

Contents

1. Introduction	1
1.1 Speech: Natural and Artificial	1
1.2 Voice Coders	2
1.3 Voiceprints for Combat and for Fighting Crime	4
1.4 The Electronic Secretary	7
1.5 The Human Voice as a Key	8
1.6 Clipped Speech	9
1.7 Frequency Division	11
1.8 The First Circle of Hell: Speech in the Soviet Union	12
1.9 Linking Fast Trains to the Telephone Network	13
1.10 Digital Decapitation	14
1.11 Man into Woman and Back	16
1.12 Reading Aids for the Blind	16
1.13 High-Speed Recorded Books	16
1.14 Spectral Compression for the Hard-of-Hearing	17
1.15 Restoration of Helium Speech	17
1.16 Noise Suppression	17
1.17 Slow Speed for Better Comprehension	19
1.18 Multiband Hearing Aids and Binaural Speech Processors	19
1.19 Improving Public Address Systems	20
1.20 Raising Intelligibility in Reverberant Spaces	20
1.21 Conclusion	21
2. A Brief History of Speech	23
2.1 Animal Talk	23
2.2 Wolfgang Ritter von Kempelen	24
2.3 From Kratzenstein to Helmholtz	26
2.4 Helmholtz and Rayleigh	27
2.5 The Bells:	
Alexander Melville and Alexander Graham Bell	28
2.6 Modern Times	29
2.7 The Vocal Tract	30
2.8 Articulatory Dynamics	31
2.9 The Vocoder and Some of Its Progeny	34

2.10	Formant Vocoder	35
2.11	Correlation Vocoder	36
2.12	The Voice-Excited Vocoder	36
2.13	Center Clipping for Spectrum Flattening	37
2.14	Linear Prediction	38
2.15	Subjective Error Criteria	38
2.16	Neural Networks	39
2.17	Wavelets	39
2.18	Conclusion	40
3.	Speech Recognition	
	and Speaker Identification	41
3.1	Speech Recognition	42
3.2	Dialogue Systems	44
3.3	Speaker Identification	45
3.4	Word Spotting	46
3.5	Pinpointing Disasters by Speaker Identification	47
3.6	Speaker Identification for Forensic Purposes	48
3.7	Dynamic Programming	49
3.8	Markov Models	49
3.9	Shannon's Outguessing Machine	
	– A Markov Model Analyzer	50
3.10	Hidden Markov Models in Speech Recognition	51
3.10.1	The model and algorithms	52
3.11	Neural Networks	55
3.11.1	The Perceptron	56
3.11.2	Multilayer Networks	56
3.11.3	Backward Error Propagation	56
3.11.4	Kohonen Self-Organizing Maps	57
3.11.5	Hopfield Nets and Associative Memory	58
3.12	Whole Word Recognition	59
3.13	Robust Speech Recognition	59
3.14	The Modulation Transfer Function	60
4.	Speech Dialogue Systems	
	and Natural Language Processing	67
4.1	The Structure of Language	67
4.1.1	From Sound to Cognition:	
	Levels of Language Analysis and Knowledge	
	Representation	68
4.1.2	Grammars	71
4.1.3	Symbolic Processing	73
4.1.4	Statistical Processing	77
4.2	Speech Dialogue Systems	86
4.2.1	Demands of a Dialogue System	87

4.2.2	Architecture and Components	89
4.2.3	How to Wreck a Nice Beach	89
4.2.4	Natural Language Processing	92
4.2.5	Discourse Engine	95
4.2.6	Response Generation	101
4.2.7	Speech Synthesis	103
4.2.8	Summary	105
5.	Speech Compression	107
5.1	Vocoders	108
5.2	Digital Simulation	109
5.3	Linear Prediction	110
5.3.1	Linear Prediction and Resonances	111
5.3.2	The Innovation Sequence	115
5.3.3	Single Pulse Excitation	116
5.3.4	Multipulse Excitation	118
5.3.5	Adaptive Predictive Coding	118
5.3.6	Masking of Quantizing Noise	119
5.3.7	Instantaneous Quantizing Versus Block Coding	120
5.3.8	Delays	122
5.3.9	Code Excited Linear Prediction (CELP)	123
5.3.10	Algebraic Codes	123
5.3.11	Efficient Coding of Parameters	124
5.4	Waveform Coding	124
5.5	Transform Coding	125
5.6	Audio Compression	126
6.	Speech Synthesis	129
6.1	Model-Based Speech Synthesis	131
6.2	Synthesis by Concatenation	132
6.3	Prosody	133
7.	Speech Production	135
7.1	Sources and Filters	136
7.2	The Vocal Source	136
7.3	The Vocal Tract	139
7.3.1	Radiation from the Lips	140
7.4	The Acoustic Tube Model of the Vocal Tract	142
7.5	Discrete Time Description	146
8.	The Speech Signal	149
8.1	Spectral Envelope and Fine Structure	150
8.2	Unvoiced Sounds	150
8.3	The Voiced–Unvoiced Classification	150
8.4	The Formant Frequencies	151

9. Hearing	153
9.1 Historical Antecedents	155
9.2 Thomas Seebeck and Georg Simon Ohm	157
9.3 More on Monaural Phase Sensitivity	157
9.4 Hermann von Helmholtz and Georg von Békésy	158
9.4.1 Thresholds of Hearing	158
9.4.2 Pulsation Threshold and Continuity Effect	159
9.5 Anatomy and Basic Capabilities of the Ear	160
9.6 The Pinnae and the Outer Ear Canal	160
9.7 The Middle Ear	160
9.8 The Inner Ear	162
9.9 Mechanical to Neural Transduction	169
9.10 Some Astounding Monaural Phase Effects	171
9.11 Masking	174
9.12 Loudness	174
9.13 Scaling in Psychology	175
9.14 Pitch Perception and Uncertainty	177
10. Binaural Hearing – Listening with Both Ears	179
10.1 Directional Hearing	179
10.2 Precedence and Haas Effects	181
10.3 Vertical Localization	183
10.4 Virtual Sound Sources and Quasi-Stereophony	185
10.5 Binaural Release from Masking	188
10.6 Binaural Beats and Pitch	189
10.7 Direction and Pitch Confused	190
10.8 Pseudo-Stereophony	194
10.9 Virtual Sound Images	196
10.10 Philharmonic Hall, New York	197
10.11 The Proper Reproduction of Spatial Sound Fields	198
10.12 The Importance of Lateral Sound	200
10.13 How to Increase Lateral Sounds in Real Halls	202
10.14 Summary	205
11. Basic Signal Concepts	207
11.1 The Sampling Theorem and Some Notational Conventions	207
11.2 Fourier Transforms	208
11.3 The Autocorrelation Function	211
11.4 The Convolution Integral and the Delta Function	213
11.5 The Cross-Correlation Function and the Cross-Spectrum	215
11.5.1 A Bit of Number Theory	217
11.6 The Hilbert Transform and the Analytic Signal	218
11.7 Hilbert Envelope and Instantaneous Frequency	220
11.8 Causality and the Kramers–Kronig Relations	224
11.8.1 Anticausal Functions	225

11.8.2 Minimum-Phase Systems and Complex Frequencies . . .	226
11.8.3 Allpass Systems	227
11.8.4 Dereverberation	228
11.9 Matched Filtering	229
11.10 Phase and Group Delay	230
11.11 Heisenberg Uncertainty and The Fourier Transform	232
11.11.1 Prolate Spheroidal Wave Functions and Uncertainty	234
11.12 Time and Frequency Windows	238
11.13 The Wigner–Ville Distribution	239
11.14 The Cepstrum: Measurement of Fundamental Frequency	241
11.15 Line Spectral Frequencies	244
A. Acoustic Theory and Modeling of the Vocal Tract	247
A.1 Introduction	247
A.2 Acoustics of a Hard-Walled, Lossless Tube	248
A.2.1 Field Equations	248
A.2.2 Time-Invariant Case	252
A.2.3 Formants as Eigenvalues	253
A.2.4 Losses and Nonrigid Walls	255
A.3 Discrete Modeling of a Tube	257
A.3.1 Time-Domain Modeling	257
A.3.2 Frequency-Domain Modeling, Two-Port Theory	260
A.3.3 Tube Models and Linear Prediction	263
A.4 Notes on the Inverse Problem	265
A.4.1 Analytic and Numerical Methods	265
A.4.2 Empirical Methods	268
B. Direct Relations	
Between Cepstrum and Predictor Coefficients	269
B.1 Derivation of the Main Result	269
B.2 Direct Computation of Predictor Coefficients from the Cepstrum	271
B.3 A Simple Check	272
B.4 Connection with Algebraic Roots and Symmetric Functions	272
B.5 Connection with Statistical Moments and Cumulants	274
B.6 Computational Complexity	274
B.7 An Application of Root-Power Sums to Pitch Detection	275
References	279
General Reading	297
Selected Journals	307
A Sampling of Societies and Major Meetings	308

XXXIV Contents

Glossary of Speech and Computer Terms	309
Name Index	339
Subject Index	349
The Author	377

Computer Speech

Recognition, Compression, Synthesis

Schroeder, M.R.

2004, XXXIV, 379 p., Hardcover

ISBN: 978-3-540-21267-6