinking now in terms of the peak signal level, the nominal level is less important than the maximum, however, PPM’s are calibrated so that Mark 4 corresponds to 0 dBu. Therefore if the PPM at the output stage on a mixing console read Mark 5 for a 1 kHz sine wave, then the output level is $1.228V_{RMS}$ between pins 2 and 3 (because Mark 5 is 4 dB higher than Mark 4, making it +4 dBu).



Figure 9.16: A photograph of a typical UK (or BBC) PPM on the output module of a Sony/MCI mixing console.

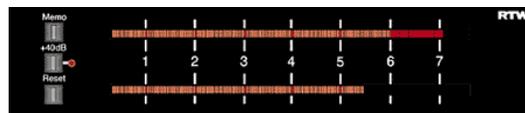


Figure 9.17: An RTW UK PPM variation.

There are a number of other PPM Scales available to the buying public. In addition to the UK PPM, there’s the EBU PPM, the DIN PPM and the Nordic PPM. Each of these has a different scale as is shown in Table 9.1 and the corresponding Figure 9.18.

| Meter | Standard | Minimum Scale and Corresponding Level | Maximum Scale and Corresponding Level | Nominal Scale and Corresponding Level |
|--------------|------------------------|---------------------------------------|---------------------------------------|---------------------------------------|
| VU | IEC 60268-17 | -20 dB = -16 dBu | +3 dB = 7 dBu | 0 dB = 4 dBu |
| UK (BBC) PPM | IEC 268-10 IIA | Mark 1 = -14 dBu | Mark 7 = 12 dBu | Mark 4 = 0 dBu |
| EBU PPM | IEC 268-10 IIB | -12 dB = -12 dBu | +12 dB = 12 dBu | Test = 0 dBu |
| DIN PPM | IEC 268-10 / DIN 45406 | -50 dB = -44 dBu | +5 dB = 11 dBu | 0 dB = 6 dBu |
| Nordic PPM | IEC 268-10 I | -42 dB = -42 dBu | +12 dB = 12 dBu | Test (or 0 dB) = 0 dBu |

Table 9.1: Various scales of analog level meters for professional recording equipment. Add 4 dB to the corresponding signal levels for professional broadcasting equipment.

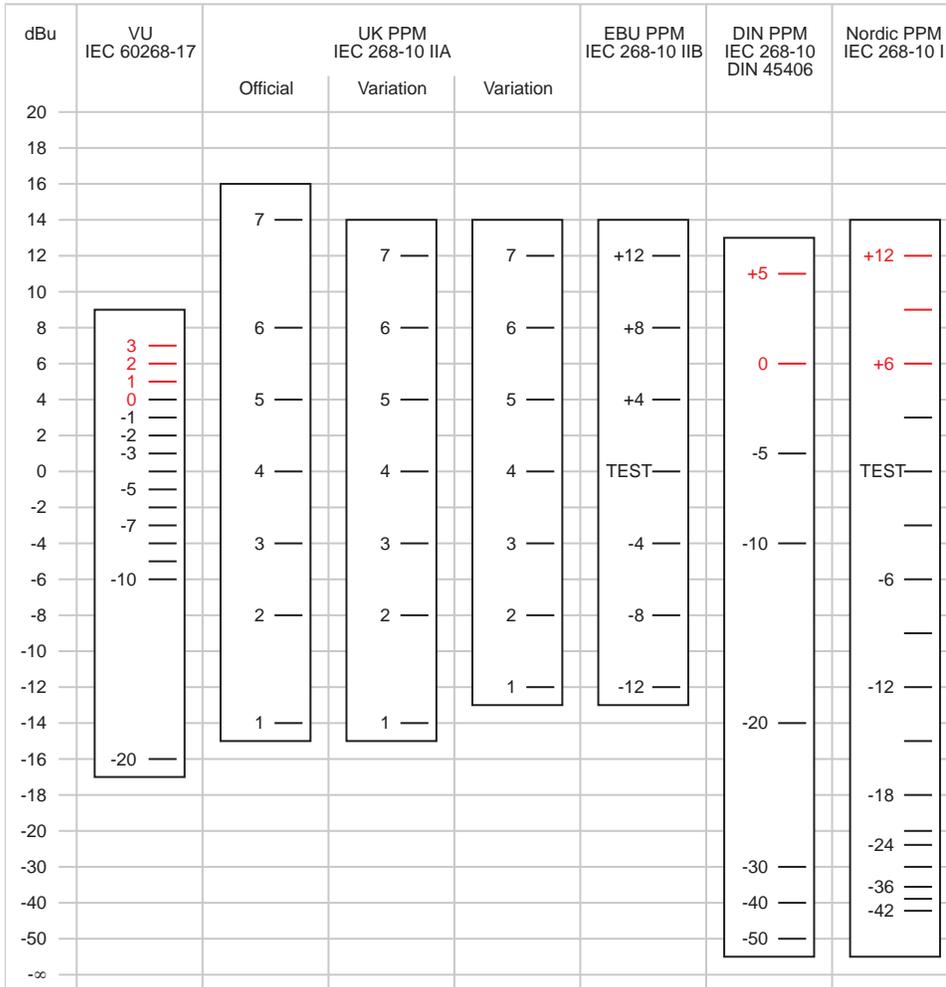


Figure 9.18: Various scales of analog level meters for professional recording equipment. Add 4 dB to the corresponding signal levels for professional broadcasting equipment.

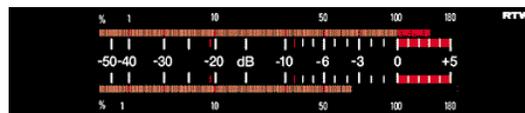


Figure 9.19: INSERT CAPTION

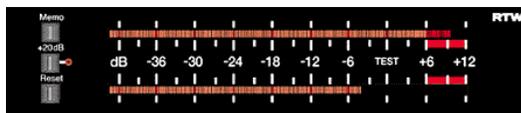


Figure 9.20: An RTW Nordic PPM.

Ballistics

Let's be complete control freaks and build the perfect PPM. It would show the exact absolute value of the voltage level of the signal all the time. The needle would dance up and down constantly and after about 3 seconds you'd have a terrible headache watching it. So, this is not the way to build a PPM. In fact, what is done is the ballistics are modified slightly so that the meter responds very quickly to a sudden increase in level, but it responds very slowly to a sudden drop in level – the decay time is much slower even than a VU Meter. You may notice that the PPM's listed in Table 1 and Figure 9.18 are grouped into two “types” Type I and Type II. These types indicate the characteristics of the ballistics of the particular meter.

Type I PPM's

The attack time of a Type I PPM is defined using an *integration time* of 5 ms – which corresponds to a time constant of 1.7 ms. Therefore, a tone burst that is 10 ms long will result in the indicator being 1 dB lower than the correct level. If the burst is 5 ms long, the indicator will be 2 dB down, a 3 ms burst will result in an indicator that is 4 dB down. The shorter the burst, the more inaccurate the reading. (Note however, that this is significantly faster than the VU Meter.)

Again, unlike the VU meter, the decay time of a Type I PPM is not the reciprocal of the attack curve. This is defined by how quickly the indicator drops – in this case, the indicator will drop 20 dB in 1.4 to 2.0 seconds.

Type II PPM's

The attack time of a Type II PPM is identical to a Type I PPM.

The decay of a Type II PPM is somewhat different from its Type I cousin. The indicator falls back at a rate of 24 dB in 2.5 to 3.1 seconds. In addition, there is a “hold” function on the peak where the indicator is held for 75 ms to 150 ms before it starts to decay.

Qualification Regarding Nominal Levels

There's one important thing to note in all of this discussion. This chapter assumes that we're talking about professional equipment in a recording studio.

Professional Broadcast Equipment

If you work with professional broadcast equipment, then the nominal level is different – in fact, it's 4 dB higher than in a recording studio. 0 dB VU corresponds to +8 dBu and all of the other scales are higher to match.

Consumer-level Equipment

If we're talking about consumer-level equipment, either for recording or just for listening to things at home on your stereo, then the nominal 0 dB VU point (and all other nominal levels) corresponds to a level of -10 dBV or $0.316V_{RMS}$.

Digital Meter

A digital meter is very similar to a PPM because, as we've already established, your biggest concern with digital audio is that the peak of the signal is never clipped. Therefore, we're most interested in the peak or the amplitude of the signal.

As we've said before, the noise floor in a PCM digital audio signal is typically determined by the dither level which is usually at approximately one half of an LSB. The maximum digital level we can encode in a PCM digital signal is determined by the number of bits. If we're assuming that we're talking about a two's complement system, then the maximum positive amplitude is a level that is expressed as a 0 followed by as many 1's as are allowed in the digital word. For example, in an 8-bit system, the maximum possible positive level (in binary) is 01111111. Therefore, in a 16-bit system with 65536 possible quantization values, the maximum possible positive level is level number 32767. In a 24-bit system, the maximum positive level is 8388607. (If you'd like to do the calculation for this, it's $\frac{(2^n)}{2-1}$ where n is the number of bits in the digital word.

Note that the negative-going signal has one extra LSB in a two's complement system as is discussed in the chapter on digital conversion.

The maximum possible value in the positive direction in a PCM digital signal is called *full scale* because a sample that has that maximum value

uses the entire scale that is possible to express with the digital word. (Note that we'll see later that this definition is actually a lie – there are a couple of other things to discuss here, but we'll get back to them in a minute.)

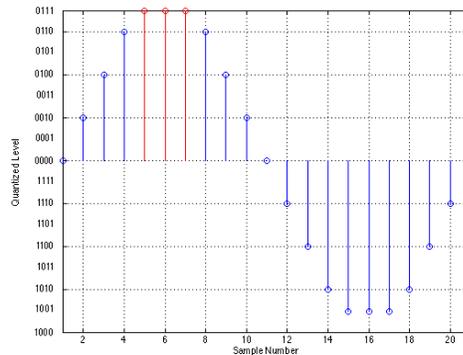


Figure 9.21: A PCM two's complement digital representation of a quantized sine wave with a frequency of 1/20th of the sampling rate. Note that three samples (numbers 5, 6 and 7) have reached full scale and are indicated in red. By comparison, the symmetrical samples (numbers 15, 16 and 17) are technically at full scale despite the extra LSB in the negative zone.

It is therefore evident that, in the digital world, there is some absolute maximum value that can be expressed, above which there is no way to describe the sample value. We therefore say that any sample that hits this maximum is “clipped” in the digital domain – however, this does not necessarily mean that we’ve clipped the audio signal itself. For example, it is highly unlikely that a single clipped sample in a digital audio signal will result in an audible distortion. In fact, it’s unlikely that two consecutively clipped samples will cause audible artifacts. The more consecutively clipped samples we have, the more audible the distortion. People tend to settle on 2 or 3 as a good number to use as a definition of a “clipped” signal.

If we look at a rectified signal in a two's complement PCM digital domain, then the amplitude of a sample can be expressed using its relationship to a sample at full scale. This level is called *dB FS* or “decibels relative to full scale” and can be calculated using the following equation:

$$\text{dB FS} = 20 * \log (\text{sample value} / \text{maximum possible value})$$

Therefore, in a 16-bit system, a sine wave that has an amplitude of 16384 (which is also the value of the sample at the positive peak of the sine wave) will have a level of -6.02 dB FS because:

$$20 * \log (16384 / 32767) = -6.02 \text{ dB FS}$$

There's just one small catch: I lied. There's one additional piece of information that I've omitted to keep things simple. Take a close look at Figure 9.21. The way I made this plot was to create a sine wave and quantize it using a 4-bit system assuming that the sampling rate is 20 times the frequency of the sine wave itself. Although this works, you'll notice that there are some quantization levels that are not used. For example, not one of the samples in the digital sine wave representation has a value of 0001, 0011 or 0101. This is because the frequency of the sine wave is harmonically related to the sampling rate. In order to ensure that more quantization levels are used, we have to use a sampling rate that is enharmonically related to the sampling rate. The technical definition of "full scale" uses a digitally-generated sine tone that has a frequency of 997 Hz. Why 997 Hz? Well, if you divide any of the standard sampling rates (32 kHz, 44.1 kHz, 48 kHz, 88.2 kHz, 96 kHz, etc...) by 997, you get a nasty number. The result is that you get a different quantization value for every sample in a second. You won't hit every quantization value because the whole system starts repeating after one second – but, if your sine tone is 997 Hz and your sampling rate is 44.1 kHz, you'll wind up hitting 44100 different quantization values. The higher the sampling rate, the more quantization values you'll hit, and the less your error from full scale.

The other reason for using this system is to avoid signals that are actually higher than Full Scale without the system actually knowing. If you have a sine tone with a frequency that is harmonically related to the sampling rate, then it's possible that the very peak of the wave is between two samples, and that it will always be between two samples. Therefore the signal is actually greater than 0 dB FS without you ever knowing it. With a 997 Hz tone, eventually, the peak of the wave will occur as close as is reasonably possible to the maximum recordable level.

This becomes part of the definition of full scale – the amplitude of a signal is compared to the amplitude of a 997 Hz sine tone at full scale. That way we're sure that we're getting as close as we can to that top quantization level.

There is one other issue to deal with: the definition of dB FS uses the RMS value of the signal. Therefore, a signal that is at 0 dB FS has the same RMS value as a 997 Hz sine wave whose peak positive amplitude reaches full scale. There are two main implications of this definition. The first has to do with the crest factor of your signal. Remember that the crest factor is a measurement of the relationship between the peak and the RMS value of the signal. In almost all cases, the peak value will be greater than RMS value (in fact, the only time this is not the case is a square wave in which they will

be equal). Therefore, if a meter is really showing you the signal strength in dB FS, then it is possible that you are clipping your signal without your meter knowing. This is because the meter would be showing you the RMS level, but the peak level is much higher. It is therefore possible that you are clipping that peak without hitting 0 dB FS. This is why digital equipment also has an *OVER* indicator (check out Figure 9.22) to tell you that the signal has clipped. Just remember that you don't necessarily have to go all the way up to 0 dB FS to clip.

Another odd implication of the dB FS definition is that, in the odd case of a square wave, you can have a level that is greater than 0 dB FS without clipping. The crest factor of a sine wave is 3.01 dB. This means that the RMS level of the sine tone is 3.01 dB less than its peak value. By comparison, the crest factor of a square wave is 0 dB, meaning that the peak and RMS values are equal. So what? Well, since dB FS is referenced to the RMS value of a sine wave whose maximum peak is at Full Scale (and therefore 3.01 dB less than Full Scale), if you put in a square wave that goes all the way up to Full Scale, it will have a level that is 3.01 dB higher than the Full Scale sine tone, and therefore a level of +3.01 dB FS. This is an odd thing for people who work a lot with digital gear. I, personally, have never seen a digital meter that goes beyond 0 dB. Then again, I don't record square waves very often either, so it doesn't really matter a great deal.

Chances are that the digital meter on whatever piece of equipment that you own really isn't telling you the signal strength in dB FS. It's more likely that the level shown is a sample-by-sample level measurement (and therefore not an RMS measurement) with a ballistic that makes the meter look like it's decaying slowly. Therefore, in such a system, 0 dB on the meter means that the sample is at Full Scale.

I'm in the process of making a series of test tones so that you can check your meters to see how they display various signal levels. Stay tuned!

Ballistics

As far as I've been able to tell, there are no standards for digital meter ballistics or appearances, so I'll just describe a typical digital meter. Most of these use what is known as a *dot bar* mode which actually shows two levels simultaneously. Looking at Figure 9.22, we can see that the meter shows a bar that extends to -24 dB. This bar shows the present level of the signal using ballistics that typically have roughly the same visual characteristics as a VU Meter. Simultaneously, there is a dot at the -8 dB mark. This indicates that the most recent peak hit -8 dB. This dot will be erased after



Figure 9.22: A photograph of a typical digital meter on a Tascam DAT machine. There are a couple of things shown here. The first is the bar graphs of the Left and Right channels just below the -24 dB mark. This is the level of the signal at the moment when the picture was taken. There are also two “dots” at the -8 dB mark. These are the level of a recent peak and will be replaced by a new peak in a couple of seconds. Finally, there is the “MARG” (for “margin”) indication of 6.7 dB – this indicates that the maximum peak of the entire program material since the recording was started hit -6.7 dB. Note that we don’t know which channel that peak was on.

approximately one second or so and be replaced by a new peak unless the signal peaks at a value greater than -8 dB in which case that value will be displayed by the dot. This is similar to a Type II PPM ballistic with the decay being replaced with simple erasure.

Many digital audio meters also include a function that gives a very accurate measurement of the maximum peak that has been hit since we’ve started recording (or playing). This value is usually called the *margin* and is typically displayed as a numerical value near the meter, but elsewhere on the display.

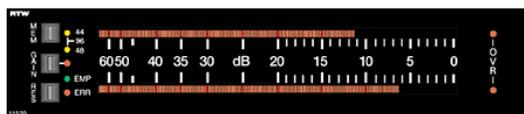


Figure 9.23: An RTW digital audio meter.

Finally, digital meters have a warning symbol to indicate that the signal has clipped. This warning is simply called *over* since all we’re concerned with is that the signal went over full scale – we don’t care how far over full scale it went. The problem here is that different meters use different definitions for the word “over.” As I’ve already pointed out, some meters keep track of the number of consecutive samples at full scale and point out when that number hits 2 or 3 (this is either defined by the manufacturer or by the user, depending on the equipment and model number – check your manual). On some equipment (particularly older gear), the “digital” meter is driven by the analog conversion of the signal and is therefore extremely inaccurate – again, check your manual. An important thing to note about

these meters is that they rarely are aware that the signal has gone over full scale when you're playing back a digital signal, or if you're using an external analog to digital to convertor – so be very careful.

9.1.6 Gain Management

From the time the sound arrives at the diaphragm of your microphone to the time the signal gets recorded, it has to travel a very perilous journey, usually through a lot of wire and components that degrade the quality of the signal every step of the way. One of the best ways to minimize this degradation is to ensure that you have an optimal gain structure throughout your recording chain, taking into account the noise and distortion characteristics of each component in the signal path. This sounds like a monumental task, but it really hinges on a couple of very simple concepts.

The first basic rule (that you'll frequently have to break but you'd better have a good reason...) is that you should make the signal as loud as you can as soon as you can. For example, consider the example of a microphone connected through a mic preamp into a DAT machine. We know that, in order to get the best quality digital conversion of the signal, its maximum should be just under 0 dB FS. Let's say that, for the particular microphone and program material, you'll need 40 dB of gain to get the signal up to that level at the DAT machine. You could apply that gain at the mic preamp or the analog input of the DAT recorder. Which is better? If possible, it's best to get all of the gain at the mic preamp. Why? Consider that each piece of equipment adds noise to the signal. Therefore, if we add the gain after the mic preamp, then we're applying that gain to the signal *and* the noise of the microphone preamp. If we add the gain at the input stage of the mic preamp, then its inherent noise is not amplified. For example, consider the following equations:

9.1.7 Phase and Correlation Meters

NOT WRITTEN YET

9.1.8 Suggested Reading List

| | | |
|--|------|--|
| $(\text{signal} + \text{noise}) * \text{gain}$ | But: | $(\text{signal} * \text{gain}) + \text{noise}$ |
| $(\text{signal} * \text{gain}) + (\text{noise} * \text{gain})$ | | High signal level, low noise level |
| High signal level, high noise level | | High signal level, high noise level |

Table 9.2: An illustration of why we should apply maximum gain as early as possible in the signal path. In this case, we're assuming that the noise is generated internally by the microphone preamplifier



Figure 9.24: A photograph of a phase meter on the output module of a Sony/MCI mixing console.

9.2 Monitoring Configuration and Calibration

9.2.1 Standard operating levels

Before we talk about the issue of how to setup a playback system, we have to discuss the issue of standard operating levels. We have already seen in Section ?? that our ears have different sensitivities to different frequencies at different levels. Basically, at low listening levels, we can't hear low-end material as easily as mid-band content. The louder the signal gets, the "flatter" the frequency response of our ears. In the practical world, this means that if I do a mix at a low level, then I'll mix the bass a little hot because I can't hear it. If I turn it up, it will sound like there's more bass in the mix, because my ears have a different response.

Therefore, in order to ensure that you (the listener) hear what I (the mixing engineer) hear, one of the first things I have to specify is how loud you should turn it up. This is the reason for a standard operating level. That way, if you say "there's not enough bass" I can say "you're listening at too low a level" – unless you aren't, then we have to talk about issues of taste. This subject will not be addressed in this book.

The film and television industries have an advantage that the music people don't. They have international standards for operating levels. What this means is that a standard operating level on the recording medium and in the equipment will result in a standard acoustic level at the listening position in the mixing studio or theatre. Of course, we typically listen to the television at lower levels than we hear at the movie theatre, so these two levels are different.

Table 9.4 shows the standard operating levels for film and television sound work. It also includes an approximate range for music mixing, although this is not a standard level.

| Medium | Signal level | Signal level | Signal level | Acoustic level |
|--------------|------------------|----------------|---------------|----------------------|
| Film | -20 dB FS | 0 dB VU | +4 dBu | 85 dBspl |
| Television | -20 dB FS | 0 dB VU | +4 dBu | 79 dBspl |
| <i>Music</i> | <i>-20 dB FS</i> | <i>0 dB VU</i> | <i>+4 dBu</i> | <i>79 - 82 dBspl</i> |

Table 9.3: Standard operating levels for mixing for an in-band pink noise signal. Note that the values for music mixing are not standardized [Owinski, 1988]. Also note that the values listed here are for a single channel. Measurements are done with an SPL meter with a C-weighting and a slow response.

It is important to note that the values in Table 9.4 are for a single

loudspeaker measured with an SPL meter with a C-weighting and a slow response. So, for example, if you're working in 5.1 surround for film, pink noise at a level of 0 dB VU sent to the centre channel only will result in a level of 85 dBspl at the mixing position. The same should be true of any other main channel.

Dolby has a slightly different recommendation in that, for film work, they suggest that each surround channel (which may be sent to multiple loudspeakers) should produce a standard level of 82 dBspl. This difference is applicable only to “film-style mixing rooms” [Dolby, 2000].

9.2.2 Channels are not Loudspeakers

Before we go any further, we have to look at a commonly-confused issue in monitoring, particularly since the popularization of so-called 5.1 systems.

In a 5.1-channel mix, we have 5 main full-range channels, Left, Centre, Right, Left Surround and Right Surround. In addition, there is a channel which is band-limited from 0 to 120 Hz called the *LFE* or *Low Frequency Effects* channel.

In the listening room in the real world, we have a number of loudspeakers:

1. The Left and Right loudspeakers are typically a pair that may not match any other loudspeaker in the room.
2. The Left Surround and Right Surround loudspeakers typically match each other, but are often smaller than the other loudspeakers in the room, particularly lacking in low end because woofers apparently cost money.
3. The Centre loudspeaker may be a third type of device, or in some cases may match the Left and Right loudspeakers. In many cases in the home situation, the Centre loudspeaker is contained in the television and may therefore, in fact, be two loudspeakers.
4. A single subwoofer.

Of course, the situation I just described for the listening environment is not the optimal situation, but it's a reasonable description of the real world. We'll look at the ideal situation below.

If we look at a very simple configuration, then the L, R, C, LS and RS signals are connected directly to the L, R, C, LS and RS loudspeakers respectively, and the LFE channel is connected to the subwoofer. In most cases, however, this is not the only configuration. In larger listening rooms,

we typically see more than two surround loudspeakers and more than one subwoofer. In smaller systems, people have been told that they don't need 5 large speakers, because all the bass can be produced by the subwoofer using a bass management system described below, consequently, the subwoofer produces more than just the LFE channel.

So, it is important to remember that delivery channels are not directly equivalent to loudspeakers. It is an LFE channel – not a subwoofer channel.

9.2.3 Bass management

Once upon a time, people who bought a stereo system bought two identical loudspeakers to make the sound they listened to. If they couldn't spend a lot of money, or they didn't have much space, they bought smaller loudspeakers which meant less bass. (This isn't necessarily a direct relationship, but that issue is dealt with in the section on loudspeaker design... We'll assume that it's the truth for this section.)

Then, one day I walked into my local stereo store and heard a demo of a new speaker system that just arrived. The two loudspeakers were tiny little things - two cubes about the size of a baseball stuck together on a stand for each side. The sound was much bigger than these little speakers could produce.. .there had to be a trick. It turns out that there was a trick... the Left and Right channels from the CD were being fed to a crossover system where all the low-frequency information was separated from the high-frequency information, summed and sent to a single low-frequency driver sitting behind the couch I was sitting on. The speakers I could see were just playing the mid- and high-frequency information... all the low-end came from under the couch.

This is the concept behind *bass management* or *bass redirection*. If you have a powerful-enough dedicated low frequency loudspeaker, then your main speakers don't need to produce that low frequency information. There are lots of arguments for and against this concept, and I'll try to address a couple of these later, but for now, let's look at the layout of a typical bass management system.

Figure 9.25 shows a block diagram for a typical bass management scheme. The five main channels are each filtered through a high-pass filter with a crossover frequency of approximately 80 Hz before being routed to their appropriate loudspeakers. These five channels are also individually filtered through low-pass filters with the same crossover frequency, and the outputs of these filters is routed to a summing buss. In addition, the LFE channel input is increased in level by 10 dB before being added to the same buss.

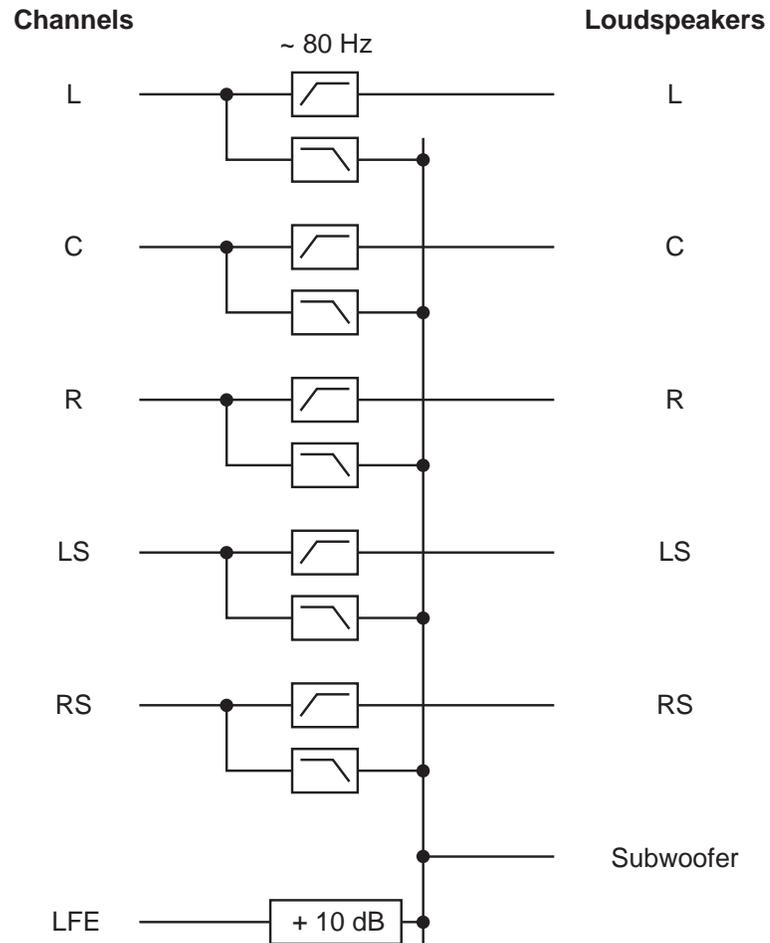


Figure 9.25: A typical monitoring path for a bass-managed system. Note the 10 dB boost on the LFE channel.

The result on this summing buss is sent to the subwoofer amplifier.

There is an important item to notice here – the 10 dB gain on the LFE channel. Why is this here? Well, consider if we send a full-scale signal to all channels. The single subwoofer is being asked to balance with 5 other almost-full-range loudspeakers, but since it is only one speaker competing with 5 others, we have to boost it to compensate. We don't need to do this to the outputs resulting from the bass management system because they five channels of low-frequency information are added, and therefore boost themselves in the process. The reason this is important will be obvious in the discussion of loudspeaker level calibration below.

9.2.4 Configuration

There are some basic rules to follow in the placement of loudspeakers in the listening space. The first and possibly most important rule of thumb is to remember that all loudspeakers should be placed at ear-level and aimed at the listening position. This is particularly applicable to the tweeters in the loudspeaker enclosure. Both of these simple rules are due to the fact that loudspeakers beam – that is to say that they are directional at high frequencies. In addition, you want your reproduced sound stage to be on your horizon, therefore the loudspeakers should be at your height. If it is required to place the loudspeakers higher (or lower) than the horizontal plane occupied by your ears, they should be angled downwards (or upwards) to point at your head.

The next issue is one of loudspeaker proximity to boundaries. As was discussed in Section ??, placing a loudspeaker immediately next to a boundary such as a wall will result in a boost of the low frequency components in the device. In addition, as we saw in Section ??, a loudspeaker placed against a wall will couple much better to room modes in the corresponding dimension, resulting in larger resonant peaks in the room response. As a result, it is typically considered good practice to place loudspeakers on stands at least 1 m from any rigid surface. Of course, there are many situations where this is simply not possible. In these cases, correction of the loudspeaker's response should be considered, either through post-crossover gain manipulation as is possible in many active monitors, or using equalization.

There are a couple of other issues to consider in this regard, some of which are covered below in Section 9.2.4.

Two-channel Stereo

A two-channel playback system (typically misnamed “stereo”) has a standard configuration. Both loudspeakers should be equidistant from the listener and at angles of -30° and 30° where 0° is directly forward of the listener. This means that the listener and the two loudspeakers form the points of an equilateral triangle as shown in Figure 9.26, producing a *loudspeaker aperture* of 60° .

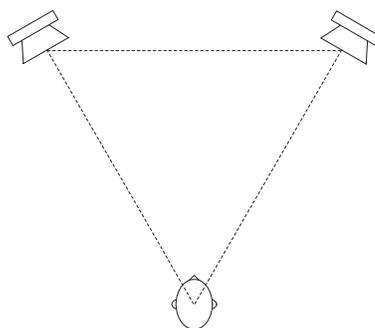


Figure 9.26: Recommended loudspeaker configuration for 2-channel stereo listening.

Note that, for all discussions in this book, all positive angles are assumed to be on the right of centre forward, and all negative angles are assumed to be left of centre forward.

5-channel Surround

In the case of 5.1 surround sound playback, we are actually assuming that we have a system comprised of 5 full-range loudspeakers and no subwoofer. This is the recommended configuration for music recording and playback[dol, 1998] whereas a true 5.1 configuration is intended only for film and television sound. Again, all loudspeakers are assumed to be equidistant from the listener and at angles of 0° , $\pm 30^\circ$ and with two surround loudspeakers symmetrically placed at an angle between $\pm 100^\circ$ and $\pm 120^\circ$. This configuration is detailed in *ITU-R BS.775.1*. [ITU, 1994] (usually called “ITU775” or just “775” in geeky conversation... say all the numbers... “seven seven five” if you want to be immediately accepted by the in-crowd) and shown in Figure 9.27. If you have 25 Swiss Francs burning a hole in your pocket, you can order this document as a pdf or hardcopy from www.itu.ch. Note that the configuration has 3 different loudspeaker apertures, 30° (with the C/L and

C/R pairs), approximately 80° (L/LS and R/RS) and approximately 140° (LS/RS).

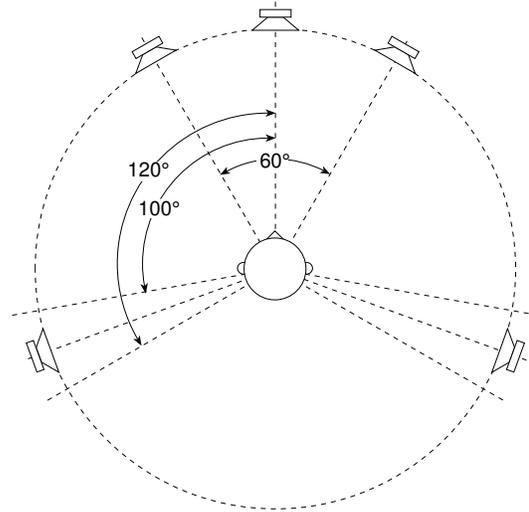


Figure 9.27: Recommended loudspeaker configuration for 5.1-channel surround listening[ITU, 1994].

How to set up a 5-channel system using only a tape measure

It's not that easy to set up a 5-channel system using only angles unless you have a protractor the size of your room. Luckily, we have trigonometry on our side, which means that we can actually do the set up without ever measuring a single angle in the room. Just follow the step-by-step instructions below.

Step 1. Mark the listener's location in the room and determine the desired distance to the loudspeakers (we'll call that distance X) Try to keep your loudspeakers at least 2 m from the listening position and no less than 1 m from any wall.

Step 2. Make an equilateral triangle marking the listener's location, the Left and the Right loudspeakers as shown in the figure on the right. See Figure 9.28.

Step 3. Find the halfway point between the L and R loudspeakers and mark it. See Figure 9.29.

Step 4. Find the location of the C speaker using the halfway mark you just made, the listener's location and the distance X . See Figure 9.30.

Step 5. Marks the locations for the LS and RS loudspeakers using the triangle measurements shown on the right. See Figure 9.31.

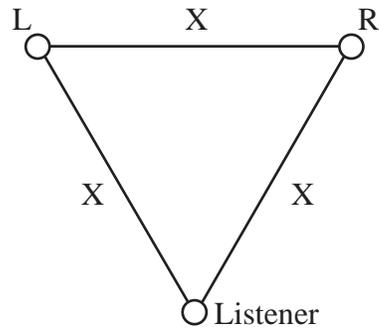


Figure 9.28: 5-channel setup: Step 1. Measure an equilateral triangle with your L and R loudspeakers and the listening position as the three corners.

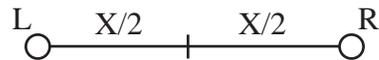


Figure 9.29: 5-channel setup: Step 2. Find the midpoint between the L and R loudspeakers.

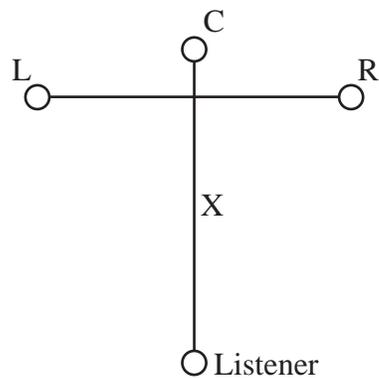


Figure 9.30: 5-channel setup: Step 3. Measure the distance between the listening position and the C loudspeaker to match the distances in Step 1.

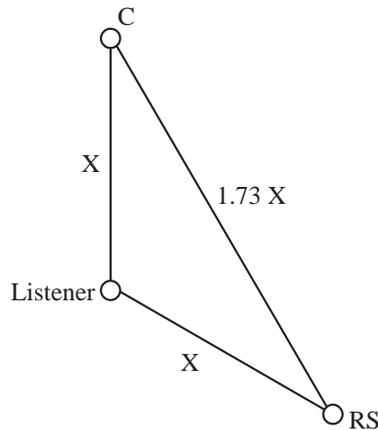


Figure 9.31: 5-channel setup: Step 4. Measure a triangle created by the C and RS loudspeakers and the listening position using the distances indicated.

Step 6. Double check your setup by measuring the distance between the LS and RS loudspeakers. It should be $1.73X$. (Therefore the C, LS and RS loudspeakers should make an equilateral triangle.) See Figure 9.32.

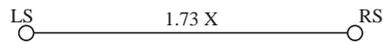


Figure 9.32: 5-channel setup: Step 5. Double check your surround loudspeaker placement by measuring the distance between them. This should be the same as either surround loudspeaker to the C.

7. If the room is small, put the sub in the corner of the room. If the room is big, put the sub under the centre loudspeaker. Alternately, you could just put the sub where you think that it sounds best.

Room Orientation

There is a minor debate between opinions regarding the placement of the monitor configuration within the listening room. Usually, unless you've spent lots of money getting a listening room or control room designed from scratch, you're probably going to be in a room that is essentially rectangular. This then raises two important questions:

1. Should you use the room symmetrically or asymmetrically?

2. Do you use the room so that it's narrow, but long, or wide but shallow?

Most people don't think twice about the answer to the first question – of course you use the room symmetrically. The argument for this logic is to ensure a number of factors:

1. The coupling of left / right pairs of loudspeakers to the room are matched.
2. The early reflection patterns from left / right pairs of loudspeakers are matched.

Therefore, your left / right pairs of speakers will “sound the same” (this also means the left surround / right surround pair) and your imaging will not pull to one side due to asymmetrical reflections.

Then again, the result of using a room symmetrically is that you are sitting in the dead centre of the room which means that you are in one of the worst possible locations for hearing room modes – the nulls are at a minimum and the antinodes are at a maximum at the centre of the room. In addition, if you listen for the fundamental axial mode in the width of the room, you'll notice that your two ears are in opposite polarities at this frequency. Moving about 15 to 20 cm to one side will alleviate this problem which, once heard once, unfortunately, cannot be ignored.

So, it is up to your logic and preference to decide on whether to use the room symmetrically.

The second question of width vs. depth depends on your requirements. Figure 9.33 shows that the choice of room orientation has implications on the maximum distance to the loudspeakers. Both floorplans in the diagram show rooms of identical size with a maximum loudspeaker distance for an ITU775 configuration laid on the diagram. As can be seen, using the room as a wide, but shallow space allows for a much larger radius for the loudspeaker placement. Of course, this is a worst-case scenario where the loudspeakers are placed against boundaries in the room, a practice which is not advisable due to low-frequency boost and improved coupling to room modes.

IMAX film sound

NOT YET WRITTEN

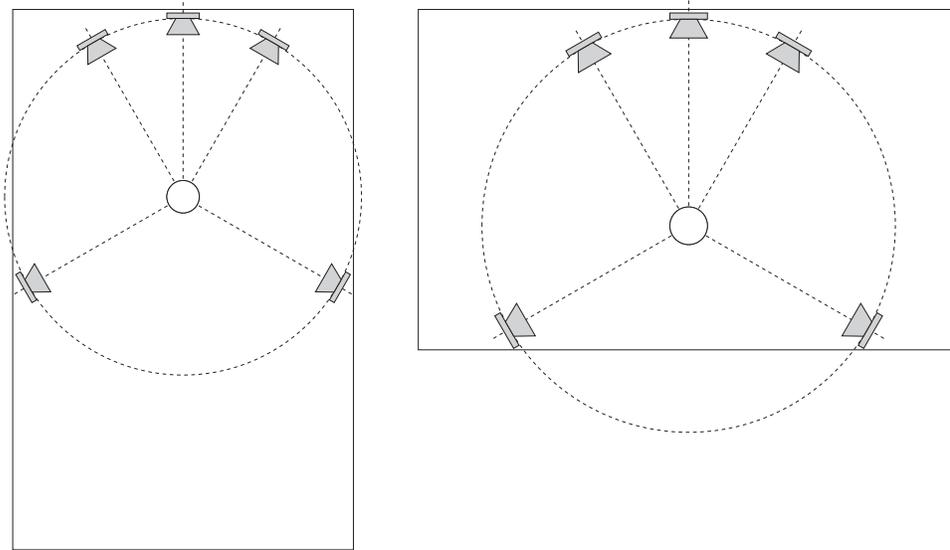


Figure 9.33: Two rectangular rooms of identical arbitrary dimensions showing the maximum possible loudspeaker distance for an ITU775 configuration. Notice that the loudspeakers can be further away when you use the room “sideways.”

10.2 Surround

From the very beginning, it was recognized that the 5.1 standard was a compromise. In a perfect system you would have an infinite number of loudspeakers, but this causes all sorts of budgetary and real estate issues... So we all decided to agree that 5 channels wasn't perfect, but it was pretty good. There are people with a little more money and loftier ideals than the rest of us who are pushing for a system based on the MIBEIYDIS system (more-is-better-especially-if-you-do-it-smartly).

One of the most popular of these systems uses the standard 5.1 system as a starting point and expands on it. Dubbed 10.2 and developed by Tomlinson Holman (the TH in **THX**) this is actually a 12.2 system that uses a total of 16 loudspeakers.

There are a couple of things to discuss about this configuration. Other than the sheer number of loudspeakers, the first big difference between this configuration and the standard ITU775 standard is the use of elevated loudspeakers. This gives the mixing engineer two possible options. If used as a stereo pair, it becomes possible to generate phantom images higher than the usual plane of presentation, giving the impression of height as in the IMAX

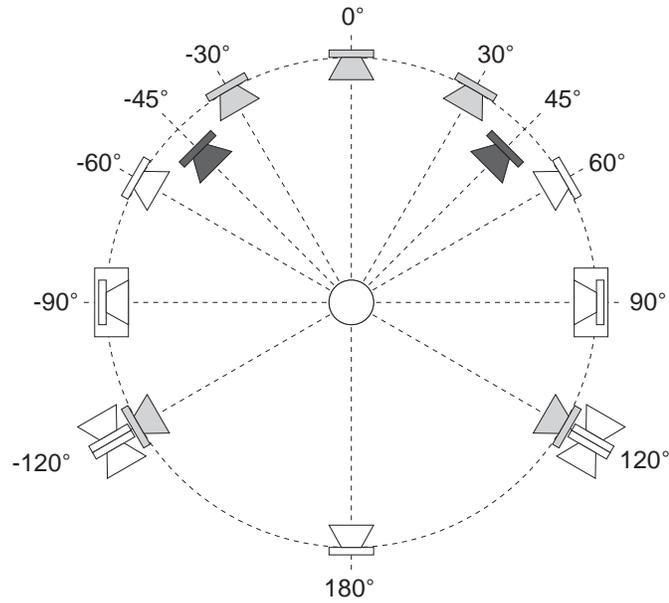


Figure 9.34: A 10.2 configuration. The light-gray loudspeakers match those in the ITU775 recommendation. The dark-gray speakers have an elevation of 45° relative to the listener as can be seen in Figure 9.35. The speakers in boxes at $\pm 90^\circ$ are subwoofers. Note that all loudspeakers are equidistant to the listener.

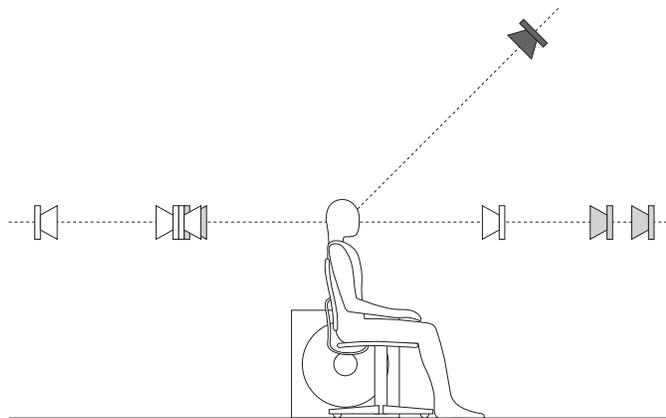


Figure 9.35: A simplified diagram of a 10.2 configuration seen from the side. The light-gray loudspeakers match those in the ITU775 recommendation. The dark-gray speakers have an elevation of 45° relative to the listener.

system. If diffuse sound is sent to these loudspeakers, then the mix relies on our impaired ability to precisely localize elevated sound sources (see Section ??) and therefore can give a better sense of envelopment than is possible with a similar number of loudspeakers distributed in the horizontal plane [].

You will also notice that there are pairs of back-to-back loudspeakers placed at the $\pm 90^\circ$ positions. These are what are called “diffuse radiators” and are actually wired to create a dipole radiator as is described in Section ?. In essence, you simply send the same signal to both loudspeakers in the pair, inverting the polarity of one of the two. This produces the dipole effect and, in theory, cancels all direct sound arriving at the listener’s location. Therefore, the listener receives only the reflected sound from the front and rear walls predominantly, creating the impression of a more diffuse sound than is typically available from the direct sound from a single loudspeaker.

Finally, you will note from the designation “10.2” that this system calls for two subwoofers. This follows the recommendations of a number of people [Martens, 1999][?] who have done research proving that uncorrelated signals from two subwoofers can result in increased envelopment at the listening position. The position of these subwoofers should be symmetrical, however more details will be discussed below.

Ambisonics

NOT WRITTEN YET

NOTES:

Radially symmetrical

Panphonic (2D) ambisonics – minimum number of speakers =

Periphonic (3D) ambisonics – minimum number of speakers =

Subwoofers

NOT WRITTEN YET

NOTES

Better to have many full-range speakers than 1 subwoofer

Floyd Toole’s idea of room mode cancellation through multiple correlated subwoofers

David Greisinger’s 2 decorrelated subwoofers driven by 1 channel

Bill Marten’s 2 subwoofer channels.

9.2.5 Calibration

The calibration of your monitoring system is possibly one of the most significant factors that will determine the quality of your mixes. As a simple example, if you have frequency-independent level differences between your two-channel monitors, then your centre position is different from the rest of the world's. You will compensate for your problem, and consequently create a problem for everyone else resulting in complaints that your lead vocals aren't centered.

Unfortunately, it is impossible to create the perfect monitor, so you have to realize the limitations of your system and learn to work within those constraints. Essentially, the better you know the behaviour of your monitoring system, the more you can trust it, and therefore the more you can be trusted by the rest of us.

There is a document available from the ITU that outlines a recommended procedure for doing listening tests on small-scale impairments in audio systems [ITU, 1997]. Essentially, this is a description of how to do the listening test itself, and how to interpret the results. However, there is a section in there that describes the minimum requirements for the reproduction system. These requirements can easily be seen as a minimum requirement for a reference monitoring system, and so I'll list them here to give you an idea of what you should have in front of you at a recording or mixing session. Note that these are not standards for recording studios, I'm just suggesting that their a good set of recommendations that can give you an idea of a "good" playback system.

Note that all of the specifications listed here are measured in a free field, 1 m from the acoustic centre of the loudspeaker.

Frequency Response

The on-axis frequency response of the loudspeaker should be measured in one-third octave bands using pink noise as a source signal. The response should not be outside the range of ± 2 dB within the frequency range of 40 Hz to 16 kHz. The frequency response measured at 10° off-axis should not differ from the on-axis response by more than 3 dB. The frequency response measured at 30° off-axis should not differ from the on-axis response by more than 4 dB [ITU, 1997].

All main loudspeakers should be matched in on-axis frequency response within 1 dB in the frequency range of 250 Hz to 2 kHz [ITU, 1997].

Directivity Index

In the frequency range of 500 Hz to 10 kHz, the directivity index, C , of the loudspeakers should be within the limit $6 \text{ dB} \leq C \leq 12 \text{ dB}$ and "should

increase smoothly with frequency” [ITU, 1997].

Non-linear Distortion

If you send a sinusoidal waveform to the loudspeaker that produces a 90 dBspl at the measurement position, the maximum limit for harmonic distortion components between 40 Hz and 250 Hz is 3% (-30 dB) and for components between 250 Hz and 16 kHz is 1% (-40 dB) [ITU, 1997].

Transient Fidelity

If you send a sine wave at any frequency to your loudspeaker and then stop the sine, it should not take more than 5 time periods of the frequency for the output to decay to $1/e$ (approximately 0.37 or -8.69 dB) of the original level [ITU, 1997].

Time Delay

The delay difference between any two loudspeakers should not exceed 100 μ s. (Note that this does not include propagation delay differences at the listening position.) [ITU, 1997]

Dynamic Range

You should be able to play a continuous signal with a level of at least 108 dBspl for 10 minutes without damaging the loudspeaker and without overloading protection circuits [ITU, 1997].

The equivalent acoustic noise produced by the loudspeaker should not exceed 10 dBspl, A-weighted [ITU, 1997].

Two-channel Stereo

So, you’ve bought a pair of loudspeakers following the recommendations of all the people you know (but you bought the ones you like anyway...) you bring them to the studio and carefully position them following all the right rules. Now you have to make sure that the outputs levels of the two loudspeakers is matched. How do you do this?

You actually have a number of different options here, but we’ll just look at a couple, based on the assumption that you don’t have access to really serious (and therefore REALLY expensive) measurement equipment.

SPL Meter Method

One of the simplest methods of loudspeaker calibration is to use pink noise as your signal and an SPL meter as your measurement device. If an SPL meter is not available (a cheap one is only about \$50 at Radio Shack... go treat yourself...) then you could even get away with an omnidirectional condenser microphone (the smaller the diaphragm, the better) and the meter bridge of your mixing console.

Send the pink noise signal to the amplifier (or the crossover input if you're using active crossovers) for one of your loudspeakers. The level of the signal should be 0 dB VU (or +4 dBu).

Place the SPL meter at the listening position pointing straight up. If you are holding the meter, hold it as far away from your body as you can and stand to the side so that the direct sound from the loudspeaker to the meter reflects off your body as little as possible (yes, this will make a difference). The SPL meter should be set to C-weighting and a slow response.

Adjust your amplifier gain so that you get 85 dBspl on the meter. (Feel free to use a different value if you think that you have a really good excuse. The 85 dBspl reference value is the one used by the film industry. Television people use 79 dBspl and music people can't agree on what value to use.)

Repeat this procedure with the other loudspeaker.

Remember that you are measuring one loudspeaker at a time – you should 85 dBspl from each loudspeaker, not both of them combined.

A word of warning: It's possible that your listening position happens to be in a particular location where you get a big resonance due to a room mode. In fact, if you have a smaller room and you've set up your room symmetrically, this is almost guaranteed. We'll deal with how to cope with this later, but you have to worry about it now. Remember that the SPL meter isn't very smart – if there's a big resonance at one frequency, that's basically what you're measuring, not the full-band average. If your two loudspeakers happen to couple differently to the room mode at that frequency, then you're going to have your speakers matched at only one frequency and possibly no others. This is not so good.

There are a couple of ways to avoid this problem. You could change the laws of physics and have room modes eliminated in your room, but this isn't practical. You could move the meter around the listening position to see if you get any weird fluctuations because many room modes produce very localized problems. However, this may not tell you anything because if the mode is a lower frequency, then the wavelength is very long and the whole area will be problematic. Your best bet is to use a measurement device that shows you the frequency response of the system at the listening position, the simplest of which is a real-time analyzer. Using this system, you'll be able to see if you have serious problems in localized frequency bands.

Real-Time Analyzer Method

If you've got a real-time analyzer (or RTA) lying around, you could be a little more precise and get a little more information about what's happening in your listening room at the listening position. Put an omnidirectional microphone with a small diaphragm at the listening position and aim it at

the ceiling. The output should go to the RTA.

Using pink noise at a level of +4 dBu sent to a single loudspeaker, you should see a level of 70 dBspl in each individual band on the RTA [Owinski, 1988]. Whether or not you want to put an equalizer in the system to make this happen is your own business (this topic is discussed a little later on), but you should come as close as you can to this ideal with the gain at the front of the amplifier.

Other methods

There are a lot of different measurement tools out there for doing exactly this kind of work, however, they're not cheap, and if they are, they may not be very reliable (although there really isn't a direct correlation between price and system reliability...) My personal favourites for electroacoustic measurements are a MLSSA system from [DRA Laboratories](#), and a number of solutions from [Brüel & Kjær](#), but there's lots of others out there.

Just be warned, if you spend a lot of money on a fancy measurement system, you should probably be prepared to spend a lot of time learning how to use it properly... My experience is that the more stuff you can measure, the more quickly and easily you can find the wrong answers and arrive at incorrect conclusions.

5.1 Surround

The method for calibrating a 5-channel system is no different than the procedure described above for a two-channel system, you just repeat the process three more times for your Centre, Left Surround and Right Surround channels. (Notice that I used the word "channels" there instead of "loudspeakers" because some studios have more than two surround loudspeakers. For example, if you do have more than one Left Surround loudspeaker, then your Left Surround loudspeakers should all be matched in level, and the total output from all of them combined should be equal to the reference value.)

The only problem that now arises is the question of how to calibrate the level of the subwoofer, but we'll deal with that below.

10.2 Surround

The same procedure holds true for calibration of a 10.2 system. All channels should give you the same SPL level (either wide band with an SPL meter or narrow band with an RTA) at the listening position. The only exception here is the diffuse radiators at $\pm 90^\circ$. You'll probably notice that you won't get as much low frequency energy from these loudspeakers at the listening

position due to the cancellation of the dipole. The easiest way to get around this problem is to band-limit your pink noise source to a higher frequency (say, 250 Hz or so...) and measure one of your other loudspeakers that you've already calibrated (Centre is always a good reference). You'll notice that you get a lower number because there's less low-end – write that number down and match the dipoles to that level using the same band-limited signal.

Ambisonics

Again, the same procedure holds for an Ambisonics configuration.

Subwoofers

Here's where things get a little ugly. If you talk to someone about how they've calibrated their subwoofer level, you'll get one of five responses:

1. "It's perfectly calibrated to +4 dB."
2. "It's perfectly calibrated to -10 dB."
3. "It's perfectly calibrated to +10 dB."
4. "I turned it up until it sounded good."
5. "Huh?"

Oddly enough, it's possible that the first three of these responses actually mean exactly the same thing. This is partly due to an issue that I pointed out earlier in Section 9.2.3. Remember that there's a 10 dB gain applied to the LFE input of a multichannel monitoring system for the remainder of this discussion.

The objective with a subwoofer is to get a low-frequency extension of your system without exaggerating the low-frequency components. Consequently, if you send a pink-noise signal to a subwoofer and look at its output level in a one-third octave band somewhere in the middle of its response, it should have the same level as a one-third octave band in the middle of the response of one of your other loudspeakers. Right? Well... maybe not.

Let's start by looking at a bass-managed signal with no signal sent to the LFE input. If you send a high-frequency signal to the centre channel and sweep the frequency down (without changing the signal level) you should see not change in sound pressure level at the listening position. This is true even after the frequency has gotten so low that it's being produced by the subwoofer. If you look at Figure 9.25 you'll see that this really is just a

matter of setting the gain of the subwoofer amplifier so that it will give you the same output as one of your main channels.

What if you are only using the LFE channel and not using bass management? In this case, you must remember that you only have one subwoofer to compete with 5 other speakers, so the signal has been boosted by 10 dB in the monitoring box. This means that if you send a pink noise to the subwoofer and monitor it in a narrow band in the middle of its range, it should be 10 dB louder than a similar measurement done with one of your main channels. This extra 10 dB is produced by the gain in the monitoring system.

Since the easiest way to send a signal to the subwoofer in your system is to use the LFE input of your monitoring box, you have to allow for this 10 dB boost in your measurements.

Again, you can do your measurements with any appropriate system, but we'll just look at the cases of an SPL meter and an RTA.

SPL Meter Method

We will assume here that you have calibrated all of your main channels to a reference level of 85 dBspl using +4 dBu pink noise.

Send pink noise at +4 dBu, band-limited from 20 to 80 Hz, to your subwoofer through the LFE input of your monitor box. Since the pink noise has been band-limited, we expect to get less output from the subwoofer than we would get from the main channels. In fact, we expect it to be about 6 dB less. However, the monitoring system adds 10 dB to the signal, so we should wind up getting a total of 89 dBspl at the listening position, using a C-weighted SPL meter set to a slow response.

Note that some CD's with test signals for calibrating loudspeakers take the 10 dB gain into account and therefore reduce the level of the LFE signal by 10 dB to compensate. If you're using such a disc instead of producing your own noise, then be sure to find out the signal's level to ensure that you're not calibrating to an unknown level...

If you choose to send your band-limited pink noise signal through your bass management circuitry instead of through the LFE input, then you'll have to remember that you do not have the 10 dB boost applied to the signal. This means that you are expecting a level of 79 dBspl at the listening position.

The same warning about SPL meters as was described for the main loudspeakers holds true here, but moreso. Don't forget that room modes are going to wreak havoc with your measurements here, so be warned. If all you have is an SPL meter, there's not really much you can do to avoid these problems... just be aware that you might be measuring something you

don't want.

Real-Time Analyzer Method

If you're using an RTA instead of an SPL meter, your goal is slightly easier to understand. As was mentioned above, the goal is to have a system where the subwoofer signal routed through the LFE input is 10 dB louder in a narrow band than any of the main channels. So, in this case, if you've aligned your main loudspeakers to have a level of 70 dBspl in each band of the RTA, then the subwoofer should give you 80 dBspl in each band of the RTA. Again, the signal is still pink noise with a level of +4 dBu and band-limited from 20 Hz to 80 Hz.

Summary

| Source | Signal level | RTA
(per band) | SPL Meter |
|--------------------------------|--------------|-------------------|-----------|
| Main channel | -20 dB FS | 70 dBspl | 85 dBspl |
| Subwoofer (LFE input) | -20 dB FS | 80 dBspl | 89 dBspl |
| Subwoofer (main channel input) | -20 dB FS | 70 dBspl | 79 dBspl |

Table 9.4: Sound pressure levels at the listening position for a standard operating level for film [Owinski, 1988]. Note that the values listed here are for a single channel. SPL Meter measurements are done with a C-weighting and a slow response.

9.2.6 Monitoring system equalization

NOT WRITTEN YET

9.2.7 Suggested Reading List

[Owinski, 1988]

9.3 Introduction to Stereo Microphone Technique

9.3.1 Panning

Before we get into the issue of the characteristics of various microphone configurations, we have to look at the general idea of panning in two-channel and five-channel systems. Typically, panning is done using a technique called *constant power pair-wise panning* which is a system whose name contains a number of different issues which are discussed in this and the next chapter.

Localization of real sources

As you walk around the world, you are able to localize sound sources with a reasonable degree of accuracy. This basically means that if your eyes are closed and something out there makes a sound, you can point at it. If you try this exercise, you'll also find that your accuracy is highly dependent on the location of the source. We are much better at detecting the horizontal angle of a sound source than its vertical angle. We are also much better at discriminating angles of sound sources in the front than at the sides of our head. This is because we are mainly relying on two basic attributes of the sound reaching our ears. These are called the *interaural time of arrival difference* (*ITD's*) and the *interaural amplitude difference* (*IAD's*).

When a sound source is located directly in front of you, the sound arrives at your two ears at the same time and at the same level. If the source moves to the right, then the sound arrives at your right ear *earlier* (*ITD*) and *louder* (*IAD*) than it does in your left ear. This is due to the fact that your left ear is farther away from the sound source and that your head gets in the way and shadows the sound on your left side.

Interchannel Differences

Panning techniques rely on these same two differences to produce the simulation of sources located between the loudspeakers at predictable locations. If we send a signal to just one loudspeaker in a two-channel system, then the signal will appear to come from the loudspeaker. If the signal is produced by both loudspeakers at the same level and the same time, then the apparent location of the sound source is at a point directly in front of the listener, halfway between the two loudspeakers. Since there is no loudspeaker at that location, we call the effect a *phantom image*.

The exact location of a phantom image is determined by the relationship of the sound produced by the two loudspeakers. In order to move

the image to the left of centre, we can either make the left channel *louder*, *earlier*, or simultaneously *louder and earlier* than the right channel. This system uses essentially the same characteristics as our natural localization system, however, now, we are talking about *interchannel time differences* and *interchannel amplitude differences*.

Almost every pan knob on almost every mixing console in the world is used to control the interchannel amplitude difference between the output channels of the mixer. In essence, when you turn the pan knob to the left, you make the left channel louder and the right channel quieter, and therefore the phantom image appears to move to the left. There are some digital consoles now being made which also change the interchannel time differences in their panning algorithms, however, these are still very rare.

9.3.2 Coincident techniques (X-Y)

This panning of phantom images can be accomplished not only with a simple pan knob controlling the electrical levels of the two or more channels, we can also rely on the sensitivity pattern of directional microphones to produce the same level differences.

Crossed cardioids

For example, let's take two cardioid microphones and place them so that the two diaphragms are vertically aligned - one directly over the other. This vertical alignment means that sounds reaching the microphones from the horizontal plane will arrive at the two microphones simultaneously - therefore there will be no time of arrival differences in the two channels. consequently we call them *coincident* Now let's arrange the microphones such that one is pointing 45° to the left and the other 45° to the right, remembering that cardioids are most sensitive to a sound source directly in front of them.

If a sound source is located at 0° , directly in front of the pair of microphones, then each microphone is pointing 45° away from the sound source. This means that each microphone is equally insensitive to the sound arriving at the mic pair, so each mic will have the same output. If each mic's output is sent to a single loudspeaker in a stereo configuration then the two loudspeakers will have the same output and the phantom image will appear dead centre between the loudspeakers.

However, let's think about what happens when the sound source is not at 0° . If the sound source moves to the left, then the microphone pointing left

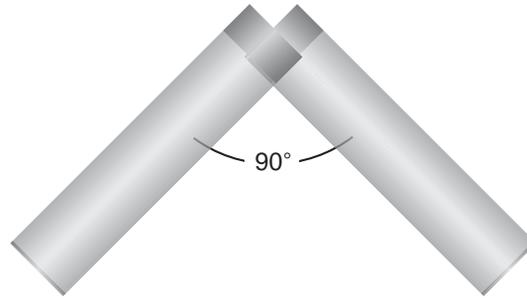


Figure 9.36: A pair of XY cardioids with an included angle of 90° .

will have a higher output because it is more sensitive to signals in directly in front of it. On the other hand, we're moving further away from the front of the right-facing microphone so its output will become quieter. The result in the stereo image is that the left loudspeaker gets louder while the right loudspeaker gets quieter and the image moves to the left.

If we had moved the sound source towards the right, then the phantom image would have moved towards the right.

This system of two coincident 90° cardioid microphones is very commonly used, particularly in situations where it is important to maintain what is called *mono compatibility*. Since the signals arriving at the two microphones are coincident, there are no phase differences between the two channels. As a result, there will be no comb filtering effects if the two channels are summed to a single output as would happen if your recording is broadcast over the radio. Note that, if a pair of 90° coincident cardioids is summed, then the total result is a single, virtual microphone with a sort of weird subcardioid-looking polar plot, but we'll discuss that later.

Another advantage of using this configuration is that, since the phantom image locations are determined only by interchannel amplitude differences, the image locations are reasonably stable and precise.

There are, however, some disadvantages to using this system. To begin with, all of your sound sources located at the centre of the stereo sound stage are located off-axis to the microphones. As a result, if you are using microphones whose off-axis response is less than desirable, then you may experience some odd colouration problems on your more important sources.

A second disadvantage to this technique is that you'll find that, when your musicians are distributed evenly in front of the pair (as in a symphony orchestra, for example), you get sources sounding like they're "clumping"

in the middle of your sound stage. There are a number of different ways of thinking of why this is the case. One explanation is presented in Section 9.4.1. A simpler explanation is given by Jörg Wuttke of [Schoeps Microphones](#). As he points out, a cardioid is one half omnidirectional and one half bidirectional. Therefore, a pair of coincident cardioids gives you a signal, half of which is a pair of coincident omnidirectionals. A pair of coincident omni's will give you a completely mono signal which will image in the dead centre of your loudspeakers – therefore instruments tend to pull to this location.

Blumlein

As we'll see in the next chapter, although a pair of coincident cardioid microphones does indeed give you good imaging characteristics, there are many problems associated with this technique. In particular, you will find that the overall sound stage in a two-channel stereo playback tends to “clump” to the centre quite a bit. In addition, there is no feeling of “spaciousness” that can be generated with phase or polarity differences between the two channels. Both of these problems can be solved by trading in your cardioids for a pair of bidirectional microphones. An arrangement of two bidirectionals in a coincident pair with one pointing 45° to the left and the other 45° to the right is commonly called a *Blumlein pair*, named after the man who patented two-channel stereo sound reproduction, Alan Blumlein.

The outputs of these two bidirectional microphones have some interesting characteristics that will be analyzed in the next chapter, however, we can look at the basic attributes of the configuration here. To begin with, in the area in front of the microphones, you have basically the same behaviour as we saw with the coincident cardioid pair. Changes in the angle of incidence of the sound source result in changes in the interchannel amplitude differences in the channels, resulting in simple pair-wise power panning. Note however, that this pair is more sensitive to changes in angle, so you will experience bigger swings in the location of sound sources with a Blumlein pair than with 90° cardioids.

Let's consider what's happening at the rear of a Blumlein pair. Since a bidirectional microphone has a rear lobe that is symmetrical to the front one, but with a negative polarity, then a Blumlein pair of microphones will have the same response in the rear as it does in the front with only two exceptions. Sources on the rear left of the pair image on the right and sources on the right image on the left. This is because the rear lobe of the left microphone is pointing towards the rear right of the pair, consequently,

the left – right orientation of the sources is flipped.

The other consequence of placing sources in the rear of the pair is that the direct sound is entering the microphones in their negative lobes. As a result, the polarity of the signals is inverted. This will not be obvious for many sources such as wind and string instruments, but you may be able to make a case for the difference being audible on more percussive sounds.

The other major difference between the Blumlein technique and coincident cardioids occurs when sound sources are located on the sides of the pair. For example, in the case of a sound source located at 90° off-axis to the pair on the left, then the source will be positioned in the front lobe of the left microphone but the rear lobe of the right microphone. As a result, the outputs of the two microphones will be matched in level, but they will be opposite in polarity. This results in the same imaging characteristics as we get when we wire one loudspeaker ‘out of phase’ with the other – a very unstable, “phasey” sound that could even be considered to be located outside the loudspeakers.

The interesting thing about this configuration, therefore, is that sound sources in front of the pair (like an orchestra) image normally; we get a good representation of the reverberation and audience coming from the front and rear lobes of the microphones; and that the early reflections from the side walls probably come in the sides of the microphone pair, thus imaging “outside” the loudspeaker aperture.

There are many other advantages to using a Blumlein pair, but these will be discussed in the next chapter.

There are some disadvantages to using this configuration. To begin with, bidirectional microphones are not as available or as cheap as cardioid microphones. The second is that, if a Blumlein pair is summed to mono, then the resulting virtual microphone is a forward-facing bidirectional microphone. This might not be a bad thing, but it does mean that you get as much from the rear of the pair as you do from the front, which might result in a sound that is a little too distant sounding in mono.

ES

There are many occasions where you will want to add some extra microphones out in the hall to capture reverberation with very little direct sound. This is helpful, particularly in sessions where you don’t have a lot of sound-check time, or when the hall has problems. (Typically, you can cover up acoustical problems by adding more microphones to smear things out.)

Many people like to use VERY widely spaced omni’s for this, but this

technique really doesn't make much sense. As we'll see later, the further apart a pair of microphones are in a diffuse field (like a reverberant concert hall) the less correlated they are. If your microphones are stuck on opposite sides of the hall (this is not an uncommon practice) you basically get two completely uncorrelated signals. The result is that the reverberation (and the audience noise, if they're there) sits in two discrete pockets in the listening room – one pocket for each loudspeaker. This gives the illusion of a very wide sound, but there is nothing holding the two sides together – it's just one big hole in the middle.

So, how can we get a nice, wide hall sound, with an even spread and avoid picking up too much direct sound?

This is the goal of the ES (or Enhanced Surround CHECK THE NAME) microphone technique developed by Wieslaw Woszczyk. The configuration is simply two cardioids with an included angle of 180° degrees (placed back-to-back) with one of the outputs reversed in polarity. Each microphone is panned completely to one channel.

This technique has a number of interesting properties.

TO DISCUSS: - SURROUND MATRIX SYSTEMS - MONO (FOR AND AGAINST) - MATRIXED VERSION WITH 1 BIDIRECTIONAL AND 1 OMNI TO GET LOW END

9.3.3 Spaced techniques (A-B)

We have seen above that left – right localization can be achieved with interchannel time differences instead of interchannel amplitude differences. Therefore, if a signal comes from both loudspeakers at the same level, but one loudspeaker is slightly delayed (no more than about a millisecond or so), then the phantom image will pull towards the earlier loudspeaker.

This effect can be achieved in DSP using simple digital delays, however, we can also get a similar effect using two omnidirectional microphones, spaced apart by some distance. Now, if a sound source is off to one side of the spaced pair, then the direct sound will arrive at one microphone before the other and therefore cause an interchannel time difference using distance as our delay line. In this case, the closer the sound source is to 90° to one side of the pair, the greater the time difference, with a maximum at 90° .

There are a number of advantages to using this technique. Firstly, since we're using omnidirectional microphones, we are probably using microphones with well-behaved off-axis frequency responses. Secondly, we will get a very extended low frequency response due to the natural frequency range characteristics of an omnidirectional microphone. Finally, the re-

sulting sound, due to the angle- and frequency-dependent phase differences between the two channels will have “fuzzy” imaging characteristics. Sources will not be precisely located between the loudspeakers, but will have wider, more nebulous locations in the stereo sound stage. Of course, this may be a disadvantage if you’re aiming for really precise imaging.

There are, of course, some disadvantages to using this technique. Firstly, the imaging characteristics are imprecise as was discussed above. In addition, if your microphones are placed too close together, you will get a very mono-like sound, with an exaggerated low frequency content in the listening room due to the increased correlation at longer wavelengths. If the microphones are placed too far apart, then there will be no stable centre image at all, resulting in a *hole-in-the-middle* effect. In this case, all sound sources appear at or around the loudspeakers with nothing in between.

9.3.4 Near-coincident techniques

Coincident microphone techniques rely solely on interchannel amplitude differences to produce the desired imaging effects. Spaced techniques, in theory, produce only interchannel time differences. There is, however, a hybrid group of techniques that produce both interchannel amplitude and time of arrival differences. These are known as *near-coincident techniques*, using a pair of directional microphones with a small separation.

ORTF

As was discussed above, a pair of coincident cardioid microphones provides good mono compatibility, good for radio broadcasts, but does not result in a good feeling of spaciousness. Spaced microphones are the opposite. Once upon a time, the French national broadcaster, the ORTF (L’Office de Radiodiffusion-Télévision Française) wanted to create a happy medium between these two worlds – to have a microphone configuration with reasonable mono compatibility and some spaciousness (or is that not much spaciousness and poor mono compatibility... I guess it depends if you’re a glass-is-half-full or a glass-is-half-empty person. I’m a glass-is-not-only-half-empty-but-now-it’s-also-dirty person, so I’m the wrong one to ask.)

The ORTF came up with a configuration that, it’s said, resembles the configuration of a pair of typical human ears. You take two cardioid microphones and place them at an angle of 110° and 17 cm apart at the capsules.

NOS

Not to be outdone, the Dutch national broadcaster, the NOS (Nederlandsche Omroep Stichting), created a microphone configuration recommendation of their own. In this case, you place a pair of cardioid microphones with an angle of 90° and a separation of 30 cm.

As you would expect from the larger separation between the microphones, your mono compatibility in this case gets worse, but you get a better sense of spaciousness from the output.

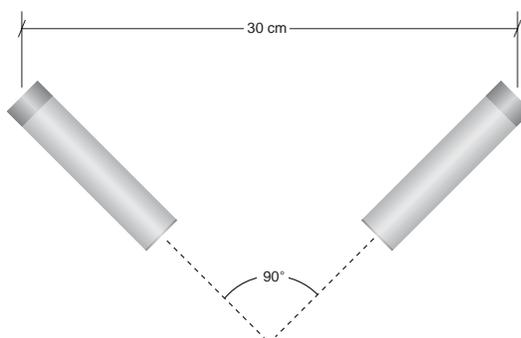


Figure 9.37: An NOS pair of cardioids with an included angle of 90° and a diaphragm separation of 30 cm.

Faulkner

One of the problems with the ORTF and NOS configurations is that they both have their microphones aimed away from the centre of the stage. As a result, what are likely the more important sound sources are subjected to the off-axis response of the microphones. In addition, these configurations may create problems in halls with very strong sidewall reflections, since there is very little attenuation of these sources as a result of directional characteristics.

One option that resolves both of these problems is known as the Faulker Technique, named after its inventor, Tony Faulkner. This technique uses a pair of bidirectional microphones, both facing directly forward and with a separation of 30 cm. In essence, you can consider this configuration to be very similar to a pair of spaced omnidirectionals but with heavy attenuation of side sources, particularly sidewall reflections.

9.3.5 More complicated techniques

Decca Tree

Spaced omnidirectional microphones have a nice, open spacious sound. The wider the separation, the more spacious, but if they get too wide, then you get a hole-in-the-middle. Also, you can wind up with some problems with sidewall reflections that are a little too strong, as was discussed in the section on the Faulker technique.

The British record label Decca came up with a technique that solves all of these problems. They start with a very widely spaced pair of Neumann M50's. This is an interesting microphone consisting of a 25 mm diameter omnidirectional capsule mounted on the face of a 50 mm plastic sphere. The omni capsule gives you a good low frequency response while the sphere gives you a more directional pattern in higher frequencies. (In recent years, many manufacturers have been trying to duplicate this by selling spheres of various sizes that can be attached to a microphone.)

The pair of omni's is placed too far apart to give a stable centre image, but they provide a very wide and spacious sound. The centre image problem is solved by placing a third M50 placed between and in front of the pair. The output of this microphone is sent to both the left and right channels.

This configuration, known as a *Decca Tree* has been used by Decca for probably all of their orchestral recordings for many many years.

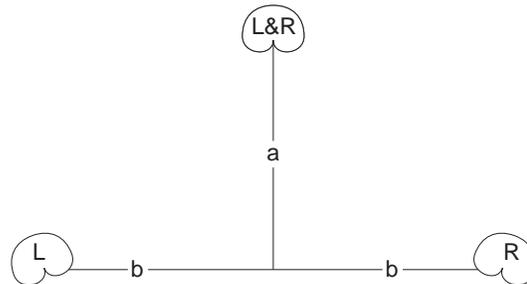


Figure 9.38: A Decca Tree configuration. The size of the array varies according to your ensemble and room, but typical spacings are around 1 m. Note that I've indicated cardioids here, however, the traditional method is with omnidirectional capsules in 50 mm spheres as in the Neumann M50. Also feel free to experiment with splaying of the L and R microphones at different angles.

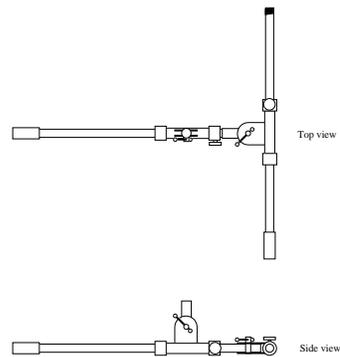


Figure 9.39: An easy way of making a Decca Tree boom using two boom stands and without going to a lot of hassle making special equipment. Note that you'll need at least 1 clamp (preferably 2) to attach your mic clips to the boom ends. This diagram was drawn to be hung (as can be seen in the side view) however, you can stand-mount this configuration as well if you have a sturdy stand. (Lighting stands with mic stand thread adapters work well.)

Binaural

NOT WRITTEN YET

OSS

NOT WRITTEN YET

9.3.6 Suggested Reading List

9.4 General Response Characteristics of Microphone Pairs

Note: Before reading this section, you should be warned that a thorough understanding of the material in Section 5.7 is highly recommended.

9.4.1 Phantom Images Revisited

It is very important to remember that the exact location of a phantom image will be different for every listener. However, there have been a number of studies done which attempt to get an idea of roughly where most people will hear a sound source as we manipulate these parameters.

For two-channel systems, the study that is usually quoted was done by Gert Simonsen during his Master's Degree at the Technical University of Denmark in Lyngby[Simonsen, 1984][Williams, 1990][Rumsey, 2001]. His thesis was constrained to monophonic sound sources reproduced through a standard two-channel setup. Modifications of the signals were restricted to basic delay and amplitude differences, and combinations of the two. According to his findings, some of which are shown in Table 9.5, in order to achieve a phantom image placement of 0° (or the centre-point between the two loudspeakers) both channels must be identical in amplitude and time. Increasing the amplitude of one channel by 2.5 dB (while maintaining the interchannel time relationship) will pull the phantom image 10° off-centre towards the louder speaker. The same result can be achieved by maintaining a 0.0 dB amplitude difference and delaying one channel by 0.22 ms. In this case, the image moves 10° away from the delayed loudspeaker. The amplitude and time differences for 20° and 30° phantom image placements are listed in Table 9.5 and represented in Figures 9.40 and 9.41.

| Image Position | $\Delta Amp.$ | $\Delta Time$ |
|----------------|---------------|---------------|
| 0° | 0.0 dB | 0.0 mS |
| 10° | 2.5 dB | 0.2 mS |
| 20° | 5.5 dB | 0.44 mS |
| 30° | 15.0 dB | 1.12 mS |

Table 9.5: Phantom image location vs. either Interchannel Amplitude Difference or Interchannel Time Difference for two-channel reproduction[Simonsen, 1984][Rumsey, 2001][Williams, 1990]. Note that the $\Delta Amp.$ column assumes a $\Delta Time$ of 0.0 ms and that the $\Delta Time$ column assumes a $\Delta Amp.$ value of 0.0 dB

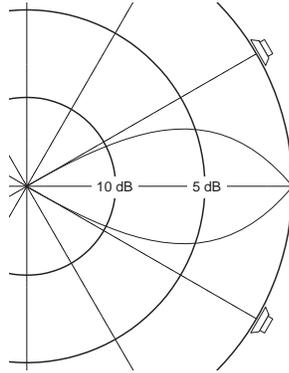


Figure 9.40: Apparent angle vs. averaged interchannel amplitude differences for pair-wise power panned sources in a standard 2-channel loudspeaker configuration. Values are based on those listed in Table 9.5 and interpolated by the author.

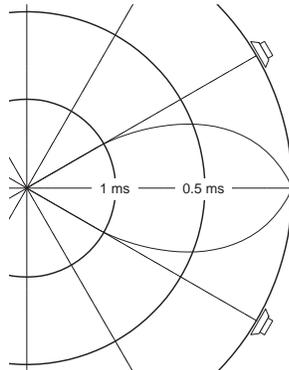


Figure 9.41: Apparent angle vs. averaged interchannel time differences for pair-wise power panned sources in a standard 2-channel loudspeaker configuration. Values are based on those listed in Table 9.5 and interpolated by the author.

A similar study was done by the author at McGill University for a 5-channel configuration [Geoff Martin, 1999]. In this case, interchannel differences were restricted to either the amplitude or time domain without combinations for adjacent pairs of loudspeakers. It was found that different interchannel differences were required to move a phantom image to the position of one of the loudspeakers in the pair as are listed in Tables 9.6 and 9.7. Note that these values should be used with caution due to the large standard deviations in the data produced in the test. One of the principal findings of this research was the large variations between listeners in the apparent location of the phantom image. This is particularly true for side locations, and moreso for images produced with time differences.

| Pair (1/2) | 1 | 2 |
|------------|-------|--------|
| C / L | 14 dB | 12 dB |
| L / LS | 9 dB | >16 dB |
| LS / RS | 9 dB | 9 dB |

Table 9.6: Minimum interchannel amplitude difference required to locate phantom image at the loudspeaker position. For example, it requires an interchannel amplitude difference of at least 14 dB to move a phantom image between the Centre and Left loudspeakers to 0° . The right side is not shown as it is assumed to be symmetrical.

| Pair (1/2) | 1 | 2 |
|------------|---------|---------|
| C / L | >2.0 ms | 2.0 ms |
| L / LS | 1.6 ms | >2.0 ms |
| LS / RS | 0.6 ms | 0.6 ms |

Table 9.7: Minimum interchannel time difference required to locate phantom image at the loudspeaker position. For example, it requires an interchannel time difference of at least 0.6 ms to move a phantom image between the Left Surround and Right Surround loudspeakers to 120° . The right side is not shown as it is assumed to be symmetrical.

Using the smoothed averages of the phantom image locations, polar plots can be generated to indicate the required differences to produce a desired phantom image position as are shown in Figures 9.42 and 9.43.

Summed power response

It is typically assumed that, when you're sitting and listening to the the sound coming out of more than one loudspeaker, the total sound level that you hear is the sum of the sound powers, and not the sound pressures from

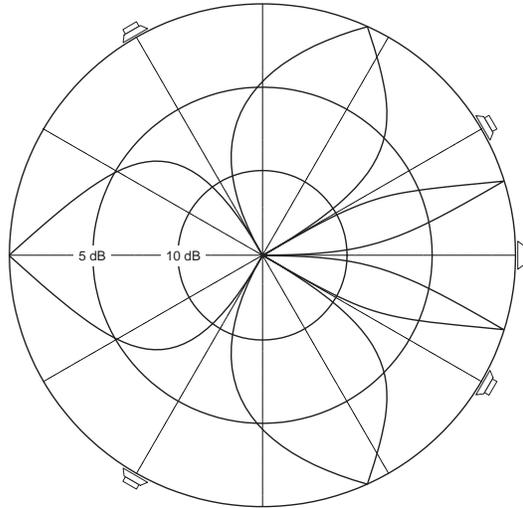


Figure 9.42: Apparent angle vs. averaged interchannel amplitude differences for pair-wise power panned sources in an ITU.R BS.775-1 loudspeaker configuration. Values taken from the raw data acquired by the author in an experiment described in [Geoff Martin, 1999].

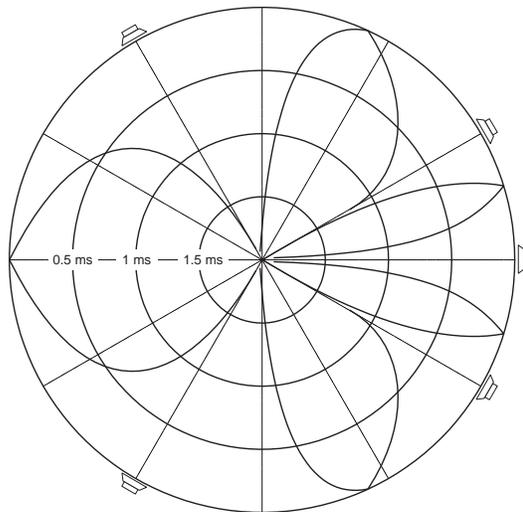


Figure 9.43: Apparent angle vs. averaged interchannel time differences for pair-wise power panned sources in an ITU.R BS.775-1 loudspeaker configuration. Values taken from the raw data acquired by the author in an experiment described in [Geoff Martin, 1999].

the individual drivers. As a result, when you pan a single sound from one loudspeaker to another, you want to maintain a constant summed power, rather than a constant summed pressure.

The top plot in Figure 9.44 shows the two gain coefficients determined by the rotation of a pan knob for two output channels. Since the sum of the two gains at any given position is 1, this algorithm is called a *constant amplitude* panning curve. It works, but, if you take a look at the bottom plot in the same figure, you'll see the problem with it. When the signal is panned to the centre position, there is a drop in the total summed power – in fact, it has dropped by half (or 3 dB) relative to an image located in one of the loudspeakers. Consequently, if this system was used for the panning in a mixing console, as you swept an image from left to centre to right, it would appear to get further away from you at the centre location because it appears to be quieter.

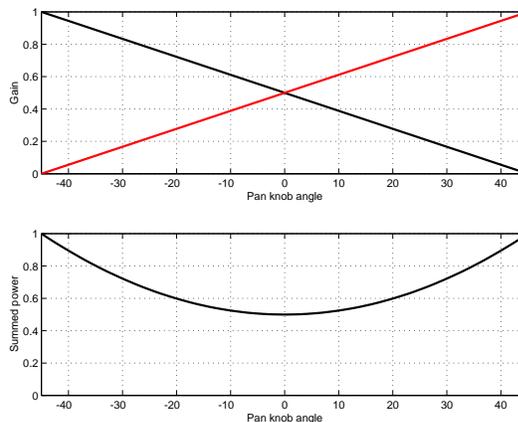


Figure 9.44: The top plot shows a linear panning algorithm where the sum of the two amplitudes will produce the same value at all rotations of the pan knob. The bottom plot shows the resulting power response vs. pan locations.

Consequently we have to use an algorithm which gives us a constant summed power as we sweep the location of the image from one loudspeaker to the other. This is accomplished by using modifying the gain coefficients as is shown in Figure 9.45.

9.4.2 Interchannel Differences

In order to understand the response of a signal from a pair of microphones (whether they're alone, or part of a larger array) we must begin by looking

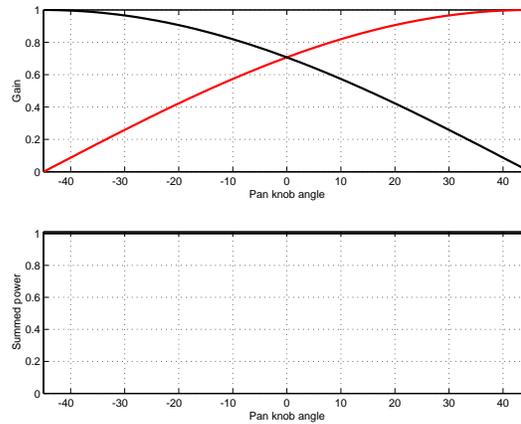


Figure 9.45: The top plot shows a constant power panning algorithm where the sum of the two powers will produce the same value at all rotations of the pan knob. The bottom plot shows the resulting power response vs. pan locations.

at the *differences* in the outputs of the two devices.

Horizontal plane

Cardioids

Unless you only record pop music *and* you never use your imagination, all of the graphs shown above don't really apply to what happens when you're recording. This is because, usually your microphone isn't pointing directly forward... you usually have more than one microphone and they're usually pointing slightly to the left or right of forward, depending on your configuration. Therefore, we have to think about what happens to the sensitivity pattern when you rotate your microphone.

Figure 9.46 shows the sensitivity pattern of a cardioid microphone that is pointing 45° to the right. Notice that this plot essentially looks exactly the same as Figure ??, it's just been pushed to the side a little bit.

Now let's consider the case of a pair of coincident cardioid microphones pointed in different directions. Figure 9.48 shows the plots of two polar patterns for cardioid microphones point at -45° and 45° , giving us an *included angle* (the angle subtended by the microphones) of 90° as is shown in Figure 9.47.

Figure 9.48 gives us two important pieces of information about how a pair of cardioid microphones with an included angle of 90° will behave. Firstly, let's look at the vertical difference between the two curves. Since

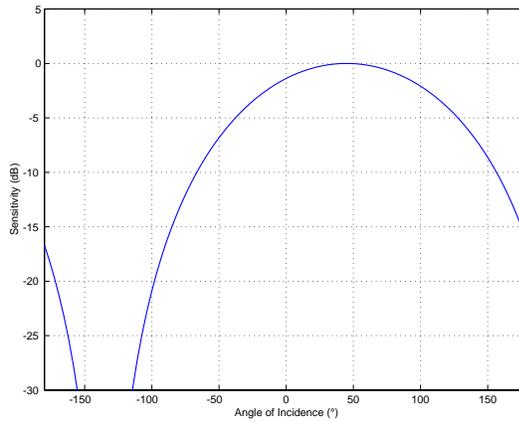


Figure 9.46: Cartesian plot of the absolute value of the sensitivity pattern of a cardioid microphone on a decibel scale turned 45° to the right.

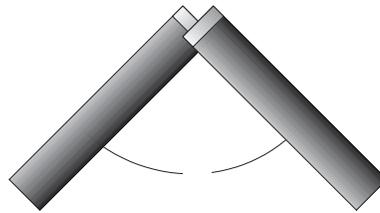


Figure 9.47: Diagram of two microphones with an included angle of 90° .

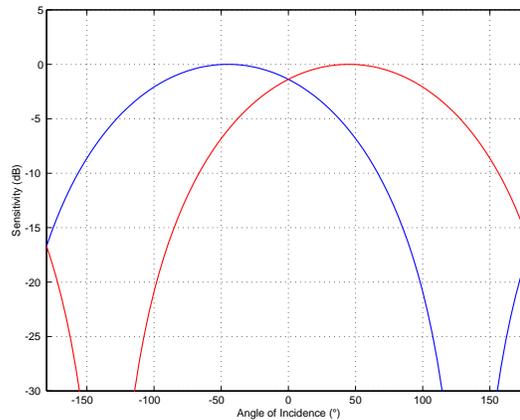


Figure 9.48: Cartesian plot of the absolute value of the sensitivity patterns of two cardioid microphones on a decibel scale turned $\pm 45^\circ$.

this plot essentially shows us the output level of each microphone for a given angle, then the distance between the two plots for that angle will tell us the interchannel amplitude difference. For example, at an angle of incidence (to the pair) of 0° , the two plots intersect and therefore the microphones have the same output level, meaning that there is an amplitude difference of 0 dB. This is also true at 180° , despite the fact that the actual output levels are different than they are at 0° – remember, we’re looking at the *difference* between the two channels and ignoring their individual output levels.

In order to calculate this, we have to find the ratio (because we’re thinking in decibels) between the sensitivities of the two microphones for all angles of incidence. This is done using Equation 9.1.

$$\Delta Amp. = 20 * \log_{10} \left(\frac{S_1}{S_2} \right) \quad (9.1)$$

where

$$S_n = P_n + G_n * \cos(\alpha + \Omega_n) \quad (9.2)$$

where Ω is the angle of rotation of the microphone in the horizontal plane.

If we plot this difference for a pair of cardioids pointing at $\pm 45^\circ$, the result will look like Figure 9.49. Notice that we do indeed have a $\Delta Amp.$ of 0 dB at 0° and 180° . Also note that the graph has positive and negative values on the right and left respectively. This is because we’re comparing

the output of one microphone with the other, therefore, when the values are positive, the right microphone is louder than the left. Negative numbers indicate that the left is louder than the right.

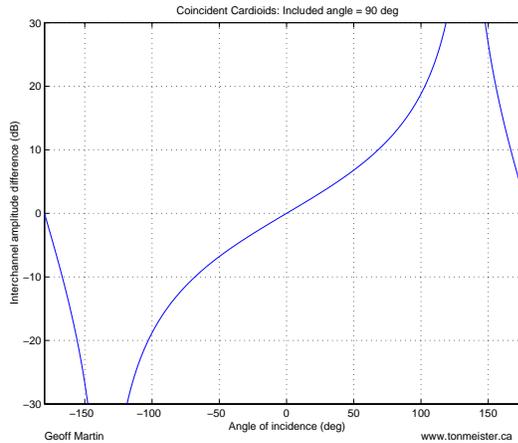


Figure 9.49: Interchannel amplitude differences for a pair of coincident cardioid microphones in the horizontal plane with an included angle of 90° .

Now, if you're using this configuration for a two-channel recording, if you like, you can feel free to try and make a leap from here back to Table 9.5 to make some conclusion about where things are going to wind up being located in the reproduced sound stage. For example, if you believe that a signal with a $\Delta Amp.$ of 2.5 dB results in a phantom image location of 10° , then you can go to the graph in Figure 9.49, find out where the graph crosses a $\Delta Amp.$ of 2.5 dB then find the corresponding angle of incidence. This then tells you that an instrument located at that angle to the microphone pair will show up at 10° off-centre between the loudspeakers. This is, of course, if you're that one person in the world for whom Table 9.5 holds true (meaning you're probably also 172.3 cm tall, you have 2.6 children and two thirds of a dog, you live in Boise, Idaho and that your life is exceedingly... well... average...)

We can do this for any included angle between the microphone pair, from 0° through to 180° . There's no point in going higher than 180° because we'll just get a mirror image. For example, the response for an included angle of 190° is exactly the same as that for 170° , just pointing towards the rear of the pair instead of the front.

Of course, if we actually do the calculation for an included angle of 0° , we're only going to find out that the sensitivities of the two microphones are

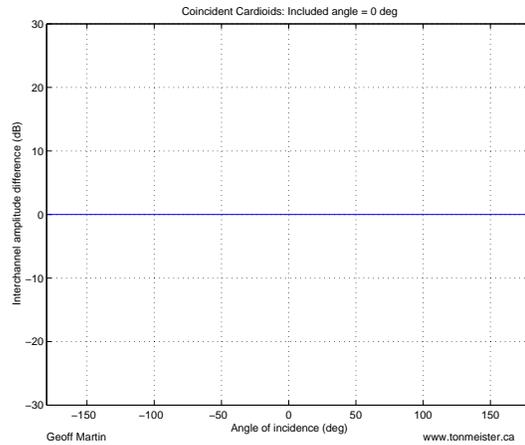


Figure 9.50: Interchannel amplitude differences for a pair of coincident cardioid microphones in the horizontal plane with an included angle of 0° .

matched, and therefore the $\Delta Amp.$ is 0 dB at all angles of incidence as is shown in Figure 9.50. This is true regardless of microphone polar pattern.

Note that, in the case of all included angles except for 0° , the plot of $\Delta Amp.$ for cardioids goes to $\pm\infty$ dB because there will be one angle where one of the microphones has no output and because, on a decibel scale, something is infinitely louder than nothing.

Also note that, in the case of cardioids, every value of $\Delta Amp.$ is duplicated at another angle of incidence. For example, in the case of an included angle of 90° , $\Delta Amp. = +10$ dB at angles of incidence of approximately 70° and 170° . This means that sound sources at these two angles of incidence to the microphone pair will wind up in the same location in the reproduced sound stage. Remember, however, that we don't know the relative levels of these two sources because, all we know is their difference. It's quite probable that one of these two locations will result in a signal that is much louder than the other, but that isn't our concern just yet.

Also note that there is one included angle (of 180°) that results in a response characteristic that is symmetrical (at least in each polarity) around the ∞ dB point.

Subcardioids

Looking back at Figures ?? through ?? we can see that the lowest sensitivity from a subcardioid microphone is 6 dB below its on-axis sensitivity. As a result, unlike a pair of cardioid microphones, the $\Delta Amp.$ of a pair of subcardioid microphones cannot exceed the ± 6 dB window, regardless of

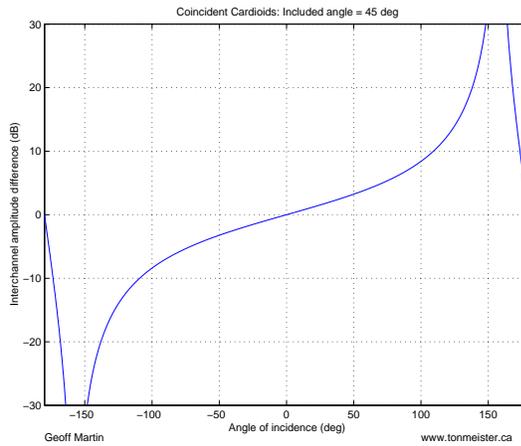


Figure 9.51: Interchannel amplitude differences for a pair of coincident cardioid microphones in the horizontal plane with an included angle of 45° .

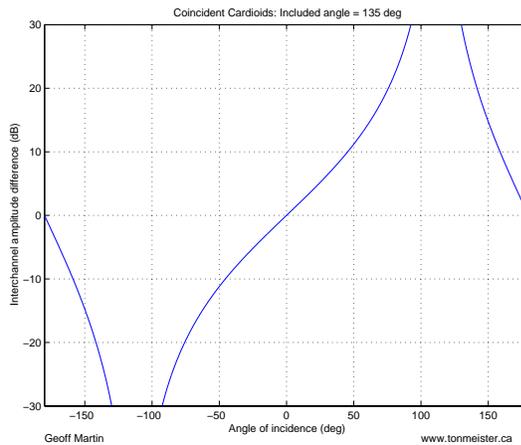


Figure 9.52: Interchannel amplitude differences for a pair of coincident cardioid microphones in the horizontal plane with an included angle of 135° .

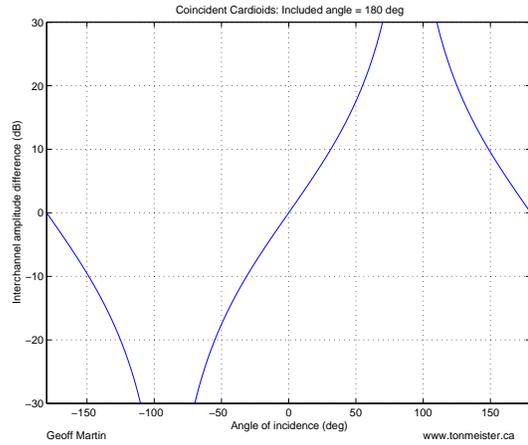


Figure 9.53: Interchannel amplitude differences for a pair of coincident cardioid microphones in the horizontal plane with an included angle of 180°.

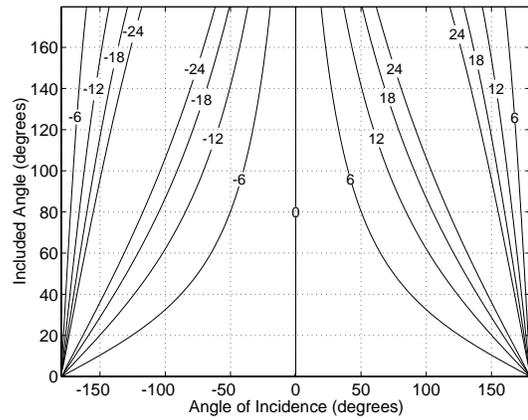


Figure 9.54: Contour plot showing the difference in sensitivity in dB between two coincident cardioid microphones with included angles of 0° to 180°, angles of rotation from -180° to 180° and a 0° angle of elevation. Note that Figure 9.49 is a horizontal “slice” of this contour plot where the included angle is 90°.

included angle or angle of incidence.

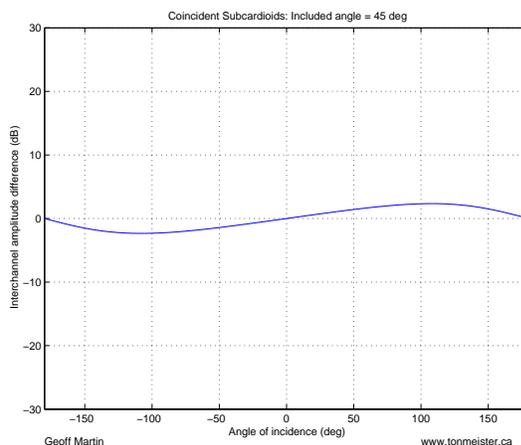


Figure 9.55: Interchannel amplitude differences for a pair of coincident subcardioid microphones in the horizontal plane with an included angle of 45° .

As can be seen in Figures 9.55 through 9.58, a pair of subcardioid microphones does have one characteristic in common with a pair of cardioid microphones in that there are two angles of incidence for every value of ΔAmp .

There is another aspect of subcardioid microphone pairs that should be remembered. As has already been discussed, in the case of subcardioids, it is impossible to have a value of ΔAmp outside the ± 6 dB window. Armed with this knowledge, take a look back at Tables 9.5, 9.6 and 9.7. You'll note that in all cases of pair-wise panning, it takes more than 6 dB to swing a phantom image all the way into one of the two loudspeakers. Consequently, it is safe to say that, in the particular case of coincident subcardioid microphones, all of the the sound stage will be confined to a width smaller than the angle subtended by the two loudspeakers reproducing the two signals. As a result, if you want an image that is at least as wide as the loudspeaker aperture, you'll have to introduce some time differences between the microphone outputs by separating them a little. This will be discussed further below.

Bidirectionals

The discussion of pairs of bidirectional microphones has to start with a reminder of two characteristics of their polar pattern. Firstly, the negative polarity of the rear lobe can never be forgotten. Secondly, as we will see below, it is significant to remember that, in the horizontal plane, this

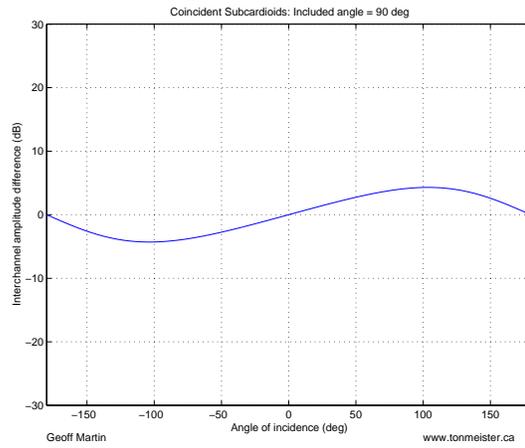


Figure 9.56: Interchannel amplitude differences for a pair of coincident subcardioid microphones in the horizontal plane with an included angle of 90°.

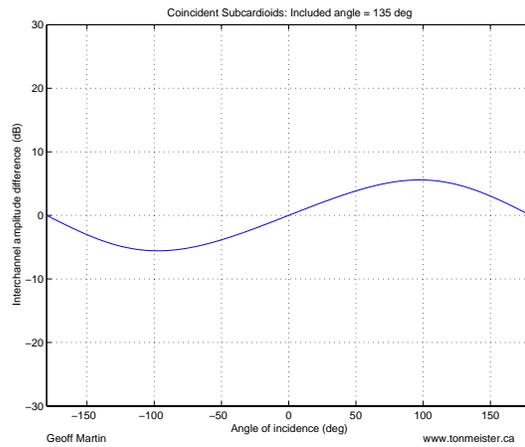


Figure 9.57: Interchannel amplitude differences for a pair of coincident subcardioid microphones with an included angle of 135° in the horizontal plane.

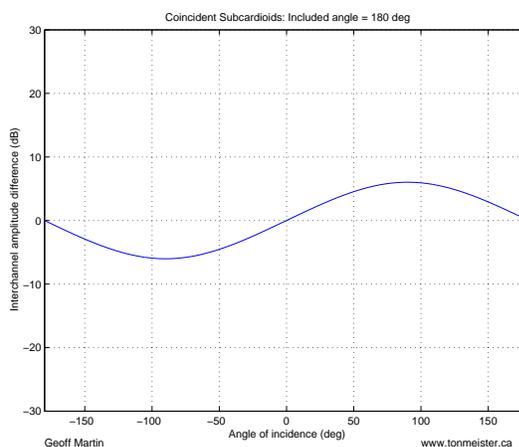


Figure 9.58: Interchannel amplitude differences for a pair of coincident subcardioid microphones with an included angle of 180° in the horizontal plane.

microphone has two angles which have a sensitivity of 0 (or $-\infty$ dB).

Figure 9.59 shows the $\Delta Amp.$ of the absolute value of a pair of bidirectional microphones with an included angle of 90° . In this case, the absolute value of the microphone's sensitivity is used in order to avoid errors when calculating the logarithm of a negative number. This negative is the result of angles of incidence which produce sensitivities of opposite polarity in the two microphones. For example, in this particular case, at an angle of incidence of $+90^\circ$ (to the right), the right microphone sensitivity is positive while the left one is negative. However, it must be remembered that the absolute value of these two sensitivities are identical at this location.

A number of significant characteristics can be seen in Figure 9.59.

Firstly, note that there are now four angles of incidence where the $\Delta Amp.$ reaches $-\infty$ dB. This is due to the fact that, in the horizontal plane, bidirectional microphones have two null points.

Secondly, note that the pattern in this plot is symmetrical around these infinite peaks, just as was the case with a pair of cardioid microphones at 180° , apparently resulting in four angles of incidence which result in sound sources located at the same phantom image location. This, however, is not the case due to polarity differences. For example, although sound sources located at 30° and 60° (symmetrical around the 45° location) appear to result in identical sensitivities, the 30° location produces similar polarity signals whereas the 60° location produces opposite polarity signals.

Finally, it is significant to note that the response of the microphone pair

is symmetrical front-to-back, with a left/right and a polarity inversion. For example, a sound source at $+10^\circ$ results in the same $\Delta Amp.$ as a sound source at -170° , however, the rear source will be have a negative polarity in both channels. Similarly, a sound source at $+60^\circ$ will have the same $\Delta Amp.$ as a sound source at -120° , however the former will be positive in the “right” channel and negative in the “left” whereas the opposite is the case for the latter.

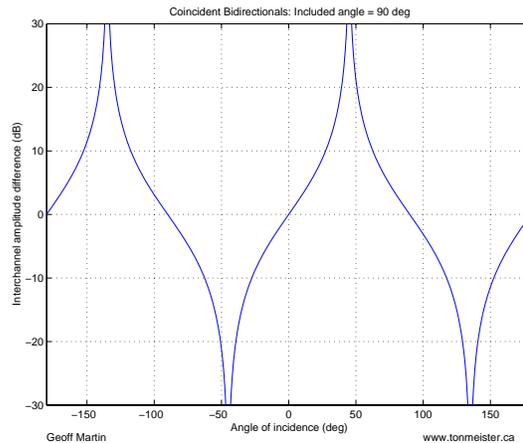


Figure 9.59: Interchannel amplitude differences for a pair of coincident bidirectional microphones in the horizontal plane with an included angle of 90° .

Various other included angles for a pair of bidirectional microphones results in a similar pattern as was seen in Figure 9.59, with a “skewing” of the response curves. This can be seen in Figures 9.60 and 9.61.

It is also important to note that Figures 9.60 and 9.61 are mirror images of each other. This, however does not simply mean that the pair can be considered to be changed from “pointing” from the front to the side in this case. This is again due to the polarity differences between the two channels for specific various angles of incidence.

There is one final configuration worth noting in the specific case of bidirectional microphones; when the included angle is 180° . As can be seen in Figure 9.62, this results in the absolute values of the sensitivities being matched at all angles of incidence. Remember however, that in this particular case, this means that the two channels are exactly matched and opposite in polarity – theoretically, you wind up with exactly the same signal as you would with one microphone split to two channels on the mixing console and the polarity switch (frequently incorrectly referred to as the “phase” switch)

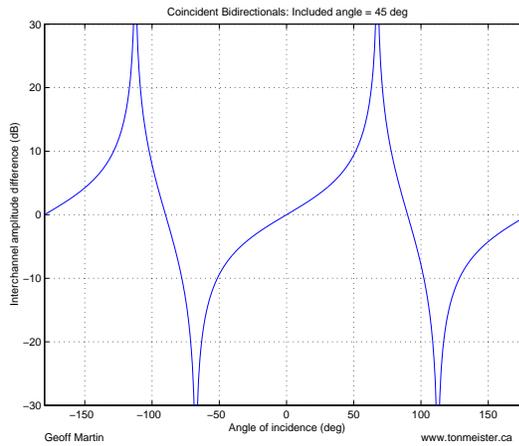


Figure 9.60: Interchannel amplitude differences for a pair of coincident bidirectional microphones in the horizontal plane with an included angle of 45° .

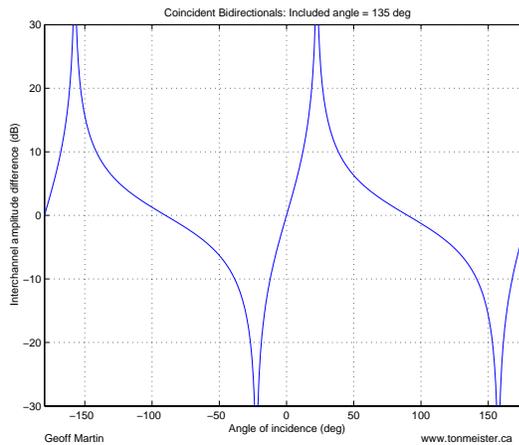


Figure 9.61: Interchannel amplitude differences for a pair of coincident bidirectional microphones in the horizontal plane with an included angle of 135° .

engaged on one of the two channels.

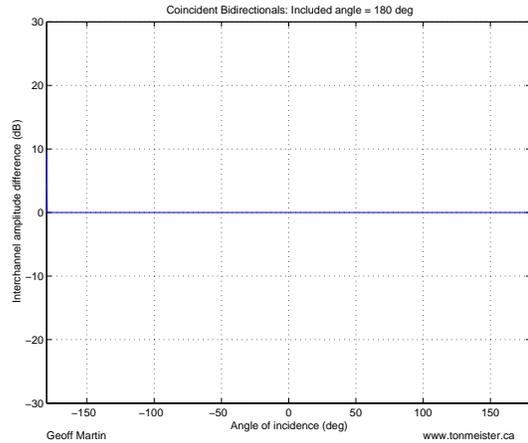


Figure 9.62: Interchannel amplitude differences for a pair of coincident bidirectional microphones in the horizontal plane with an included angle of 180° .

Hypercardioids

Not surprisingly, the response of a pair of hypercardioid microphones looks like a hybrid of the bidirectional and cardioid pairs. As can be seen in Figure 9.64, there are four infinite peaks in the value of $\Delta Amp.$, similar to bidirectional pairs, however the slope of the peaks are skewed further left and right as in the case of cardioids.

Again, similar to the case of bidirectional microphones, changing the included angle of the hypercardioids results in a further skewing of the response curve to one side or the other as can be seen in Figures 9.65 and 9.66.

Figure 9.67 shows the interesting case of hypercardioid microphones with an included angle of 180° . In this case the maximum sensitivity point in the rear lobe of each microphone is perfectly aligned with the maximum sensitivity point in the other microphone's front lobe. However, since the rear lobe has a sensitivity with an absolute value that is 6 dB lower than the front lobe, the value of $\Delta Amp.$ remains outside the ± 6 dB window for the larger part of the 360° rotation.

Spaced omnidirectionals

In the case of spaced omnidirectional microphones, it is commonly assumed that the distance to the sound source is adequate to ensure that the impinging sound can be considered to be a plane wave. In addition, it is also assumed that there is no difference in signal levels due to differences

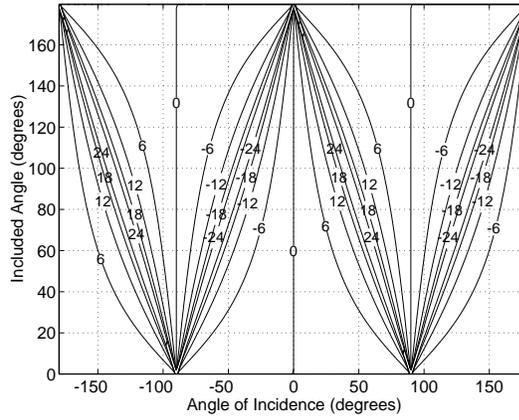


Figure 9.63: Contour plot showing the difference in sensitivity in dB between two coincident bidirectional microphones with included angles of 0° to 180°, angles of rotation from -180° to 180° and a 0° angle of elevation.

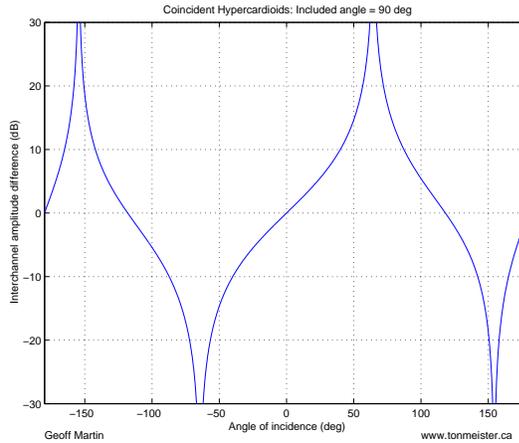


Figure 9.64: Interchannel amplitude differences for a pair of coincident hypercardioid microphones in the horizontal plane with an included angle of 90°.

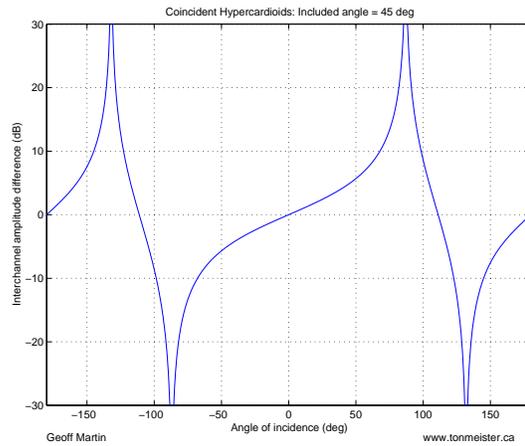


Figure 9.65: Interchannel amplitude differences for a pair of coincident hypercardioid microphones in the horizontal plane with an included angle of 45° .

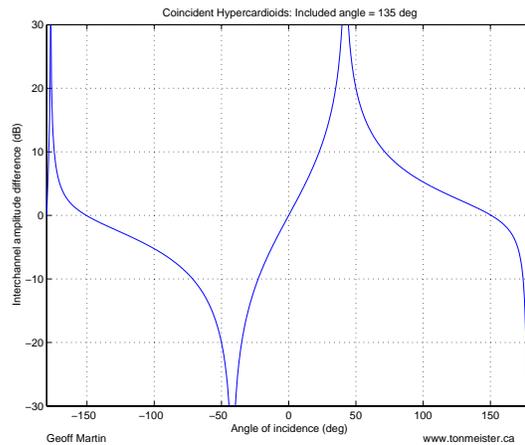


Figure 9.66: Interchannel amplitude differences for a pair of coincident hypercardioid microphones in the horizontal plane with an included angle of 135° .

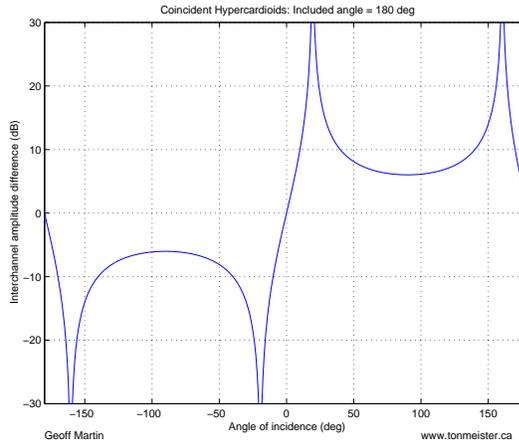


Figure 9.67: Interchannel amplitude differences for a pair of coincident hypercardioid microphones in the horizontal plane with an included angle of 180°.

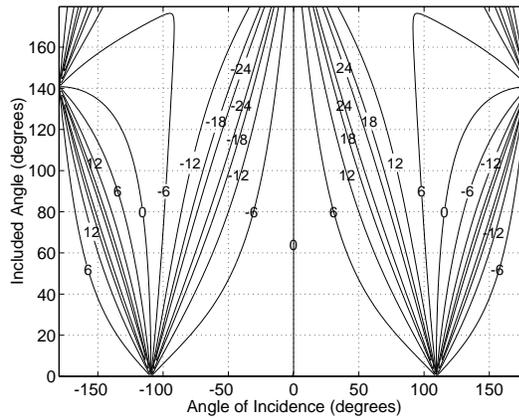


Figure 9.68: Contour plot showing the difference in sensitivity in dB between two coincident hypercardioid microphones with included angles of 0° to 180°, angles of rotation from -180° to 180° and a 0° angle of elevation.

in propagation distance to the transducers. In reality, for widely spaced microphones and/or for sound sources closely located to any microphone, neither of these assumptions is correct, however they will be used for this discussion.

The difference in time of arrival of a sound at two spaced microphones is dependent both on the separation of the transducers d and the angle of rotation around the pair ϑ .

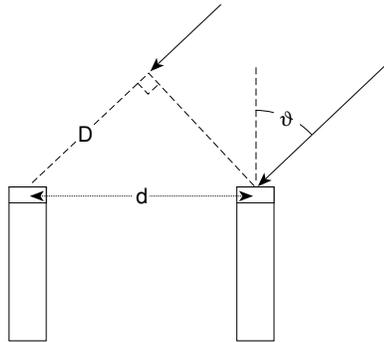


Figure 9.69: Spaced omnidirectional microphones showing the microphone separation d , the angle of rotation ϑ and the resulting extra distance D to the further microphone.

The additional distance, D , travelled by the sound wave to the further of the two microphones, shown in Figure 9.69, can be calculated using Equation 9.3.

$$D = d \sin \vartheta \quad (9.3)$$

where d is the distance between the microphone capsules in cm.

The additional time $\Delta Time$ required for the sound to travel this distance is calculated using Equation 9.4.

$$\Delta Time = \frac{10D}{c} \quad (9.4)$$

where $\Delta Time$ is the interchannel time difference in ms, ϑ is the angle of incidence of the sound source to the pair, and c is the speed of sound in m/s.

This time of arrival difference is plotted for various microphone separations in Figures 9.70 through 9.73. Note that the general curve formed by this calculation is a simple sine wave, scaled by the separation between the microphones. Also note that the value of $\Delta Time$ is 0 ms for sound sources located at 0° and 180° and a maximum for sound sources at 90° and -90° .

As was mentioned early in the section on interchannel amplitude differences between coincident directional microphones, one might be tempted to draw conclusions and predictions regarding image locations based on the values of $\Delta Time$ and the values listed in the tables and figures in Section 9.3.1. Again, one shouldn't be hasty in this conclusion unless you consider your listeners to be average.

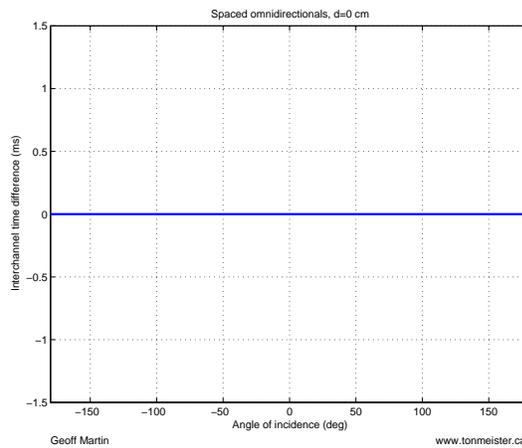


Figure 9.70: Interchannel time differences for a pair of spaced microphones in the horizontal plane with a separation of 0 cm.

Three-dimensional analysis

One of the big problems with looking at microphone polar responses in only the horizontal plane is that we usually don't only have sound sources restricted to those two dimensions. Invariably, we tend to raise or lower the microphone stand to obtain a different direct-reverberant ratio, for example, without considering that we're also changing the vertical angle of incidence to the microphone pair. In almost all cases, this has significant effects on the response of the pair which can be seen in a three-dimensional analysis.

In order to include vertical angles in our calculations of microphone sensitivity, we need to use a little spherical trigonometry – not to worry though. The easiest way to do this is to follow the instructions below:

1. Put your two index fingers side by side pointing forwards.
2. Rotate your right index finger 45° to the right in the horizontal plane. Your fingers should now be at an angle of 45° .

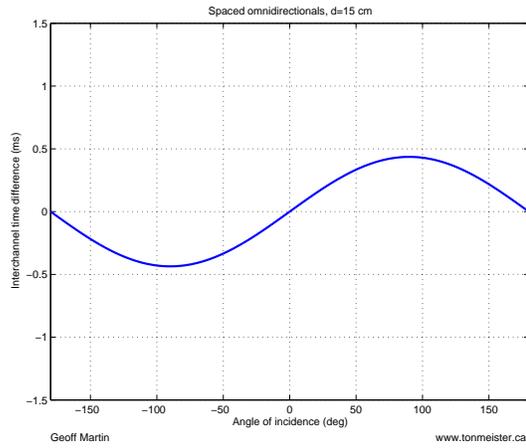


Figure 9.71: Interchannel time differences for a pair of spaced microphones in the horizontal plane with a separation of 15 cm.

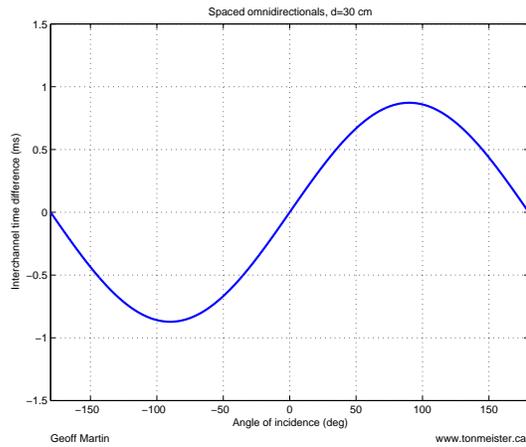


Figure 9.72: Interchannel time differences for a pair of spaced microphones in the horizontal plane with a separation of 30 cm.

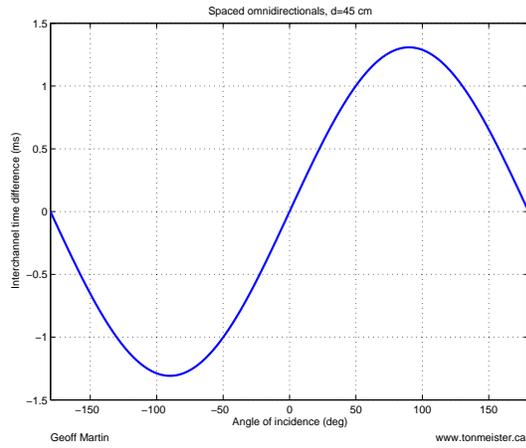


Figure 9.73: Interchannel time differences for a pair of spaced microphones in the horizontal plane with a separation of 45 cm.

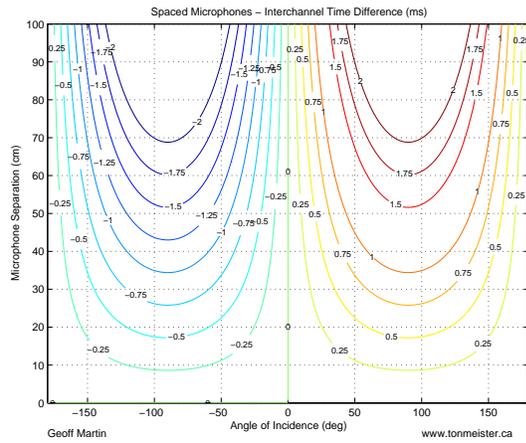


Figure 9.74: Interchannel time differences vs. microphone separation for a pair of spaced microphones in the horizontal plane.

3. Rotate your right index finger 45° upwards. Make sure that this new angle of rotation goes 90° off the horizontal plane.
4. The question we now need to ask is, “What is the angle subtended by your two fingers?”

The answer to this last question is actually pretty easy. If we call the horizontal angle of rotation ϑ , matching the angle in the horizontal plane we talked about earlier, and the vertical angle of rotation ϕ , then the total resulting angle γ can be calculated using Equation 9.5.

$$\gamma = \arccos(\cos(\vartheta) \cos(\phi)) \quad (9.5)$$

Now, there’s one nice little thing about this equation. Since we’re talking about microphone sensitivity patterns, and since the only part of this pattern is dependent on the Pressure Gradient component of the microphone, and since this component only relies on the cosine of the angle of incidence, then we don’t need to do as much math. For example, what we’re really interested in is $\cos(\gamma)$ but we can only calculate γ by doing an arccos function. So, instead of using Equation 9.5, we can simplify it to Equation 9.7.

$$\cos \gamma = \cos(\arccos(\cos(\vartheta) \cos(\phi))) \quad (9.6)$$

$$= \cos(\vartheta) \cos(\phi) \quad (9.7)$$

There’s one more thing. In order to simplify our lives a bit, I’m going to restrict the included angle between the microphones to the horizontal plane. Basically, this means that, for all three dimensional analyses in this paper, we’re thinking that we have a pair of microphones that’s set up parallel to the floor, and that the instrument can be at any angle of incidence to the pair. That angle of incidence is comprised of an angle of rotation and an angle of elevation. We’re not going to have the luxury of tilting the microphone pair on its side (unless you’re able to just think of that as moving the instrument to a new position...).

So, in order to convert Equation 9.2 into a three-dimensional version, we combine it with Equation 9.7, resulting in Equation 9.8.

$$S = P + G(\cos(\vartheta) \cos(\phi)) \quad (9.8)$$

This can include the horizontal rotation of the microphone, Ω as follows:

$$S_n = P_n + G_n(\cos(\vartheta + \Omega_n) \cos(\phi)) \quad (9.9)$$

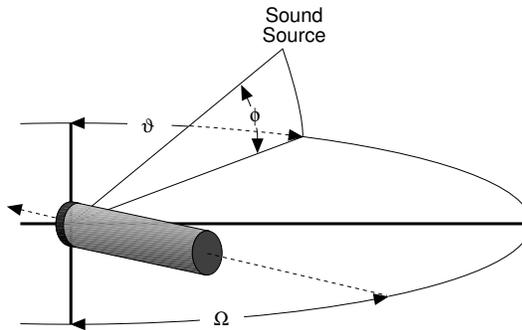


Figure 9.75: A three-dimensional view of a microphone showing the microphone angle of rotation, Ω , the angle of rotation of the sound source to the pair ϑ , and the elevation angle of the sound source ϕ .

3-D Polar Coordinate Systems

Some discussion should be made here regarding the issue of different three-dimensional polar coordinate systems. Microphone polar patterns in three dimensions are typically described using the *spherical coordinate system* which uses two angles referenced to the origin on the surface of the sphere at the location $(0, 0)$. The first, α , is an angle of rotation in the horizontal axis around the centre of the sphere. The second, δ , is an angle of rotation around the axis intersecting the centre of the sphere and the origin. This is shown on the right in Figure 9.76. In the case of microphones, the origin is thought of as being located at the centre of the diaphragm of the microphone and the axis for the rotation δ is perpendicular to the diaphragm.

The *geographic coordinate system* also uses two angles of rotation. Similar to the spherical coordinate system, the first, ϑ , is a rotation around the centre of a sphere in the horizontal plane from the origin at the location $(0,0)$. In geographic terms, this would be the measurement of longitude around the equator. The second angle, ϕ , is slightly different in that it is an orthogonal rotation off the equator, also rotating around the sphere's centre. The geographic equivalent of this vertical rotation is the latitude of a location as can be seen on the left in Figure 9.76.

This investigation uses the geographic coordinate system for its evaluation in order to make the explanations and figures more intuitive. For example, when a recording engineer places a microphone pair in front of an ensemble and raises the microphone stand, the resulting change in polar location of the instruments relative to the array is a change in elevation in

the geographic coordinate system. These changes in the angle of elevation of the microphone pair can correspond to changes in ϕ up to $\pm 90^\circ$.

One issue to note in the use of the geographic coordinate system is the location of positions with values of ϕ outside the $\pm 90^\circ$ window. It must be kept in mind that these positions produce and alternation from left to right and vice versa. Therefore $\vartheta = 45^\circ, \phi = 180^\circ$ is in the same location as $\vartheta = -135^\circ, \phi = 0^\circ$

Although the use of the geographic coordinate system makes the figures and discussion of microphone pair characteristics more intuitive, this unfortunately comes at the expense of a somewhat increased level of complexity in the calculations.

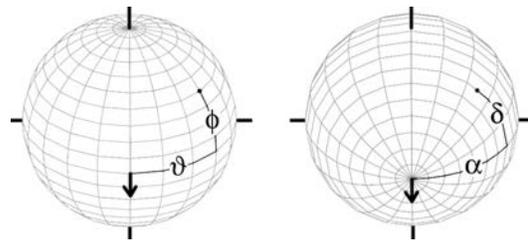


Figure 9.76: A comparison of the geographic and spherical coordinate systems. The point (ϑ, ϕ) in the geographic coordinate system shown on the left is identical to the point (α, δ) in the spherical coordinate system on the right.

Cardioids

We'll begin with a pair of coincident cardioid microphones. Figures 9.77 and 9.78 show plots of the interchannel amplitude difference of a pair of cardioids with an included angle of 90° . (Unfortunately, if you're looking at a black and white printout of this paper, you're going to have a bit of a problem seeing things... sorry...)

Take a look at the spherical plot in Figure 9.77. If we look at the value in dB plotted on this sphere around its equator, we'd get exactly the same graph as is shown in Figure 9.48. Now, however, we can see that changes in the vertical angle might have an effect on the way things sound. For example, if we have a sound source with a vertical elevation of 0° and a horizontal rotation of 90° , then the interchannel amplitude difference is about 18 dB, give or take (to get this number, I cheated and looked back at Figure 9.49). Now, if we maintain that angle of rotation and change the vertical angle, we reduce this difference until, with a vertical angle of 90° , the interchannel amplitude difference is 0 dB – in other words, if the sound source is directly above the pair, the outputs of the two mic's are the same.

What effect does this have on our phantom image location in the real world? Let's go do a live-to-two-track recording where we set up a 90° pair of cardioids at eye level in front of an orchestra at the conductor's position. This means that your principal double bass player is at a vertical angle of 0° and a horizontal angle of 90° , so that instrument will be about 18 dB louder in the right channel than in the left. This means that its phantom image will probably be parked in the right speaker. We start to listen to the sound and we decide that we're too close to the orchestra – and since we read a rule from an unreliable source that concert halls always sound much better if you go up about 4 or 5 m, we'll raise the mic stand way up. This means that the double bass is still at an angle of rotation of 90° , but now at a vertical angle of about 60° (we went *really* high). This means that the interchannel amplitude difference for the instrument has dropped to about 6 dB, which would put it well inside the right loudspeaker in the recording.

So, the moral of the story is, if you have a pair of coincident cardioids and you raise your mic stand without pointing the mic's downwards to compensate, your orchestra gets narrower in the stereo image. Also, don't forget that this doesn't just apply to two-channel stuff. We could just as easily be talking about the image between your surround loudspeakers.

Figure 9.78 shows a contour map of the same plot shown in Figure 9.77 which makes it a little easier to read. Now, the two angles are the X- and Y-axes of the graph, and the lines indicate the angles where a given interchannel amplitude difference occurs. From this, you can see that, unless your rotation is 0° , then any change in the vertical angle, upwards or downwards, will reduce your interchannel difference. This is true for all included angles (except for 0°) of a pair of coincident cardioids.

There's one more thing to consider here, and that's the fact that the microphones are not just recording an instrument – they're recording the room as well. So, you have to keep in mind that, in the case of coincident cardioids, all reflections and reverberation that come from above or below the microphones will tend to pull towards the centre of your loudspeaker pair. Again, the greater the angle of elevation, the more the image collapses.

Subcardioids

Subcardioids have a pretty predictable behaviour, now that we've looked at the response patterns of cardioid microphones. The only big difference is that, as we've seen before, the interchannel amplitude difference never goes outside the ± 6 dB window. Apart from that, the responses are not too different.

Bidirectionals

Now, let's compare those results with a pair of coincident bidirectional

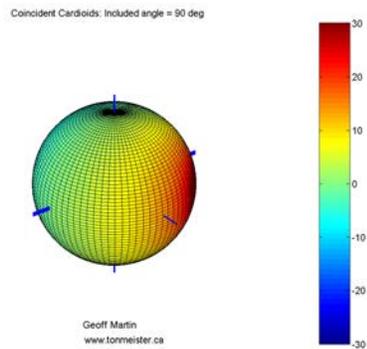


Figure 9.77: Interchannel amplitude difference response (in dB) for a pair of coincident cardioid microphones with an included angle of 90° .

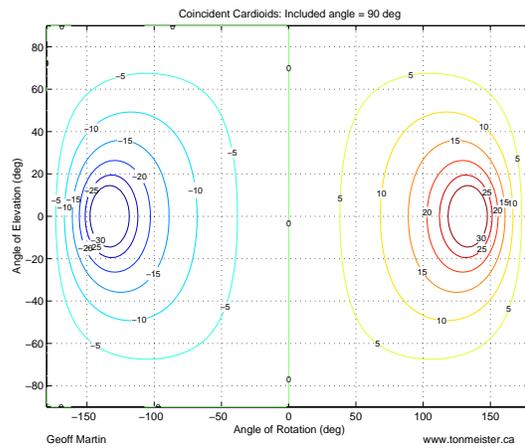


Figure 9.78: Interchannel amplitude difference response for a pair of coincident cardioid microphones with an included angle of 90° .

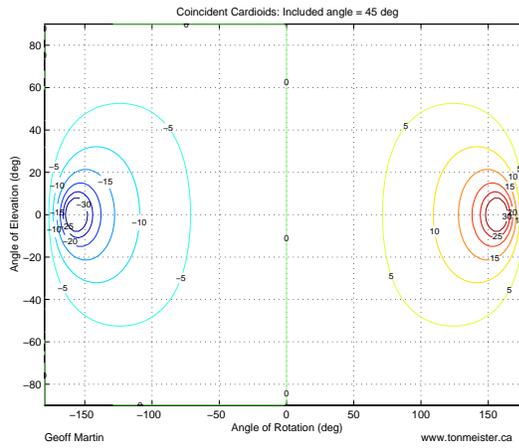


Figure 9.79: Interchannel amplitude difference response for a pair of coincident cardioid microphones with an included angle of 45° .

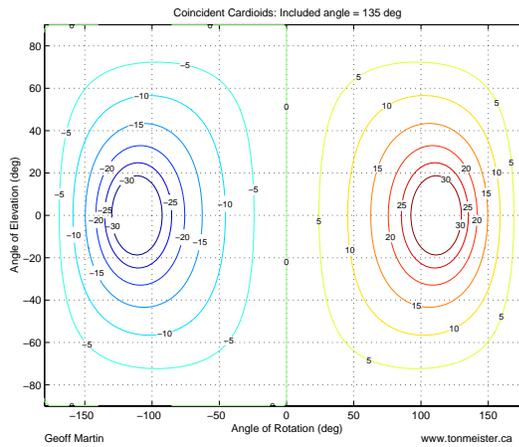


Figure 9.80: Interchannel amplitude difference response for a pair of coincident cardioid microphones with an included angle of 135° .

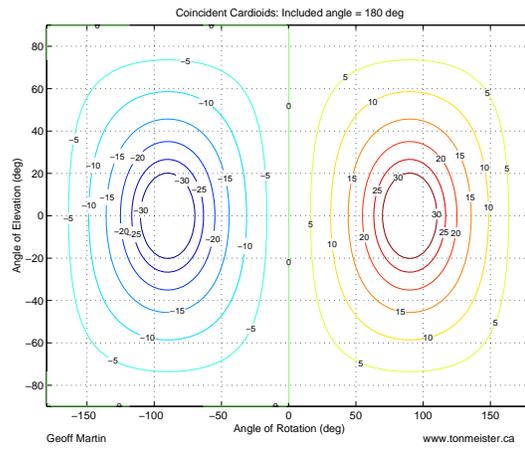


Figure 9.81: Interchannel amplitude difference response for a pair of coincident cardioid microphones with an included angle of 180°.

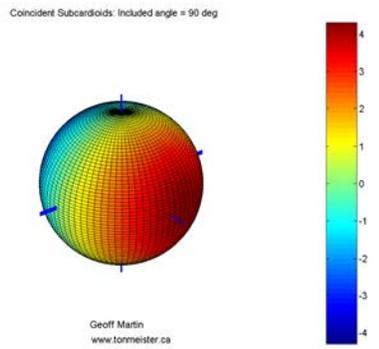


Figure 9.82: Interchannel amplitude difference response (in dB) for a pair of coincident subcardioid microphones with an included angle of 90°.

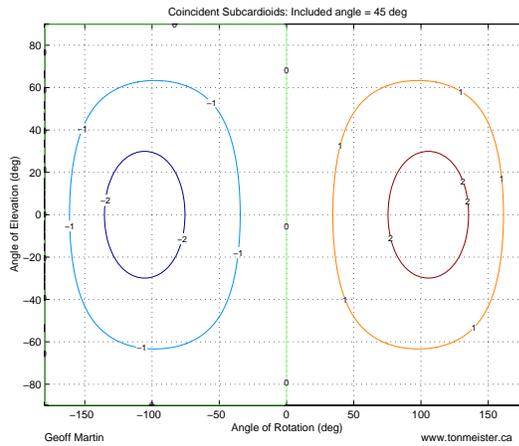


Figure 9.83: Interchannel amplitude difference response for a pair of coincident subcardioid microphones with an included angle of 45° .

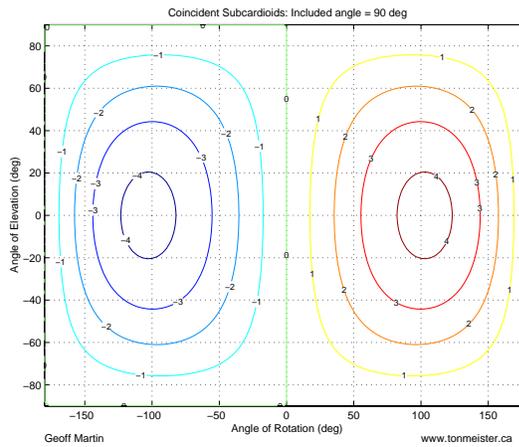


Figure 9.84: Interchannel amplitude difference response for a pair of coincident subcardioid microphones with an included angle of 90° .

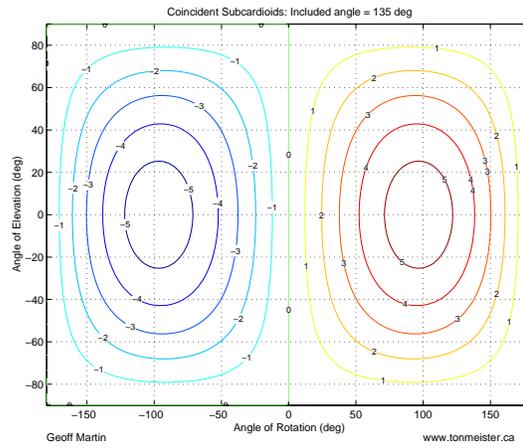


Figure 9.85: Interchannel amplitude difference response for a pair of coincident subcardioid microphones with an included angle of 135°.

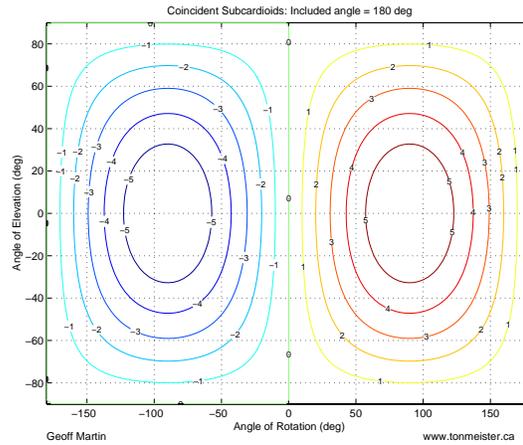


Figure 9.86: Interchannel amplitude difference response for a pair of coincident subcardioid microphones with an included angle of 180°.

microphones instead. One caveat here before we begin, because we're looking at decibel scales, and because calculators don't like being asked to find the logarithm of a negative number, we're looking at the *absolute value* of the sensitivity of the microphones again.

This time, the plots tell a very different story. Notice how, in the spherical plot in Figure 9.87 and in the contour plots in Figures 9.88 through 9.90, we get a bunch of vertical lines. The summary of what that means is that the interchannel amplitude difference of a pair of bidirectional microphones doesn't change with changes in the angle of elevation. So you can raise and lower the mic stand all you want without collapsing the image of the orchestra. As we get further off axis (because we've changed the vertical angle), the orchestra will get quieter, but it won't pull to the centre of the loudspeakers. This is a good thing.

There is a small drawback here, though. Remember that if you have a big vertical angle, then a small horizontal movement of the instrument corresponds to a large change in the angle of rotation, so you can get a violent swing in the image location if you're not careful. For example, if your bidirectional pair is really high and you're recording a singer that likes to sway back and forth, you might wind up with a phantom image that is always swinging back and forth between the two speakers, making your listeners seasick.

Also, note with a pair of bidirectional microphones with an included angle of 180 degrees that all angles of incidence produce the same sensitivity – just remember that the two signals are of opposite polarity. If you want to do this, use your “phase flip” button. It's cheaper than a second microphone.

Hypercardioids

Once again, hypercardioids exhibit properties that are recognizable as being something between a cardioid and a bidirectional. If we look at the spherical plot of a pair of coincident hypercardioids with an included angle of 90° shown in Figure 9.92, we can see that there is a dividing line along the side of the pair, similar to that found in a bidirectional pair. Just like the bidirectionals, this follows the null point in one of the two microphones, the dividing line between the front and rear lobes. However, like the cardioid pair, notice that vertical changes alter the interchannel amplitude difference. There is one big difference from the cardioids, however. In the case of cardioids, a vertical change always results in a reduction in the interchannel amplitude difference whereas, in the case of a hypercardioid pair, it is possible to have a vertical change that produces an increase in the interchannel amplitude difference. This is most easily visible in the contour plot in Figure 9.94. Notice that if you start with a horizontal angle of 100°

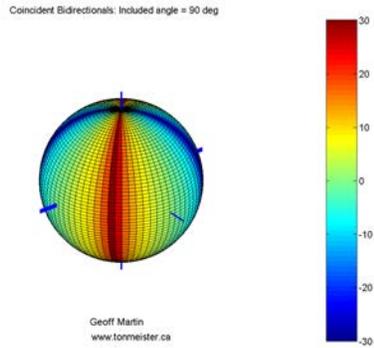


Figure 9.87: Interchannel amplitude difference response (in dB) for a pair of coincident bidirectional microphones with an included angle of 90° .

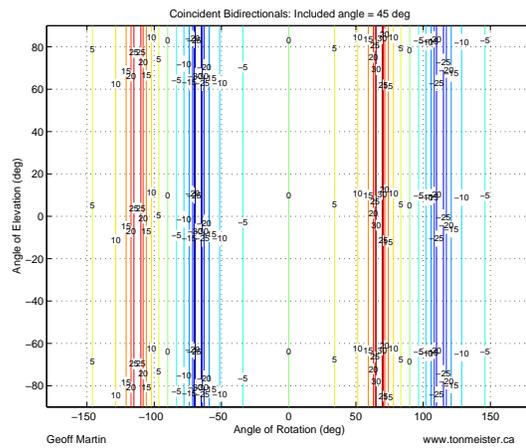


Figure 9.88: Interchannel amplitude difference response for a pair of coincident bidirectional microphones with an included angle of 45° .

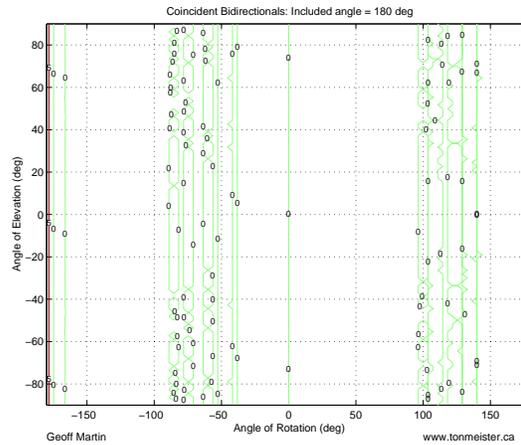


Figure 9.91: Interchannel amplitude difference response for a pair of coincident bidirectional microphones with an included angle of 180° .

degrees, then a vertical change off the equator will cause the interchannel amplitude difference to increase to ∞ dB before it reduces back down to 0 dB at the north or south pole.

There are three principal practical issues to consider here. Firstly, remember that a change in the height of your mic stand with a pair of hypercardioids will change the apparent width of your sound stage. Unlike the case of cardioids, however, the change might wind up increasing the width of some components while simultaneously decreasing the width of others. So you wind up squeezing together the centre part of the orchestra while you pull apart the sides.

The second issue to consider is similar to that with cardioids. Don't forget that you've got sound coming in from all angles at the same time - so it's possible that some parts of your room sound will be pulled wider while others are pushed closer together.

Thirdly, there's just the now-repetitious reminder that a lot of the signals coming into the pair are arriving at the rear lobes of the microphones, so you're going to have signals that are either in opposite polarity in the two channels, or similar, but inverted polarities.

Spaced omnidirectionals

Calculation of the interchannel time of arrival differences for a pair of spaced microphones in a three-dimensional world requires only a small change to Equation 9.3 as can be seen in Equation 9.10.

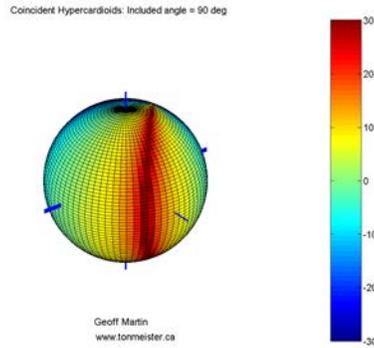


Figure 9.92: Interchannel amplitude difference response (in dB) for a pair of coincident hypercardioid microphones with an included angle of 90° .

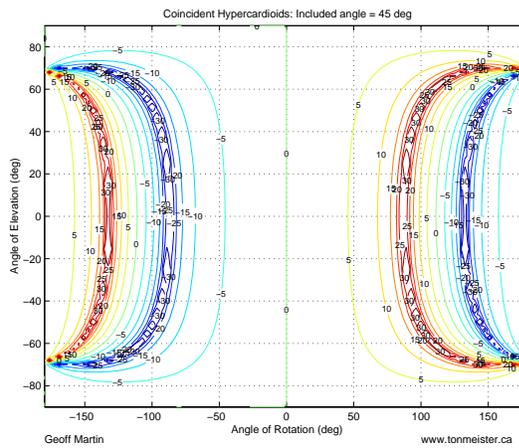


Figure 9.93: Interchannel amplitude difference response for a pair of coincident hypercardioid microphones with an included angle of 45° .

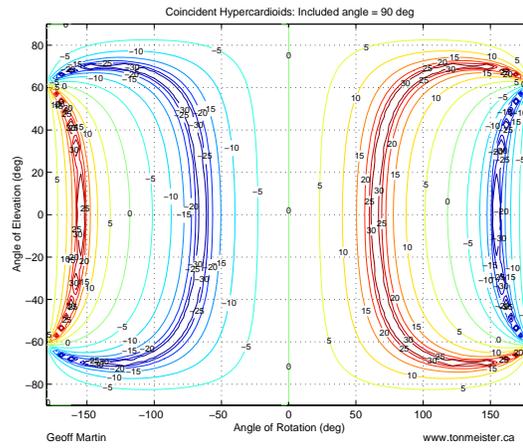


Figure 9.94: Interchannel amplitude difference response for a pair of coincident hypercardioid microphones with an included angle of 90° .

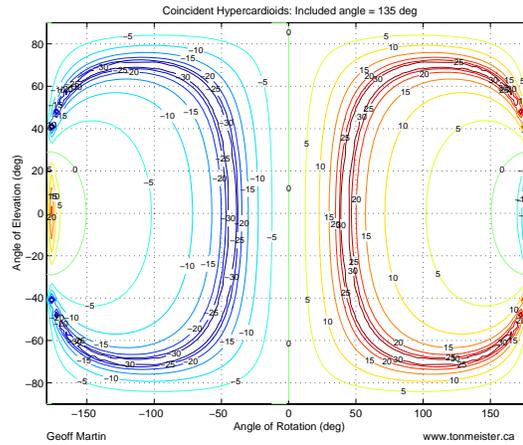


Figure 9.95: Interchannel amplitude difference response for a pair of coincident hypercardioid microphones with an included angle of 135° .

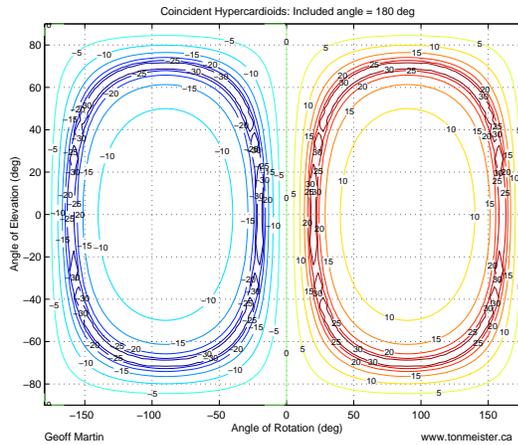


Figure 9.96: Interchannel amplitude difference response for a pair of coincident hypercardioid microphones with an included angle of 180° .

$$D = d \sin \vartheta \cos \phi \quad (9.10)$$

Consider that a change in elevation in the geographic coordinate system means that we are heading away from the “equator” towards the “north pole”, relative to the microphones. When the sound source is located at any angle of horizontal rotation and an angle of elevation $\phi = 90^\circ$, it is equidistant from the two microphones, therefore the time of arrival difference is 0 ms. Consequently, we can intuitively see that the greatest time of arrival difference is for sources where $\phi = 0^\circ$, and that any change in elevation away from this plane will result in a reduced value of $\Delta Time$.

This behaviour can be seen in Figures 9.97 through 9.99 as well as Figure 9.100.

9.4.3 Summed power response

The previous chapter deals only with the interchannel differences between the two microphones in a pair. This information gives us an idea of the general placement of phantom images between pairs of loudspeakers in a playback system, but there are a number of other issues to consider. Back in the discussion on panning in Chapter 9.3.1, something was mentioned that has not yet been discussed. As was pointed out, pan knobs on consoles work by changing the amplitude difference between the output channels, but the issue of why they are typically *constant power* panners was not mentioned.

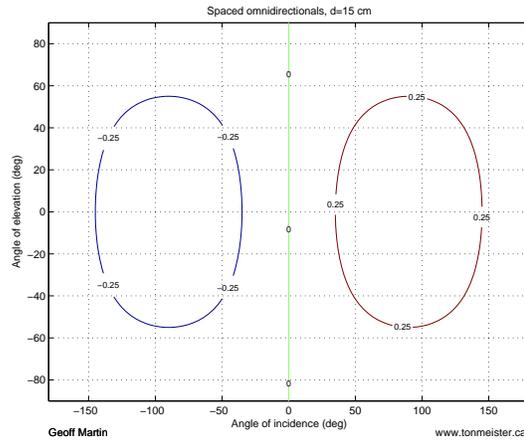


Figure 9.97: Interchannel time differences in ms for a pair of spaced microphones with a separation of 15 cm.

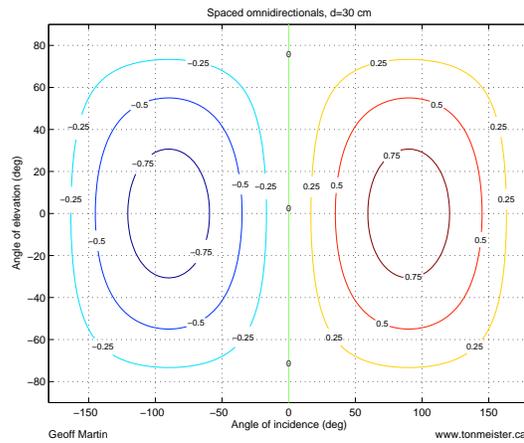


Figure 9.98: Interchannel time differences in ms for a pair of spaced microphones with a separation of 30 cm.

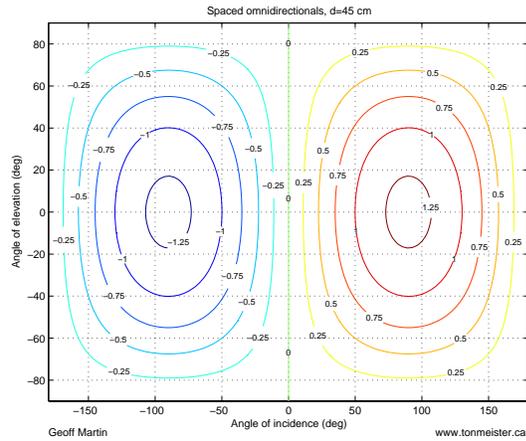


Figure 9.99: Interchannel time differences in ms for a pair of spaced microphones with a separation of 45 cm.

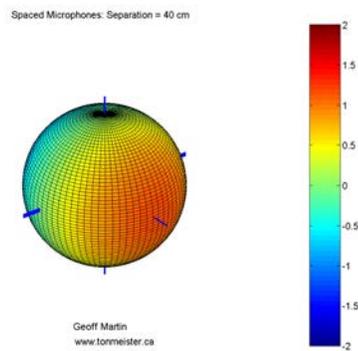


Figure 9.100: Interchannel time differences in ms for a pair of spaced microphones with a separation of 40 cm.

Since we've been thinking of microphone pairs as panning algorithms, we'll continue to do so, but this time looking at the summed power output of the pair. Since the power of a signal is proportional to the square of its amplitude, this can be done using Equation 9.11.

$$O = S_1^2 + S_2^2 \quad (9.11)$$

where O is the total output power of the two microphones.

In order to calculate this power response curve on a decibel scale, the following equation is used:

$$O = 10 * \log_{10} (S_1^2 + S_2^2) \quad (9.12)$$

The question, of course, is "What will this tell us?" I'll answer that question using the example of a pair of coincident cardioid microphones.

Horizontal plane

Cardioids

Figure 9.101 shows the summed power response of a pair of coincident cardioid microphones with an included angle of 90° . as you can see, the total power for sources with an angle of incidence of 0° is about 2 dB. As you rotate away from the front of the pair, the summed power drops to a minimum of about -12 dB directly behind. Remember from the previous chapter that the 0° and 180° locations in the horizontal plane are the two positions where the interchannel amplitude difference is 0 dB, therefore instruments in these two locations will result in phantom images between the two loudspeakers, however, we can now see that, although this is true, the two images will differ in power by approximately 15 dB, with sources in the front of the microphone pair being louder than those behind.

The range of the summed power for a pair of cardioids changes with the included angle as is shown in Figures 9.101 through 9.104. In fact, the smaller the included angle, the greater the range. As can be seen in Figure 9.102, the range of summed power is approximately 28 dB compared to only 3 dB for an included angle of 180° . Also note that for larger included angles, there are two symmetrical peaks in the power response rather than one at an angle of incidence of 0° .

Each of these analyses, in conjunction with their corresponding inter-channel amplitude difference plot for the same included angle, gives us an indication of the general distribution of energy across the reproduced sound stage. For example, if we look at a pair of coincident cardioids with an

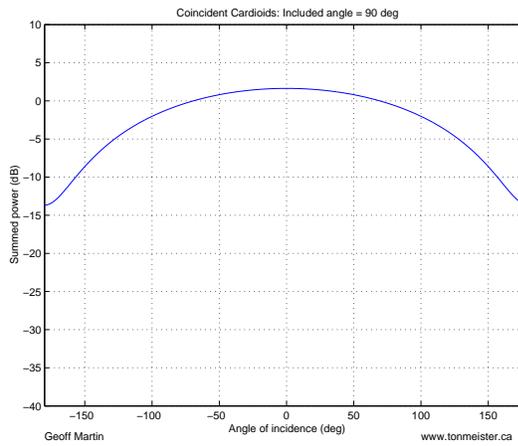


Figure 9.101: Summed power response for a pair of coincident cardioid microphones in the horizontal plane with an included angle of 90° .

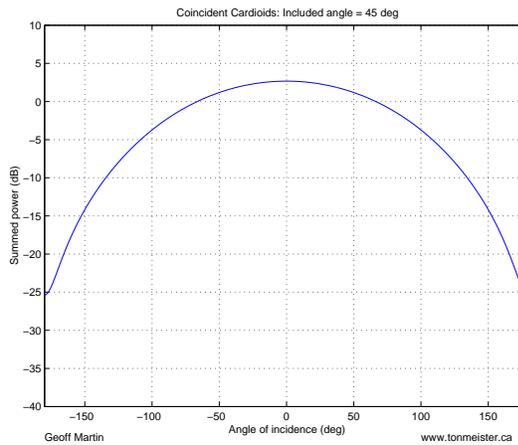


Figure 9.102: Summed power response for a pair of coincident cardioid microphones in the horizontal plane with an included angle of 45° .

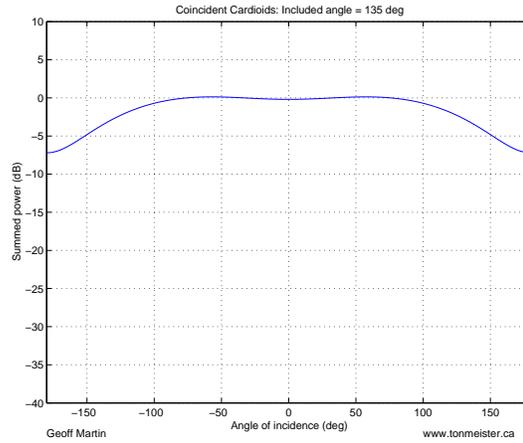


Figure 9.103: Summed power response for a pair of coincident cardioid microphones in the horizontal plane with an included angle of 135°.

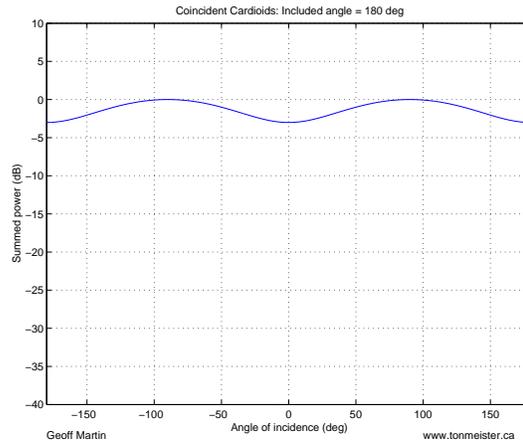


Figure 9.104: Summed power response for a pair of coincident cardioid microphones in the horizontal plane with an included angle of 180°.

included angle of 45° , we can see that instruments, reflections and reverberation with an angle of incidence of 0° are much louder than those away from the front of the pair. In addition, we can see from the $\Delta Amp.$ plot that a large portion of sources around the pair will image near the centre position between the loudspeakers. Consequently, the resulting sound stage appears to “clump” in the middle rather than being spread evenly across the playback room.

In addition, for smaller included angles, it can be seen that much more emphasis is placed on sound sources and reflections in the front of the pair with sources to the rear attenuated.

Subcardioids

Figures 9.105 through 9.108 show the summed power plots for the horizontal plane of a pair of subcardioid microphones. As is evident, there is a much smaller range of values than is seen in the cardioid microphones, however the general shape of the curves are similar. As can be seen in these plots, there is a more evenly distributed sensitivity to sound sources and reflections around the microphone pair, however, due to the limited range of values for $\Delta Amp.$, these sources typically image between the loudspeakers as well.

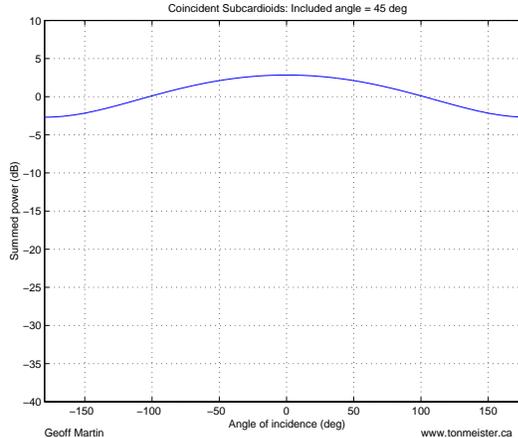


Figure 9.105: Summed power response for a pair of coincident subcardioid microphones in the horizontal plane with an included angle of 45° .

Bidirectionals

Due to the symmetrical double lobes of bidirectional microphones, they exhibit a considerably different power response as can be seen in Figures 9.109 through 9.112. When the included angle of the microphones is 90° ,

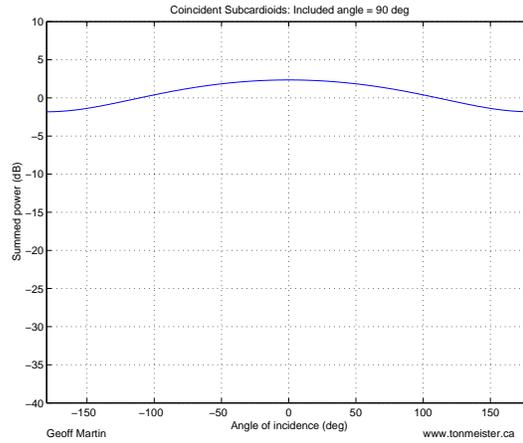


Figure 9.106: Summed power response for a pair of coincident subcardioid microphones in the horizontal plane with an included angle of 90°.

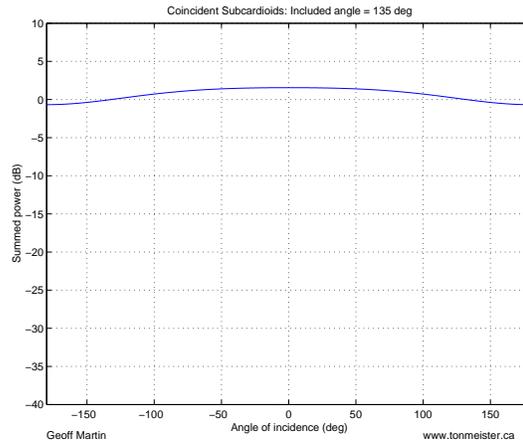


Figure 9.107: Summed power response for a pair of coincident subcardioid microphones in the horizontal plane with an included angle of 135°.

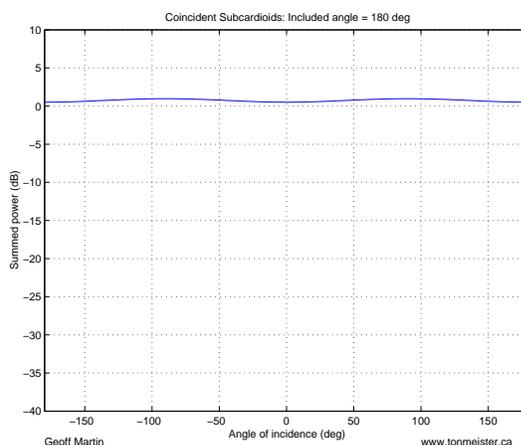


Figure 9.108: Summed power response for a pair of coincident subcardioid microphones in the horizontal plane with an included angle of 180° .

there is a constant summed power throughout all angles of incidence. Consequently, all sources around the pair appear to have the same level at the listening position, regardless of angle of incidence. If the included angle of the pair is reduced as is shown in Figure 9.109, then we reduce the apparent level of sources to the side of the microphone pair. When the included angle is greater than 90° , the dips in the power response happen directly in front of and behind the microphone pair.

Notice as well that bidirectional pairs differ from cardioids in that the high and low points in the summed power response are always in the same locations – 0° , 90° , 180° and 270° . They do not move with changes in included angle, they simply change power level.

Note now that we are not talking about the absolute value of the sensitivity of the microphones. This is because the calculation of power automatically squares the sensitivity, thus making all values positive and therefore making the logarithm happy...

Hypercardioids

Once again, hypercardioids behave like a hybrid between cardioids and bidirectionals as can be seen in Figures 9.113 through 9.116.

Spaced omnidirectionals

It is typically assumed that the outputs levels of omnidirectional microphones are identical, differing only in time of arrival. This assumption is incorrect for sources whose distance to the pair is similar to the distance between the mic's, or when you're using omnidirectionals that aren't really

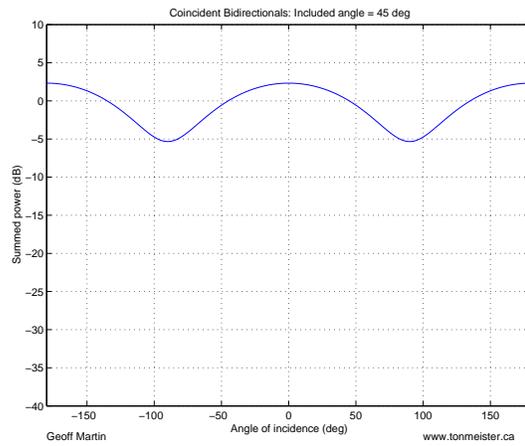


Figure 9.109: Summed power response for a pair of coincident bidirectional microphones in the horizontal plane with an included angle of 45° .

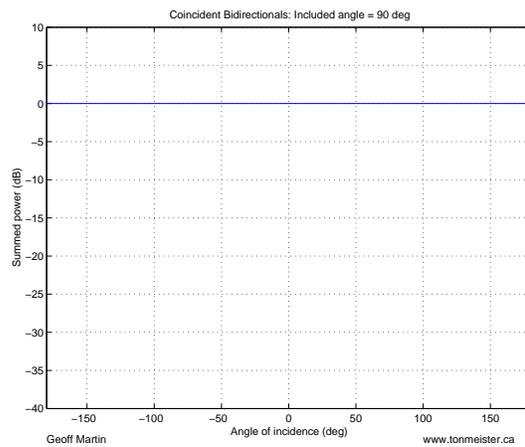


Figure 9.110: Summed power response for a pair of coincident bidirectional microphones in the horizontal plane with an included angle of 90° .

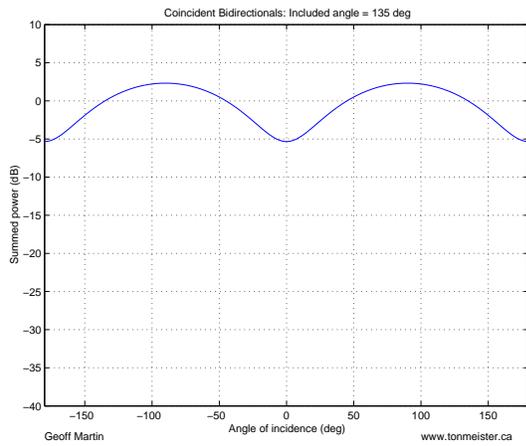


Figure 9.111: Summed power response for a pair of coincident bidirectional microphones in the horizontal plane with an included angle of 135° .

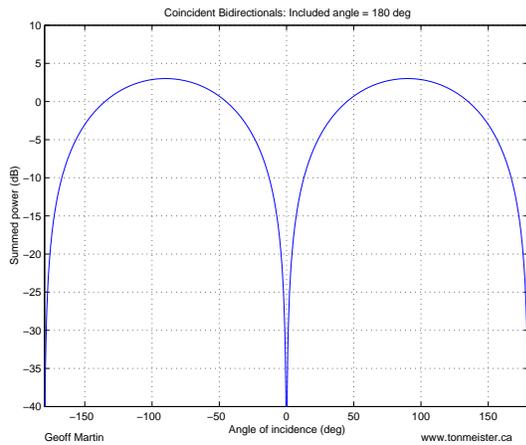


Figure 9.112: Summed power response for a pair of coincident bidirectional microphones in the horizontal plane with an included angle of 180° .

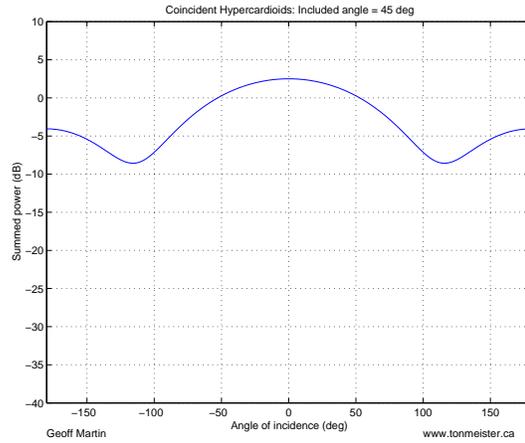


Figure 9.113: Summed power response for a pair of coincident hypercardioid microphones in the horizontal plane with an included angle of 45°.

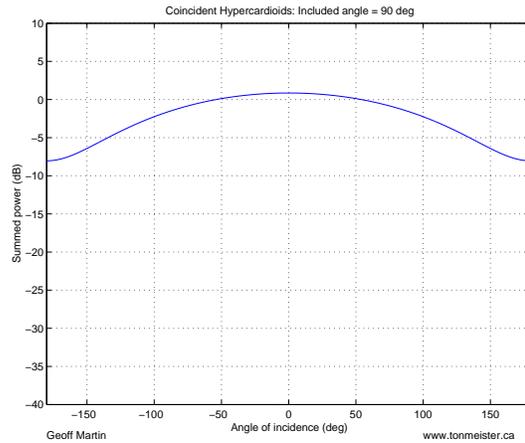


Figure 9.114: Summed power response for a pair of coincident hypercardioid microphones in the horizontal plane with an included angle of 90°.

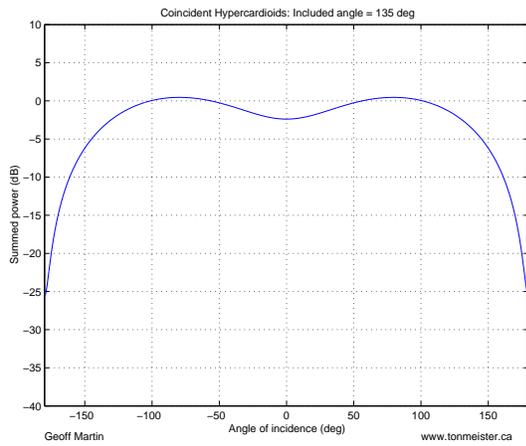


Figure 9.115: Summed power response for a pair of coincident hypercardioid microphones in the horizontal plane with an included angle of 135° .

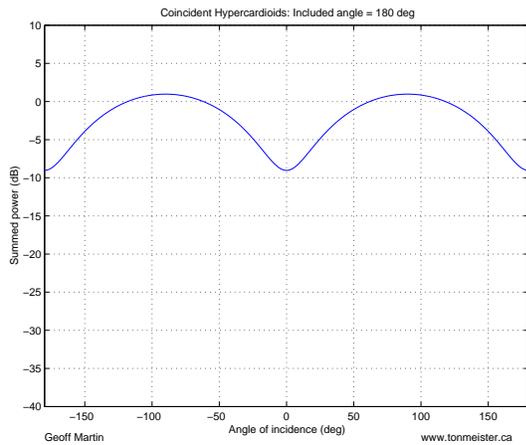


Figure 9.116: Summed power response for a pair of coincident hypercardioid microphones in the horizontal plane with an included angle of 180° .

omnidirectional. Both cases happen frequently, but we'll stick with the assumption for this paper. As a result, we'll assume that the summed power output for a pair of omnidirectional microphones is +3 dB relative to either of the microphones for all angles of rotation and elevation.

Three-dimensional analysis

As before, we can only get a complete picture of the response of the microphones by looking at a three-dimensional response plot.

Cardioids

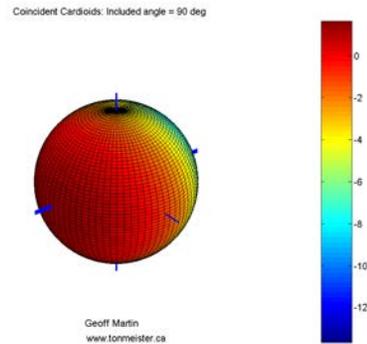


Figure 9.117: Summed power response for a pair of coincident cardioid microphones with an included angle of 90° .

The three dimensional plots for cardioids hold no surprises. As we can see, for smaller included angles as is shown in Figure 9.118, the range of values for the summed power is quite wide, with the high and low points being at the front and rear of the pair respectively, on the “equator.”

If the included angle is increased beyond 90° , then two high points in the power response appear on either side of the front of the pair. Meanwhile, as was seen in the two-dimensional analyses, the overall range is reduced with a smaller attenuation behind the pair.

Subcardioids

Again, subcardioid microphones behave similar to cardioids with a smaller range in their summed power response.

Bidirectionals

Bidirectional microphone pairs have a rather interesting property at 90° . Note in Figure 9.128 that the contour plot for the summed power for the

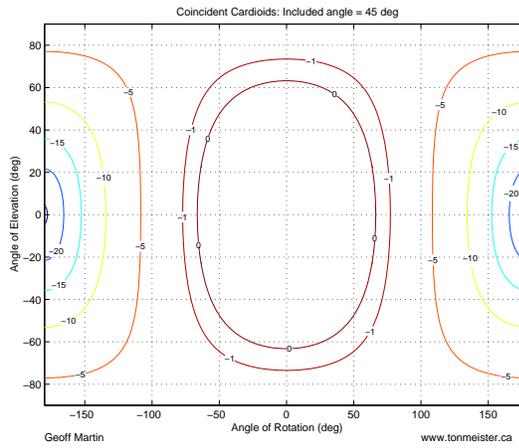


Figure 9.118: Summed power response for a pair of coincident cardioid microphones with an included angle of 45° .

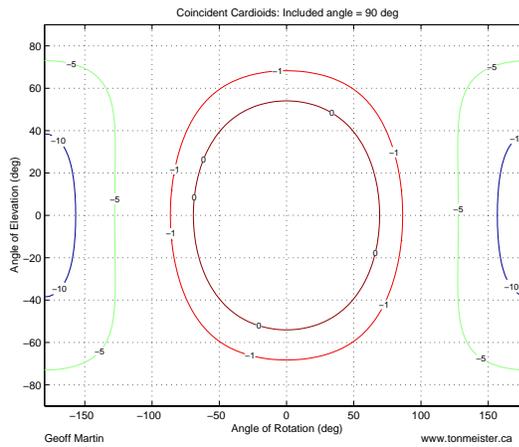


Figure 9.119: Summed power response for a pair of coincident cardioid microphones with an included angle of 90° .

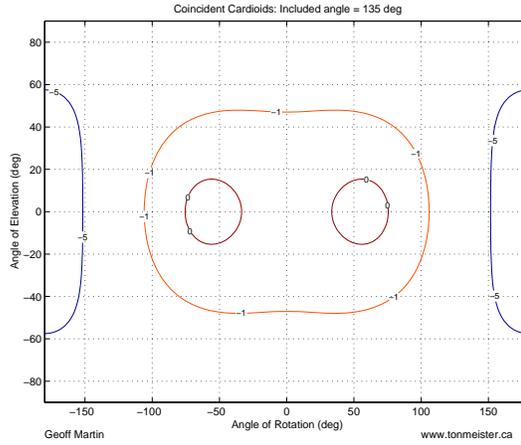


Figure 9.120: Summed power response for a pair of coincident cardioid microphones with an included angle of 135°.

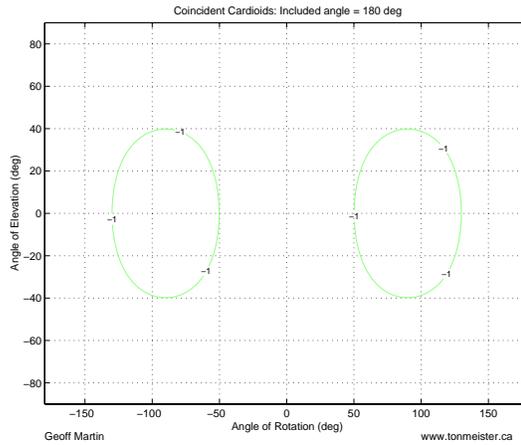


Figure 9.121: Summed power response for a pair of coincident cardioid microphones with an included angle of 180°.

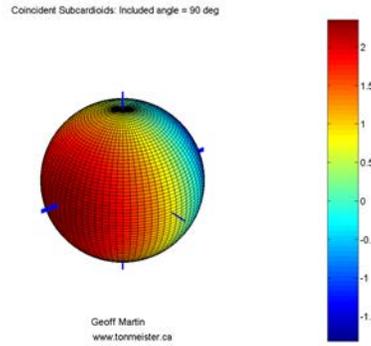


Figure 9.122: Summed power response for a pair of coincident subcardioid microphones with an included angle of 90° .

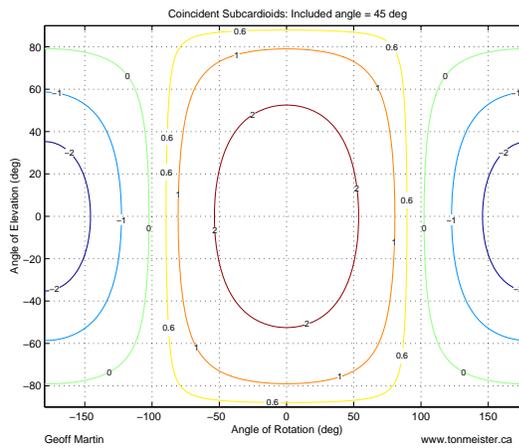


Figure 9.123: Summed power response for a pair of coincident subcardioid microphones with an included angle of 45° .

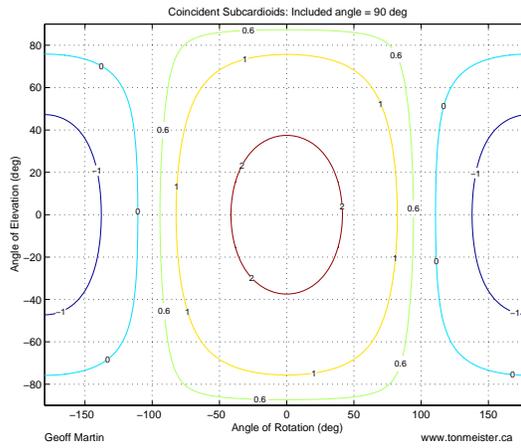


Figure 9.124: Summed power response for a pair of coincident subcardioid microphones with an included angle of 90° .

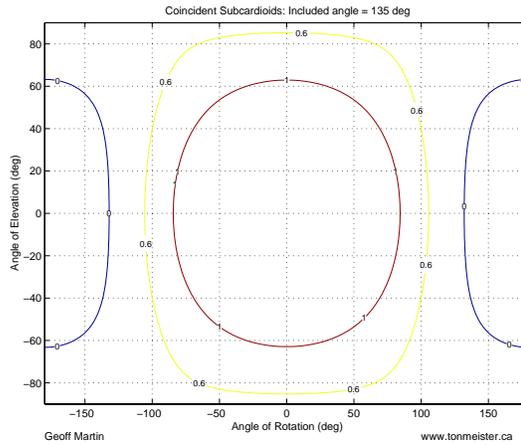


Figure 9.125: Summed power response for a pair of coincident subcardioid microphones with an included angle of 135° .

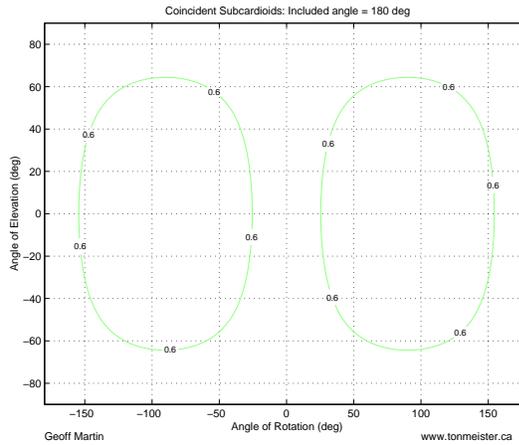


Figure 9.126: Summed power response for a pair of coincident subcardioid microphones with an included angle of 180° .

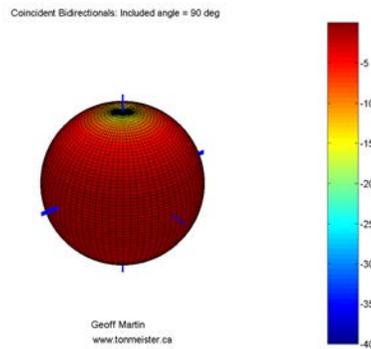


Figure 9.127: Summed power response for a pair of coincident bidirectional microphones with an included angle of 90° .

microphone pair shows a number of horizontal lines. This means that if you have a sound source at a given angle of elevation, changes in the angle of rotation will have no effect on the apparent level of the sound source. This, in turn, means that all sources at a given angle of elevation have the same apparent total gain.

Additionally, notice that the point where the total power is the greatest is the horizontal plane, with a value of 0 dB with decreasing level as we move away from the equator.

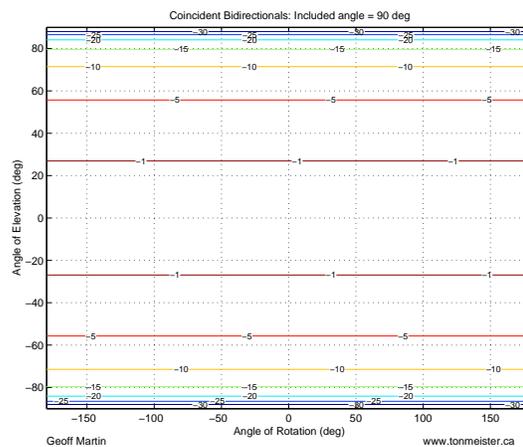


Figure 9.128: Summed power response for a pair of coincident bidirectional microphones with an included angle of 90° .

As was seen in the two-dimensional plots, included angles of less than 90° cause the high points in the power plots to occur at the front and rear of the pairs as is shown in Figure 9.129. Notice that, unlike the cardioid and subcardioid pairs, the minimum points (with a value of $-\infty$ dB) are directly above and below the microphone pair. This is true for bidirectional pairs with any included angle.

As can be seen in Figures 9.130 and 9.131, when the included angle is greater than 90° , the behaviour of the pair is exactly the same as the symmetrical included angle less than 90° , with a 90° rotation in the behaviour. For example, pairs with included angles of 70° and 110° (symmetrical around 90°) have identical responses, but where the 70° pair is “pointing” forward, the 110° pair is “pointing” sideways.

Hypercardioids

Finally, as we have seen before, the hypercardioid pairs exhibit a response pattern that is a combination of the cardioid and bidirectional patterns. One

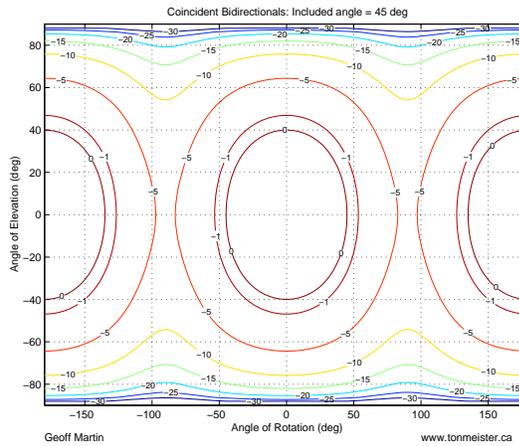


Figure 9.129: Summed power response for a pair of coincident bidirectional microphones with an included angle of 45° .

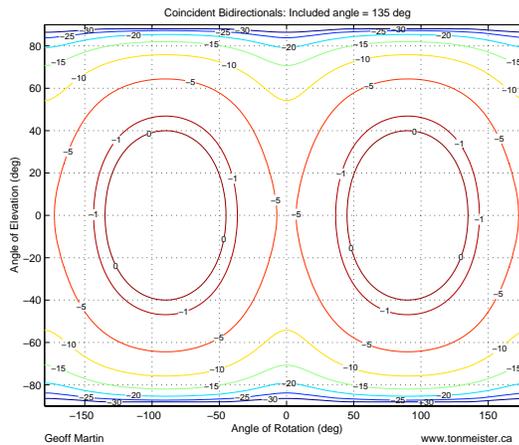


Figure 9.130: Summed power response for a pair of coincident bidirectional microphones with an included angle of 135° .

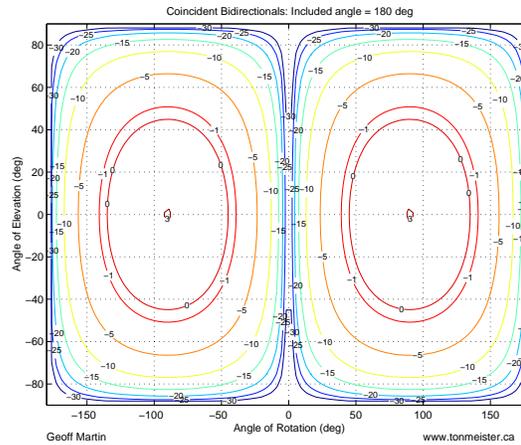


Figure 9.131: Summed power response for a pair of coincident bidirectional microphones with an included angle of 180° .

important thing to note here is that, although the location with the highest summed power value is on the horizontal plane as in the case of the cardioid microphones, the point of minimum power is between the equator and the poles in all cases but an included angle of 180° .

9.4.4 Correlation

Finally, we'll take a rather holistic view of the pair of microphones and take a look at the correlation coefficient of their outputs. This can give us a general idea of the similarity of the two signals which could be interpreted as a sensation of spaciousness. We have to be very careful here in making this jump between correlation and spaciousness as will be discussed below, but first, we'll look at exactly what is meant by the term "correlation."

Correlation Coefficient

The term "correlation" is one that is frequently misused and, as a result, misunderstood in the field of audio engineering. Consequently, some discussion is required to define the term. Generally speaking, the *correlation* of two audio signals is a measure of the relationship of these signals in the time domain [Fahy, 1989]. Specifically, given two two-dimensional variables (in the case of audio, the two dimensions are amplitude and time), their *correlation coefficient*, r is calculated using their covariance $\overline{s_{xy}}$ and their standard deviations

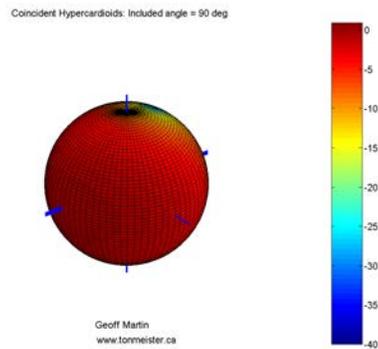


Figure 9.132: Summed power response for a pair of coincident hypercardioid microphones with an included angle of 90° .

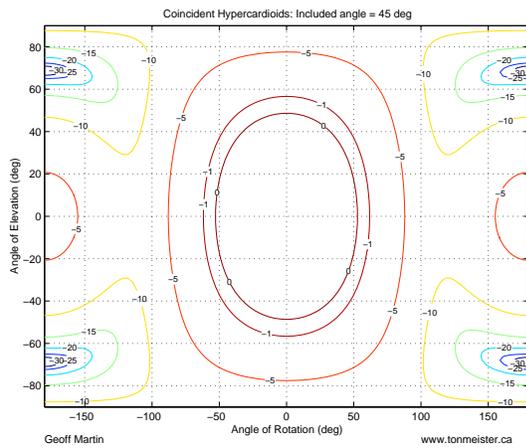


Figure 9.133: Summed power response for a pair of coincident hypercardioid microphones with an included angle of 45° .

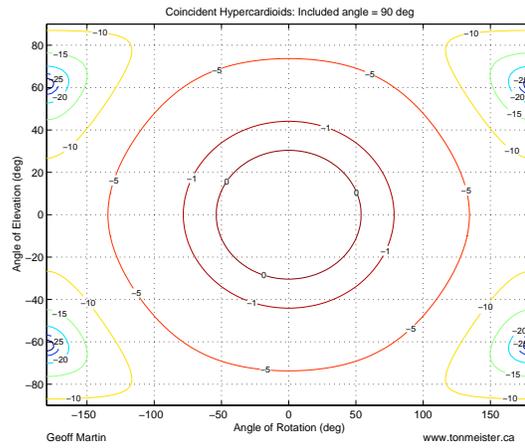


Figure 9.134: Summed power response for a pair of coincident hypercardioid microphones with an included angle of 90° .

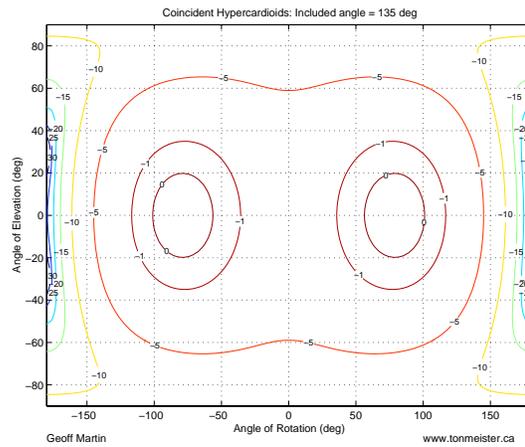


Figure 9.135: Summed power response for a pair of coincident hypercardioid microphones with an included angle of 135° .

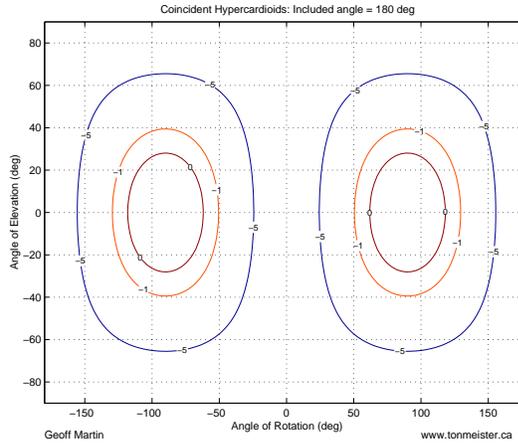


Figure 9.136: Summed power response for a pair of coincident hypercardioid microphones with an included angle of 180°.

$\overline{s_x}$ and $\overline{s_y}$ as is shown in Equation 9.13 [Wasserman and Whitmore, 1992]. The line over these three components indicates a time average, as is discussed below.

$$r = \frac{\overline{s_{xy}}}{\overline{s_x} * \overline{s_y}} \quad (9.13)$$

The *standard deviation* of a series of values is an indication of the average amount the individual values are different from the total average for all values. Specifically, it is the square root of the average of the squares of the differences between the average of all values and each individual value. For example, in order to find the standard deviation of a PCM digital audio signal, we begin by finding the average of all sample values. This will likely be 0 since audio signals typically do not contain a DC offset. Each sample is then individually subtracted from this average and each result is squared. The average of these squares is calculated and its square root is the standard deviation. When there is no DC component in an audio signal, its standard deviation is equal to its RMS value. In such a situation, it can be considered the square root of the average power of the signal.

The *covariance* of two series of values is an indication of whether they are interrelated. For example, if the average temperature for today's date is 19° C and the average humidity is 50%, yet today's actual temperature and humidity are 22° C and 65%, we can find whether there is an interdependent relationship between these two values, called the *covariation*

[Wasserman and Whitmore, 1992]. This is accomplished by multiplying the differences of today's values from their respective averages, therefore $(19 - 22) * (50 - 65) = 45$. The result of this particular calculation, being a positive number, indicates that there is a positive relationship between the temperature and humidity today – in other words, if one goes up, the other does also. Had the covariation been negative, then the relationship would indicate that the two variables had behaved oppositely. If the result is 0, then at least one of the variables equalled the average value. The covariance is the average of the covariations of two variables measured over a period of time. The difficulty with this measurement is that its scale changes according to the scale of the two variables being measured. Consequently, covariance values for different statistical samples cannot be compared. For example, we cannot tell whether the covariance of air temperature and humidity is greater or less than the covariance of the left and right channels in a stereo audio recording if both have the same polarity.

Fortunately, if the standard deviations of the two signals are multiplied, the scale is identical to that of the covariance. Therefore, the *correlation coefficient* (the covariance divided by the product of the two standard deviations) can be considered to be a normalised covariance. The result is a value that can range from -1 to 1 where 1 indicates that the two signals have a positive linear relationship (in other words, their slopes always have the same polarity). A correlation coefficient of -1 indicates that the two signals are negatively linearly related (therefore, their slopes always have opposite polarities). In the case of wide-band signals, a correlation of 0 usually indicates that the two signals are either completely unrelated or separated in time by a delay greater than the averaging time.

In the particular case of two sinusoidal waveforms with identical frequency and a constant phase difference $\omega\tau$, Equation 9.13 can be simplified to Equation 9.14 [Morfey, 2001].

$$r = \cos(\omega\tau) \quad (9.14)$$

where the *radian frequency* ω is defined in Equation 9.15 [Strawn, 1985] and where τ is the time separation of the two sinusoids.

$$\omega \triangleq 2\pi f \quad (9.15)$$

where the symbol \triangleq denotes “is defined as” and f is the frequency in Hz.

Further investigation of the topic of correlation highlights a number of interesting points. Firstly, two signals of identical frequency and phase have

a correlation of 1. It is important to remember that this is true regardless of the amplitudes of the two signals [Welle,]. Two signals of identical frequency and with a phase difference of 180° have a correlation of -1. Signals of identical frequency and with a phase difference of $\pm 90^\circ$ have a correlation of 0. Finally, it is important to remember that the correlation between two signals is highly dependent on the amount of time used in the averaging process.

So what?

In the field of perception of concert hall acoustics, it has long been known that there is a link between *Interaural Cross Correlation (IACC)* and a perceived sensation of diffuseness and auditory source width (ASW)[Ando, 1998]. (IACC is a measure of the cross correlation between the signals at your two ears.) The closer the IACC approaches 1, the lower the subjective impression of diffuseness and ASW. The lower the IACC, the more diffuse and wider the perceived sound field.

One of the nice things about recording is that you can control the IACC at the listening position by controlling the interchannel correlation coefficient. Although the interchannel correlation coefficient doesn't directly correspond to the IACC, they are related. In order to figure out the exact relationship between these two measurements, you'll also need to know a little bit about the room that the speakers are in.

There are a couple of things to remember about interchannel correlation that we have to remember before we start talking about microphone response characteristics.

1. When the correlation coefficient is 1, this is an indication that the information in the two channels is identical in every respect with the possible exception of level. This also means that, if the two channels are added, none of the information will be reduced due to destructive interference.
2. When the correlation coefficient is 0, this is an indication of one of three situations. Either (1) you have signal in one channel and none in the other, (2) you have two *completely* different signals, or (3) you have two sinusoids that are separated by 90° .
3. When the correlation coefficient is -1, you have two signals that are identical in every respect except polarity and possibly level.

4. When the correlation coefficient is positive or negative, but not 1 or -1, this is an indication that the two signals are partially alike, but differ in one or more of a number of ways. This will be discussed as is required below.

Free field

A free-field situation is one where the waveform is free to expand outwards forever. This typically only happens in thought experiments, however we typically assume that the situation exists in anechoic chambers and on the top of poles outdoors. To extend our assumptions even further, we can simplify the direct sound from a sound source received at a microphone to be considered as a free field source. Consequently, the analysis of a microphone pair in a free field becomes applicable to real life.

Coincident pairs

If we convert Equation 9.13 to something that represents the signal at the two microphones, then we wind up with Equation 9.16 below.

$$r_{\{\alpha,\phi\}} = \frac{S_1 S_2}{\sqrt{S_1^2} \sqrt{S_2^2}} \quad (9.16)$$

where r is the correlation coefficient of the outputs of the two microphones with sensitivities S_1 and S_2 for the angles of rotation α and elevation ϕ .

Luckily, this can be simplified to Equation 9.17.

$$r_{\{\alpha\phi\}} = \text{sign}(S_1 S_2) \quad (9.17)$$

where the function $\text{sign}(x)$ indicates the polarity of x . $\text{sign}(x) = 1$ for all $x > 0$, $\text{sign}(x) = -1$ for all $x < 0$ and $\text{sign}(x) = 0$ for all $x = 0$

This means that, for coincident omnidirectional, subcardioid and cardioid microphones, all free field sources have a correlation of 1 all the time. This is because the only theoretical difference between the outputs of the two microphones is in their level. The only exception here is the particular case of a sound source located exactly at the null point for one of the cardioids in a pair. In this location, the correlation coefficient of the two outputs will be 0 because one of the channels contains no signal.

In the case of hypercardioids and bidirectionals, the value of the correlation coefficient for various free field sources will be either 1, 0 or -1. In locations where the polarities of the two signals are the same, either both positive or both negative, then the correlation coefficient will be 1. Sources

located at the null point of at least one of the microphones will produce a correlation coefficient of 0. Sources located in positions where the receiving lobes have opposite polarities (for example, to the side of a Blumlein pair of bidirectionals), the correlation coefficient will be -1.

Spaced omnidirectionals

As was discussed above, spaced omnidirectional microphones are used under the (usually incorrect) assumption that the only difference between the two microphones in a free field situation will be their time of arrival. As was shown in Figure 9.69, this time separation is dependent on the spatial separation of the two microphones and the angles of rotation and elevation of the source to the pair.

The result of this time of arrival difference caused by the extra propagation distance is a frequency-dependent phase difference between the two channels. This interchannel phase difference $\omega\tau$ can be calculated for a given additional propagation distance using Equation 9.18.

$$\omega\tau = \frac{2\pi D}{\lambda} \quad (9.18)$$

where λ is the acoustic wavelength in air.

This can be further adapted to the issue of microphone separation and angle of incidence by combining Equations 9.3 and 9.18, resulting in Equation 9.19.

$$\omega\tau_{\{\vartheta,\phi\}} = kd \sin \vartheta \cos \phi \quad (9.19)$$

where $\omega\tau_{\{\vartheta,\phi\}}$ is the frequency-dependent phase difference between the channels for a sound source at angle of rotation ϑ , angle of elevation ϕ , and k is the acoustic wavenumber, defined in Equation 9.20 [Morfe, 2001].

$$k \triangleq \frac{\omega}{c} \quad (9.20)$$

where c is the speed of sound in air.

Consequently, we can calculate the correlation coefficient for the outputs of two omnidirectional microphones with a free field source using the model of a single signal correlated with a delayed version of itself. This is calculated using Equation 9.21.

$$r = \cos(\omega\tau) \quad (9.21)$$

All we need to do is to insert the phase difference of the two microphones, calculated using Equation 9.19 into this equation and we'll see that the correlation coefficient is a complete mess.

So now the question is “do we care?” The answer is “probably not.” Why not? Well, the correlation coefficient in this case is dependent on the phase relationship between the two signals. Typically, if you present a signal to a pair of loudspeakers where the only difference is their time of arrival at the listening position, then you probably aren’t paying attention to their phase difference – it’s more likely that you’re more interested in their time difference. This is because the interaural time of arrival difference is such a strong cue for localization of sources in the real world. Even for something we’d consider to be a “steady state” source like a melodic instrument playing slow music, the brain is grabbing each little cue it can to determine the time relationship. The only time the interchannel phase information (and therefore the free field correlation coefficient) is going to be the predominant cue is if you listen to sinusoidal waves – and nobody wants to do that...

Diffuse field

Now we have to talk about what a diffuse field is. If we get into the official definition of a diffuse field, then we have to have a talk about things like infinity, plane waves, phase relationships and probability distribution... maybe some other time... Instead, let’s think about a diffuse field in a couple of different, equally acceptable ways. One way is to think that you have sound coming from everywhere simultaneously. Another way is that you have sound coming from different directions in succession with no time inbetween their arrival.

If we think of reverberation as a very, very big number of reflections coming from all directions in fast succession, then we can start to think of what a diffuse field is like. Typically, we like to think of reverberation as a diffuse field – this is particularly true for the people that make digital reverb units because it’s much easier to create random messes that sort of sound like reverb than it is to calculate everything that happens to sound as it bounces around a room for a couple of seconds.

We need to pay a lot of attention to the correlation coefficient of the diffuse component of the recorded signal. This can be used as a rough guide to the overall sense of “spaciousness” (or whatever word you wish to use – this area creates a lot of discussion) in your recording. If you have a correlation coefficient of 1, this will probably mean that you have a reverberant sound that is completely clumped into one location between the two loudspeakers. The only possible exception to this is if your signals are going to the adjacent pair of front and surround loudspeakers (i.e. Left and Left Surround) where you’ll find it very difficult to obtain a single phantom

location.

If your correlation coefficient is -1, then you have what most people call two “out of phase” signals, but what they really are is identical signals with opposite polarity.

If your correlation coefficient is 0, then there could be a number of different explanations behind the result. For example, a pair of coincident bidirectionals with an included angle of 90° will have a correlation coefficient of 0. If we broke the signals hitting the two diaphragms into individual sounds from an infinite number of sources, then each one would have a correlation coefficient of either 1 or -1, but since there are as many 1's as -1's, the whole diffuse field averages out to a total correlation of 0. Although the two signals appear to be completely uncorrelated according to the math, there will be an even distribution of sound between the speakers (because there are some components in there that have a correlation of 1, remember...)

On the other hand, if we take two omnidirectional microphones and put them very, very far apart – let's put them in completely different rooms to start, then the two signals are completely unrelated, therefore the correlation coefficient will be 0 and you'll get an image with no phantom sources at all – just two loudspeakers producing a pocket of sound. The same is true if you place the omni's very far apart in the same concert hall (you'll sometimes see engineers doing this for their ambience microphones). The resulting correlation coefficient, as we'll see below, will also be 0 because the sound fields at the two locations will sound similar, but they'll be completely unrelated. The result is a hall with a very large hole in the middle – because there are no correlated components in the two signals, there cannot be an even spread of energy between the loudspeakers.

The moral of the story here is that, in order to keep a “spacious” sound for your reverb, you have to keep your correlation coefficient close or equal to 0, but you can't just rely on that one number to tell you everything. Spacious isn't necessarily pretty, or believable...

Coincident pairs

Calculating the correlation of the outputs of a pair of coincident microphones is somewhat less than simple. In fact, at the moment, I have to confess that I really don't know the correct equation for doing this. I've searched for this piece of information in all of my books, and I've asked everyone that I think would know the answer, and I haven't found it yet. So, I wrote some MATLAB code to model the situation instead of doing the math the right way. In other words, I did a numerical calculation to produce the plots in Figures 9.137 and 9.138, but this should give us the right answer.

Some of the characteristics seen in Figure 9.137 should be intuitive. For example, if you have a pair of coincident omnidirectional microphones in a diffuse field, then the correlation coefficient of their outputs will be 1 regardless of their included angle. This is because the outputs of the two mic's will be identical no matter what the angle.

Also, if you have any matched pair of microphones with an included angle of 0° , then their outputs will also be identical and the correlation coefficient will be 1.

Finally, if you have a pair of matched bidirectional microphones with an included angle of 180° , then their outputs will be identical but opposite in polarity, therefore their correlation coefficient will be -1.

Everything else on that plot will be less obvious.

Just in case you're wondering, here's how I calculated the two graphs in Figures 9.137 and 9.138.

If you have a pair of coincident bidirectional microphones with an included angle of 90° in a diffuse field, then the correlation coefficient of their outputs will be 0. This is somewhat intuitive if we think that half of the field entering the microphone pair will have a correlation of 1 (all of the sound coming from the front and rear of the pair where the lobes have matched polarities) while the other half of the sound field results in a correlation of -1 (because it's entering the sides of the pair where you have opposite polarity lobes.) Since the area producing the correlation of 1 is identical to the area producing the correlation of -1, then the two cancel each other out and produce a correlation of 0 for the whole.

Similarly, if we have a coincident bidirectional and an omnidirectional in a diffuse field, then the correlation coefficient of their outputs will also be 0 for the same reason.

As we'll see in Section 9.5, if you have a coincident trio of microphones consisting of two bidirectionals at 90° and an omni, then you can create a microphone pointing in any direction in the horizontal plane with any polar pattern you wish – you just need to know what the relative mix of the three mic's should be.

Using MATLAB, I produced three uncorrelated vectors containing a bunch of 10000 random numbers, each vector representing the output of each of the three microphones in that magic array described in the previous paragraph sitting in a noisy diffuse field. I then made two mixes of the three vectors to produce a simulation of a given pair of microphones. I then simply asked MATLAB to give me the correlation coefficient of these two simulated outputs.

If someone could give me the appropriate equation to do this the right

way, I would be very grateful.

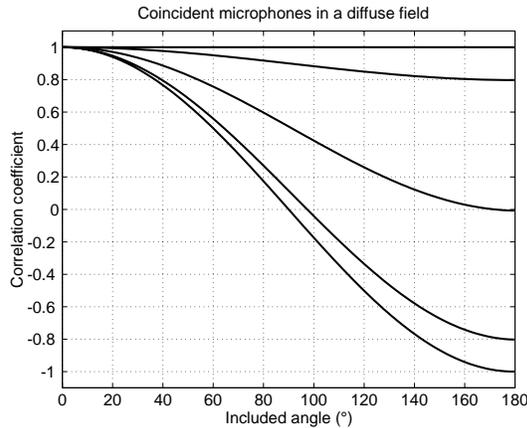


Figure 9.137: Correlation coefficients in the horizontal plane of a diffuse field for coincident omnidirectionals (top), subcardioids, cardioids, hypercardioids and bidirectionals (bottom) with included angles from 0° through 180° .

Spaced omnidirectionals

If we have a pair of omnidirectionals spaced apart in a diffuse field, then we can intuitively get an idea of what their correlation coefficient will be. At 0 Hz, the pressure at the two locations of the microphones will be the same. This is because the sound pressure variations in the room are all varying the day's barometric pressure which is, for our purposes, 0 Hz. At very low frequencies, the wavelengths of the sound waves going past the microphones will be longer than the distance between the mic's. As a result, the two signals will be very correlated because the phase difference between the mic's is small. As we go higher and higher in frequency, then the correlation should be less and less, until, at some high frequency, the wavelengths are much smaller than the microphone separation. This means that the two signals will be completely unrelated and the correlation coefficient goes to 0.

In fact, the relationship is a little more complicated than that, but not much. According to Kuttruff [Kuttruff, 1991], the correlation coefficient of two spaced locations in a theoretical diffuse field can be calculated using Equation 9.22.

$$r = \frac{\sin(kd)}{kd} \quad (9.22)$$

where k is the “wave number.” This is a slightly different way of saying

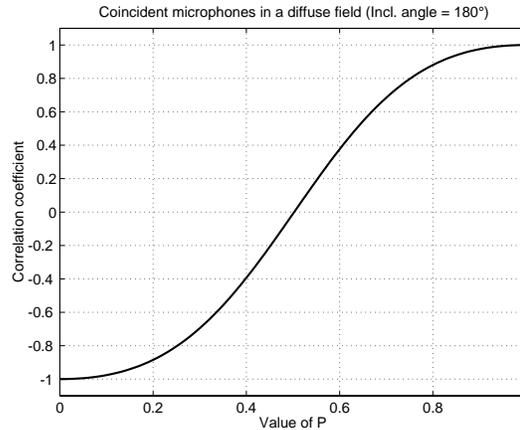


Figure 9.138: Correlation coefficients in the horizontal plane of a diffuse field for coincident microphones with an included angle of 180° and various values of P (remember that $G = 1 - P$).

“frequency” as can be seen in Equation 9.23 below (also see Equation 9.20).

$$k = \frac{2\pi f}{c} \quad (9.23)$$

Note that k is proportional to frequency and therefore inversely proportional to wavelength.

If we were to calculate the correlation coefficient for a given microphone separation and all frequencies, the plot would look like Figure 9.139. Note that changes in the distance between the mic’s will only change the frequency scale of the plot – the closer the mic’s are to each other, the higher the frequency range of the area of high correlation.

9.4.5 Conclusions

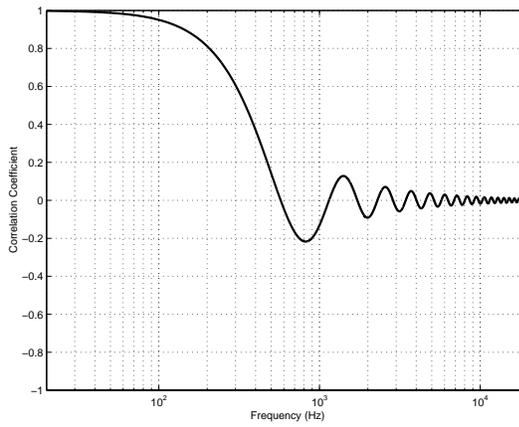


Figure 9.139: Correlation coefficient vs. Frequency for a pair of omnidirectional microphones with a separation of 30 cm in a diffuse field.

9.5 Matrixed Microphone Techniques

9.5.1 MS

Also known as *Mid-Side* or *Mono-Stereo*

We have already seen in Section ?? that any microphone polar pattern can be created by mixing an omnidirectional with a bidirectional microphone. Let's look at what happens when we mix other polar patterns.

Virtual Blumlein

We'll begin with a fairly simple case – two bidirectional microphones, one facing forward towards the middle of the stage (the “mid” microphone) and the other facing to the right (the “side” microphone). If we plot these two individual polar patterns on a cartesian plot the result will look like Figure 9.140.

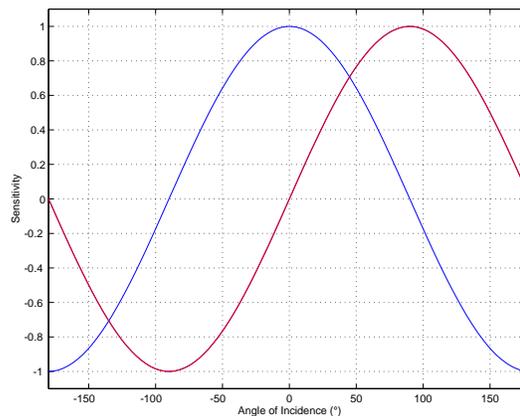


Figure 9.140: Two bidirectional microphones with an included angle of 90° , with one facing forward (blue line) and the other facing to the right (red line).

The same can be shown as the more familiar polar plot, displayed in Figure 9.141.

Now, let's take those two microphone outputs and, instead of sending them to the left and right outputs as we normally do with stereo microphone configurations, we'll send them both to the right output by panning both channels to the right. We'll also drop their two levels by 3 dB while we're at it (we'll see why later...)

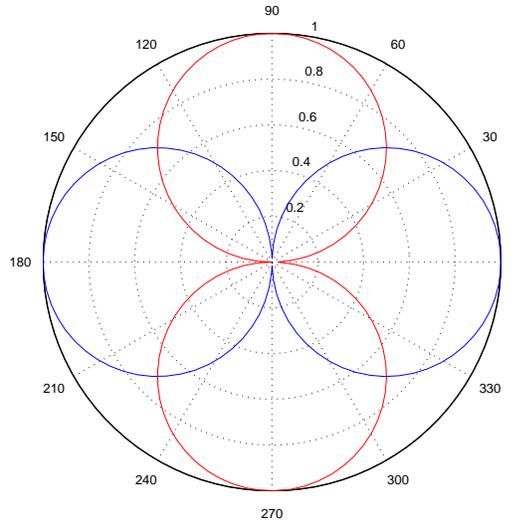


Figure 9.141: A polar plot showing the same information as is shown in Figure 9.140

What will be the result? We can figure this out mathematically as is shown in the equation below.

$$S_{total} = 0.707S_{mid} + 0.707S_{side} \quad (9.24)$$

$$= 0.707 \left(\cos(\vartheta) + \cos\left(\vartheta + \frac{\pi}{2}\right) \right) \quad (9.25)$$

$$= \cos\left(\vartheta + \frac{\pi}{4}\right) \quad (9.26)$$

If we were to plot this, it would look like Figure 9.142.

You will notice that the result is a bidirectional microphone aimed 45° to the right. This shouldn't really come as a big surprise, based on the assumption that you've read Section 1.4 way back at the beginning of this book.

In fact, using the two bidirectional microphones arranged as shown in Figure 9.140, you can create a "virtual" bidirectional microphone facing in any direction, simply by adding the outputs of the two microphones with carefully-chosen gains calculated using Equations 9.27 and 9.28.

$$M = \cos \sigma \quad (9.27)$$

$$S = \sin \sigma \quad (9.28)$$

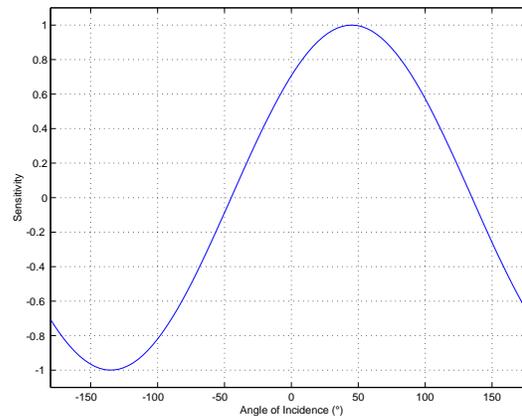


Figure 9.142: The resulting sensitivity pattern of the sum of the two bidirectional microphones shown in Figure 9.140 with levels dropped by 3 dB.

where M is the gain applied to the mid bidirectional, S is the gain applied to the side bidirectional and σ is the desired on-axis angle of the virtual bidirectional.

One important thing to notice here is that, for some desired angles of the virtual bidirectional microphone, you're going to have a negative gain on at least one of your microphones – possibly both of them. This, however, is easy to accomplish on your average mixing console. You just have to hit the polarity flip switch.

So, now we've seen that, using only two bidirectional microphones and a little math, you can create a bidirectional microphone aimed in any direction. This might be particularly useful if you don't have time to do a sound check before you have to do a recording (yes... it does happen occasionally). If you set up a pair of bidirectionals, one mid and one side and record their outputs straight to a two-track, you can do the appropriate summing later with different gains for your two stereo outputs to create a virtual pair of bidirectional microphones with an included angle that is completely adjustable in post-production. The other beautiful thing about this technique is that, if you are using bidirectional microphones whose front and back lobes are matched to each other on each microphone, your resulting matrixed (summed) outputs will be a *perfectly* matched pair of bidirectionals – even if your two original microphones are not matched... they don't even have to be the same brand name or model... Let's say that you really like a Beyers M130 ribbon microphone for its timbre, but the SNR is too

low to use to pick up the sidewall reflections, you can use it for the mid, and something like a Sennheiser MKH30 for the side bidirectional. Once they're matrixed, your resulting virtual pair of microphones (assuming that you have symmetrical gains on your two outputs) will be perfectly matched. Cool huh?

Traditional MS

You are not restricted to using two similar polar patterns when you use matrixing in your microphone techniques. For example, most people when they think of MS, think of a cardioid microphone for the mid and a bidirectional for the side. These two are shown in Figures 9.143 and 9.144.

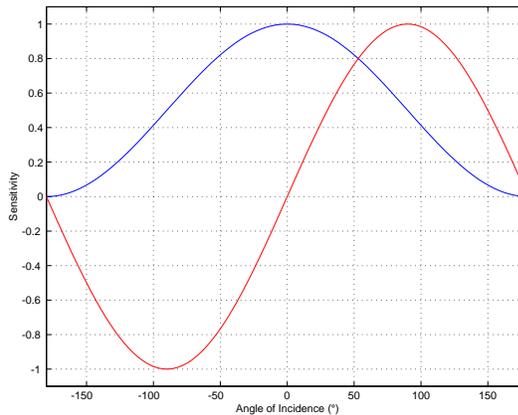


Figure 9.143: Two microphones with an included angle of 90° , with one forward-facing cardioid (blue line) and a side-facing bidirectional (red line).

What happens when we mix the outputs of these two microphones? Well, in order to maintain a constant on-axis response for the virtual microphone that results, we know that we're going to have to attenuate the outputs of the real microphones before mixing them. So, let's look at an example of the cardioid reduced by 6.02 dB and the bidirectional reduced by 3.01 dB (we'll see why I chose these particular values later). If we were to express the sensitivity of the resulting virtual mic as an equation it would look like Equation 9.29

$$S_{virtual} = 0.5(0.5 + 0.5 \cos(\vartheta)) + 0.707(\cos(\vartheta - \frac{\pi}{4})) \quad (9.29)$$

What does all this mean? $S_{virtual}$ is the sensitivity of the virtual micro-

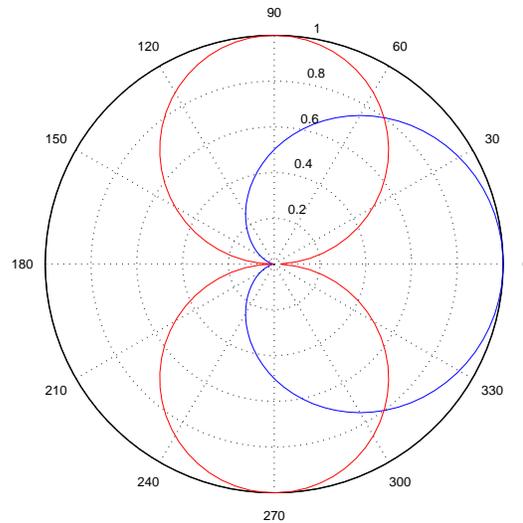


Figure 9.144: Two microphones with an included angle of 90° , with one forward-facing cardioid (blue line) and a side-facing bidirectional (red line).

The first 0.5 is there because we're dropping the level of the cardioid by 6 dB, similarly the 0.707 is there to drop the output of the bidirectional by 3 dB. The output of the cardioid should be recognizable as the $0.5 + 0.5 \cos(\vartheta)$. The bidirectional is the part that says $\cos(\vartheta - \frac{\pi}{4})$. Note that the $\frac{\pi}{4}$ is there because we've turned the bidirectional 90° . (An easier way to say this is to use $\sin(\vartheta)$ instead – it will give you the same results.

If we graph the result of Equation 9.29 it will look like Figure 9.145. Note that the result is a traditional hypercardioid pattern “aimed” at 71° .

There is a common misconception that using a cardioid and a bidirectional as an MS pair will give you a pair of virtual cardioids with a controllable included angle. This is not the case. The polar pattern of the virtual microphones will change with the included angle. This can be seen in Figure 9.146 which shows 10 different balances between the cardioid mid microphone and the bidirectional side.

[LINK TO ANIMATION](#)

How do you know what the relative levels of the two microphones should be? Let's look at the theory and then the practice.

Theory

If you wanted to maintain a constant on-axis sensitivity for the virtual microphone as it rotates, then we could choose an arbitrary number n and

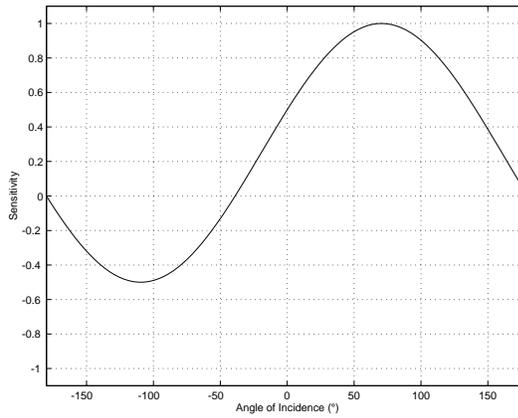


Figure 9.145: The resulting sensitivity of a forward-facing cardioid with an attenuation of 6.02 dB and a side-facing bidirectional with an attenuation of 3.01 dB. Note that the result is a traditional hypercardioid pattern “aimed” at 71°.

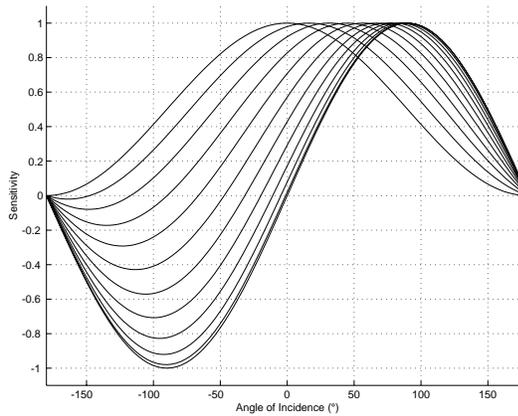


Figure 9.146: The result of the sum of a forward-facing cardioid and a side-facing bidirectional for ten different balances. Notice that the polar pattern changes from cardioid to bidirectional as the angle of rotation changes from 0° to 90°.

do the math in Equations 9.30 and 9.31:

$$M = 0.5 + 0.5 \cos(n) \quad (9.30)$$

$$S = \sin\left(\frac{n}{2}\right) \quad (9.31)$$

This M and S are the gains to be applied to the cardioid and bidirectional feeds, respectively. If you were to graph this relationship it would look like Figure 9.147.

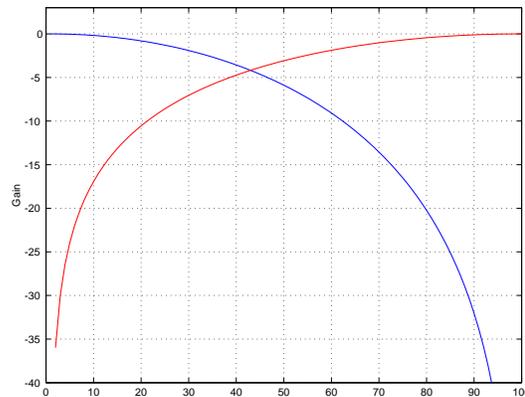


Figure 9.147: The gains applied to the mid cardioid (blue) and the side bidirectional (red) required to maintain a constant on-axis sensitivity for the virtual microphone as it rotates from 0° to 90° . Note that the x-axis of the graph does not directly correspond directly with any angle, however it is non-linearly related to the direction in which the virtual microphone is aimed.

You'll notice that I haven't spent much time on this theoretical side. This is because it has so little to do with the practical usage of MS mic'ing. If you wanted to really do this the right way, then I'd suggest that you do things a slightly different way. Use the math in Section 9.5.1 to create a virtual bidirectional pointing in the desired direction, then add some omni to produce the desired polar pattern. This idea will be elaborated in Section 9.5.2.

Practice

Okay, if you go to do an MS recording, I'll bet money that you don't bring your calculator along to do any of the math I just described. Use the following steps...

1. Arrange your microphones such that the cardioid is pointing forwards and the positive lobe of the bidirectional is pointing to the right.

2. Pan your cardioid to the centre.
3. Split the output of your bidirectional to two male XLR connectors using a splitter cable.
4. Pan the two parallel outputs of your bidirectional to hard left and hard right.
5. Flip the polarity of the left-panned bidirectional channel.
6. While monitoring the output of the mixer...
7. Bring up just the cardioid channel. Listen to this. It should be a mono signal in the centre of your stereo image.
8. While dropping the level of the cardioid, bring up the two bidirectional channels. The image should get wider and farther away. When you have bidirectional-only, you should hear an “out-of-phase” (i.e. opposite polarity) signal of mostly sidewall reflections.
9. Adjust the balance of the cardioid and the bidirectional channels to the desired image spread.

Don't feel obliged to keep your two bidirectional gains equal. Just remember that if they're not, then the polar patterns of your two virtual microphones are not matched. Also, they will not be aimed symmetrically, but that might be a good thing... People listening to the recording at home won't know whether you used a matched pair of microphones or not.

9.5.2 Ambisonics

To a large portion of the population who have encountered the term, “Ambisonics” generally means one of two things :

1. A method of quadaphonic playback which was as unsuccessful as the QD systems in the 1970's.
2. A strange British recording technique which uses a soundfield microphone which may or may not be reproduced using 4 loudspeakers in the 4 corners of the room.

In actual fact, is really neither of these things. Ambisonics is more of a mathematical concept and technique which is an attempt to reproduce

a recorded soundfield. This idea differs from most stereo and surround recordings in that the intention is to re-create the acoustic wavefronts which existed in the space at the time of the recording rather than to synthesize an interpretation of an acoustic event.

Theory

Go to a room (you may already be in one...) and put a perfect omnidirectional microphone in it. As we discussed in Section ??, an omnidirectional microphone is also known as a *pressure transducer* which means that it responds to the changes in air pressure at the diaphragm of the microphone. If you make a perfect recording of the output of the perfect omnidirectional microphone when stuff is happening in the room, you have captured a record (here, I'm using the word "record" as in a historical record, not as in a record that you buy at a record shop from a record lady[Lovett, 1994]) of the change in pressure over time at that location in that room on that day. If, at a later date, you play back that perfect recording over a perfect loudspeaker in a perfectly anechoic space, then you will hear a perfect representation (think "re-presentation") of that historical record. Interestingly, if you have a perfect loudspeaker and you're in a perfectly anechoic space, then what you hear from the playback is exactly what the microphone "heard" when you did the recording.

This is a good idea, however, let's take it a step farther. Since a pressure transducer has an omnidirectional polar pattern, we don't have any information regarding the direction of travel of the sound wavefront. This information is contained in the velocity of the pressure wave (which is why a single directional microphone of any sort must have a velocity component). So, let's put up a perfect velocity microphone in the same place as our perfect pressure microphone. As we saw in Section ?? a velocity microphone (if we're talking about directional characteristics and not transducer design) is a bidirectional microphone. Great, so we put a bidirectional mic facing forward so we can tell if the wave is coming from the front or the rear. If the outputs of the omni and the bidirectional have the same polarity, then the sound source is in the front. If they're opposite polarity, then the sound source is in the rear. Also, we can see from the relative levels of the two mic outputs what the angle to the sound source is, because we know the relative sensitivities of the two microphones. For example, if the level is 3 dB lower in the bidirectional than the omni and both have the same polarity, then the sound source must be 45° away from directly forward. The problem is that we don't know if it's to the left or the right. This problem is easily solved by

putting in another bidirectional microphone facing to the side. Now we can tell, using the relative polarities and outputs of the three microphones where the sound source is... but we can't tell what the sound source elevation is. Again, no problem, we'll just put in a bidirectional facing upwards.

So, with a single omni and three bidirectionals facing forward, to the right and upwards, we can derive all sorts of information about the location of the sound source. If all four microphones have the same polarity, and the outputs of the three bidirectionals are each 3 dB below the output of the omni, then the sound source must be 45° to the right and 45° up from the microphone array.

Take a look at the top example in Figure 9.148. We can see here that if the sound source is directly in front of the microphone array, then we get equal positive outputs from the omni (we'll call this the W channel) and the forward-facing bidirectional (we'll call that one the Y channel) and nothing from the side-facing bidirectional (the X channel). Also, if I had been keen and did a 3D diagram and drawn the upwards-facing bidirectional (the Z channel), we'd see that there was no signal from that one either if the sound source is on the same horizontal plane as the microphones.

Let's record the pressure wave (using the omni mic) and the velocity, and therefore the directional information (using the three bidirectional mic's) on a perfect four-channel recorder. Can we play these channels back to reproduce all of that information in our anechoic listening room? Take a look at Figure 9.149.

Let's think about what happens to the sound source in the top example in Figure 9.148 if we play back the W , X , and Y channels through the system in Figure 9.149. In this system, we have four identical loudspeakers placed at 0° , 180° , and $\pm 90^\circ$. These loudspeakers are all identical distances from the sweet spot.

The top example in Figure 9.148 results in a positive spike in both the W and Y channels, and nothing in the X channel. As a result, in the playback system:

- the front loudspeaker produces a positive pressure.
- The two side speakers produce equal positive pressures that are one-third the outputs of the front (because there's nothing in the X channel and they don't play the Y channel).
- Finally, the rear speaker produces a negative pressure at one-third the output of the front loudspeaker because the information in the W and

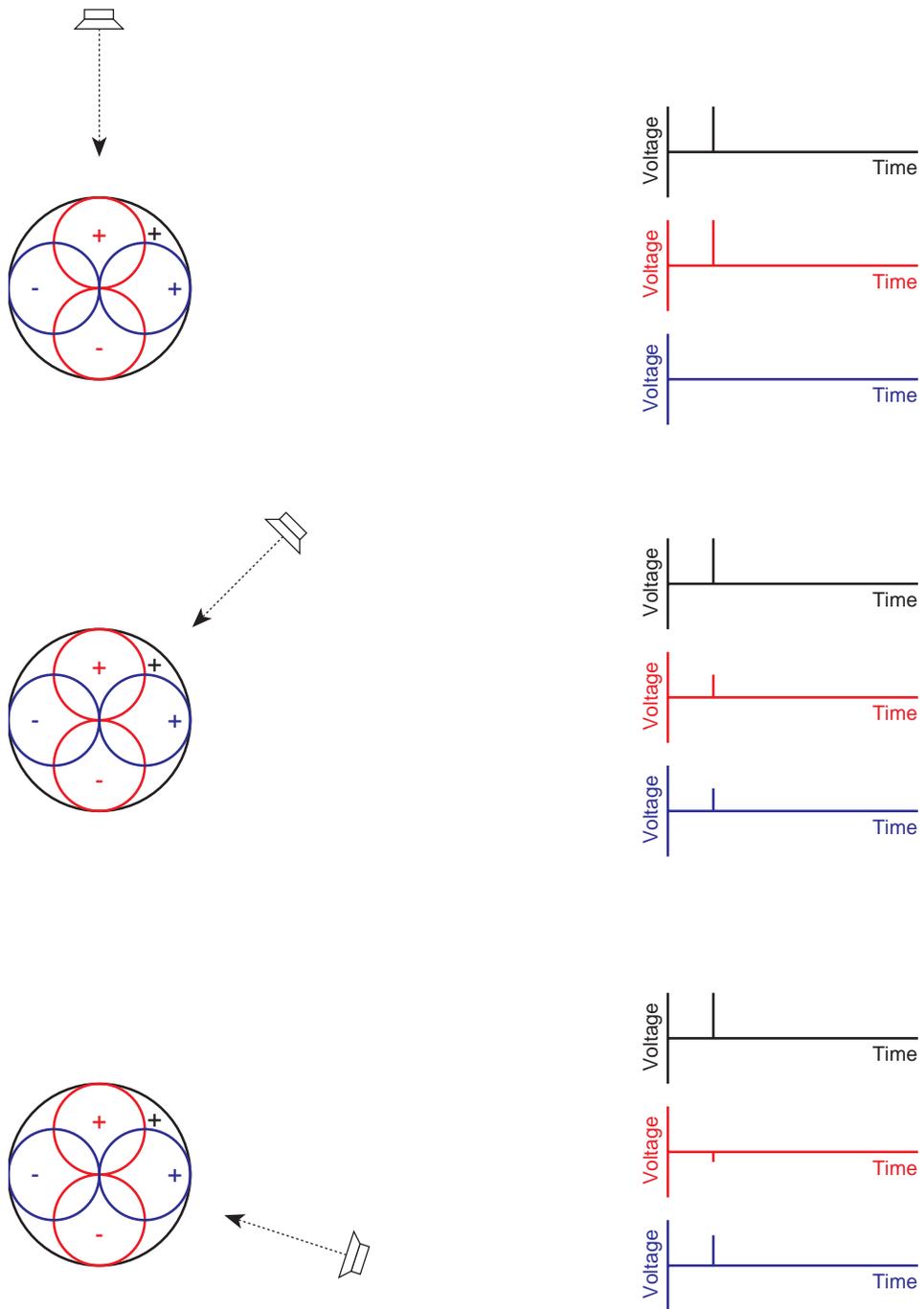


Figure 9.148: Top views of a two-dimensional version of the system described in the text. These are three examples showing the relationship between the outputs of the omnidirectional and two of the bidirectional microphones for sound sources in various locations producing a positive impulse.

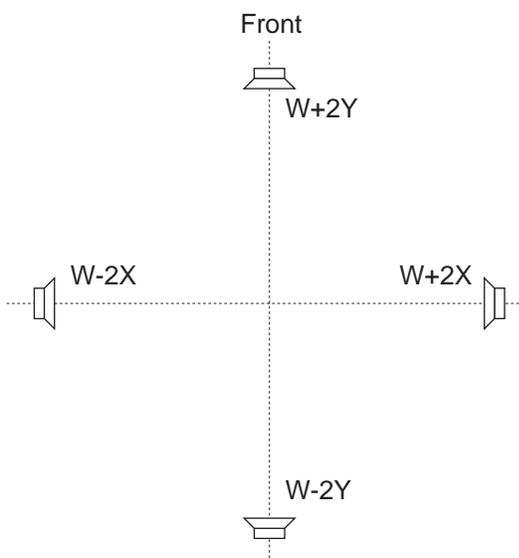


Figure 9.149: A simple configuration for playing back the information captured by the three microphones in Figure 9.148.

the negative Y channels cancel each other a little when they're mixed together at the speaker, but the negative signal is louder.

The loudspeakers produce a signal at exactly the same time, and the different waves will propagate towards the sweet spot at the same speed. At the sweet spot, the waves all add together (think of adding vectors together) to produce a resulting pressure wave that has a velocity that is moving towards the rear loudspeaker (because the two side speakers push equally against each other, so there's no sideways velocity, and because the front speaker is pushing towards the rear one which is pulling the wave towards itself).

If we used perfect microphones and a perfect recording system and perfect loudspeakers, the result, at the sweet spot in the listening room, is that the sound wave has exactly the same pressure and velocity components as the original wave that existed at the microphones' position at the time of the recording.

Consequently, we say that we have re-created the *soundfield* in the recording space. If we pretend that the sound wave has only two components, the pressure and the velocity, then our perfect system perfectly duplicates reality.

As an exercise, before you keep reading, you might want to consider what will come out of the loudspeakers for the other two examples in Figure 9.148.

So far, what we've got here is a simple first-order Ambisonics system. The collective outputs of the four microphones is what is known as an Ambisonics *B-Format* signal. Notice that the B-Format signal contains all four channels. If we wanted to restrict ourselves to just the horizontal plane, we can legally leave out the *Z*-channel (the upwards-facing bidirectional). This is legal because most people don't have loudspeakers in their ceiling and floors... not good ones anyways... The Ambisonics people have fancy names for their two versions of the system. If we include the height information with the *Z*-channel, then we call it a *periphonic* system (think *periscope* and you'll remember that there's stuff above you...). If we leave out the *Z*-channel and just capture and playback the horizontal plan directional information, then we call it a *panphonic* system (think stereo *panning* or *panoramic*).

Let's take this a step further. We begin by mathematically describing the relationship between the angle to the sound source and the sensitivity patterns (and therefore the relative outputs) of the B-format channels. Then we define the mix of each of these channels for each loudspeaker. That mix is determined by the angle of the loudspeaker in your setup. It's generally assumed that you have a circle of loudspeakers with equal apertures (meaning that they are all equally spaced around the circle). Also, notice that there is an equation to define the minimum number of loudspeakers required to accurately reproduce the Ambisonics signal. These equations are slightly different for panphonic and periphonic systems

One important thing to notice in the following equations is that, at its most complicated (meaning a periphonic system), a first-order Ambisonics system has only 4 channels of recorded information within the B-format signal. However, you can play that signal back over any number of loudspeakers. This is one of the attractive aspects of Ambisonics – unlike traditional two-channel stereo, or discrete 5.1, the number of recording channels is not defined by the number of output channels. You always have the same number of recording channels whose mix is changed according to the number of playback channels (loudspeakers).

First-order panphonic

$$W = P_{\Psi} \tag{9.32}$$

$$X = P_{\Psi} \cos \Psi \tag{9.33}$$

$$Y = P_{\Psi} \sin \Psi \quad (9.34)$$

Where W , X and Y are the amplitudes of the three ambisonics B-format channels, P_{Ψ} is the pressure of the incident sound wave and Ψ is the angle to the sound source (where 0° is directly forward) in the horizontal plane. Notice that these are just descriptions of an omnidirectional microphone and two bidirectionals. The bidirectionals have an included angle of 90° – hence the cosine and sine (these are the same function, just 90° apart – $Y = P_{\Psi} \sin \Psi$ is shorter than writing $Y = P_{\Psi} \cos(\Psi + 90^{\circ})$).

$$P_n = \frac{W + 2X \cos \varphi_n + 2Y \sin \varphi_n}{N} \quad (9.35)$$

Where

P_n is the amplitude of the n^{th} loudspeaker, φ_n is the angle of the n^{th} loudspeaker in the listening room, and N is the number of loudspeakers.

The decoding algorithm used here is one suggested by Vanderkooy and Lipshitz which differs from Gerzons original equations in that it uses a gain of 2 on the X and Y channels rather than the standard $\sqrt{2}$. This is due to the fact that this method omits the 1 gain from $\frac{1}{\sqrt{2}}$ the W channel in the encoding process for simpler analysis [Bamford and Vanderkooy, 1995].

$$B = 2m + 1 \quad (9.36)$$

Where B is the minimum number of loudspeakers required to accurately produce the panphonic ambisonics signal and m is the order of the system. (So far, we have only discussed 1st-order Ambisonics in this book.)

First-order periphonic

NOT YET WRITTEN

$$B = ? \quad (9.37)$$

Practical Implementation

Practically speaking, it is difficult to put four microphones (the omnidirectional and the three bidirectionals) in a single location in the recording space. If you're doing a panphonic recording, you can make a vertical array with the omni in between the two bidirectionals and come pretty close. There is also the small problem of the fact that the frequency responses of the bidirectionals and the omni won't match perfectly. This will make the contributions of the pressure and the velocity components frequency-dependent when you mix them to send to the loudspeakers.

So, what we need is a smarter microphone arrangement, and at the same time (if we're smart enough), we need to match the frequency responses of the pressure and velocity components. It turns out that both of these goals are achievable (within reason).

We start by building a tetrahedron. If you're not sure what that looks like, don't panic. Imagine a pyramid made of 4 equilateral triangles (normal pyramids have a square base – a tetrahedron has a triangular base). Then we make each face of the tetrahedron the diaphragm of a cardioid microphone. Remember that a cardioid microphone is one-half pressure and one-half velocity, therefore we have matched our components (in theory, at least...).

This arrangement of four cardioid microphones in a single housing is what is typically called a *Soundfield microphone*. Various versions of this arrangement have been made over the years by different companies.

The signal consisting of the outputs of the four cardioids in the soundfield microphone make up what is commonly called an *A-Format* Ambisonics signal. These are typically converted into the B-Format using a standard set of equations given below.

EQUATIONS HERE TO CONVERT A-FORMAT INTO B-FORMAT

There are some problems with this implementation. Firstly, we are oversimplifying a little too much when we think that the four cardioid capsules can be combined to give us a perfect B-Format signal. This is because the four capsules are just too far apart to be effective at synthesizing a virtual omni or bidirectional at high frequencies. We can't make the four cardioids coincident (they can't be in exactly the same place) so the theory falls apart a little bit here – but only for high frequencies. Secondly, nobody has ever built a perfect cardioid that is *really* a cardioid at all frequencies. Consequently, our dream of matching the pressure and velocity components falls apart a bit as well.

Finally, there's a strange little problem that typically gets glossed over a bit in most discussions of Soundfield microphones. You'll remember from earlier in this book that the output of a velocity transducer has a naturally rising response of 6 dB per octave. In other words, it has no low-end. In order to make bidirectional (as well as all other directional) microphones sound better, the manufacturers boost the low end using various methods, either acoustical or electrical. Therefore, an off-the-shelf bidirectional (or cardioid) microphone really doesn't behave "correctly" to be used in a theoretically "correct" Ambisonics system – it simply has too much low-end (and a messed-up low-frequency phase response to go with it...). The strange thing is that, if you build a system that uses "correct" velocity components with the rising 6 dB per octave response and listen to the output, it will

sound too bright and thin. In order to make the Ambisonics output sound warm and fuzzy (and therefore good) you have to boost the low-frequency components in your velocity channels. Technically, this is incorrect, however, it sounds better, so people do it.

Higher orders

Ambisonics is a systems that works on “orders” – the higher the order of the system, the more accurate the reproduction of the sound field.

- If we just use the W -channel (the omnidirectional component) then we just get the pressure information and we consider it to be a 0th- (zeroth) order system. This gives us the change in pressure over time and nothing else.
- If we add the X -, Y - and Z -channels, we get the velocity information as well. As a result we can tell not only the change in pressure of the sound source, but also its direction relative to the microphone array. This gives us a 1st-order system.
- A 2nd-order Ambisonics system adds information about the curvature of the sound wave. This information is captured by a microphone that doesn't exist yet. It has a strange four-leaved clover shaped pattern with four lobes.

Second-order periphonic

$$U = P_{\Psi} \cos 2\Psi \quad (9.38)$$

$$V = P_{\Psi} \sin 2\Psi \quad (9.39)$$

$$W = P_{\Psi} \quad (9.40)$$

$$X = P_{\Psi} \cos \Psi \quad (9.41)$$

$$Y = P_{\Psi} \sin \Psi \quad (9.42)$$

$$P_n = \frac{2U \cos 2\varphi_n + 2V \sin 2\varphi_n + W + 2X \cos \varphi_n + 2Y \sin \varphi_n}{N} \quad (9.43)$$

FIX THE ABOVE EQUATION TO ADD THE HEIGHT INFORMATION

NOT YET WRITTEN

$$B = ? \quad (9.44)$$

Why Ambisonics cannot work

Let's take a simple 1st-order panphonic Ambisonics system. We can use the equations given above to think of the system in a more holistic way. If we combine the sensitivity equations for the B-format signal with the mix equation for a loudspeaker in the playback system, we can make a plot of the gain applied to a signal as a function of the relationship between the angle to the sound source and the angle to the loudspeaker. That function looks like the graph in Figure 9.150.

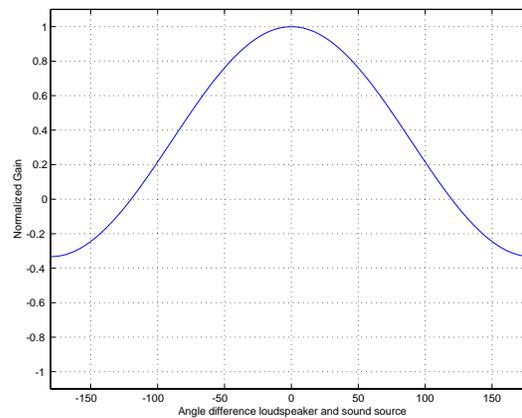


Figure 9.150: The sensitivity function combining the sensitivities of the B-format channels with the mix for the loudspeaker.

So far, we have been considering the sound from a source recorded by a microphone at a single point in space, played back over loudspeakers and analyzed at a single point in space (the sweet spot). In doing this, we have found out that the soundfield at the sweet spot exactly matches the soundfield at the sweet spot within the constraints of the order of the Ambisonics system.

Let's now change the analysis to consider what you actually hear. Instead of a single-point microphone, you have two ears, one on either side of your head. Let's look at two situations, a sound source directly in front of you and a sound source directly to the side. To simplify things, we'll put ourselves in an anechoic world.

As we saw in Section ?? a Head Related Transfer Function (or HRTF) is a description of what your head and ears do to a sound signal before hitting your eardrum. These HRTF's can be used in a number of ways, but for our purposes, we'll stick to impulse responses, showing what's happening in the

time domain. The analysis you're about to read uses the HRTF database measured at MIT using a KEMAR dummy head. This is a public database available for download via the Internet[Gardner and Martin, 1995].

We'll begin by looking at the HRTF's of two sound sources, one directly in front of the listener and one directly to the right. The impulse responses for the resulting HRTF's for these two locations are shown in Figures 9.151 and 9.152 respectively.

There are two things to notice about the two impulse responses shown in Figure 9.151 for a frontal sound source. Firstly, the times of arrival of the impulses at the two ears are identical. Secondly, the impulse responses themselves are identical throughout the entire length of the measurement.

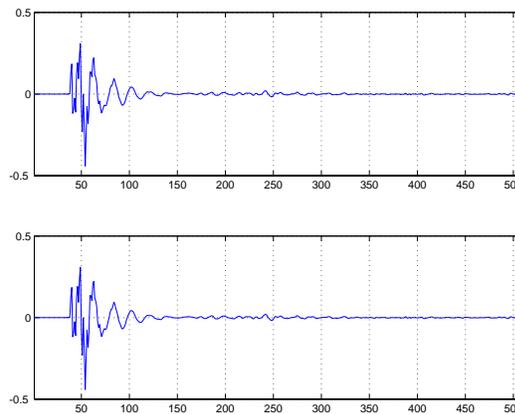


Figure 9.151: The impulse responses measured at the two ears of a KEMAR dummy head for a sound source directly in front[Gardner and Martin, 1995]. The top plot is the left ear and the bottom plot is the right ear. The x-axes are time, measured in samples.

Let's consider the same two aspects for Figure 9.152 which shows the HRTF's for a sound source on the side of a listener. Notice in this case that the times of arrival of the impulses at the two ears differ. Since the sound source is on the right side of the listener, the impulse arrives at the right ear before the left. This makes sense since the right ear is closer to sound sources on the right side of your head. Now take a look at the impulse response over time. The first big spike in the right ear goes positive. Similarly, the first big spike in the left ear also goes positive. This should not come as a surprise, since your eardrums are not bidirectional transducers. These interaural time differences (ITD's) are *very* significant components that our brains use in determining where a sound source is.

Let's now consider a source directly in front of a soundfield microphone,

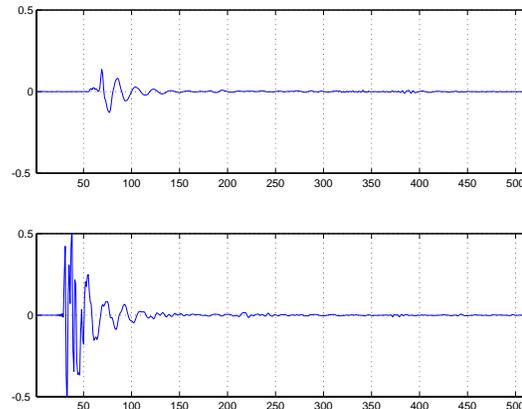


Figure 9.152: The impulse responses measured at the two ears of a KEMAR dummy head for a sound source directly in to the right[Gardner and Martin, 1995]. The top plot is the left ear and the bottom plot is the right ear. The x-axes are time, measured in samples.

recorded in 1st-order Ambisonics and played over an 8-channel loudspeaker configuration shown in Figure 9.153.

If we assume that the sound source emits a perfect impulse and is recorded by a perfect soundfield microphone and subsequently reproduced by 8 perfect loudspeakers, we can use the same HRTF measurements to determine the resulting signal that arrives at the ears of the dummy head. Figure 9.154 shows the HRTF's for a sound source recorded and reproduced through such a system. Again, let's look at the same two characteristics of the impulse responses. The times of arrival of the impulse at the two ears are identical, as we would expect for a frontal sound source. Also, the two impulse responses are identical, also expected for a frontal sound source. So far, Ambisonics seems to be working... however, you may notice that the impulse responses in Figure 9.154 aren't identical to those in Figure 9.151. Frankly, however, this doesn't worry me too much. We'll move on...

Figure 9.155 shows the HRTF's for a sound source 90° off to the side of a soundfield microphone and reproduced through the same 8-channel Ambisonics system. Again, we'll look at the same two characteristics of the impulse responses. Firstly, notice that the times of arrival of the pressure waves at the two ears is identical. This contrasts with the impulse responses in Figure 9.152. The interaural time differences that occur with real sources are eliminated in the Ambisonics system. This is caused by the fact that the soundfield microphone cannot detect time of arrival differences because it is in one location.

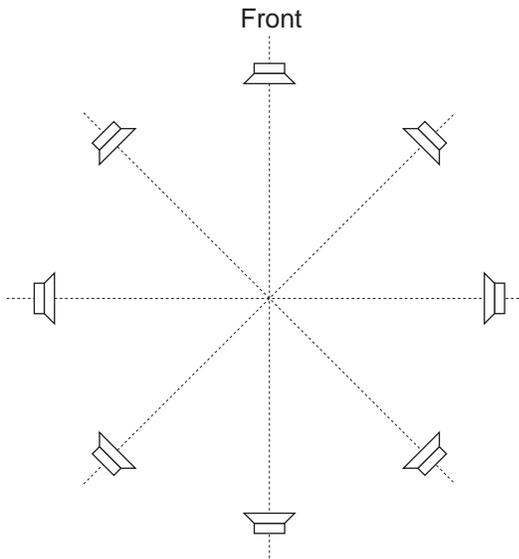


Figure 9.153: The 8-channel Ambisonics loudspeaker configuration used in this analysis.

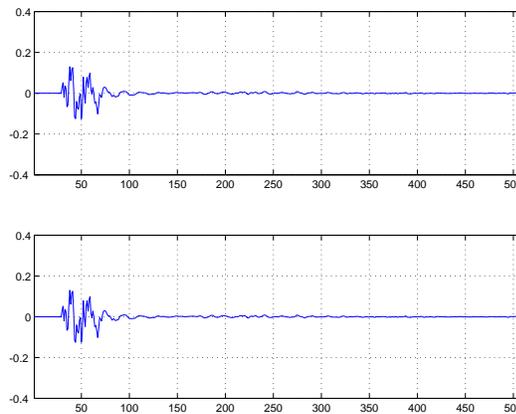


Figure 9.154: FRONT AMBISONICS

Secondly, notice the differences in the impulse responses at the two ears. The initial spike in the right ear is positive whereas the first spike in the left ear is negative. This is caused by the fact that loudspeakers that are opposite each other in the listening space in a 1st-order Ambisonics system are opposite in polarity. This can be seen in the sensitivity function shown in Figure 9.150. The result of this opposite polarity is that sound sources on the sides sound similar to a stereo signal normally described as being “out of phase” where the two channels are opposite in polarity [Geoff Martin, 1999].

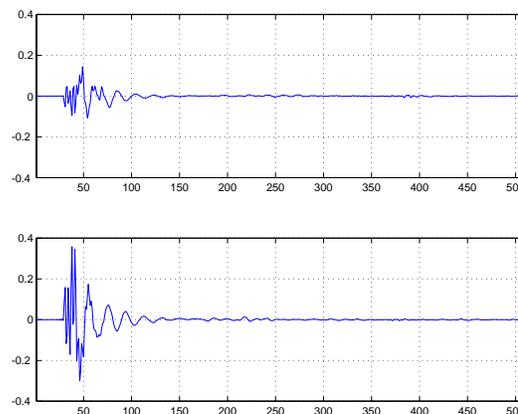


Figure 9.155: SIDE AMBISONICS

In the interest of fairness, a couple of things should be pointed out here. The first is that these problems are most evident in 1st-order Ambisonics systems, The higher the order, the less problematic they are. However, for the time being, it is impossible to do a recording in a real space in any Ambisonics system higher than 1st-order. Work has been done to develop coefficients that avoid polarity differences in the system [?] and people are developing fancy ways of synthesizing higher-order directional microphones using multiple transducers[], however, these systems have associated problems that will not be discussed here.

9.5.3 Suggested Reading List

9.6 Introduction to Surround Microphone Technique

Microphone techniques for surround sound are still in their infancy. I suppose that, if you really wanted to argue about it, you could say that stereo microphone techniques are as well, but at least we've all had a little more experience with two channels than five.

As a result, Most of what is contained in this chapter is basically just a list of some suggested configurations from various people along with a lot of my own opinions. Someday, in the future, I plan on adding an additional chapter to this book that is the surround equivalent to Section 9.4 (Although you should note that there is some information in there as well on surround.).

The goal in surround recording is basically the same as it is for two-channel. You have to get a pretty (or at least desired) timbral and spatial representation of your ensemble to the end listener. The good thing about surround is that your possibilities for spatial effects are much greater than they are with stereo. We could make a case for realism – a “you-are-there” representation of the ensemble, but I'm not one of the people that belongs to this camp, so I won't go there.

9.6.1 Advantages to Surround

If you've one of the many people that has only heard a couple of surround sound recordings, you are probably questioning whether it's really worth all the trouble to record in surround. Many, if not most of the surround recordings available for consumption are bad examples of the capabilities of the system. Don't despair – the problem lies in the recordings, not surround itself. You have to remember that everyone's just in the learning process of what to do with surround. Also, keep in mind that surround is still new, so everyone is doing fairly gimmicky recordings. Things that show off individual characteristics of the system instead of trying to do a good recording. (For example, it's easy to tell when a studio buys a new reverb unit because their next two releases have *tons* of reverb in them.) There are many exceptions to this broad criticism of surround recordings, but I won't mention any names, just in case I get myself in trouble.

There are many reasons to upgrade to surround recordings. the biggest reason is freedom – surround allows you much more room to use spatial characteristics of the distribution medium as a creative tool in your recordings.

The centre channel

There is a lot of debate in the professional community regarding whether or not to use the centre channel. Most people think that it's a necessary component in a surround system in order to have a larger listening area. There is a small, but very vocal group that it absolutely dead set against the centre channel, arguing that a phantom centre sounds better (for various reasons, the most important of which is usually timbre).

Here I would like to categorically state a personal opinion that I belong to both camps. The centre channel can very easily be mis-used and, as a result, ruin an otherwise good recording. On the other hand, I have heard some excellent surround recordings that use the centre channel. My personal feeling is that the centre speaker should be used as a spatial tool. A phantom centre and a real centre are very different beasts – use the one you prefer at the moment in your recording where it is most appropriate.

It surrounds...

The title of this is pretty obvious... One of the best reasons to use surround sound is that you can have sound that surrounds. I don't know if I have to say anything else. It is, of course, extremely difficult to have a completely enveloping soundfield for a listener using 2-channel stereo. It's not easy to do this in 5-channel surround, but it's easier than it is with stereo.

Playback configuration

Believe it or not, if you do a surround recording, it is more likely that your end listeners have a reasonably correct speaker configuration than if you're working in 2-channel stereo. Think of all your friends (this is easier to do if you think of them one-by-one, particularly if some of them don't get along...). Of these people, think of the ones with a stereo system. How many of these people have a "correct" loudspeaker configuration? I will bet that the number is very close to none. Now think of people with a surround sound system. In most cases, they have this system for watching movies – so they've got a television which is the centre channel (or there's a small speaker on top), a speaker on either side and a couple of speakers in the rear. Okay, so the configuration probably doesn't conform to the ITU-775[?] standard, but it's better than having a stereo system where the left speaker is in the living room and the right speaker is in the kitchen.

Binaural cues: Leaving things to the listener

Go to an orchestra concert and sit in a front row. Try to listen to the oboe when everyone else is playing and you'll probably find that you're able to do this pretty easily. If you could wind back time, you would find that you could have listened to the clarinet instead. (If you don't like orchestral music, go to a bar and eavesdrop on people's conversations – you can do this too. If you get caught, tell them it's research.) You're able to do this because you are able to track both the timbre and the location of a sound source. (Check out a phenomenon called the *cocktail party effect* for more information on this.)

If you record in mono, you have to be *very* careful about your mix. You have to balance all the components very precisely to ensure that people can hear what is needed to be heard at that particular moment. This is because people are unable to use spatial cues to determine where instruments are and therefore devote more concentration to them.

If we graduate to 2-channel stereo, life gets a little easier. By panning sources across the stereo sound stage between the two speakers, people are able to concentrate on one instrument and effectively attenuate others within their brains.

The more spatial cues you give to the listener, the better able they are to effectively zero in on whatever component of the recording they like. This doesn't necessarily mean that their perception can't be manipulated but it also means that you don't have to do as much manipulation in your mixes. Mastering engineers such as Bob Ludwig also report that they are finding that less compression and sweetening is required in surround sound media for the same reasons.

Back in the old days, we used level almost exclusively to manipulate people's attention in a mix. Nowadays, with surround, you can use level, but also spatial cues to draw attention towards or away from components.

9.6.2 Common pitfalls

Of course, every silver cloud has a dark lining... Recording in surround doesn't automatically fix all of your problems and make you a great recording engineer. In fact, if you're a bad engineer in 2-channel stereo, you'll be 2.5 times worse in 5-channel (2.55 times worse in 5.1) so be careful.

I would also go so far as to create an equation which I, in a pathetic stab at immortality, will dub *Martin's Law* which states that the difficulty of doing a recording and mix in surround sound can be calculated from the

number of channels in your media, using Equation 9.45.

$$D_{ris} = \left(\frac{n}{ref} \right)^n \quad (9.45)$$

where D_{ris} is the Difficulty of Recording in Surround, n is the number of channels in your recording and ref is the number of channels in the media to which you're comparing. For example it is 118.4 times more difficult to record in 5.1 than stereo (*sic*).

The centre channel

You may notice that I used this same topic as one of the advantages for recording in surround. I put it here to make you aware that you shouldn't just go putting things in the centre channel all willy-nilly. Use the centre channel with caution. Always remember that it's much easier to localize a real source (like a loudspeaker) in a real room than it is to localize a phantom source. This means that if you want to start playing with people's perception of distance to the sound source, you might want to carefully consider the centre channel.

Another possible problem that the centre channel can create is in timbre. Take a single channel of pink noise and send it to your Left and Right speakers. If all goes well, you'll get a phantom centre. Now, move around the sweet spot a bit and listen to the timbre of the noise. You should hear some comb filtering, but it really shouldn't be too disturbing. (Now that you've heard that, listen for it on the lead vocals of every CD you own... Sorry to ruin your life...) Repeat this experiment, but this time, send the pink noise to the Centre and Right channels simultaneously. Now you should get a phantom image somewhere around 15° off-centre, but if you move around the sweet spot you'll hear much more serious problems with your comb filtering. This is a bigger problem than in stereo because your head isn't getting in the way of the signal from the Centre speaker. In the case of Left interfering with Right, you have a reasonably high degree of attenuation in the crosstalk (Left speaker getting to the right ear and vice versa.) In the case of the Centre channel, this attenuation is reduced, so you get a big interference between the Centre and Right channels in your right ear. The left ear isn't a problem because the Right channel is attenuated. So, the moral of this story is to be careful with sources that are panned between speakers – decide whether you want the comb filtering. It's okay to have it as long as it's intentional [Martin, 2002a][Martin, 2002b].

So, going back to my statement in the previous section... The centre channel itself is not a problem, it's in the way you use it. Centre channels don't kill recordings, recording engineers kill recordings.

Localization

Don't expect perfect localization for all sources, 360° around the listener. A single, focused phantom image on the side is probably impossible to achieve. Phantom images behind the listener appear to get very close, sometimes causing in-head localization. (Remember that the surround channels are basically a giant pair of headphones.) Rear images are highly unstable and dependent on the listener's movements due to the wide separation of the loudspeakers.

Note that if you want to search for the holy grail of stable, precise and accurate side images, you'll probably have to start worrying about the spatial distribution and timing of your early reflections [?].

Soundfield continuity

If you watch a movie mixed by a bad *re-recording engineer* (the film world's equivalent of a mixing engineer in the music world), you'll notice a couple of obvious things. All of the dialog and *foley* (all of the little extra sound effects like zipping zippers, stepping foot steps and shutting doors) comes from the Centre speaker, the music comes from the Left and Right speakers, and the Surround speakers are used for the occasional special effect like rain sounds or crowd noises. Essentially, you're presented with three completely unrelated soundfields. You can barely get away with this independence of signal in a movie because people are busy using their eyes watching beautiful people in car chases. In music-only recordings, however, we don't have the luxury of this distraction, unfortunately.

Listen to a poorly-recorded or mixed surround recording and you'll notice a couple of obvious, but common, mistakes. There is no connection between the front and surround speakers – instruments in the front, reverb in the surround is a common presentation that comes from the film world. Don't get me wrong here, I'm not saying that you shouldn't have a presentation where the instruments are in the front only – if that's what you want, that's up to you. What I'm saying is, if you're going for that spatial representation, it's a bad idea to use your surrounds as the only reverb in the mix. They'll sound completely disconnected with your instruments. You can correct this by making some of the signals in the front and the rear the same – either

send instruments to the rear or send reverb to the front. What you do is up to you, but please be careful to not have a large wall between your front and your rear. This is the surround equivalent of some of the early days of stereo where the lead vocals were on the left and the guitar on the right. (Not that I don't like the Beatles, but their early stereo recordings weren't exactly sophisticated, spatially speaking...)

How big is your sweet spot?

This may initially seem like a rather personal question, but it really isn't, I assure you.

Take some pink noise and send it to all 5 channels in your surround system. Start at the sweet spot and move around just a little – you'll notice some pretty horrendous comb filtering. If you move more, you'll notice that the noise collapses pretty quickly to the nearest speaker. This is normal – and the more absorptive your room, the worse it will be. Now, send a different pink noise generator to each of your five channels. (If you don't have five pink noise generators lying around, you can always just use big delays on the order of a half second or so. So you have the pink noise going to Left, then Centre a half-second later, and Right a half-second after that and so on around the room...) Now you'll notice that if you move around the sweet spot, you don't get any comb filtering. This is because the signals are too different to interfere with each other. Also, big movements will result in less collapse to a speaker. This is because your brain is getting multiple different signals so it's not trying to put them all together into one single signal.

The moral of this story is that, if you want a bigger sweet spot, you're going to have to make the signals in the speakers different. (In technical terms, you're looking for *decorrelation* between your channels. This is particularly true of reverberation. The easiest way to achieve this with microphone technique is to space your microphones. The farther apart they are, the less alike their signals as is discussed in Section ??.

“Surround” and “Rear” are not synonyms

This is a common error that many people make. If you take a look at the official standard for a 5-channel loudspeaker configuration shown in Section ??, you'll see that the surround loudspeakers are to be placed in the 100° - 120° zone. Typically, people like to use the middle of this zone - 110°, but lots of pop people like pushing the surrounds a little to the rear, keeping

them at 120° .

Be careful not to think of the surround loudspeakers as rear loudspeakers. They're really out to the sides, and a little to the rear, they're not directly behind you. This becomes really obvious if you try to create a phantom centre rear image using only the surround channels. You'll find in this case that the apparent distance to the sound source is quite close to the back of your head, but this is not surprising if you draw a straight line between your two surround loudspeakers... In theory, of course, the apparent distance to the source should remain constant as it pans from LS to RS, but this is not the case, probably because the speakers are so far apart,

9.6.3 Fukada Tree

This configuration, developed by Akira Fukada at NHK Japan was one of the first published recommendations for a microphone technique for ITU-775 surround [Tsujiimoto and Akita, 1997]. As can be seen in Figure 9.156, it consists of five widely spaced cardioids, each sending a signal to a single channel. In addition, two omnidirectional microphones are placed on the sides with each signal routed to two channels.

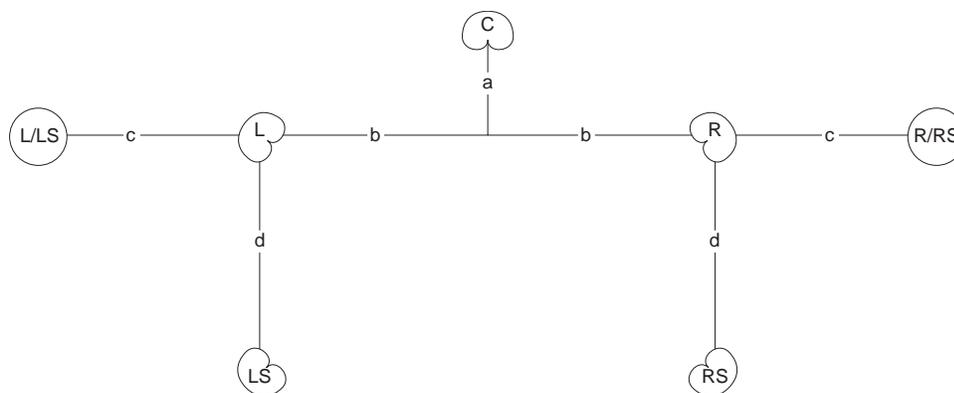


Figure 9.156: Fukada Tree: $a = b = c = 1 - 1.5$ m, $d = 0 - 2$ m, L/R angle = $110^\circ - 130^\circ$, LS/RS angle = $60^\circ - 90^\circ$.

This is a very useful technique, particularly in larger halls with big ensembles. The large separation of the front three cardioids prevents any detrimental comb filtering effects in the listening room on the direct sound of the ensemble (this problem is discussed above). One interesting thing to try with this configuration is to just listen to the five cardioid outputs with

a large distance to the rear microphones. You will notice that, due to the large separation between the front and rear signals in the recording space, the perceived sound field in the listening room has two separate areas – that is to say that the frontal sound stage appears to be separate from the surround with nothing connecting them. This is caused by the low correlation between the front and rear signals. Fukada cures this problem by sending the outputs of the omnis to front and surround. The result is a very spacious soundfield, but with reasonably reliable imaging characteristics. You may notice some comb filtering effects caused by having identical signals in the L/LS and R/RS pairs, but you will have to listen carefully for them...

Notice that the distance between the front array and rear pair of microphones can be as little as 0 m, therefore, there may be situations where all microphones but the centre are placed on a single boom.

9.6.4 OCT Surround

NOT YET WRITTEN

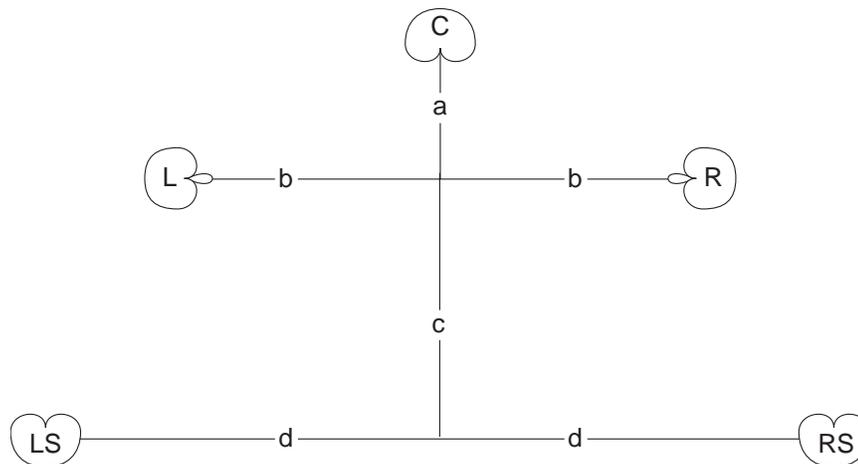


Figure 9.157: OCT Surround: $a = 8$ cm, $b = 40 - 90$ cm, $c = 40$ cm, $d = 10 - 100$ cm.

9.6.5 OCT Front System + IRT Cross

NOT YET WRITTEN

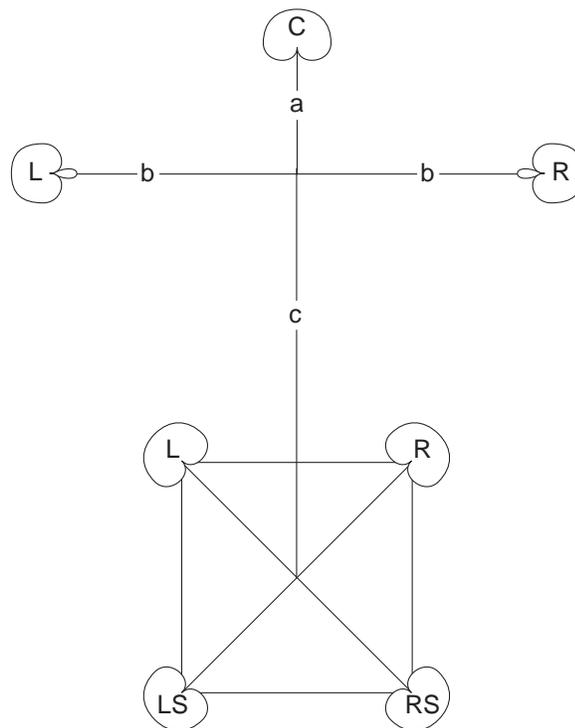


Figure 9.158: OCT Front + IRT Cross: $a = 8$ cm, $b = 40 - 90$ cm, $c \approx 100$ cm, cross side = $20 - 25$ cm

9.6.6 OCT Front System + Hamasaki Square

NOT YET WRITTEN

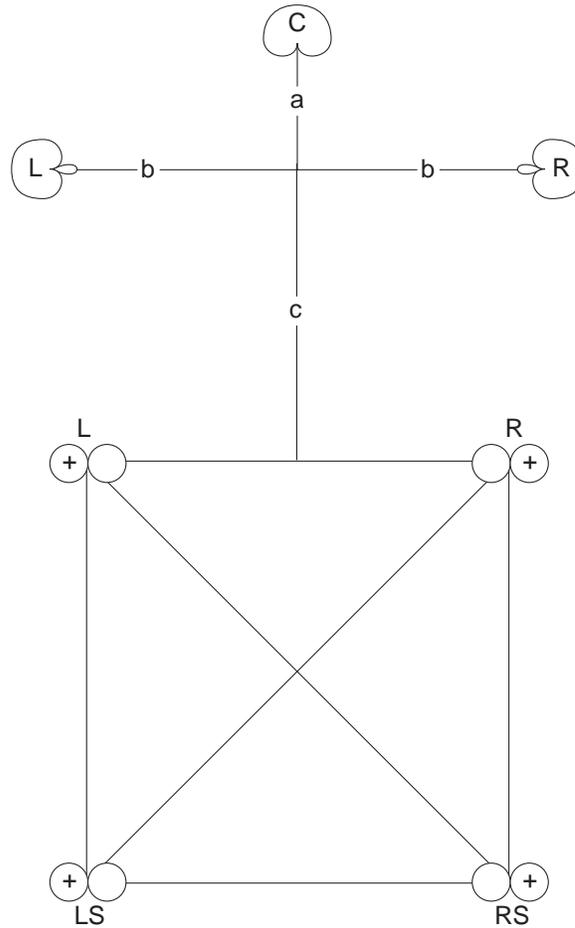


Figure 9.159: OCT Front + Hamasaki Square: $a = 8$ cm, $b = 40 - 90$ cm, $c \approx 100$ cm, cross side = 2 - 3 m

9.6.7 Klepko Technique

John Klepko developed an interesting surround microphone technique as part of his doctoral work at McGill University [?]. He suggests that the surround loudspeakers, due to their inherent low interaural crosstalk, can be considered as a very large pair of headphones and could therefore be used

to reproduce a binaural signal. Consequently, he suggests that the surround microphones be replaced by a dummy head placed approximately XXX metres behind the front microphone array of three HYPERCARDIOIDS?.

This is an interesting technique in its ability to create a very coherent soundfield that envelops the listener, both left to right and front to back. Part of the reason for this is that the binaural signal played through the surround channels is able to deliver a frontal image in many listeners, even in the absence of signals in the front loudspeakers.

Of course, there are some problems with this configuration, in particular caused by the reliance on the binaural signal. In order to provide the desired spatial cues, the listener must be located exactly in the sweet spot between the surround loudspeakers. Movements away from this position will cause the binaural cues to collapse. In addition, there are timbral issues to consider – dummy heads typically have very strong timbral effects on your signal which are frequently not desirable.

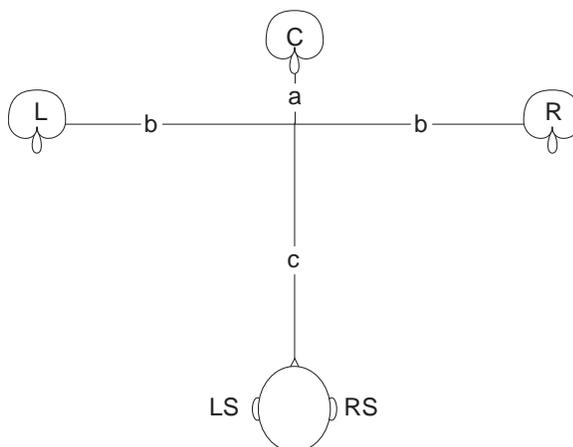


Figure 9.160: Klepko Tree: GET THE DIMENSIONS

9.6.8 Corey / Martin Tree

NOT YET WRITTEN

9.6.9 Michael Williams

NOT YET WRITTEN

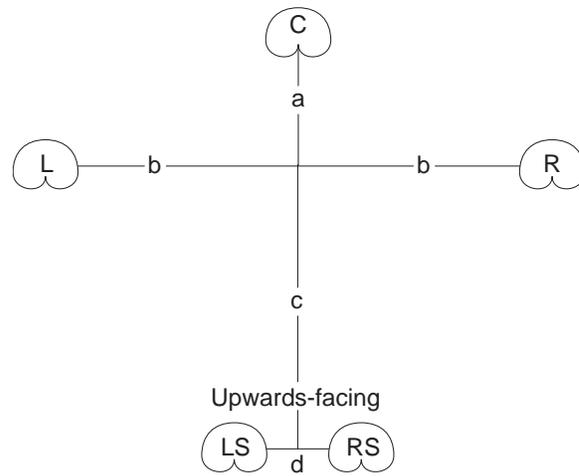


Figure 9.161: Corey / Martin Tree: $a = 0 - 15$ cm, $b = 60 - 80$ cm, $c = 60 - 90$ cm, $d = 30$ cm. L, C and R are subcardioids; LS and RS cardioids aimed towards ceiling.

9.6.10 Suggested Reading List

9.7 Introduction to Time Code

9.7.1 Introduction

When you tape a show on television, you don't need to worry about how the sound and the video both wind up on your tape – or how they stay synchronized once they're both on there – but if you're working in the television or film industry, this is a very big issue. Usually, in film recording, the audio gets recorded on a completely different machine than the film. In video editing, you need a bunch of video playback and recording units all “synchronized” so that some central machine knows where on the tapes all the machines are – down to the individual frame. How is this done? Well, the location of the tape – the “time” the tape is at – is recorded on the tape itself, so that no matter which machine you play it on, you can tell where you are in the show – to the nearest frame. This is done with what's called *time code* of which there are various species to talk about...

9.7.2 Frame Rates

Before discussing time code formats, we have to talk about the different frame rates of the various media which use time code. In spite of the original name “moving pictures,” all video and film that we see doesn't really use pictures that move. We are shown a number of still pictures in quick succession that fool our eyes into thinking that we're seeing movement. The rate at which these pictures (or “frames”) are shown is called the *frame rate*. This rate varies depending on what you're watching and where you live.

Film

The film industry uses a standard frame rate of 24 fps (Frames per Second). This is just about the slowest frame rate you can get away with and still have what appears to be smooth motion. (though some internet video-streaming people might argue that point in their advertising brochures...) The only exception to this rule is the IMAX format which runs at twice this rate – 48 fps.

Television – North America, Black and White

a.k.a NTSC (National Television System Committee)

(although some people think it stands for “Never The Same Colour” ...)

In North America (or at least most of it...) the AC power that runs our curling irons and golf ball washers (and televisions...) has a fundamental frequency of 60 Hz. As a result, the people that invented black and white television, thinking that it would be smart to have a frame rate that was compatible with this frequency, set the rate to 30 fps.

I should mention a little bit about the way televisions work. They're slightly different from films in that a film shows you 24 actual pictures on the screen each second. A television has a single gun that creates a line of varying intensity on the screen. This gun traces a bunch of lines, one on top of each other on the screen, that, when seen together, look like a single picture. There are 525 lines in each frame – but each frame is divided into two “fields” – the TV shows you the odd-numbered lines, and then goes back to do the even-numbered ones. This system (of alternating between the odd and even number lines using two fields) is called *interlacing*. The important thing to remember is that since we have 30 fps and 2 fields per frame, this means that there are 60 fields per second.

Television – North America, Colour (NTSC)

It would be nice if B and W and colour could run at the same rate – this would make life simple... unfortunately, however, this can't be done because colour TV requires that a little more information be sent to your unit than B and W does. They're close – but not exactly the same. Colour NTSC runs at a frame rate of 29.97 fps. This will cause us some grief later.

Television – Europe, Colour and Black and White

a.k.a PAL (Phase Alternating Line)

and SECAM (Sequential Couleur Avec Memoire or Sequential Colour with Memory)

The Europeans have got this figured out far more intelligently than the rest of us. Both PAL and SECAM, whether you're watching colour OR B and W, run at 25 fps. (There is one exception to this rule called PAL M – but we'll leave that out...) The difference between PAL and SECAM lies in the methods by which the colour information is broadcast – but we'll ignore that too. Both systems are interlaces – so you're seeing 50 fields per second.

It's interesting that this rate is so close to the film rate of 24 fps. In fact, it's close enough that some people, when they're showing movies on television in PAL or SECAM just play them at 25 fps to simplify things. The audio has to be pitch shifted a bit, but the difference is close enough

that most people don't notice the artifacts.

9.7.3 How to Count in Time Code

Time code assumes a couple of things. Firstly, it assumes that there are 24 hours in a day, and that your film or television program won't last longer than a day. So, you can count up to 24 hours (almost...) before you start counting at 0 again. The time is divided in a similar way to the system we use to tell time, with the exception of the sub-division of the seconds. This is divided in frames instead of fractions, since that's the way the film or video unit works.

So, a typical time code *address* will look like this:

01:23:35:19

or

HH:MM:SS:FF

Which means 01 hours, 23 minutes, 35 seconds and 19 frames. The number of frames in each second depends on the time code format, which in turn, depends on the medium for which it's being used, which is discussed in the next section.

9.7.4 SMPTE/EBU Time Code Formats

Given the number of different "standard" frame rates, there must be a number of different time code formats correspondingly. Each one is designed to match its corresponding rate, however, there are two that stand out as being the most widely used. These formats have been standardized by two organizations, **SMPTE** (the Society of Motion Picture and Television Engineers – pronounced SIM-tee) and the **EBU** (the European Broadcasting Union – pronounced ee-be-you)

24 fps

As you'll probably guess, this time code is used in the film industry. The system counts 24 frames in each second, as follows:

00:00:00:00

00:00:00:01

00:00:00:02

.

.

.

00:00:00:22

00:00:00:23

00:00:01:00

00:00:01:01

and so on. Notice that the first frame is labelled “00” therefore we count up to 23 frames and then skip to second 01, frame 00. As was previously mentioned, there are a maximum of 24 hours in a day, therefore we roll back around to 0 as follows:

23:59:59:21

23:59:59:22

23:59:59:23

00:00:00:00

00:00:00:01

and so on.

Each of these addresses corresponds to a frame in the film, so while the film is rolling, our time code reader will display a new address every 24th of a second.

In theory, if our film started exactly at midnight, and it ran for 24 hours, then the time code would display the time of day, all day.

25 fps

This time code is used with PAL and SECAM television formats. The system counts 25 frames in each second, as follows:

00:00:00:00

00:00:00:01

00:00:00:02

.

.

.

00:00:00:23

00:00:00:24

00:00:01:00

00:00:01:01

and so on. Again, the first frame is labelled “00” but we count to 24 frames before skipping to second 01, frame 00. The roll around to 0 after 24 hours happens the same as the 24 fps counterpart, with the obvious exception that we get to 23:59:59:24 before rolling back to 00:00:00:00.

Again, each of these addresses corresponds to a frame in the video, so while the program is playing, our time code reader will display a new address every 25th of a second.

And again, the time code exactly corresponds to the time of day.

30 fps “non-drop”

This time code is designed to be used in black and white NTSC television formats, however it’s rarely used these days except in the occasional commercial. The system counts 30 frames in each second corresponding with the frame rate of the format, as follows:

```
00:00:00:00
00:00:00:01
00:00:00:02
.
.
.
00:00:00:28
00:00:00:29
00:00:01:00
00:00:01:01
and so on.
```

The question you’re probably asking is ‘why is it called “non-drop?”’ ‘but that question is probably best answered by explaining one more format called ‘30 fps “drop-frame” ’

30 fps “drop frame” (aka 29.97 fps “drop frame”)

This time code is the one that’s used most frequently in North America – and the one that takes the most dancing around to understand. Remember that NTSC started as a 30 fps format until they invented colour – then things slowed down to a frame rate of 29.97 fps when they upgraded. The problem is, how do you count up to 29.97 every second? Well, obviously, you don’t. So, what they came up with was a system where they’d keep counting with the old 30 fps system, pretending that there were 30 frames every second. This means that, at the end of the day, your clock is wrong – you’re counting too slowly (0.03 fps too slowly, to be precise...), so there’s some time left over to make up for at the end of the day – in fact, when the clock on the wall strikes midnight, 24 hours after you started counting, you’re going to have 2592 frames left to count.

The committee that was deciding on this format elected to leave out these frames in something approaching a systematic fashion. Rather than leave out the 2592 frames at the end of the day, they decided to distribute

the omissions evenly throughout the day (that way, television programs of less than 24 hours would come close to making sense...) So, this means that we have to omit 108 frames every hour. The way they decided to do this was to omit 2 frames every minute on the minute. The problem with this idea is that omitting 2 frames a minute means losing too many frames – so they had to add a couple back in. This is accomplished by NOT omitting the two frames if the minute is at 0, 10, 20, 30, 40 or 50.

So, now the system counts like this:

```
00:00:00:00
00:00:00:01
00:00:00:02
.
.
.
00:00:58:29
00:00:59:00
00:00:59:01
.
.
.
00:00:59:28
00:00:59:29
00:01:00:02 (Notice that we've skipped two frame numbers)
00:01:00:03
and so on... but...
00:09:59:00
00:09:59:01
00:09:59:02
.
.
.
00:09:59:28
00:09:59:29
00:10:00:00 (Notice that we did not skip two frame numbers)
00:10:00:01
```

It's important to keep in mind that we're not actually leaving out frames in the video – we're just skipping numbers... like when you were counting to 100 while playing hide and seek as a kid... 21, 22, 25, 26, 30... You didn't count any faster, but you started looking for your opponents sooner.

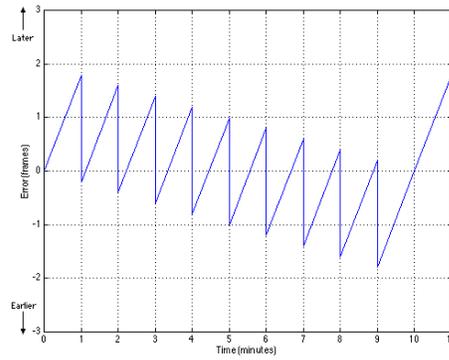


Figure 9.162: The accumulated error in drop frame time code. At time 0, the time code reads 00:00:00:00 and is therefore correct, so there is no error. As time increase up to the 1-minute mark, the time code is increasingly incorrect, displaying a time that is increasingly later than the actual time. At the 1-minute mark, two frames are dropped from the count, making the display slightly earlier than the actual time. This trend is repeated until the 9th minute, where the error becomes increasingly less early until the display shows the correct time at the 10-minute mark.

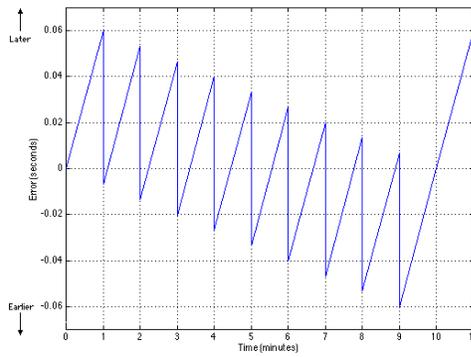


Figure 9.163: This shows exactly the same information as the plot in Figure 1, showing the error in seconds rather than frames.

Since we're dropping out the numbers for the occasional frame, we call this system "drop-frame," hence the designation "non-drop" when we don't leave things out.

Look back to the explanation of 30 fps "non-drop" and you'll see that we said that pretty well the only place it's used nowadays is for television commercials. This is because the first frame to get dropped in drop-frame happens 1 minute after the time code starts running. Most TV commercials don't go for longer than 30 seconds, so for that amount of time, drop-frame and non-drop are identical. There is an error accumulated in the time code relative to the time of day, but the error wouldn't be fixed until the clock read 1 minute anyway... we'll never get there on a 30-second commercial.

29.97 fps "non-drop"

This is a bit of an odd one that's the result of one small, but very important corner of the video industry – commercials. Commercials for colour NTSC run at a frame rate of 29.97 frames per second, just like everything else on TV – but they typically only last 30 seconds or less. Since the timecode never reaches the 1 minute mark, there's no need to drop any frames – so you have a frame rate of 29.97 fps, but it's non-drop.

9.7.5 Time Code Encoding

There are two methods by which the time code signal is recorded onto a piece of videotape, which in turn determine their names. In order to get a handle on why they're called what they're called and why they have different advantages and disadvantages, we have to look a bit at how any signal is recorded onto videotape in the first place.

We said that your average NTSC television is receiving 525 lines of information every second – each of these lines is comprised of light and dark areas (or different colours, if your TV is newer than mine...) This means that a LOT of information must get recorded very quickly on the tape – essentially, we require a high bandwidth. This means that we have to run the tape very quickly across the head of the video recorder in order to get everything on the tape. Well, for a long time, we had to do without recorded video because no one could figure out a way of getting the tape-to-head speed high enough without requiring more tape than even a small country could afford... Then one day, someone said "What if we move the head quickly instead of the tape?" BINGO! They put a head on a drum, tilted the drum on an angle relative to the tape and spun the drum while they

moved the tape. This now meant that the head was going past the tape quickly, making diagonal streaks across it while the tape just crept along at a very slow speed indeed.

(Sidebar: the guy in this story was named Alexander M. Poniatoff who was given some cash to figure all this out by another guy named Bing Crosby back in the 40's... see Bing wanted to tape his shows and then sit at home relaxing with the family while he watched himself on TV... Now, look at Alexander's initials and tack on an "excellence" and you get AMPEX.)

Back to videotape design... The system with the rotating heads is still used today in your VCR (except that we call it "helical scanning" to sound fancy) and this head on the drum is used to record and play both the video information as well as the "hi-fi" audio (if you own a hi-fi VHS machine). The tape is moving just fast enough to put another head in there which isn't on the drum – it records the mono audio on your VCR at home, but it could be used for other low-bandwidth signals. It can't handle high-bandwidth material because the tape is moving so slowly relative to the head that physics just doesn't allow it.

The important thing to remember from all this is that there are two ways of putting the signal (the time code, in our case) on the tape – *longitudinally* with the stationary head (because it runs parallel with the tape direction) and *vertically* with the rotating head (well, it's actually not completely vertical, but it's getting close, depending on how much the drum is tilted).

Keep in mind that what we're discussing here now is how the numbers for the hours, minutes, seconds and frames actually get stored on a piece of tape or transmitted across a wire.

Longitudinal Time Code

aka LTC (say each letter)

LTC was developed by the SMPTE which explains the fact that it is occasionally called "SMPTE Time Code" in older books). It is a continuous serial signal with a clock rate of 2400 bits per second and can be recorded on or transmitted through a system with a bandwidth as low as 10 kHz.

The system uses something called bi-phase modulation to encode the numbers into what essentially becomes an audio signal. The numbers denoting the time code address are converted into binary code and then that is converted into the bi-phase mark. "What's a bi-phase mark?" I hear you cry...

There are a number of ways we can transmit 1's and 0's on a wire using voltages. The simplest method would be to assign a voltage to each and send

the appropriate voltage at the appropriate time (if I send a 0 V signal – that means 0, but if it’s 1 V, that means 1...) This is a nice system until someone uses a transmission cable that inverts the polarity of the system, then the voltages become 0 V and –1 V – which could be confused for a high and low, then the whole system is screwed up. In order to make the transmission (and recording) system a little more idiot-proof, we use a different system. We’ll keep a high and low voltage, but alternate between them at a pre-determined rate (think square wave). Now the rule is, every transition of the square wave is a new “bit” or number – either a 1 or a 0. Each bit is divided into two “cells” – if there is a voltage transition between cells, then the value of the bit is a 1, if there is no transition between cells, then the bit is a 0 (see the following diagram).

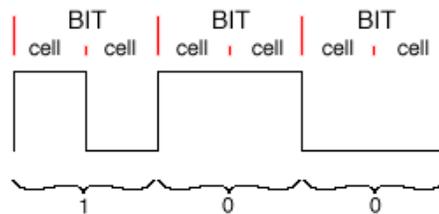


Figure 9.164: A bi-phase mark used to designate a 0 from a 1 by the division of each bit into two cells. If the cells are the same, then the bit is a 0, if they are different, then the bit is a 1.

This allows us to invert the polarity and make the voltage of the signal independent of the value of the bit, essentially, making the signal more robust. There is no DC content in the signal (so we don’t have to worry about the signal going through DC-blocking capacitors) and it’s self-clocking (that is to say, if we build a smart receiving device, it can figure out how fast the signal is coming at it). In addition, if we make the receiver a little tolerant, we can change the rate of the signal (the tape speed, when shuttling, for example) and still derive a time code address from it.

Each time code address required one *word* to define all of its information. This word is comprised of 80 bits (which, at 30 fps means 2400 bits per second). All 80 bits are not required for telling the machines what frame we’re on – in fact only 26 bits are used for this. The rest of the word is divided up as follows:

There are a number of texts which discuss exactly how these are laid out in the signal – we won’t get into that. But we should take a quick glance at what the additional parts of the TC word are used for.

| Information | Number of Bits |
|--------------------|----------------|
| Time Address | 26 |
| User Information | 32 |
| Sync Word | 16 |
| Status Information | 5 |
| Unassigned | 1 |

Table 9.8: INSERT CAPTION HERE

Time Address – 26 bits

The time address information uses 4 bits to encode each of the decimal numbers in the time code address. That is to say, for example, to encode the designation of “12 frames,” the numbers 1 (0001) and 2 (0010) are used sequentially instead of encoding the number 12 as a binary number. This means, in theory, that we require 32 bits to store or transmit the time code address, four bits each for 8 digits (HH:MM:SS:FF). This is not really the case, however, since we don’t count all the way up to 9 with all of the digits. In fact, we only require 2 bits each for the tens of hours (because it never goes past “2”) and tens of frames, and 3 bits for each of the tens of minutes and tens of seconds. This frees up 6 bits which are used for status information, meaning 26 bits are used for the time address information.

User Information – 32 bits

There are 32 bits in the time code word reserved for storing what’s called “user information.” These 32 bits are divided into eight 4-bit words which are generally devoted to recording or transmitting things like reel numbers, or the date on which the recording was made – things which don’t change while the time code rolls past. There are two options that are not used as frequently which are:

- encoding ASCII characters to send secret messages (like song lyrics? credits?... be creative...) This would require 8-bit bytes instead of 4-bit words, but we’ll come back to that in the Status Information

- since we have 32 bits to work with, we could lay down a second time code address... though I don’t know what for offhand.

Sync Word – 16 bits

In order for the receiving device to have any clue as to what's going on, it has to know when the time code word starts and stops, otherwise, its guess is as good as anybody's. The way the beginning of the word is marked is by putting in a string of bits that cannot happen anywhere else, irrespective of the data that's being transmitted. This string of data (0011111111111101 to be precise) tells the receiver two things – firstly, where the subsequent word starts (the string is put at the *end* of each word) and whether the signal is going forwards or backwards. (Notice that the string is two 0's, twelve 1's, a 0 and a 1. The receiver sees the twelve 1's, and looks to see if that string was preceded by two 0's or a 1 and a 0... that's how it figures it out.)

Status Information – 5 bits

The five status bits are comprised of “flags” which provide the receiving device with some information about the signal coming in. I'll omit explanations for some of these.

Drop Frame Flag

If this bit is a 1, then the time code is drop-frame. If it's a 0, then it's non-drop.

Colour Frame Flag**Bi-Phase Correction**

This is a bit which will change between 0 and 1 depending on the content of the remainder of the word. It's just there to ensure that the transition at the beginning of every word goes in the same direction.

User Flags (2)

We said earlier that the User Information is 32 bits divided into eight 4-bit words. However, if it's being used for ASCII information, then it has to be divided differently into four 8-bit bytes. These User Flags in the Status Information section tell the receiver which system you're using in the User Information.

Unassigned

This is a bit that has no determined designation.

Other Information

There's a couple of little things to know about LTC what might come in useful some day.

Firstly, a time code reader should be able to read time code at speeds between 1/40th of the normal rate and 40 times the normal rate. This would put the maximum bandwidth up to 96000 bits per second (2400 * 40)

Secondly, remember that LTC is recorded using the stationary head on the video recorder. This means a couple of things:

1) as you slow down, you get more inaccurate. If you're stopped (or paused) you don't read a signal – therefore no time code.

2) as you slow down, the signal gets lower in amplitude (because the voltage produced by the read head is proportional to the change in magnetism across its gap). Therefore, the slower the tape, the more difficult to read.

Lastly, if you're using LTC to lock two machines together, keep in mind that the “slave” machine is not locked to every single frame to the closest 80th of a frame to the “master.” In fact, more likely than not, the slave is running on its own internal clock (or some external “house sync” signal) and checking the incoming time code every once and a while to make sure that things are on schedule. (this explains the “forward/backward” info stored in the Sync Word) This can explain why, when you hit STOP on your master machine, it might take a couple of moments for the slave device to realize that it's time to hit the brakes...

Vertical Interval Time Code

aka VITC (say VIT–see)

There are two major problems associated with longitudinal time code. Firstly, it is not frame-accurate (not even close...) when you're moving at very slow speeds, or stopped, as would be the case when shuttling between individual frames (remember that it's possible to view a stopped frame because the head keeps rotating and therefore moving relative to the tape). Secondly, the LTC takes up too much physical space on the videotape itself.

The solution to both of these problems was presented by Sony back in 1979 in the form of Vertical Interval Time Code or VITC. This information is stored in what is known as the “vertical blanking interval” in each field (we won't get into what that means... but it's useful to know that there is one in each *field* rather than each *frame*, increasing our resolution over LTC by a factor of 2. It does, however, mean that an extra flag will be required

to indicate which field we're on.

Since the VITC signal is physically located in the vertical tracks on the tape (because it's recorded by the rotating head instead of the stationary one) it has a number of characteristics which make it differ from LTC.

- Firstly, it cannot be played backwards. Although the video can be played backwards, each field of each frame (and therefore each word of the VITC code) is being shown forwards (in reverse order).

- Secondly, we can trust that the phase of the signal is reliable, so a different binary encoding system is used called Non-Return to Zero or NRZ. In this system, a value of 80 IRE (an IRE is a unit equal to 1/140th of the peak-peak amplitude of a video signal, usually 1 IRE = 7.14 mV, therefore 140 IRE = 1 V) is equal to a binary "1" and 0 IRE is a 0.

- Thirdly, as was previously mentioned, the VITC signal contains the address of a field rather than a frame.

- Finally, we don't need to indicate when the word starts (as is done by the sync bits in LTC) since every field has only 1 frame associated with it. since both the field and the word are read simultaneously (they cannot be read separately) we don't need to know when the word starts... it's obvious.

Unlike LTC, there are a total of 90 bits in the VITC word, with the following assignments:

| Information | Number of Bits |
|-------------------------------|----------------|
| Time Address | 26 |
| User Information | 32 |
| Sync Groups | 18 |
| Cyclic Redundancy Code | 8 |
| Status Information | 5 |
| Unassigned (Field Mark Flag?) | 1 |

Table 9.9: INSERT CAPTION HERE

In many respects, these bits are the same as their LTC counterparts, but we'll go through them again anyway.

Time Address – 26 bits

This is exactly the same as in LTC.

User Information – 32 bits

This is exactly the same as in LTC.

Sync Groups – 18 bits

Instead of a single big string of binary bits stuck in one part of the word, the Sync Groups in VITC are fit into the information throughout the word. There is a binary “10” located at the beginning of every 10 bits. This information is used to synchronize the word only. It doesn’t need to indicate either the beginning of the word (since this is obvious due to the fact that there is one word “locked” to each field) nor does it need to indicate direction (we’re always going forwards – even when we’re going backwards...)

Cyclic Redundancy Code – 8 bits

This is a method of error detection (not error correction).

Status Information – 5 bits

There are five status information flags, just as with LTC. All are identical to their LTC counterparts with one exception, the Field Mark. Since we don’t worry about phase in the VITC signal, we can leave out the bi-phase correction.

Field Mark Flag

This flag indicates the field of the address.

One extra thing:

There is one disadvantage to VITC that makes use of LTC more versatile. Since VITC is buried in the video information recorded by the helical head, it *must* be recorded simultaneously with the video. It cannot be recorded afterwards. Since LTC is located on a different section of the tape, it can be *striped* (recorded) onto the tape later without disturbing the video information.

9.7.6 Annex – Time Code Bit Assignments**Longitudinal Time Code**

See the information in Table 9.10.

| Frame Number | Bit Assignment |
|--------------|----------------|
| 0 | Frame Units |
| 1 | Frame Units |

| Frame Number | Bit Assignment |
|--------------|------------------|
| 2 | Frame Units |
| 3 | Frame Units |
| 4 | User Bit |
| 5 | User Bit |
| 6 | User Bit |
| 7 | User Bit |
| 8 | Frame Tens |
| 9 | Frame Tens |
| 10 | Drop Frame Bit |
| 11 | Colour Frame Bit |
| 12 | User Bit |
| 13 | User Bit |
| 14 | User Bit |
| 15 | User Bit |
| 16 | Second Units |
| 17 | Second Units |
| 18 | Second Units |
| 19 | Second Units |
| 20 | User Bit |
| 21 | User Bit |
| 22 | User Bit |
| 23 | User Bit |
| 24 | Second Tens |
| 25 | Second Tens |
| 26 | Second Tens |
| 27 | Bi-Phase Bit |
| 28 | User Bit |
| 29 | User Bit |
| 30 | User Bit |
| 31 | User Bit |
| 32 | Minute Units |
| 33 | Minute Units |
| 34 | Minute Units |
| 35 | Minute Units |
| 36 | User Bit |
| 37 | User Bit |
| 38 | User Bit |
| 39 | User Bit |
| 40 | Minute Tens |

| Frame Number | Bit Assignment |
|--------------|----------------|
| 41 | Minute Tens |
| 42 | Minute Tens |
| 43 | User Flag Bit |
| 44 | User Bit |
| 45 | User Bit |
| 46 | User Bit |
| 47 | User Bit |
| 48 | Hour Units |
| 49 | Hour Units |
| 50 | Hour Units |
| 51 | Hour Units |
| 52 | User Bit |
| 53 | User Bit |
| 54 | User Bit |
| 55 | User Bit |
| 56 | Hour Tens |
| 57 | Hour Tens |
| 58 | Unassigned |
| 59 | User Flag Bit |
| 60 | User Bit |
| 61 | User Bit |
| 62 | User Bit |
| 63 | User Bit |
| 64 | Sync Word |
| 65 | Sync Word |
| 66 | Sync Word |
| 67 | Sync Word |
| 68 | Sync Word |
| 69 | Sync Word |
| 70 | Sync Word |
| 71 | Sync Word |
| 72 | Sync Word |
| 73 | Sync Word |
| 74 | Sync Word |
| 75 | Sync Word |
| 76 | Sync Word |
| 77 | Sync Word |
| 78 | Sync Word |

| Frame Number | Bit Assignment |
|--------------|----------------|
| 79 | Sync Word |

Table 9.10: A bit-by-bit listing of the information contained in Longitudinal Time Code.

Vertical Interval Time Code

See the information in Table 9.11.

| Frame Number | Bit Assignment |
|--------------|-------------------|
| 0 | Sync Bit (1) |
| 1 | Sync Bit (0) |
| 2 | Frame Units |
| 3 | Frame Units |
| 4 | Frame Units |
| 5 | Frame Units |
| 6 | User Bit |
| 7 | User Bit |
| 8 | User Bit |
| 9 | User Bit |
| 10 | Sync Bit (1) |
| 11 | Sync Bit (0) |
| 12 | Frame Tens |
| 13 | Frame Tens |
| 14 | Drop Frame Flag |
| 15 | Colour Frame Flag |
| 16 | User Bit |
| 17 | User Bit |
| 18 | User Bit |
| 19 | User Bit |
| 20 | Sync Bit (1) |
| 21 | Sync Bit (0) |
| 22 | Second Units |
| 23 | Second Units |
| 24 | Second Units |
| 25 | Second Units |
| 26 | User Bit |
| 27 | User Bit |
| 28 | User Bit |

| Frame Number | Bit Assignment |
|--------------|-------------------|
| 29 | User Bit |
| 30 | Sync Bit (1) |
| 31 | Sync Bit (0) |
| 32 | Second Tens |
| 33 | Second Tens |
| 34 | Second Tens |
| 35 | Field Mark |
| 36 | User Bit |
| 37 | User Bit |
| 38 | User Bit |
| 39 | User Bit |
| 40 | Sync Bit (1) |
| 41 | Sync Bit (0) |
| 42 | Minute Units |
| 43 | Minute Units |
| 44 | Minute Units |
| 45 | Minute Units |
| 46 | User Bit |
| 47 | User Bit |
| 48 | User Bit |
| 49 | User Bit |
| 50 | Sync Bit (1) |
| 51 | Sync Bit (0) |
| 52 | Minute Tens |
| 53 | Minute Tens |
| 54 | Minute Tens |
| 55 | Binary Group Flag |
| 56 | User Bit |
| 57 | User Bit |
| 58 | User Bit |
| 59 | User Bit |
| 60 | Sync Bit (1) |
| 61 | Sync Bit (0) |
| 62 | Hour Units |
| 63 | Hour Units |
| 64 | Hour Units |
| 65 | Hour Units |
| 66 | User Bit |
| 67 | User Bit |

| Frame Number | Bit Assignment |
|--------------|----------------|
| 68 | User Bit |
| 69 | User Bit |
| 70 | Sync Bit (1) |
| 71 | Sync Bit (0) |
| 72 | Hour Tens |
| 73 | Hour Tens |
| 74 | Hour Tens |
| 75 | Unassigned |
| 76 | User Bit |
| 77 | User Bit |
| 78 | User Bit |
| 79 | User Bit |
| 80 | Sync Bit (1) |
| 81 | Sync Bit (0) |
| 82 | CRC Bit |
| 83 | CRC Bit |
| 84 | CRC Bit |
| 85 | CRC Bit |
| 86 | CRC Bit |
| 87 | CRC Bit |
| 88 | CRC Bit |
| 89 | CRC Bit |

Table 9.11: A bit-by-bit listing of the information contained in Vertical Interval Time Code.

9.7.7 Suggested Reading List

Chapter 10

Conclusions and Opinions

A lot of what I've presented in this book relates to the theory of recording and very little to do with aesthetics. Please don't misinterpret this balance of information – I'm not one of those people who believes that a recording should be done using math instead of ears.

One of my favourite movies is *My Dinner with André*. If you haven't seen this film, you should. It takes about two hours to watch, and in those two hours, you see two people having dinner and a very interesting conversation. One of my other favourite movies is *Cinema Paradiso* in which, in two hours, approximately 40 or 50 years pass. Let's look at the difference between these two concepts.

In the first, the idea is to present time in real time – watching the movie is almost the same as sitting at the table with the two people having the conversation. If there's a break in the conversation because the waiter brings the dessert, then you have to wait for the conversation to continue.

In the second, time is compressed. In order to tell the story, we need to see a long stretch of time packed into two hours. Waiting 10 seconds for a waiter to bring dessert would not be possible in this movie because there isn't time to wait.

The first movie is a lot like real life. The second movie is not, although it shows the story of real lives. Most movies are like the latter – essentially, the story is better than real life, because it leaves out the boring bits.

My life is full of normal-looking people doing mundane things, there's very little conflict, and therefore little need for resolution, not many car chases, gun fights or martial arts. There are no robots from the future and none of the people I know are trying to overthrow governments or large corporations (assuming that these are different things...). This is real life.

All of the things that are in real life are the things that aren't in movies.

Most people don't go to movies to see real life – they go for the entertainment value. Robots from the future that look like beautiful people, who, using their talents in car chases, gun fights and martial arts, are trying to overthrow governments and large corporations.

Basically speaking, movies are better than life.

Recordings are the same. Listen to any commercial recording of an orchestra and you'll notice that there is lots of direct sound from the violins but they also have lots of reverb on them. You're simultaneously next to and far from the instruments in the best hall in the world. This doesn't happen in real life. Britney Spears (or whoever is popular when you're reading this) can't really sing like that without hundreds of takes and a lot of processing gear. Commercial recordings are better than life. Or at least, that's usually the goal.

So, the moral of the story is that your goal on every recording you do is to make it sound as good as possible. Of course, I can't define "good" – that's up to your expectations (which is in turn dependent on your taste and style) and your desired audience's expectations.

So, to create that recording, whether it's classical, jazz, pop or sound effects in mono, stereo or multichannel, you do whatever you can to make it sound better.

In order to do this, however, you need to know what direction to head in. How do you move the microphone to make the sound wider, narrower, drier. What will adding reverb, chorus, flanging or delay do to the signal? You need to know in advance what direction you want to head in based on aesthetics, and you need the theory to know how to achieve that goal.

10.0.8 Suggested Reading List

Chapter 11

Reference Information

11.1 ISO Frequency Centres

This list, shown in Table 11.1 is a standard list of frequencies that can be used for various things such as the centre frequencies for graphic equalizers. The numbers are used as multipliers for the decades within the audio range. For example, if you were building an equalizer with half-octave frequency centres, the filters would have centre frequencies of 25 Hz, 40 Hz, 63 Hz, 100 Hz, 160 Hz, 250 Hz, 400 Hz, 630 Hz, 1000 Hz, 1600 Hz, 2500 Hz, 4000 Hz, 6300 Hz, 10000 Hz, and 16000 Hz.

Use the numbers listed in the table as multipliers. For example, if you were building a 1-Octave equalizer, then its frequency centres would use the right-most column of numbers. In Hz, these would be 10 Hz, 12.5 Hz, 16 Hz, 20 Hz, 25 Hz, 31.5 Hz, 40 Hz, 50 Hz, 63 Hz, 80 Hz, 100 Hz, 125 Hz, 160 Hz ... 630 Hz, 800 Hz, 1 kHz, 1.25 kHz, 1.6 kHz ... 6.3 kHz, 8 kHz, 10 kHz, 12.5 kHz and so on.

| $\frac{1}{12}^{th}$ Oct | $\frac{1}{6}^{th}$ Oct | $\frac{1}{3}^{th}$ Oct | $\frac{1}{2}^{th}$ Oct | $\frac{2}{3}^{th}$ Oct | 1 Oct |
|-------------------------|------------------------|------------------------|------------------------|------------------------|-------|
| 1 | 1 | 1 | 1 | 1 | 1 |
| 1.06 | | | | | |
| 1.12 | 1.12 | | | | |
| 1.18 | | | | | |
| 1.25 | 1.25 | 1.25 | | | |
| 1.32 | | | | | |
| 1.4 | 1.4 | | 1.4 | | |
| 1.5 | | | | | |
| 1.6 | 1.6 | 1.6 | | 1.6 | |
| 1.7 | | | | | |
| 1.8 | 1.8 | | | | |
| 1.9 | | | | | |
| 2 | 2 | 2 | 2 | | 2 |
| 2.12 | | | | | |
| 2.24 | 2.24 | | | | |
| 2.36 | | | | | |
| 2.5 | 2.5 | 2.5 | | 2.5 | |
| 2.65 | | | | | |
| 2.8 | 2.8 | | 2.8 | | |
| 3 | | | | | |
| 3.15 | 3.15 | 3.15 | | | |
| 3.35 | | | | | |
| 3.55 | 3.55 | | | | |
| 3.75 | | | | | |
| 4 | 4 | 4 | 4 | 4 | 4 |
| 4.25 | | | | | |
| 4.5 | 4.5 | | | | |
| 4.75 | | | | | |
| 5 | 5 | 5 | | | |
| 5.3 | | | | | |
| 5.6 | 5.6 | | 5.6 | | |
| 6 | | | | | |
| 6.3 | 6.3 | 6.3 | | 6.3 | |
| 6.7 | | | | | |
| 7.1 | 7.1 | | | | |
| 7.5 | | | | | |
| 8 | 8 | 8 | 8 | | 8 |
| 8.5 | | | | | |
| 9 | 9 | | | | |
| 9.5 | | | | | |

Table 11.1: ISO Frequency Centres

11.2 Scientific Notation and Prefixes

11.2.1 Scientific notation

Scientific notation is a method used to write very small or very big numbers without using up a lot of ink on the page. In order to write the number “one-million,” it is faster to write 10^6 than it is to write 1,000,000. If the number is “two-million,” then we just multiply the appropriate number – $2 * 10^6$.

The same system is used for very small numbers. 0.000,000,35 is more easily written as $3.5 * 10^{-7}$.

A good trick to remember is that the exponent of the 10 shows you how far to move the decimal place to get the number. For example, $4.3 * 10^3$ means that you just move the decimal three places to the right to make the number 4300. If the number is $8.94 * 10^{-5}$ you move the decimal five places to the left (negative exponents move to the left) making the number 0.0000894.

If you’d like to use scientific notation on your calculator, you use the button that’s marked “E” (on some calculators it’s marked “EXP”). So, $6 * 10^5$ is entered by pressing 6 E 5.

11.2.2 Prefixes

Similarly, prefixes make our lives much easier by acting as built-in multipliers in our spoken numbers. It’s too difficult to talk about very short distances in meters – so we use millimeters. A large resistance is more easily expressed in $k\Omega$ than Ω .

So, for example, if you see something written in kilometers (abbreviated km), then you know that you must multiply the number you see by 10^3 to convert to metres. $5 \text{ km} = 5 * 10^3 \text{ m} = 5,000 \text{ m}$. If the unit is milligrams (mg), then you multiply by 10^{-3} to convert to grams (g). $15 \text{ mg} = 15 * 10^{-3} \text{ g} = 0.015 \text{ g}$.

If you’re dealing with computer memory, however, you use the other column. A computer with 20 Gigabytes (GB) of hard disk space has $20 * 2^{30}$ bytes = 21,474,836,480 bytes.

Note that, if the prefix denotes a multiplier that is less than 1, it starts with a small letter (i.e. milli- or m). If the multiplier is greater than one, then a capital letter is used (i.e. mega- and M). The odd exception to this case is that of kilo- or k. I don’t know why this is different, I guess somebody has to be...

| Prefix | Abbrev. | Normal Multiplier | Computer Multiplier |
|--------|---------|-------------------|---------------------|
| atto- | a | 10^{-18} | |
| femto- | f | 10^{-15} | |
| pico- | p | 10^{-12} | |
| nano- | n | 10^{-9} | |
| micro- | μ | 10^{-6} | |
| milli- | m | 10^{-3} | |
| centi- | c | 10^{-2} | |
| deci- | d | 10^{-1} | |
| | | 1 | 1 |
| deca- | da | 10^1 | |
| hecto- | h | 10^2 | |
| kilo- | k | 10^3 | 2^{10} |
| mega- | M | 10^6 | 2^{20} |
| giga- | G | 10^9 | 2^{30} |
| tera- | T | 10^{12} | 2^{40} |
| peta- | P | 10^{15} | 2^{50} |

Table 11.2: Prefixes with corresponding multipliers, both for the real world and for computer people.

11.3 Resistor Colour Codes

If you go to the resistor store, you can buy different types of resistors in different sizes. Some fancy types will have the resistance stamped right on them, but most just have bands of different colours around them as is shown in Figure 11.1.

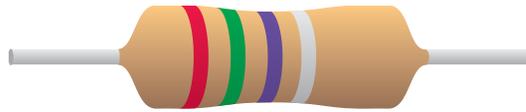


Figure 11.1: A typical moulded composition resistor.

The bands are used to indicate a couple of pieces of information about the resistor. In particular, you can find out both the resistance of the device, as well as its *tolerance*.

Firstly, let's talk about what the tolerance of a resistor is. Unfortunately, it's practically impossible to make a resistor that has *exactly* the resistance that you're looking for in your circuit. Instead of guaranteeing a particular value such as 1 k Ω , the manufacturer tells you that the resistance is approximately 1 k Ω , \pm some percentage. For example, you'll see something like 1 k Ω , $\pm 20\%$. Therefore, if you measure the resistor, you'll see that it has a value between 800 Ω and 1.2 k Ω . If you want better precision than that, you'll have to do one of two things. You'll either have to spend more money to get resistors with a tighter tolerance, or you'll have to hand-pick them yourself by measuring one by one.

So, how do we tell what the specifications of a resistor are just by looking at it? Take a look at Figure 11.2. You'll see that each band has a label attached to it. Reading from left to right (the right side of the resistor doesn't have a colour band) we have the 1st and 2nd digits, the multiplier and the tolerance. Using Table 11.3 we can figure out what the shown resistor is.

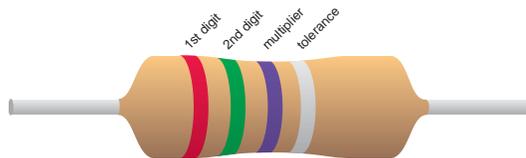


Figure 11.2: A typical moulded composition resistor showing the meanings of the coloured bands.

| Colour | Digit | Multiplier | Carbon \pm Tolerance | Film-type Tolerance |
|-----------|-------|-------------|------------------------|---------------------|
| Black | 0 | 1 | 20% | 0 |
| Brown | 1 | 10 | 1% | 1% |
| Red | 2 | 100 | 2% | 2% |
| Orange | 3 | 1,000 | 3% | |
| Yellow | 4 | 10,000 | GMV | |
| Green | 5 | 100,000 | 5% (alt) | 0.5% |
| Blue | 6 | 1,000,000 | 6% | 0.25% |
| Violet | 7 | 10,000,000 | 12.5% | 0.1% |
| Gray | 8 | 0.01 (alt) | 30% | 0.05% |
| White | 9 | 0.1 (alt) | 10% (alt) | |
| Silver | | 0.01 (pref) | 10% (pref) | 10% |
| Gold | | 0.1 (pref) | 5% (pref) | 5% |
| No colour | | | 20% | |

Table 11.3: The corresponding meanings of the colour bands on a resistor [Mandl, 1973]. (GMV stands for Guaranteed minimum value)

The colour bands on the resistor in Figure 11.2 are red, green, violet and silver. Table 11.4 shows how this combination is translated into a value.

| 1st digit | 2nd digit | Multiplier | Tolerance |
|-----------|-----------|------------|-----------|
| red | green | violet | silver |
| 2 | 5 | 10,000,000 | 10% |

Table 11.4: Using Table 11.3, the colour bands can be used to find that the resistor in Figure 11.1 has a value of 250,000,000 Ω or 250 M Ω , $\pm 10\%$.

Chapter 12

Hints and tips

This list is just a bunch of things to keep in mind when you're doing a recording. It is by no means a complete list, just a collection of things that I think about when I'm doing a recording. Also note that some of the items in the list should be taken with a grain of salt...

- Spike everything. Masking tape is your friend – if your recording is going to run over multiple sessions, put it on the floor under your microphone stands, the music stands, the instruments and the amplifiers. Digital photos help too...
- When you're not the only person around the gear, make sure that's it's damned near impossible to trip in any cables. Tape everything down. On remote recording gigs, it's always a good idea to run cables over door frames rather than across the threshold...
- If you have a temporary setup, or there is the possibility of someone tripping in a cable, leave lots of slack at both ends of the cable. I always leave a couple of loops of mic cable at the base of the mic stand, and at the other end, usually on the floor under the mic preamp. That way, if someone does trip in a cable, they just drag the cable a bit without your gear crashing to the floor.
- On a remote recording gig, make friends with the caretaker, the cleaning staff, the secretary, the stagehands... anyone that is in a position to help you out of a jam. It's always a good idea to bring along a couple of CD's to give away to people as presents to make friends quicker. I admit that this is manipulative, but it works, and it pays off.

- Put your gain as early in the signal path as is possible. (See Section 9.1.6)
- When you're recording to a digital medium, try to get the peak level as close as you can to 0 dBFS without going over. (See Section ??)
- Never use a boost in the EQ if you can use a cut instead.
- Never use EQ if the problem can be solved with a different microphone, microphone position, or microphone orientation.
- Usually, the best microphone position looks really strange. My personal feeling is that, if a microphone looks like it's in the right place, it probably isn't. Always remember that a person listening to a CD can't see where the microphones were.
- No one that buys a CD cares how tired you were at the end of the session – they expect \$20 worth of perfection. In other words, fix everything.
- If you're recording a group with a drum kit, consider your drum overheads as a wide stereo pair. Using them, listen to the pan location of all other instruments in the group (including each individual drum). Do not try to override that location by panning the instrument's own microphone to a different location. If you want a different left-right arrangement than you're getting in the overheads, move the instruments or the microphones. (If you want to get really detailed about this, you have to consider every pair of microphones as a stereo pair with the resulting imaging issues.)
- Monitor on as many playback systems as is possible/feasible. At the very least, you should know what your mix sounds like on loudspeakers and headphones, and know how the monitoring systems behave (i.e. your mix will sound closer and wider on headphones than on loudspeakers.)
- If you're doing a remote recording, get used to your control room. Set up your loudspeakers first, and play CD's that you know well while you're setting up everything else.
- Don't piss off your musicians in your efforts to find the perfect sound. Better to get a perfect performance from happy performers than a perfect recording of a bad performance. Of course, if you can get a

perfect recording of a perfect performance, then all the better. Just be aware of the feelings on the other side of the glass.

- Don't get too excited about new gear. Just because you just bought a fancy new reverb unit doesn't mean that you have to put tons of fancy new reverb on every track on the CD you're recording.
- Louder isn't always better – usually, louder is just louder.
- Don't over-compress unless you *really* mean it. This is particularly true if you're mastering. I have spoken with a number of professional mastering engineers who have told stories of sending tracks back to amateur mixing engineers because they (the mastering engineers) simply can't undo the excessive compression on the tracks. The result is that the mixing engineer has to go back and do it all again. It's not necessarily a good idea to keep a compressor (multi-band or otherwise) as a permanent fixture on the 2-mix output of your mixer...
- Don't believe everything that you read or hear. (Particularly gear reviews in magazines. Ever notice how an advertisement for that same piece of gear is very near the review? Suspicions abound...) Sometimes, the cheapest device works the best. (For example, in a very carefully calibrated very fair blind comparison between various mic pre-amp's, a friend of mine in Boston, along with a number of professional recording engineers and mic pre-manufacturers, found that the second best one they heard was a \$200 box, sounding better than other fancy tube devices for thousands of dollars. They threw in the cheap pre as a joke, and it wound up surprising everyone.)
- A laser-flat frequency response is not necessarily a good thing.
- Microphones are like paintbrushes. Use the characteristics of the microphone for the desired effect. In other words, the output of the microphone should never be considered the ultimate Truth. Instead, it is an interpretation of what is happening to the sound of the instrument in the room.
- Record your room as well as your instrument. Never forget that you're very rarely recording in an anechoic environment.
- If you're using spot microphone, use stereo pairs instead of mono spots whenever possible. The reason for this is directly related to the previous point. If you put a single mic on a guitar amp and pan the output

to the desired location, then you have the sound of the amp *as well as* the sound of the room, all clumped into one mono location in your mix. If you close-mic'ed with a stereo pair instead, your panning can be the same (with the right orientation of the microphones and the amp) but the room is spread out over the full image.

- Always trust your ears, but ask people for their opinions to see if they might be hearing something that you're not. There are times when the most inexperienced listener can identify problems that a professional (that's you...) miss for one reason or another.
- Always experiment. Don't use a technique just because it worked last time, or because you read in a magazine that someone else uses it.
- Wherever possible, keep your audio signal cables away from all other wires, particularly AC mains cables. If you have to cross wires, do so at right angles. (See Section ??)

Bibliography

- [dol, 1998] (1998). Some guidelines for producing music in 5.1 channel surround. Technical report, Dolby Laboratories.
- [Ando, 1998] Ando, Y. (1998). *Architectural acoustics: blending sound sources, sound fields, and listeners*. AIP series in modern acoustics and signal processing. AIP Press: Springer, New York.
- [Bamford and Vanderkooy, 1995] Bamford, J. S. and Vanderkooy, J. (1995). Ambisonic sound for us. In *99th Convention of the Audio Engineering Society*, volume 4138, New York. Audio Engineering Society.
- [Dempster, 1995] Dempster, S. (1995). Underground overlays from the cistern chapel.
- [Dolby, 2000] Dolby (2000). *5.1-Channel Production Guidelines*. Dolby Laboratories Inc., San Francisco, 1st edition.
- [Fahy, 1989] Fahy, F. J. (1989). *Sound intensity*. Elsevier Applied Science, London; New York.
- [Gardner and Martin, 1995] Gardner, W. and Martin, K. (1995). HRTF measurements of a KEMAR. *Journal of the Acoustical Society of America*, 97:3907–3908.
- [Geoff Martin, 1999] Geoff Martin, Wieslaw Woszczyk, J. C. R. Q. (1999). Controlling phantom image focus in a multichannel reproduction system. In *107th Convention of the Audio Engineering Society, preprint no. 4996*, volume 4996, New York, USA. Audio Engineering Society.
- [Golomb, 1967] Golomb, S. W. (1967). *Shift Register Sequences*. Holden Day, San Francisco.
- [Isaacs, 1990] Isaacs, A., editor (1990). *A Concise Dictionary of Physics*. Oxford University Press, New York, new edition edition.

- [ITU, 1994] ITU (1994). Multichannel stereophonic sound system with and without accompanying picture. Standard BS.775-1, International Telecommunication Union.
- [ITU, 1997] ITU (1997). Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. Standard BS.1116-1, International Telecommunication Union.
- [Kinsler and Frey, 1982] Kinsler, L. E. and Frey, A. R. (1982). *Fundamentals of acoustics*. Wiley, New York, 3rd edition.
- [Kuttruff, 1991] Kuttruff, H. (1991). *Room Acoustics*. Elsevier Applied Science. Elsevier Science Publishers, Essex, 3rd edition.
- [Lovett, 1994] Lovett, L. (1994). I love everybody.
- [Mandl, 1973] Mandl, M. (1973). *Handbook of Modern Electronic Data*. Reston Publishing Company, Inc., Reston, Virginia.
- [Martens, 1999] Martens, W. L. (1999). The impact of decorrelated low-frequency reproduction on auditory spatial imagery: Are two subwoofers better than one? In *AES 16th International Conference*, pages 87–77, Rovaniemi, Finland. Audio Engineering Society.
- [Martin, 2002a] Martin, G. (2002a). Interchannel interference at the listening position in a five-channel loudspeaker configuration. In *113th Convention of the Audio Engineering Society*, Los Angeles. Audio Engineering Society.
- [Martin, 2002b] Martin, G. (2002b). The significance of interchannel correlation, phase and amplitude differences on multichannel microphone techniques. In *113th Convention of the Audio Engineering Society*, Los Angeles. Audio Engineering Society.
- [Morfey, 2001] Morfey, C. L. (2001). *Dictionary of Acoustics*. Academic Press, San Diego.
- [Nuttgens, 1997] Nuttgens, P. (1997). *The Story of Architecture*. Phaidon Press, London, 2nd edition.
- [Olson, 1957] Olson, H. F. (1957). *Acoustical engineering*. Van Nostrand, Princeton, N.J.
- [Owinski, 1988] Owinski, B. (1988). Calibrating the 5.1 system. *Surround Professional*.

- [Rumsey, 2001] Rumsey, F. (2001). *Spatial Audio*. Music Technology Series. Focal Press, Oxford.
- [Sanchez and Taylor, 1998] Sanchez, C. and Taylor, R. (Feb. 1998). Overview of digital audio interface data structures. Technical report, Cirrus Logic Inc.
- [Schroeder, 1975] Schroeder, M. R. (1975). Diffuse reflection by maximum-length sequences. *Journal of the Acoustical Society of America*, 57(1):149–150.
- [Schroeder, 1979] Schroeder, M. R. (1979). Binaural dissimilarity and optimum ceilings for concert halls: More lateral sound diffusion. *Journal of the Acoustical Society of America*, 65(4):958–963.
- [Simonsen, 1984] Simonsen, G. (1984). Master’s thesis, Technical University of Denmark, Lyngby.
- [Steiglitz, 1996] Steiglitz, K. (1996). *A DSP primer: with applications to digital audio and computer music*. Addison-Wesley Pub., Menlo Park.
- [Strawn, 1985] Strawn, J., editor (1985). *Digital Audio Signal Processing: An Anthology*. William Kaufmann, Inc., Los Altos.
- [Svensson, 1994] Svensson, B.-I. D. M. K. P. (1994). A macroscopic view of diffuse reflection. *Journal of the Audio Engineering Society*, 42(10):793–805.
- [Tsujiimoto and Akita, 1997] Tsujimoto, A. F. K. and Akita, S. (1997). Microphone techniques for ambient sound on a music recording. In *103rd Conference of the Audio Engineering Society*, volume 4540, New York. Audio Engineering Society.
- [Wasserman and Whitmore, 1992] Wasserman, J. N. W. and Whitmore, G. (1992). *Applied Statistics*. Allyn and Bacon, Boston, USA, 4th edition.
- [Watkinson, 1988] Watkinson, J. (1988). *The Art of Digital Audio*. Focal Press, London, first edition.
- [Welle,] Welle, D. Stereo recording techniques. www.dwelle.de/rtc/infotheque.
- [Williams, 1990] Williams, M. (1990). The stereophonic zoom - a variable dual microphone system for stereophonic sound recording. Bru sur Marne, France.

- [Woram, 1989] Woram, J. M. (1989). *Sound Recording Handbook*. Howard W. Sams and Co., Indianapolis, 1st edition.

Index

- P*, 153
- R*, 183
- RT*₆₀, 246
- T*, 183
- U*, 159
- X*_C, 86
- α , 184
- λ , 164
- ω , 162
- ρ_o , 151
- φ_o , 151
- k*, 165
- k*₀, 165
- u*, 159
- 3 dB down point, 274

- A-Format, 739
- A-weighting, 262
- absorption coefficient, 440
- absorption, air, 184
- AC, 62
- acceptance angle, 376
- active filters, 145
- additive inverse, 24
- adiabatic, 183
- aliasing, 447
- alternate, 419
- alternating current, 62
- amp, 58
- ampere, 58
- Ampex, 766
- amplitude, 159
- amplitude, instantaneous, 153
- anti-aliasing filter, 449
- antinode, 201
- asymmetrical filter, 281
- attack time, 315, 316
- audio path, 318
- audio sample, 475
- auxiliary data, 475

- B-Format, 737
- B-weighting, 262
- back electromotive force, 104
- back EMF, 104
- band-reject filter, 279
- bandpass filter, 276
- Bandpass filters, 149
- bandstop filter, 279
- bandwidth, 276, 429
- base, 35
- bass management, 621
- bass redirection, 621
- beating, 167
- Bel, 70
- bell curve, 460
- Bi-Phase Correction, 769
- bi-phase mark, 766
- bi-phase mark, 472
- bias, forward, 116
- bias, reverse, 116
- binary, 35
- binary word, 37
- binaural beating, 168

- biquad, 554
biquadratic filter, 554
bits, 37
blocks, 473
Blumelein microphone technique, 642
Blumlein, Alan, 642
boost, 296
breakdown voltage, 121
bridge, 199
bridge rectifier, 125
Butterworth filter, 145
- c, 163
C-weighting, 262
capacitive reactance, 86
capacitor, 82
Cash, Johnny, 574
cent, 240
centre frequency, 280
channel code, 474
Chop, 419
CMRR, 142, 433
cocktail party effect, 748
coefficient, absorption, 184
coincident microphones, 640
colour frame flag, 769
comb filter, FIR, 538
comb filter, IIR, 549
combination tones, 168
combining filters, 286
common mode rejection ratio, 142, 433
comparator, 132
complex conjugate, 25
complex numbers, 21
compression ratio, 302
compression, acoustic, 158
cone of confusion, 266
consonance, 233
constant equilibrium density, 151
constant Q, 281
continuity tester, 413
continuous time, 442
control path, 318
control voltage, 319
conventional current, 58
convolution, 564
convolution, fast, 566
convolution, real, 565
correlation, 711
cosine, 11
crest factor, 65, 314
critical distance, 251
Crosby, Bing, 766
cross-correlation, 436
cross-correlation function, 423
crosstalk, 436
current, 58
cutoff frequency, 91, 274
CV, 319
- damped oscillator, 155
dB, 70
dB FS, 75
dBm, 73
dBspl, 72
dBV, 74
DC, 61
DC offset, 436
de-essing, 299
decade, 161, 274
Decca Tree, 647
decibel, 70
decimal, 34, 35
denominator, 579
derivative, 45
device under test, 420, 427
DFT, 516
difference tone, 167

- differential amplifier, 139
- diffuse field, 173, 245
- digit, 34
- digital multimeter, 411
- Digital Signal Processing, 509
- diode, zener, 120
- Dirac impulse, 546
- direct current, 61
- direct sound, 241
- Directivity Factor, 381
- discontinuity, 526
- discrete Fourier transform, 516
- discrete time, 442
- dissonance, 233
- Distance Factor, 382
- distortion, loudspeaker, 438
- dither, 457
- division, 415
- DMM, 411
- DRF, 381
- drop frame, 762
- drop frame flag, 769
- DSF, 382
- DSP, 509
- DUT, 420, 427
- duty cycle, 421
- dynamic range, 432

- e, 31
- effective pressure, 153
- EFM, 498
- eight to fourteen modulation, 498
- EIN, 432
- electromotive force, 58
- enharmonic, 157
- equal loudness contour, 259
- equalizers, 273
- equation, 41
- equivalent input noise, 432
- equivalent noise level, 439

- Euler's formula, 31
- Euler's identity, 31
- Euler, Leonhard, 31
- evil scientists, 415
- exponent, 5
- Eyring Equation, 247

- F, 104
- factorial, 31
- Farad, 82
- Farads, 104
- fast Fourier transform, 514
- FFT, 514
- filter, allpass, 559
- filter, Butterworth, 145
- filters, 273
- Fletcher, 259
- FM, 481
- foley, 750
- Fourier, 513
- Fourier transform, 514
- Fourier, Jean Baptiste, 514
- frame, 473
- frame rate, 758
- free field, 172
- frequency, 161
- frequency modulation, 481
- frequency response, 273, 429, 439
- frequency response, loudspeaker, 438
- frequency, angular, 162
- frequency, negative, 162
- frequency, normalized, 511
- frequency, radian, 162
- friction coefficient, 227
- full-wave rectifier, 125
- function, 41
- function generator, 420
- fundamental, 156

- Gain, 280

- gain, 69, 428
gain bandwidth product, 143
gain before compression, 306
gain linearity, 428
galvanometer, 411
GBP, 143
ground bus, 334
group delay, 429, 430
- Haas effect, 267
half-power point, 274
half-wave rectifier, 123
half-wavelength resonators, 216
Hamming, 534
Hanning, 532
harmonic, 156
Helmholtz resonator, 220
Hertz, 161
hexadecimal, 37
hole-in-the-middle, 645
horizontal position, 419
hypotenuse, 2
Hz, 161
- IIR, 549
imaginary numbers, 20
IMD, 436
impedance, 87, 439
impedance curve, loudspeaker, 438
impedance, acoustic, 174
impedance, characteristic, 175
impedance, specific acoustic, 175
impulse response, 205, 436
index, 151
inductive reactance, 104
inductor, 104
Infinite Impulse Response, 549
input impedance, 427
input resistance, 141
input voltage range, 142
- integer, 19
integer delay, 510, 561
integral, 52
intelligibility, 440
intensity, 178
interaural cross-correlation, 436
interference, constructive, 165
interference, destructive, 166
interlacing, 759
intermodulation distortion, 168, 436
Ipanema, Girl From, 221
IRE, 771
irrational numbers, 19
isophon, 261
ITU-R BS.775.1., 624
- just intonation, 236
Just Temperament, 236
- knee, 303
- Lamberts Law, 188
LFE, 620
limit, 42, 43
linear gain, 302
linear phase filter, 293
Linear Time Invariant, 571
Lissajous pattern, 420
log, 6
logarithm, 6
logarithm, natural, 7
loudness switch, 296
loudspeaker aperture, 624
Low Frequency Effects, 620
low-pass filter, 274
LTC, 766
LTI, 571
- magnetic lines of force, 96
Martin's Law, 748
masking, 264

- maximum differential input voltage, 141
- maximum length sequence, 422
- maximum output, 430
- Maximum Supply Voltage, 141
- mean, geometric, 277
- mid-side, 725
- minimum phase, 290
- mixer, 138
- MLS, 422
- modal density, 250
- mode, 200
- modes, 247
- modes, axial, 248
- modes, oblique, 249
- modes, tangential), 249
- modulus, 26
- mono compatibility, 641
- mono-stereo, 725
- MS, 725
- multiplicative inverse, 24
- Munson, 259

- N-type, 111
- natural logarithm, 31
- near-coincident techniques, 645
- node, 201
- noise level, 440
- noise, blue, 170
- noise, pink, 170
- noise, purple, 171
- noise, red, 171
- noise, white, 170
- non-drop frame, 765
- non-combining filters, 286
- non-symmetrical filter, 281
- normal distribution, 460
- notch filter, 279
- NTSC, 758
- numbers, whole, 19

- numerator, 579
- nut, 199
- Nyquist frequency, 447

- octave, 161
- off-axis response, 439
- off-axis response, loudspeaker, 438
- Ohm's Law, 59
- op amp, 132
- open loop voltage gain, 143
- Operating Common-Mode Range, 142
- operational amplifier, 132
- operator, 574
- order, filter, 274
- oscilloscope, 415
- output impedance, 427
- output resistance, 142
- output short-circuit duration, 141
- output voltage swing, 142
- overlap-add, 567
- overtones, 156

- p, 153
- P-type, 111
- PAL, 759
- panphonic, 737
- paragraphic equalizer, 288
- parallel, 79
- parity bit, 475
- passband, 276
- PDF, 457
- peak, 311
- period, 161
- periodic, 164
- periphonic, 737
- phase distortion, 290
- phase response, 429, 439
- phase response, loudspeaker, 438
- phase, unwrapped, 430

- phase, wrapped, 430
phon, 260
polar pattern, 439
pole, 587
Poniatoff, Alexander M., 766
power, 60
power response, loudspeaker, 438
power, acoustic, 177
ppm, 163
preamble, 474
precedence effect, 267
precision rectifier, 324
pressure, acoustic, 153
pressure, effective, 160
pressure, excess, 153
pressure, instantaneous, 153
pressure, peak, 160
pressure, peak to peak, 160
pressure, stasis, 151
probability density function, rectangular, 460
probability density functions, 457
probability distribution function, Gaussian, 460
probability distribution function, triangular, 460
proximity effect, 375
pseudo-random signal, 425
pumping, 309
pure temperament, 236
Pythagorean Comma, 236
Pythagorean Theorem, 3
- Q, filter, 278, 436
quack, duck, 267
quadratic residue diffuser, 195
quality, filter, 278
quantization, 444
quantization error, 444
quantization noise, 444
- quasi-parametric equalizer, 289
- radians, 15
Random-Energy Efficiency, 379
Random-Energy Response, 377
rational number, 19
re-recording engineer, 750
reactance, acoustic, 175
real numbers, 20
reconstruction filter, 445
REE, 379
reflection coefficient, 183, 440
reflection, order, 242
reflections, 440
refraction, acoustic, 158
regulator, 127
release time, 316, 317
RER, 377
resistance, acoustic, 175
resistor, 59
resultant tones, 168
reverb, 245
reverberation, 245
reverberation time, 246, 440
reverse breakdown voltage, 120
Riemann Sum, 50
right angle, 1
right triangle, 1
ringing, 294
RMS, 62, 160, 311
room mode, 247
room modes, 440
room radius, 251
root mean square (RMS), 62
rosin, 227
rotation point, 302
RPDF, 460
RT60, 440
- S/N ratio, 432

- Sabine Equation, 246
Sabine, Wallace Clement, 246
sample and hold, 442
sampling, 442
Schroeder diffuser, 195
Schroeder frequency, 250
SECAM, 759
semi-parametric filter, 289
semiconductors, 110
sensitivity, 439
sensitivity, loudspeaker, 438
series, 77
shelving filter, 280
side chain, 299, 318
signal generator, 420
signal to noise ratio, 432
simple harmonic motion, 11, 153
sine, 9
slew rate, 143, 430
slope, 3, 418
slope, filter, 274
smoothing filter, 445
Snells law, 187
SNR, 432
soft knee, 311
sone, 261
sound pressure level, 153
sound transmission, 440
soundfield, 736
Soundfield microphone, 739
source code, 475
speed of sound, 163
Spinal Tap, 69
SPL, 153
status bit, 476
stop frequency, 287
sub-frame, 473
symmetrical filter, 284
sync code, 474
syntonic comma, 237
tangent, 43
tau, 84
temperaments, equal, 238
temperaments, just, 236
temperaments, meantone, 238
temperaments, Pythagorean, 234
termination, 198
THD, 434
THD+N, 435
thermal noise, 431
threshold, 302
threshold of hearing, 153, 257
threshold of pain, 257
Time / div., 416
time code, 758
time constant, 66, 315
time constant, RC, 85
tolerance, resistor, 785
toroidal, 108
total harmonic distortion, 434
total harmonic distortion + noise, 435
TPDF, 460
trace rotation, 416
transfer function, 300, 575
transformer, 108
transition ratio, 287
transmission coefficient, 183
trigger, 418
true RMS meter, 414
turnover frequency, 287
unassigned bits, 769
undamped oscillator, 155
unterminated, 432
user bit, 476
user flags (2), 769
valence shell, 55
validity bit, 475

- VCA, 319
- velocity, instantaneous, 159
- Vertical position, 417
- VITC, 770
- voltage, 58
- voltage controlled amplifier, 319
- voltage regulator, 127
- voltmeter, 411
- volume density, 151

- Watt's Law, 60
- wave number, 165
- wave, longitudinal, 157
- wave, plane, 175
- wave, torsional, 157
- wave, transverse, 157
- waveguide, 210
- wavelength, 164
- wavenumber, 165
- wavenumber, acoustic, 165
- weighting curve, 262
- weighting filter, 262
- window, Hamming, 534
- window, Hanning, 532
- window, rectangular, 530
- windowing, 524
- windowing function, 527
- wolf fifth, 235

- X-Y mode, 419

- Zen, 554
- zero, 582