
**THE KLUWER INTERNATIONAL SERIES
IN ENGINEERING AND COMPUTER SCIENCE**

APPLICATIONS OF DIGITAL SIGNAL PROCESSING TO AUDIO AND ACOUSTICS

edited by

Mark Kahrs

Rutgers University

Piscataway, New Jersey, USA

Karlheinz Brandenburg

Fraunhofer Institut Integrierte Schaltungen

Erlangen, Germany

KLUWER ACADEMIC PUBLISHERS

NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 0-3064-7042-X
Print ISBN 0-7923-8130-0

©2002 Kluwer Academic Publishers
New York, Boston, Dordrecht, London, Moscow

All rights reserved

No part of this eBook may be reproduced or transmitted in any form or by any means, electronic, mechanical, recording, or otherwise, without written consent from the Publisher

Created in the United States of America

Visit Kluwer Online at: <http://www.kluweronline.com>
and Kluwer's eBookstore at: <http://www.ebooks.kluweronline.com>

Contents

List of Figures	xiii
List of Tables	xxi
Contributing Authors	xxiii
Introduction	xxix
<i>Karlheinz Brandenburg and Mark Kahrs</i>	
1	
Audio quality determination based on perceptual measurement techniques	1
<i>John G. Beerends</i>	
1.1 Introduction	1
1.2 Basic measuring philosophy	2
1.3 Subjective versus objective perceptual testing	6
1.4 Psychoacoustic fundamentals of calculating the internal sound representation	8
1.5 Computation of the internal sound representation	13
1.6 The perceptual audio quality measure (PAQM)	17
1.7 Validation of the PAQM on speech and music codec databases	20
1.8 Cognitive effects in judging audio quality	22
1.9 ITU Standardization	29
1.9.1 ITU-T, speech quality	30
1.9.2 ITU-R, audio quality	35
1.10 Conclusions	37
2	
Perceptual Coding of High Quality Digital Audio	39
<i>Karlheinz Brandenburg</i>	
2.1 Introduction	39

2.2	Some Facts about Psychoacoustics	42
2.2.1	Masking in the Frequency Domain	42
2.2.2	Masking in the Time Domain	44
2.2.3	Variability between listeners	45
2.3	Basic ideas of perceptual coding	47
2.3.1	Basic block diagram	48
2.3.2	Additional coding tools	49
2.3.3	Perceptual Entropy	50
2.4	Description of coding tools	50
2.4.1	Filter banks	50
2.4.2	Perceptual models	59
2.4.3	Quantization and coding	63
2.4.4	Joint stereo coding	68
2.4.5	Prediction	72
2.4.6	Multi-channel: to matrix or not to matrix	73
2.5	Applying the basic techniques: real coding systems	74
2.5.1	Pointers to early systems (no detailed description)	74
2.5.2	MPEG Audio	75
2.5.3	MPEG-2 Advanced Audio Coding (MPEG-2 AAC)	79
2.5.4	MPEG-4 Audio	81
2.6	Current Research Topics	82
2.7	Conclusions	83
3	Reverberation Algorithms	85
	<i>William G. Gardner</i>	
3.1	Introduction	85
3.1.1	Reverberation as a linear filter	86
3.1.2	Approaches to reverberation algorithms	87
3.2	Physical and Perceptual Background	88
3.2.1	Measurement of reverberation	89
3.2.2	Early reverberation	90
3.2.3	Perceptual effects of early echoes	93
3.2.4	Reverberation time	94
3.2.5	Modal description of reverberation	95
3.2.6	Statistical model for reverberation	97
3.2.7	Subjective and objective measures of late reverberation	98
3.2.8	Summary of framework	100
3.3	Modeling Early Reverberation	100
3.4	Comb and Allpass Reverberators	105
3.4.1	Schroeder's reverberator	105
3.4.2	The parallel comb filter	108
3.4.3	Modal density and echo density	109
3.4.4	Producing uncorrelated outputs	111
3.4.5	Moorer's reverberator	112
3.4.6	Allpass reverberators	113
3.5	Feedback Delay Networks	116

3.5.1	Jot's reverberator	119
3.5.2	Unitary feedback loops	121
3.5.3	Absorptive delays	122
3.5.4	Waveguide reverberators	123
3.5.5	Lossless prototype structures	125
3.5.6	Implementation of absorptive and correction filters	128
3.5.7	Multirate algorithms	128
3.5.8	Time-varying algorithms	129
3.6	Conclusions	130
4		
	Digital Audio Restoration	133
	<i>Simon Godsill, Peter Rayner and Olivier Cappé</i>	
4.1	Introduction	134
4.2	Modelling of audio signals	135
4.3	Click Removal	137
4.3.1	Modelling of clicks	137
4.3.2	Detection	141
4.3.3	Replacement of corrupted samples	144
4.3.4	Statistical methods for the treatment of clicks	152
4.4	Correlated Noise Pulse Removal	155
4.5	Background noise reduction	163
4.5.1	Background noise reduction by short-time spectral attenuation	164
4.5.2	Discussion	177
4.6	Pitch variation defects	177
4.6.1	Frequency domain estimation	179
4.7	Reduction of Non-linear Amplitude Distortion	182
4.7.1	Distortion Modelling	183
4.7.2	Non-linear Signal Models	184
4.7.3	Application of Non-linear models to Distortion Reduction	186
4.7.4	Parameter Estimation	188
4.7.5	Examples	190
4.7.6	Discussion	190
4.8	Other areas	192
4.9	Conclusion and Future Trends	193
5		
	Digital Audio System Architecture	195
	<i>Mark Kahrs</i>	
5.1	Introduction	195
5.2	Input/Output	196
5.2.1	Analog/Digital Conversion	196
5.2.2	Sampling clocks	202
5.3	Processing	203
5.3.1	Requirements	204
5.3.2	Processing	207
5.3.3	Synthesis	208

5.3.4	Processors	209
5.4	Conclusion	234
6	Signal Processing for Hearing Aids	235
	<i>James M. Kates</i>	
6.1	Introduction	236
6.2	Hearing and Hearing Loss	237
6.2.1	Outer and Middle Ear	238
6.3	Inner Ear	239
6.3.1	Retrocochlear and Central Losses	247
6.3.2	Summary	248
6.4	Linear Amplification	248
6.4.1	System Description	249
6.4.2	Dynamic Range	251
6.4.3	Distortion	252
6.4.4	Bandwidth	253
6.5	Feedback Cancellation	253
6.6	Compression Amplification	255
6.6.1	Single-Channel Compression	256
6.6.2	Two-Channel Compression	260
6.6.3	Multi-Channel Compression	261
6.7	Single-Microphone Noise Suppression	263
6.7.1	Adaptive Analog Filters	263
6.7.2	Spectral Subtraction	264
6.7.3	Spectral Enhancement	266
6.8	Multi-Microphone Noise Suppression	267
6.8.1	Directional Microphone Elements	267
6.8.2	Two-Microphone Adaptive Noise Cancellation	268
6.8.3	Arrays with Time-Invariant Weights	269
6.8.4	Two-Microphone Adaptive Arrays	269
6.8.5	Multi-Microphone Adaptive Arrays	271
6.8.6	Performance Comparison in a Real Room	273
6.9	Cochlear Implants	275
6.10	Conclusions	276
7	Time and Pitch scale modification of audio signals	279
	<i>Jean Laroche</i>	
7.1	Introduction	279
7.2	Notations and definitions	282
7.2.1	An underlying sinusoidal model for signals	282
7.2.2	A definition of time-scale and pitch-scale modification	282
7.3	Frequency-domain techniques	285
7.3.1	Methods based on the short-time Fourier transform	285
7.3.2	Methods based on a signal model	293
7.4	Time-domain techniques	293

7.4.1	Principle	293
7.4.2	Pitch independent methods	294
7.4.3	Periodicity-driven methods	298
7.5	Formant modification	302
7.5.1	Time-domain techniques	302
7.5.2	Frequency-domain techniques	302
7.6	Discussion	303
7.6.1	Generic problems associated with time or pitch scaling	303
7.6.2	Time-domain vs frequency-domain techniques	308
8		
	Wavetable Sampling Synthesis	311
	<i>Dana C. Massie</i>	
8.1	Background and introduction	311
8.1.1	Transition to Digital	312
8.1.2	Flourishing of Digital Synthesis Methods	313
8.1.3	Metrics: The Sampling - Synthesis Continuum	314
8.1.4	Sampling vs. Synthesis	315
8.2	Wavetable Sampling Synthesis	318
8.2.1	Playback of digitized musical instrument events.	318
8.2.2	Entire note - not single period	318
8.2.3	Pitch Shifting Technologies	319
8.2.4	Looping of sustain	331
8.2.5	Multi-sampling	337
8.2.6	Enveloping	338
8.2.7	Filtering	338
8.2.8	Amplitude variations as a function of velocity	339
8.2.9	Mixing or summation of channels	339
8.2.10	Multiplexed wavetables	340
8.3	Conclusion	341
9		
	Audio Signal Processing Based on Sinusoidal Analysis/Synthesis	343
	<i>T.F. Quatieri and R. J. McAulay</i>	
9.1	Introduction	344
9.2	Filter Bank Analysis/Synthesis	346
9.2.1	Additive Synthesis	346
9.2.2	Phase Vocoder	347
9.2.3	Motivation for a Sine-Wave Analysis/Synthesis	350
9.3	Sinusoidal-Based Analysis/Synthesis	351
9.3.1	Model	352
9.3.2	Estimation of Model Parameters	352
9.3.3	Frame-to-Frame Peak Matching	355
9.3.4	Synthesis	355
9.3.5	Experimental Results	358
9.3.6	Applications of the Baseline System	362
9.3.7	Time-Frequency Resolution	364
9.4	Source/Filter Phase Model	366

9.4.1	Model	367
9.4.2	Phase Coherence in Signal Modification	368
9.4.3	Revisiting the Filter Bank-Based Approach	381
9.5	Additive Deterministic/Stochastic Model	384
9.5.1	Model	385
9.5.2	Analysis/Synthesis	387
9.5.3	Applications	390
9.6	Signal Separation Using a Two-Voice Model	392
9.6.1	Formulation of the Separation Problem	392
9.6.2	Analysis and Separation	396
9.6.3	The Ambiguity Problem	399
9.6.4	Pitch and Voicing Estimation	402
9.7	FM Synthesis	403
9.7.1	Principles	404
9.7.2	Representation of Musical Sound	407
9.7.3	Parameter Estimation	409
9.7.4	Extensions	411
9.8	Conclusions	411
10		
	Principles of Digital Waveguide Models of Musical Instruments	417
	<i>Julius O. Smith III</i>	
10.1	Introduction	418
10.1.1	Antecedents in Speech Modeling	418
10.1.2	Physical Models in Music Synthesis	420
10.1.3	Summary	422
10.2	The Ideal Vibrating String	423
10.2.1	The Finite Difference Approximation	424
10.2.2	Traveling-Wave Solution	426
10.3	Sampling the Traveling Waves	426
10.3.1	Relation to Finite Difference Recursion	430
10.4	Alternative Wave Variables	431
10.4.1	Spatial Derivatives	431
10.4.2	Force Waves	432
10.4.3	Power Waves	434
10.4.4	Energy Density Waves	435
10.4.5	Root-Power Waves	436
10.5	Scattering at an Impedance Discontinuity	436
10.5.1	The Kelly-Lochbaum and One-Multiply Scattering Junctions	439
10.5.2	Normalized Scattering Junctions	441
10.5.3	Junction Passivity	443
10.6	Scattering at a Loaded Junction of N Waveguides	446
10.7	The Lossy One-Dimensional Wave Equation	448
10.7.1	Loss Consolidation	450
10.7.2	Frequency-Dependent Losses	451
10.8	The Dispersive One-Dimensional Wave Equation	451
10.9	Single-Reed Instruments	455

10.9.1 Clarinet Overview	457
10.9.2 Single-Reed Theory	458
10.10 Bowed Strings	462
10.10.1 Violin Overview	463
10.10.2 The Bow-String Scattering Junction	464
10.11 Conclusions	466
References	467
Index	535

List of Figures

1.1	Basic philosophy used in perceptual audio quality determination	4
1.2	Excitation pattern for a single sinusoidal tone	9
1.3	Excitation pattern for a single click	10
1.4	Excitation pattern for a short tone burst	11
1.5	Masking model overview	12
1.6	Time-domain smearing as a function of frequency	15
1.7	Basic auditory transformations used in the PAQM	18
1.8	Relation between MOS and PAQM, ISO/MPEG 1990 database	19
1.9	Relation between MOS and PAQM, ISO/MPEG 1991 database	21
1.10	Relation between MOS and PAQM, ITU-R 1993 database	22
1.11	Relation between MOS and PAQM, ETSI GSM full rate database	23
1.12	Relation between MOS and PAQM, ETSI GSM half rate database	24
1.13	Basic approach used in the development of PAQMC	25
1.14	Relation between MOS and PAQMC, ISO/MPEG 1991 database	28
1.15	Relation between MOS and PAQMC, ITU-R 1993 database	29
1.16	Relation between MOS and PAQMC, ETSI GSM full rate database	30
1.17	Relation between MOS and PAQMC, ETSI GSM half rate database	31
1.18	Relation between MOS and PSQM, ETSI GSM full rate database	32
1.19	Relation between MOS and PSQM, ETSI GSM half rate database	33
1.20	Relation between MOS and PSQM, ITU-T German speech database	34
1.21	Relation between MOS and PSQM, ITU-T Japanese speech database	35
1.22	Relation between Japanese and German MOS values	36
2.1	Masked thresholds: Masker: narrow band noise at 250 Hz, 1 kHz, 4 kHz	44
2.2	Example of pre-masking and post-masking	45

2.3	Masking experiment as reported in [Spille, 1992]	46
2.4	Example of a pre-echo	47
2.5	Block diagram of a perceptual encoding/decoding system	48
2.6	Basic block diagram of an n-channel analysis/synthesis filter bank with downsampling by k	51
2.7	Window function of the MPEG-1 polyphase filter bank	54
2.8	Frequency response of the MPEG-1 polyphase filter bank	55
2.9	Block diagram of the MPEG Layer 3 hybrid filter bank	57
2.10	Window forms used in Layer 3	58
2.11	Example sequence of window forms	59
2.12	Example for the bit reservoir technology (Layer 3)	67
2.13	Main axis transform of the stereo plane	69
2.14	Basic block diagram of M/S stereo coding	70
2.15	Signal flow graph of the M/S matrix	70
2.16	Basic principle of intensity stereo coding	71
2.17	ITU Multichannel configuration	73
2.18	Block diagram of an MPEG-1 Layer 3 encode	77
2.19	Transmission of MPEG-2 multichannel information within an MPEG-1 bitstream	78
2.20	Block diagram of the MPEG-2 AAC encoder	80
2.21	MPEG-4 audio scaleable configuration	82
3.1	Impulse response of reverberant stairwell measured using ML sequences.	90
3.2	Single wall reflection and corresponding image source A' .	91
3.3	A regular pattern of image sources occurs in an ideal rectangular room.	91
3.4	Energy decay relief for occupied Boston Symphony Hall	96
3.5	Canonical direct form FIR filter with single sample delays.	101
3.6	Combining early echoes and late reverberation	102
3.7	FIR filter cascaded with reverberator	102
3.8	Associating absorptive and directional filters with early echoes.	103
3.9	Average head-related filter applied to a set of early echoes	104
3.10	Binaural early echo simulator	104
3.11	One-pole, DC-normalized lowpass filter.	104
3.12	Comb filter response	106
3.13	Allpass filter formed by modification of a comb filter	106
3.14	Schroeder's reverberator consisting of a parallel comb filter and a series allpass filter [Schroeder, 1962].	108
3.15	Mixing matrix used to form uncorrelated outputs	112

3.16	Controlling IACC in binaural reverberation	112
3.17	Comb filter with lowpass filter in feedback loop	113
3.18	Lattice allpass structure.	115
3.19	Generalization of figure 3.18.	115
3.20	Reverberator formed by adding absorptive losses to an allpass feedback loop	115
3.21	Dattorro's plate reverberator based on an allpass feedback loop	117
3.22	Stautner and Puckette's four channel feedback delay network	118
3.23	Feedback delay network as a general specification of a reverberator containing N delays	120
3.24	Unitary feedback loop	121
3.25	Associating an attenuation with a delay.	122
3.26	Associating an absorptive filter with a delay.	123
3.27	Reverberator constructed with frequency dependent absorptive filters	124
3.28	Waveguide network consisting of a single scattering junction to which N waveguides are attached	124
3.29	Modification of Schroeder's parallel comb filter to maximize echo density	126
4.1	Click-degraded music waveform taken from 78 rpm recording	138
4.2	AR-based detection, $P=50$. (a) Prediction error filter (b) Matched filter	138
4.3	Electron micrograph showing dust and damage to the grooves of a 78rpm gramophone disc.	139
4.4	AR-based interpolation, $P=60$, classical chamber music, (a) short gaps, (b) long gaps	147
4.5	Original signal and excitation ($P=100$)	150
4.6	LSAR interpolation and excitation ($P = 100$)	150
4.7	Sampled AR interpolation and excitation ($P=100$)	151
4.8	Restoration using Bayesian iterative methods	155
4.9	Noise pulse from optical film sound track ('silent' section)	157
4.10	Signal waveform degraded by low frequency noise transient	157
4.11	Degraded audio signal with many closely spaced noise transients	161
4.12	Estimated noise transients for figure 4.11	161
4.13	Restored audio signal for figure 4.11 (different scale)	162
4.14	Modeled restoration process	164
4.15	Background noise suppression by short- time spectral attenuation	165
4.16	Suppression rules characteristics	168
4.17	Restoration of a sinusoidal signal embedded in white noise	169
4.18	Probability density of the relative signal level for different mean values	172

4.19	Short-time power variations	175
4.20	Frequency tracks generated for example ‘Viola’	179
4.21	Estimated (full line) and true (dotted line) pitch variation curves generated for example ‘Viola’	180
4.22	Frequency tracks generated for example ‘Midsum’	180
4.23	Pitch variation curve generated for example ‘Midsum’	181
4.24	Model of the distortion process	184
4.25	Model of the signal and distortion process	186
4.26	Typical section of AR-MNL Restoration	191
4.27	Typical section of AR-NAR Restoration	191
5.1	DSP system block diagram	196
5.2	Successive Approximation Converter	198
5.3	16 Bit Floating Point DAC (from [Kriz, 1975])	202
5.4	Block diagram of Moore’s FRMbox	210
5.5	Samson Box block diagram	211
5.6	diGiugno 4A processor	213
5.7	IRCAM 4B data path	214
5.8	IRCAM 4C data path	215
5.9	IRCAM 4X system block diagram	216
5.10	Sony DAE-1000 signal processor	217
5.11	Lucasfilm ASP ALU block diagram	218
5.12	Lucasfilm ASP interconnect and memory diagram	219
5.13	Moorer’s update queue data path	219
5.14	MPACT block diagram	222
5.15	Rossum’s cached interpolator	226
5.16	Sony OXF DSP block diagram	227
5.17	DSP.* block diagram	228
5.18	Gnusic block diagram	229
5.19	Gnusic core block diagram	230
5.20	Sony SDP-1000 DSP block diagram	232
5.21	Sony’s OXF interconnect block diagram	233
6.1	Major features of the human auditory system	238
6.2	Features of the cochlea: transverse cross-section of the cochlea	239
6.3	Features of the cochlea: the organ of Corti	240
6.4	Sample tuning curves for single units in the auditory nerve of the cat	241
6.5	Neural tuning curves resulting from damaged hair cells	242
6.6	Loudness level functions	244
6.7	Mean results for unilateral cochlear impairments	246

6.8	Simulated neural response for the normal ear	247
6.9	Simulated neural response for impaired outer cell function	248
6.10	Simulated neural response for 30 dB of gain	249
6.11	Cross-section of an in-the-ear hearing aid	250
6.12	Block diagram of an ITE hearing aid inserted into the ear canal	251
6.13	Block diagram of a hearing aid incorporating signal processing for feedback cancellation	255
6.14	Input/output relationship for a typical hearing-aid compression amplifier	256
6.15	Block diagram of a hearing aid having feedback compression	257
6.16	Compression amplifier input/output curves derived from a simplified model of hearing loss.	260
6.17	Block diagram of a spectral-subtraction noise-reduction system.	265
6.18	Block diagram of an adaptive noise-cancellation system.	268
6.19	Block diagram of an adaptive two-microphone array.	270
6.20	Block diagram of a time-domain five-microphone adaptive array.	271
6.21	Block diagram of a frequency-domain five-microphone adaptive array.	274
7.1	Duality between Time-scaling and Pitch-scaling operations	285
7.2	Time stretching in the time-domain	293
7.3	A modified tape recorder for analog time-scale or pitch-scale modification	294
7.4	Pitch modification with the sampling technique	295
7.5	Output elapsed time versus input elapsed time in the sampling method for Time-stretching	296
7.6	Time-scale modification of a sinusoid	297
7.7	Output elapsed time versus input elapsed time in the optimized sampling method for Time-stretching	300
7.8	Pitch-scale modification with the PSOLA method	301
7.9	Time-domain representation of a speech signal showing shape invariance	305
7.10	Time-domain representation of a speech signal showing loss of shape-invariance	306
8.1	Expressivity vs. Accuracy	316
8.2	Sampling tradeoffs	316
8.3	Labor costs for synthesis techniques	317
8.4	Rudimentary sampling	320
8.5	“Drop Sample Tuning” table lookup sampling playback oscillator	323
8.6	Classical sample rate conversion chain	325
8.7	Digital Sinc function	326

8.8	Frequency response of at linear interpolation sample rate converter	327
8.9	A sampling playback oscillator using high order interpolation	329
8.10	Traditional ADSR amplitude envelope	331
8.11	Backwards forwards loop at a loop point with even symmetry	333
8.12	Backwards forwards loop at a loop point with odd symmetry	333
8.13	Multisampling	337
9.1	Signal and spectrogram from a trumpet	345
9.2	Phase vocoder based on filter bank analysis/synthesis.	349
9.3	Passage of single sine wave through one bandpass filter.	350
9.4	Sine-wave tracking based on frequency-matching algorithm	356
9.5	Block diagram of baseline sinusoidal analysis/synthesis	358
9.6	Reconstruction of speech waveform	359
9.7	Reconstruction of trumpet waveform	360
9.8	Reconstruction of waveform from a closing stapler	360
9.9	Magnitude-only reconstruction of speech	361
9.10	Onset-time model for time-scale modification	370
9.11	Transitional properties of frequency tracks with adaptive cutoff	372
9.12	Estimation of onset times for time-scale modification	374
9.13	Analysis/synthesis for time-scale modification	375
9.14	Example of time-scale modification of trumpet waveform	376
9.15	Example of time-varying time-scale modification of speech waveform	376
9.16	KFH phase dispersion using the sine-wave preprocessor	380
9.17	Comparison of original waveform and processed speech	381
9.18	Time-scale expansion ($\times 2$) using subband phase correction	383
9.19	Time-scale expansion ($\times 2$) of a closing stapler using filter bank/overlap-add	385
9.20	Block diagram of the deterministic plus stochastic system.	389
9.21	Decomposition example of a piano tone	391
9.22	Two-voice separation using sine-wave analysis/synthesis and peak-picking	393
9.23	Properties of the STFT of $x(n) = x_a(n) + x_b(n)$	396
9.24	Least-squared error solution for two sine waves	397
9.25	Demonstration of two-lobe overlap	400
9.26	H matrix for the example in Figure 9.25	401
9.27	Demonstration of ill conditioning of the H matrix	402
9.28	FM Synthesis with different carrier and modulation frequencies	405
9.29	Spectral dynamics of FM synthesis with linearly changing modulation index	406

9.30	Comparison of Equation (9.82) and (9.86) for parameter settings $\omega_c = 2000$, $\omega_m = 200$, and $I = 5.0$	407
9.31	Spectral dynamics of trumpet-like sound using FM synthesis	408
10.1	The ideal vibrating string.	423
10.2	An infinitely long string, “plucked” simultaneously at three points.	427
10.3	Digital simulation of the ideal, lossless waveguide with observation points at $x = 0$ and $x = 3X = 3cT$.	429
10.4	Conceptual diagram of interpolated digital waveguide simulation.	429
10.5	Transverse force propagation in the ideal string.	433
10.6	A waveguide section between two partial sections, a) Physical picture indicating traveling waves in a continuous medium whose wave impedance changes from R_0 to R_1 to R_2 . b) Digital simulation diagram for the same situation.	437
10.7	The Kelly-Lochbaum scattering junction.	439
10.8	The one-multiply scattering junction.	440
10.9	The normalized scattering junction.	441
10.10	A three-multiply normalized scattering junction	443
10.11	Four ideal strings intersecting at a point to which a lumped impedance is attached.	446
10.12	Discrete simulation of the ideal, lossy waveguide.	449
10.13	Discrete-time simulation of the ideal, lossy waveguide.	450
10.14	Section of a stiff string where allpass filters play the role of unit delay elements.	453
10.15	Section of a stiff string where the allpass delay elements are consolidated at two points, and a sample of pure delay is extracted from each allpass chain.	454
10.16	A schematic model for woodwind instruments.	455
10.17	Waveguide model of a single-reed, cylindrical-bore woodwind, such as a clarinet.	457
10.18	Schematic diagram of mouth cavity, reed aperture, and bore.	458
10.19	Normalised reed impedance $G(p_\Delta) \triangleq R_b u_m(p_\Delta)$ overlaid with the “bore load line” $p_\Delta^+ - p_\Delta = R_b u_b$.	459
10.20	Simple, qualitatively chosen reed table for the digital waveguide clarinet	461
10.21	A schematic model for bowed-string instruments.	463
10.22	Waveguide model for a bowed string instrument, such as a violin.	464
10.23	Simple, qualitatively chosen bow table for the digital waveguide violin	465

List of Tables

2.1	Critical bands according to [Zwicker, 1982]	43
2.2	Huffman code tables used in Layer 3	66
5.1	Pipeline timing for Samson box generators	212
6.1	Hearing thresholds, descriptive terms, and probable handicaps (after Goodman, 1965)	236

Acknowledgments

Mark Kahrs would like to acknowledge the support of J.L. Flanagan. He would also like to acknowledge the the assistance of Howard Trickey and S.J. Orfanidis. Jean Laroche has helped out with the production and served as a valuable forcing function. The patience of Diane Litnran has been tested numerous times and she has offered valuable advice.

Karlheinz Brandenburg would like to thank Mark for his patience while he was always late in delivering his parts.

Both editors would like to acknowledge the patience of Bob Holland, our editor at Kluwer.

Contributing Authors

John G. Beerends was born in Millicent, Australia, in 1954. He received a degree in electrical engineering from the HTS (Polytechnic Institute) of The Hague, The Netherlands, in 1975. After working in industry for three years he studied physics and mathematics at the University of Leiden where he received the degree of M.Sc. in 1984. In 1983 he was awarded a prize of DF1 45000,- by Job Creation, for an innovative idea in the field of electro-acoustics. During the period 1984 to 1989 he worked at the Institute for Perception Research where he received a Ph.D. from the Technical University of Eindhoven in 1989. The main part of his Ph.D. work, which deals with pitch perception, was patented by the NV. Philips Gloeilampenfabriek. In 1989 he joined the audio group of the KPN research lab in Leidschendam where he works on audio quality assessment. Currently he is also involved in the development of an objective video quality measure.

Karlheinz Brandenburg received M.S. (Diplom) degrees in Electrical Engineering in 1980 and in Mathematics in 1982 from Erlangen University. In 1989 he earned his Ph.D. in Electrical Engineering, also from Erlangen University, for work on digital audio coding and perceptual measurement techniques. From 1989 to 1990 he was with AT&T Bell Laboratories in Murray Hill, NJ, USA. In 1990 he returned to Erlangen University to continue the research on audio coding and to teach a course on digital audio technology. Since 1993 he is the head of the Audio/Multimedia department at the Fraunhofer Institute for Integrated Circuits (FhG-IIS). Dr. Brandenburg is a member of the technical committee on Audio and Electroacoustics of the IEEE Signal Processing Society. In 1994 he received the ASE Fellowship Award for his work on perceptual audio coding and psychoacoustics.

Olivier Cappé was born in Villeurbanne, France, in 1968. He received the M.Sc. degree in electrical engineering from the Ecole Supérieure d'Electricité (ESE), Paris in 1990, and the Ph.D. degree in signal processing from the Ecole Nationale Supérieure des Télécommunications (ENST), Paris, in 1993. His Ph.D. thesis dealt with noise-reduction for degraded audio recordings. He is currently with the Centre National de la Recherche Scientifique (CNRS) at ENST, Signal department. His research interests are in statistical signal processing for telecommunications and speech/audio processing. Dr. Cappé received the IEE Signal Processing Society's Young Author Best Paper Award in 1995.

Bill Gardner was born in 1960 in Meriden, CT, and grew up in the Boston area. He received a bachelor's degree in computer science from MIT in 1982 and shortly thereafter joined Kurzweil Music Systems as a software engineer. For the next seven years, he helped develop software and signal processing algorithms for Kurzweil synthesizers. He left Kurzweil in 1990 to enter graduate school at the MIT Media Lab, where he recently completed his Ph.D. on the topic of 3-D audio using loudspeakers. He was awarded a Motorola Fellowship at the Media Lab, and was recipient of the 1997 Audio Engineering Society Publications Award. He is currently an independent consultant working in the Boston area. His research interests are spatial audio, reverberation, sound synthesis, realtime signal processing, and psychoacoustics.

Simon Godsill studied for the B.A. in Electrical and Information Sciences at the University of Cambridge from 1985-88. Following graduation he led the technical development team at the newly-formed CEDAR Audio Ltd., researching and developing DSP algorithms for restoration of degraded sound recordings. In 1990 he took up a post as Research Associate in the Signal Processing Group of the Engineering Department at Cambridge and in 1993 he completed his doctoral thesis: *The Restoration of Degraded Audio Signals*. In 1994 he was appointed as a Research Fellow at Corpus Christi College, Cambridge and in 1996 as University Lecturer in Signal Processing at the Engineering Department in Cambridge. Current research topics include: Bayesian and statistical methods in signal processing, modelling and enhancement of speech and audio signals, source signal separation, non-linear and non-Gaussian techniques, blind estimation of communications channels and image sequence analysis.

Mark Kahrs was born in Rome, Italy in 1952. He received an A.B. from Revelle College, University of California, San Diego in 1974. He worked intermittently for Tymshare, Inc. as a Systems Programmer from 1968 to 1974. During the summer of 1975 he was a Research Intern at Xerox PARC and then from 1975 to 1977 was a Research Programmer at the Center for Computer Research in Music and

Acoustics (CCRMA) at Stanford University. He was a chercheur at the Institut de Recherche et Coordination Acoustique Musique (IRCAM) in Paris during the summer of 1977. He received a PhD. in Computer Science from the University of Rochester in 1984. He worked and consulted for Bell Laboratories from 1984 to 1996. He has been an Assistant Professor at Rutgers University from 1988 to the present where he taught courses in Computer Architecture, Digital Signal Processing and Audio Engineering. In 1993 he was General Chair of the *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* ("Mohonk Workshop"). Since 1993 he has chaired the Technical Committee on Audio And Electroacoustics in the Signal Processing Society of the IEEE.

James M. Kates was born in Brookline, Massachusetts, in 1948. He received the degrees of BSEE and MSEE from the Massachusetts Institute of Technology in 1971 and the professional degree of Electrical Engineer from MIT in 1972. He is currently Senior Scientist at AudioLogic in Boulder, Colorado, where he is developing signal processing for a new digital hearing aid. Prior to joining AudioLogic, he was with the Center for Research in Speech and Hearing Sciences of the City University of New York. His research interests at CUNY included directional microphone arrays for hearing aids, feedback cancellation strategies, signal processing for hearing aid test and evaluation, procedures for measuring sound quality in hearing aids, speech enhancement algorithms for the hearing-impaired, new procedures for fitting hearing aids, and modeling normal and impaired cochlear function. He also held an appointment as an Adjunct Assistant Professor in the Doctoral Program in Speech and Hearing Sciences at CUNY, where he taught a course in modeling auditory physiology and perception. Previously, he has worked on applied research for hearing aids (Siemens Hearing Instruments), signal processing for radar, speech, and hearing applications (SIGNATRON, Inc.), and loudspeaker design and signal processing for audio applications (Acoustic Research and CBS Laboratories). He has over three dozen published papers and holds eight patents.

Jean Laroche was born in Bordeaux, France, in 1963. He earned a degree in Mathematics and Sciences from the Ecole Polytechnique in 1986, and a Ph.D. degree in Digital Signal Processing from the Ecole Nationale des Télécommunications in 1989. He was a post-doc student at the Center for Music Experiment at UCSD in 1990, and came back to the Ecole Nationale des Télécommunications in 1991 where he taught audio DSP, and acoustics. Since 1996 he has been a researcher in audio/music DSP at the Joint Emu/Creative Technology Center in Scotts Valley, CA.

Robert J. McAulay was born in Toronto, Ontario, Canada on October 23, 1939. He received the B.A.Sc. degree in Engineering Physics with honors from the University of Toronto, in 1962; the M.Sc. degree in Electrical Engineering from the University of Illinois, Urbana in 1963; and the Ph.D. degree in Electrical Engineering from the University of California, Berkeley, in 1967. He joined the Radar Signal Processing Group of the Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, MA, where he worked on problems in estimation theory and signal/filter design using optimal control techniques. From 1970 until 1975, he was a member of the Air Traffic Control Division at Lincoln Laboratory, and worked on the development of aircraft tracking algorithms, optimal MTI digital signal processing and on problems of aircraft direction finding for the Discrete Address Beacon System. On a leave of absence from Lincoln Laboratory during the winter and spring of 1974, he was a Visiting Associate Professor at McGill University, Montreal, P.Q., Canada. From 1975 until 1996, he was a member of the Speech Systems Technology Group at Lincoln Laboratory, where he was involved in the development of robust narrowband speech vocoders. In 1986 he served on the National Research Council panel that reviewed the problem of the removal of noise from speech. In 1987 he was appointed to the position of Lincoln Laboratory Senior Staff. On retiring from Lincoln Laboratory in 1996, he accepted the position of Senior Scientist at Voxware to develop high-quality speech products for the Internet. In 1978 he received the M. Barry Carlton Award for the best paper published in the IEEE Transactions on Aerospace and Electronic Systems for the paper "Interferometer Design for Elevation Angle Estimation". In 1990 he received the IEEE Signal Processing Society's Senior Award for the paper "Speech Analysis/Synthesis Based on a Sinusoidal Representation", published in the IEEE Transactions on Acoustics, Speech and Signal Processing.

Dana C. Massie studied electronic music synthesis and composition at Virginia Commonwealth University in Richmond Virginia, and electrical engineering at Virginia Polytechnic Institute and State University in Blacksburg, VA. He worked in professional analog recording console and digital telecom systems design at Datatronix, Inc., in Reston, VA from 1981 through 1983. He then moved to E-mu Systems, Inc., in California, to design DSP algorithms and architectures for electronic music. After brief stints at NeXT Computer, Inc. and WaveFrame, Inc., developing MultiMedia DSP applications, he returned to E-mu Systems to work in digital filter design, digital reverberation design, and advanced music synthesis algorithms. He is now the Director of the Joint E-mu/Creative Technology Center, in Scotts Valley, California. The "Tech Center" develops advanced audio technologies for both E-mu Systems and Creative Technology, Limited in Singapore, including VLSI designs, advanced music synthesis algorithms, 3D audio algorithms, and software tools.

Thomas F. Quatieri was born in Somerville, Massachusetts on January 31, 1952. He received the B.S. degree from Tufts University, Medford, Massachusetts in 1973, and the SM., E.E., and Sc.D. degrees from the Massachusetts Institute of Technology (M.I.T.), Cambridge, Massachusetts in 1975, 1977, and 1979, respectively. He is currently a senior research staff member at M.I.T. Lincoln Laboratory, Lexington, Massachusetts. In 1980, he joined the Sensor Processing Technology Group of M.I.T., Lincoln Laboratory, Lexington, Massachusetts where he worked on problems in multi-dimensional digital signal processing and image processing. Since 1983 he has been a member of the Speech Systems Technology Group at Lincoln Laboratory where he has been involved in digital signal processing for speech and audio applications, underwater sound enhancement, and data communications. He has contributed many publications to journals and conference proceedings, written several patents, and co-authored chapters in numerous edited books including: *Advanced Topics in Signal Processing* (Prentice Hall, 1987), *Advances in Speech Signal Processing* (Marcel Dekker, 1991), and *Speech Coding and Synthesis* (Elsevier, 1995). He holds the position of Lecturer at MIT where he has developed the graduate course *Digital Speech Processing*, and is active in advising graduate students on the MIT campus. Dr. Quatieri is the recipient of the 1982 Paper Award of the IEEE Acoustics, Speech and Signal Processing Society for the paper, "Implementation of 2-D Digital Filters by Iterative Methods". In 1990, he received the IEEE Signal Processing Society's Senior Award for the paper, "Speech Analysis/Synthesis Based on a Sinusoidal Representation", published in the *IEEE Transactions on Acoustics, Speech and Signal Processing*, and in 1994 won this same award for the paper "Energy Separation in Signal Modulations with Application to Speech Analysis" which was also selected for the 1995 IEEE W.R.G. Baker Prize Award. He was a member of the IEEE Digital Signal Processing Technical Committee, from 1983 to 1992 served on the steering committee for the bi-annual Digital Signal Processing Workshop, and was Associate Editor for the *IEEE Transactions on Signal Processing* in the area of nonlinear systems.

Peter J.W. Rayner received the M.A. degree from Cambridge University, U.K., in 1968 and the Ph. D. degree from Aston University in 1969. Since 1968 he has been with the Department of Engineering at Cambridge University and is Head of the Signal Processing and Communications Research Group. In 1990 he was appointed to an ad-hominem Readership in Information Engineering. He teaches course in random signal theory, digital signal processing, image processing and communication systems. His current research interests include image sequence restoration, audio restoration, non-linear estimation and detection and time series modelling and classification.

Julius O. Smith received the B.S.E.E. degree from Rice University, Houston, TX, in 1975. He received the M.S. and Ph.D. degrees from Stanford University, Stanford, CA,

in 1978 and 1983, respectively. His Ph.D. research involved the application of digital signal processing and system identification techniques to the modeling and synthesis of the violin, clarinet, reverberant spaces, and other musical systems. From 1975 to 1977 he worked in the Signal Processing Department at ESL in Sunnyvale, CA, on systems for digital communications. From 1982 to 1986 he was with the Adaptive Systems Department at Systems Control Technology in Palo Alto, CA, where he worked in the areas of adaptive filtering and spectral estimation. From 1986 to 1991 he was employed at NeXT Computer, Inc., responsible for sound, music, and signal processing software for the NeXT computer workstation. Since then he has been an Associate Professor at the Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, teaching courses in signal processing and music technology, and pursuing research in signal processing techniques applied to musical instrument modeling, audio spectral modeling, and related topics.

INTRODUCTION

Karlheinz Brandenburg and Mark Kahrs

With the advent of multimedia, digital signal processing (DSP) of sound has emerged from the shadow of bandwidth-limited speech processing. Today, the main applications of audio DSP are high quality audio coding and the digital generation and manipulation of music signals. They share common research topics including perceptual measurement techniques and analysis/synthesis methods. Smaller but nonetheless very important topics are hearing aids using signal processing technology and hardware architectures for digital signal processing of audio. In all these areas the last decade has seen a significant amount of application oriented research.

The topics covered here coincide with the topics covered in the biannual workshop on “Applications of Signal Processing to Audio and Acoustics”. This event is sponsored by the IEEE Signal Processing Society (Technical Committee on Audio and Electroacoustics) and takes place at Mohonk Mountain House in New Paltz, New York.

A short overview of each chapter will illustrate the wide variety of technical material presented in the chapters of this book.

John Beerends: Perceptual Measurement Techniques. The advent of perceptual measurement techniques is a byproduct of the advent of digital coding for both speech and high quality audio signals. Traditional measurement schemes are bad estimates for the subjective quality after digital coding/decoding. Listening tests are subject to statistical uncertainties and the basic question of repeatability in a different environment. John Beerends explains the reasons for the development of perceptual measurement techniques, the psychoacoustic fundamentals which apply to both perceptual measurement and perceptual coding and explains some of the more advanced techniques which have been developed in the last few years. Completed and ongoing standardization efforts concludes his chapter. This is recommended reading not only to people interested in perceptual coding and measurement but to anyone who wants to know more about the psychoacoustic fundamentals of digital processing of sound signals.

Karlheinz Brandenburg: Perceptual Coding of High Quality Digital Audio. High quality audio coding is rapidly progressing from a research topic to widespread applications. Research in this field has been driven by a standardization process within the Motion Picture Experts Group (MPEG). The chapter gives a detailed introduction of the basic techniques including a study of filter banks and perceptual models. As the main example, MPEG Audio is described in full detail. This includes a description of the new MPEG-2 Advanced Audio Coding (AAC) standard and the current work on MPEG-4 Audio.

William G. Gardner: Reverberation Algorithms. This chapter is the first in a number of chapters devoted to the digital manipulation of music signals. Digitally generated reverb was one of the first application areas of digital signal processing to high quality audio signals. Bill Gardner gives an in depth introduction to the physical and perceptual aspects of reverberation. The remainder of the chapter treats the different types of artificial reverberators known today. The main quest in this topic is to generate natural sounding reverb with low cost. Important milestones in the research, various historic and current types of reverberators are explained in detail.

Simon Godsill, Peter Rayner and Olivier Cappé: Digital Audio Restoration. Digital signal processing of high quality audio does not stop with the synthesis or manipulation of new material: One of the early applications of DSP was the manipulation of sounds from the past in order to restore them for recording on new or different media. The chapter presents the different methods for removing clicks, noise and other artifacts from old recordings or film material.

Mark Kahrs: Digital Audio System Architecture. An often overlooked part of the processing of high quality audio is the system architecture. Mark Kahrs introduces current technologies both for the conversion between analog and digital world and the processing technologies. Over the years there is a clear path from specialized hardware architectures to general purpose computing engines. The chapter covers specialized hardware architectures as well as the use of generally available DSP chips. The emphasis is on high throughput digital signal processing architectures for music synthesis applications.

James M. Kates: Signal Processing for Hearing Aids. A not so obvious application area for audio signal processing is the field of hearing aids. Nonetheless this field has seen continuous research activities for a number of years and is another field where widespread application of digital technologies is under preparation today. The chapter contains an in-depth treatise of the basics of signal processing for hearing aids including the description of different types of hearing loss, simpler amplification

and compression techniques and current research on multi-microphone techniques and cochlear implants.

Jean Laroche: Time and Pitch Scale Modification of Audio Signals. One of the conceptionally simplest problems of the manipulation of audio signals is difficult enough to warrant ongoing research for a number of years: Jean Laroche explains the basics of time and pitch scale modification of audio signals for both speech and musical signals. He discusses both time domain and frequency domain methods including methods specially suited for speech signals.

Dana C. Massie: Wavetable Sampling Synthesis. The most prominent example today of the application of high quality digital audio processing is wavetable sampling synthesis. Tens of millions of computer owners have sound cards incorporating wavetable sampling synthesis. Dana Massie explains the basics and modern technologies employed in sampling synthesis.

T.F. Quatieri and R.J. McAulay: Audio Signal Processing Based on Sinusoidal Analysis/Synthesis. One of the basic paradigms of digital audio analysis, coding (i.e. analysis/synthesis) and synthesis systems is the sinusoidal model. It has been used for many systems from speech coding to music synthesis. The chapter contains the unified view of both the basics of sinusoidal analysis/synthesis and some of the applications.

Julius O. Smith III: Principles of Digital Waveguide Models of Musical Instruments. This chapter describes a recent research topic in the synthesis of music instruments: Digital waveguide models are one method of physical modeling. As in the case of the Vocoder for speech, a model of an existing or hypothetical instrument is used for the sound generation. In the tutorial the vibrating string is taken as the principle illustrative example. Another example using the same underlying principles is the acoustic tube. Complicated instruments are derived by adding signal scattering and reed-bore or bow-string interactions.

Summary This book was written to serve both as a text book for an advanced graduate course on digital signal processing for audio or as a reference book for the practicing engineer. We hope that this book will stimulate further research and interest in this fascinating and exciting field.