

Contents

Foreword

page xv

Preface

xix

Part I Sound Analysis and Representation Overview

1	Introduction	5
1.1	On Vision and Hearing <i>à la</i> David Marr	8
1.2	Top-Down versus Bottom-Up Analysis	11
1.3	The Neuromimetic Approach	13
1.4	Auditory Images	14
1.5	The Ear as a Frequency Analyzer?	16
1.6	The Third Sound	18
1.7	Sound Understanding and Extraction of Meaning	18
1.8	Leveraging Techniques from Machine Vision and Machine Learning	19
1.9	Machine Hearing Systems “by the Book”	20
2	Theories of Hearing	23
2.1	A “New” Theory of Hearing	23
2.2	Newer Theories of Hearing	26
2.3	Active and Nonlinear Theories of Hearing	27
2.4	Three Auditory Theories	28
2.5	The Auditory Image Theory of Hearing	29
3	On Logarithmic and Power-Law Hearing	33
3.1	Logarithms and Power Laws	33
3.2	Log Frequency	35
3.3	Log Power	37
3.4	Bode Plots	38
3.5	Perceptual Mappings	41
3.6	Constant- Q Analysis	44
3.7	Use Logarithms with Caution	44

4	Human Hearing Overview	46
4.1	Human versus Machine	46
4.2	Auditory Physiology	46
4.3	Key Problems in Hearing	48
4.4	Loudness	50
4.5	Critical Bands, Masking, and Suppression	52
4.6	Pitch Perception	56
4.7	Timbre	65
4.8	Consonance and Dissonance	66
4.9	Speech Perception	69
4.10	Binaural Hearing	72
4.11	Auditory Streaming	74
4.12	Nonlinearity	75
4.13	A Way Forward	76
5	Acoustic Approaches and Auditory Influence	78
5.1	Sound, Speech, and Music Modeling	78
5.2	Short-Time Spectral Analysis	79
5.3	Smoothing and Transformation of Spectra	83
5.4	The Source–Filter Model and Homomorphic Signal Processing	85
5.5	Backing Away from Logarithms	88
5.6	Auditory Frequency Scales	88
5.7	Mel-Frequency Cepstrum	89
5.8	Linear Predictive Coding	91
5.9	PLP and RASTA	92
5.10	Auditory Techniques in Automatic Speech Recognition	93
5.11	Improvements Needed	94
Part II	Systems Theory for Hearing	95
6	Introduction to Linear Systems	97
6.1	Smoothing: A Good Place to Start	98
6.2	Linear Time-Invariant Systems	99
6.3	Filters and Frequencies	101
6.4	Differential Equations and Homogeneous Solutions	103
6.5	Impulse Responses	103
6.6	Causality and Stability	105
6.7	Convolution	106
6.8	Eigenfunctions and Transfer Functions	107
6.9	Frequency Response	111
6.10	Transforms and Operational Methods	113
6.11	Rational Functions, and Their Poles and Zeros	116

6.12	Graphical Computation of Transfer Function Gain and Phase	119
6.13	Convolution Theorem	120
6.14	Interconnection of Filters in Cascade, Parallel, and Feedback	121
6.15	Summary and Next Steps	125

7 Discrete-Time and Digital Systems 126

7.1	Simulating Systems in Computers	126
7.2	Discrete-Time Linear Shift-Invariant Systems	126
7.3	Impulse Response and Convolution	127
7.4	Frequency in Discrete-Time Systems	127
7.5	Z Transform and Its Inverse	128
7.6	Unit Advance and Unit Delay Operators	129
7.7	Filters and Transfer Functions	131
7.8	Sampling and Aliasing	134
7.9	Mappings from Continuous-Time Systems	136
7.10	Filter Design	138
7.11	Digital Filters	138
7.12	Multiple Inputs and Outputs	141
7.13	Fourier Analysis and Spectrograms	142
7.14	Perspective and Further Reading	144

8 Resonators 145

8.1	Bandpass Filters	145
8.2	Four Resonant Systems	149
8.3	Resonator Frequency Responses	152
8.4	Resonator Impulse Responses	154
8.5	The Complex Resonator and the Universal Resonance Curve	157
8.6	Complex Zeros from a Parallel System	159
8.7	Keeping It Real	163
8.8	Digital Resonators	165

9 Gammatone and Related Filters 169

9.1	Compound Resonators as Auditory Models	169
9.2	Multiple Poles	170
9.3	The Complex Gammatone Filter	172
9.4	The Real Gammatone Filter	175
9.5	All-Pole Gammatone Filters	178
9.6	Gammachirp Filters	181
9.7	Variable Pole Q	184
9.8	Noncoincident Poles	184
9.9	Digital Implementations	185

10	Nonlinear Systems	189
10.1	Volterra Series and Other Descriptions	189
10.2	Essential Nonlinearity	191
10.3	Hopf Bifurcation	192
10.4	Distributed Bandpass Nonlinearity	194
10.5	Response Curves of Nonlinear Systems	195
10.6	Two-Tone Responses	198
10.7	Nonlinearity and Aliasing	199
10.8	Cautions	201
11	Automatic Gain Control	202
11.1	Input–Output Level Compression	202
11.2	Nonlinear Feedback Control	204
11.3	AGC Compression at Equilibrium	205
11.4	Multiple Cascaded Variable-Gain Stages	207
11.5	Gain Control via Damping Control in Cascaded Resonators	209
11.6	AGC Dynamics	210
11.7	AGC Loop Stability	215
11.8	Multiple-Loop AGC	218
12	Waves in Distributed Systems	219
12.1	Waves in Uniform Linear Media	221
12.2	Transfer Functions from Wavenumbers	226
12.3	Nonuniform Media	230
12.4	Nonuniform Media as Filter Cascades	234
12.5	Impulse Responses	235
12.6	Group Velocity and Group Delay	235
Part III	The Auditory Periphery	237
13	Auditory Filter Models	239
13.1	What Is an Auditory Filter?	241
13.2	From Resonance to Gaussian Filters	243
13.3	Ten Good Properties for Auditory Filter Models	244
13.4	Representative Auditory Filter Models	246
13.5	Complications: Time-Varying and Nonlinear Auditory Filters	252
13.6	Fitting Parameters of Filter Models	255
13.7	Suppression	257
13.8	Impulse Responses from Physiological Data	260
13.9	Summary and Application to Cochlear Models	264

14	Modeling the Cochlea	265
14.1	On the Structure of the Cochlea	266
14.2	The Traveling Wave	268
14.3	1D, 2D, and 3D Hydrodynamics	273
14.4	Long Waves, Short Waves, and 2D Models	276
14.5	Active Micromechanics	279
14.6	Scaling Symmetry and the Cochlear Map	280
14.7	Filter-Cascade Cochlear Models	281
14.8	Outer Hair Cells as Active Gain Elements	284
14.9	Dispersion Relations from Mechanical Models and Experiments	287
14.10	Inner Hair Cells as Detectors	288
14.11	Adaptation to Sound via Efferent Control	288
14.12	Summary and Further Reading	291
15	The CARFAC Digital Cochlear Model	293
15.1	Putting the Pieces Together	293
15.2	The CARFAC Framework	294
15.3	Physiological Elements	294
15.4	Analog and Bidirectional Models	297
15.5	Open-Source Software	298
15.6	Detailing the CARFAC	298
16	The Cascade of Asymmetric Resonators	299
16.1	The Linear Cochlear Model	299
16.2	Coupled-Form Filter Realization	300
17	The Outer Hair Cell	309
17.1	Multiple Effects in One Mechanism	309
17.2	The Nonlinear Function	311
17.3	AGC Effect of DOHC	313
17.4	Typical Distortion Response Patterns	315
17.5	Completing the Loop	319
18	The Inner Hair Cell	320
18.1	Rectification with a Sigmoid	322
18.2	Adaptive Hair-Cell Models	324
18.3	A Digital IHC Model	328
19	The AGC Loop Filter	331
19.1	The CARFAC's AGC Loop	331
19.2	AGC Filter Structure	332

19.3	Smoothing Filter Pole–Zero Analysis	332
19.4	AGC Filter Temporal Response	335
19.5	AGC Filter Spatial Response	337
19.6	Time–Space Smoothing with Decimation	338
19.7	Adapted Behavior	341
19.8	Binaural or Multi-Ear Operation	341
19.9	Coupled and Multistage AGC in CARFAC and Other Systems	342

Part IV The Auditory Nervous System 345

20 Auditory Nerve and Cochlear Nucleus 347

20.1	From Hair Cells to Nerve Firings	347
20.2	Tonotopic Organization	350
20.3	Fine Time Structure in Cochleagrams	351
20.4	Cell Types in the Cochlear Nucleus	352
20.5	Inhibition and Other Computation	353
20.6	Spike Timing Codes	354

21 The Auditory Image 355

21.1	Movies of Sound	355
21.2	History	356
21.3	Stabilizing the Image	357
21.4	Triggered Temporal Integration	360
21.5	Conventional Short-Time Autocorrelation	365
21.6	Asymmetry	367
21.7	Computing the SAI	367
21.8	Pitch and Spectrum	369
21.9	Auditory Images of Music	369
21.10	Auditory Images of Speech	369
21.11	Summary SAI Tracks: Pitchograms	371
21.12	Cochleagram from SAI	373
21.13	The Log-Lag SAI	376

22 Binaural Spatial Hearing 379

22.1	Rayleigh's Duplex Theory: Interaural Level and Phase	379
22.2	Interaural Time and Level Differences	385
22.3	The Head-Related Transfer Function	386
22.4	Neural Extraction of Interaural Differences	389
22.5	The Role of the Cochlear Nucleus and the Trapezoid Body	392
22.6	Binaural Acoustic Reflex and Gain Control	394
22.7	The Precedence Effect	395

	22.8	Completing the Model	397
	22.9	Interaural Coherence	397
	22.10	Binaural Applications	398
23		The Auditory Brain	400
	23.1	Scene Analysis: ASA and CASA	400
	23.2	Attention and Stream Segregation	402
	23.3	Stages in the Brain	407
	23.4	Higher Auditory Pathways	410
	23.5	Prospects	415
Part V		Learning and Applications	417
24		Neural Networks for Machine Learning	419
	24.1	Learning from Data	419
	24.2	The Perceptron	420
	24.3	The Training Phase	421
	24.4	Nonlinearities at the Output	423
	24.5	Nonlinearities at the Input	426
	24.6	Multiple Layers	428
	24.7	Neural Units and Neural Networks	428
	24.8	Training by Error Back-Propagation	429
	24.9	Cost Functions and Regularization	432
	24.10	Multiclass Classifiers	434
	24.11	Neural Network Successes and Failures	436
	24.12	Statistical Learning Theory	437
	24.13	Summary and Perspective	439
25		Feature Spaces	441
	25.1	Feature Engineering	442
	25.2	Automatic Feature Optimization by Deep Networks	443
	25.3	Bandpass Power and Quadratic Features	444
	25.4	Quadratic Features of Cochlear Filterbank Outputs	445
	25.5	Nonlinearities and Gain Control in Feature Extraction	446
	25.6	Neurally Inspired Feature Extraction	448
	25.7	Sparsification and Winner-Take-All Features	448
	25.8	Which Approach Will Win?	449
26		Sound Search	450
	26.1	Modeling Sounds	451
	26.2	Ranking Sounds Given Text Queries	457
	26.3	Experiments	461

26.4	Results	463
26.5	Conclusions and Followup	465
27	Musical Melody Matching	467
27.1	Algorithm	469
27.2	Experiments	475
27.3	Discussion	478
27.4	Summary and Conclusions	480
28	Other Applications	481
28.1	Auditory Physiology and Psychoacoustics	481
28.2	Audio Coding and Compression	482
28.3	Hearing Aids and Cochlear Implants	483
28.4	Visible Sound	489
28.5	Diagnosis	491
28.6	Speech and Speaker Recognition	493
28.7	Music Information Retrieval	493
28.8	Security, Surveillance, and Alarms	494
28.9	Diarization, Summarization, and Indexing	495
28.10	Have Fun	495
	<i>Bibliography</i>	497
	<i>Author Index</i>	545
	<i>Subject Index</i>	557