

CONTENTS

1	Introduction	1
1.1	What Is Speech Communication?	1
1.2	History of Developments in Speech Communication	1
1.3	Outline of the Book	2
	Production	3
	Perception	3
	Analysis	4
	Coding	5
	Synthesis	6
	Recognition	6
1.4	Other Topics	7
2	Review of Mathematics for Speech Processing	9
2.1	Mathematical Preliminaries	9
	Number Representations	9
	Matrix Arithmetic	10
2.2	Signals and Linear Systems	12
	Simple Signals	13
	Filtering and Convolution	16
2.3	Frequency Analysis	17
	Fourier Transform	18
	Spectra and Correlation	19
	Laplace Transform: Poles and Zeros	19
2.4	Circuits	20
2.5	Discrete-Time Signals and Systems	21
	Sampling	22
	Frequency Transforms of Discrete-Time Signals	23
	Decimation and Interpolation	26
2.6	Filters	27
	Bandpass Filters	28
	Digital Filters	29
	Difference Equations and Filter Structures	30

2.7	Probability and Statistics	33
	Probability Densities and Histograms	34
	Averages and Variances	35
	Gaussian Probability Density	36
	Joint Probability	36
	Noise	37
2.8	Summary	38
3	Speech Production and	
	Acoustic-Phonetics	39
3.1	Introduction	39
3.2	Anatomy and Physiology of Speech Organs	42
	The Lungs and the Thorax	42
	Larynx and Vocal Folds	43
	Vocal Tract	50
3.3	Articulatory Phonetics	55
	Manner of Articulation	57
	Structure of the English Syllable	58
	Voicing	59
	Place of Articulation	60
	Phonemes in Other Languages	62
3.4	Acoustic-Phonetics	63
	Spectrograms	64
	Vowels	65
	Diphthongs	68
	Glides and Liquids	68
	Nasals	70
	Fricatives	71
	Stops (Plosives)	73
	Variants of Normal Speech	78
3.5	Acoustic Theory of Speech Production	79
	Acoustics of the Excitation Source	79
	Acoustics of the Vocal Tract	81
	Transmission Line Analog of the Vocal Tract	91
	Effects of Losses in the Vocal Tract	99
	Radiation at the Lips	99
	Model of Glottal Excitation	101
	Quantal Theory of Speech Production	102

3.6	Practical Vocal Tract Models for Speech Analysis and Generation	103
	Articulatory Model	103
	Terminal-Analog Model	109
3.7	Coarticulation	111
	Where Does Coarticulation Occur?	112
	Coarticulation Effects for Different Articulators	113
	Invariant Features	115
	Effects of Coarticulation on Duration	117
	Models for Coarticulation	117
3.8	Prosody (Suprasegmentals)	118
	Duration	119
	Effects of Stress and Speaking Rate	120
	Fundamental Frequency (F0)	121
	Problems	123
4	Hearing	128
4.1	Introduction	128
4.2	Anatomy and Physiology of the Ear	128
	Outer Ear	129
	Middle Ear	130
	Inner Ear	130
	Basilar Membrane Behavior	132
	Electrical Activity in the Auditory Neurons	134
4.3	Sound Perception	140
	Auditory Psychophysics	141
	Thresholds	141
	Just-noticeable Differences	143
	Pitch Perception	145
	Masking	146
	Critical Bands	148
	Nonsimultaneous or Temporal Masking (‡)	151
	Masking Due to Complex Stimuli	153
	Origins of Masking	154
4.4	Response of the Ear to Complex Stimuli	154
	Speech Stimuli (‡)	155
	Adaptation (‡)	156
	Just-noticeable Differences in Speech	157

	Timing	159
4.5	Summary	161
	Problems	161
5	Speech Perception	164
5.1	Introduction	164
5.2	Perceptually Important Features of Speech Signals	165
	Synthetic vs. Natural Speech	166
	Redundancy in Speech	167
5.3	Models of Speech Perception	168
	Categorical Perception	168
	Distinctive Features	169
	Active Models	171
	Passive Models	173
5.4	Vowel Perception	175
	Perceived Formant Location in Synthetic Vowels	175
	Context Normalization	176
	Coarticulation Effects	177
5.5	Consonant Perception	178
	Perception of the Manner of Articulation Feature	178
	Perception of Place of Articulation	180
	Perception of Voicing in Obstruents	185
5.6	Duration as a Phonemic Cue	188
	Manner Cues	188
	Place Cues	189
	Speaking Rate Effects	190
5.7	Intonation: Perception of Prosody	191
	Stress: Lexical and Sentential	192
	Acoustic Correlates of Stress	193
	Perception of Syntactic Features	195
	Perceptually Relevant Pitch Movements	198
5.8	Other Aspects of Speech Perception (‡)	199
	Adaptation Studies	199
	Dichotic Studies	199
	Phase Effects	200
	Word and Syllable Effects	200
	Perception of Distorted Speech	201

Speech Perception by the Handicapped	202
5.9 Summary	202
Problems	203
6 Speech Analysis	204
6.1 Introduction	204
6.2 Short-Time Speech Analysis	205
Windowing	206
Spectra of Windows	208
6.3 Time-Domain Parameters	211
Time-Domain Analysis	211
Short-Time Average Energy and Magnitude	213
Short-Time Average Zero-Crossing Rate	213
Short-Time Autocorrelation Function	216
6.4 Frequency-Domain Parameters	219
Filter Bank Analysis	219
Short-Time Fourier Analysis	220
Spectral Displays	222
Formant Estimation and Tracking	224
6.5 Homomorphic (Cepstral) Analysis (‡)	226
Mathematical Details	227
Applications	229
6.6 F0 (“Pitch”) Estimation	232
Time-Domain Techniques	233
Short-Time Spectral Techniques	234
6.7 Reduction of Information	237
Taking Advantage of Gradual Vocal Tract Motion	238
Smoothing: Linear and Nonlinear	239
6.8 Summary	240
Problems	241
7 Coding of Speech Signals	243
7.1 Introduction	243
7.2 Quantization	246
Quantization Error or Noise	247
Bit Protection	250
Signal-to-Noise Ratio (SNR)	251

	Nonuniform Quantization	252
	Relationship of Bandwidth and Noise to Coding Rate.....	256
7.3	Speech Redundancies and Quality Measures	256
7.4	Time-Adaptive Waveform Coding	258
	Design Choices	258
	Feedforward vs. Feedback Adaptation	259
7.5	Exploiting Properties of the Spectral Envelope	262
	A Measure of Spectral Redundancy	264
	DPCM Details	265
	Linear Prediction	266
	Linear Delta Modulation (LDM)	268
	Adaptive Delta Modulation (ADM)	271
	Adaptive Differential Pulse-Code Modulation (ADPCM)	273
7.6	Exploiting the Periodicity of Voiced Speech	274
	Pitch-Adaptive Predictive Coders (APC)	276
	Higher-Order Pitch Prediction (‡).....	277
	Combining Pitch and Envelope Prediction.....	278
7.7	Exploiting Auditory Limitations	279
7.8	Spectral (Frequency-Domain) Coders	281
	Filter Bank Analysis	282
	Subband Vocoders (SBC)	283
	Adaptive Transform Coders (ATC).....	288
	Harmonic Coding	293
7.9	Vocoders	302
	Phase Vocoder.....	303
	Evaluating Quality	305
	Channel Vocoders	305
	Excitation for Vocoders	307
	Homomorphic (Cepstral) Vocoder (‡)	309
	Mozier Coding (‡)	310
	Other Vocoders (‡).....	311
7.10	Vector Quantization (VQ)	313
	Codebooks	314
	Codebook Design	315
	Searching the Codebook	316
	Gain/Shape Vector Quantization.....	318

Other Types of Vector Quantization	320
7.11 Network Considerations	323
Packet Transmission	324
Time Assignment Speech Interpolation (TASI)	324
Embedded Coding	325
Tandeming of Coders	327
7.12 Hardware Implementation: Integrated Circuits	328
7.13 Summary	330
Problems	331
8 Linear Predictive Coding	336
8.1 Introduction	336
8.2 Basic Principles	337
LPC Model	337
Least-Squares Autocorrelation Method	339
Least-Squares Covariance Method	342
Computational Considerations	343
8.3 Spectral Estimation via LPC	345
Pre-Emphasis	345
Order of the LPC Model	346
8.4 Updating the LPC Model Sample by Sample(‡)	349
Transversal Predictors	349
Lattice LPC Models	350
8.5 Window Considerations	354
8.6 Equivalent Forms for LPC Coefficients	355
Line Spectrum Pair (‡)	357
8.7 Parameter Updating and Transmission	358
Variable-Frame-Rate Transmission	358
Transmission Details	359
Time-Varying LPC Coefficients (‡)	359
8.8 Modifications to Standard LPC(‡)	360
Emphasizing Low Frequencies	361
Pole-Zero LPC Models	363
8.9 Different Excitation Models	364
Residual-Excited Linear Predictive (RELP) Vocoders	365
Multipulse-Excited LPC	370

8.10	Vector Quantization (VQ)	374
	Very-Low-Rate LPC Vocoders	375
8.11	Hardware Implementation: Integrated Circuits	376
8.12	Summary	377
	Problems	378
9	Speech Synthesis	380
9.1	Introduction	380
9.2	Principles of Speech Synthesis	381
	Size of the Stored Speech Unit	381
	Synthesis Method	386
	Limited Text (Voice Response) Systems	386
	Unrestricted Text (Text-to-Speech) Systems	387
9.3	Synthesizer Operation	389
	Articulatory Synthesis (‡)	389
	Formant Synthesis	392
	Specifying the Parameter Trajectories	397
	Intraframe Parameter Updating (‡)	400
	Linear Predictive Coding (LPC) Synthesis	402
	Excitation Modeling (‡)	402
	Synthesis of Intonation	406
9.4	Speech Synthesis in Other Languages(‡)	409
9.5	Hardware for Speech Synthesis(‡)	410
9.6	Summary	411
	Problems	411
10	Speech Recognition	413
10.1	Introduction	413
	Performance Evaluation	414
	System Evaluation	415
10.2	Basic Pattern Recognition Approach	415
10.3	Preprocessing	419
10.4	Parametric Representation	420
	Typical Parameters Used in Recognition	422
	Feature Extraction	423
10.5	Similarity and Distance Measures	424

Euclidean and Mahalanobis Distances (‡)	425
LPC Distance Measures (‡)	427
Decision Rules	429
Thresholds	430
Timing Considerations	430
10.6 Segmentation	432
Endpoint Detection	433
Segmentation of Connected-Word Speech into Words (‡)	434
Segmentation of Continuous Speech into Syllables	435
Segmentation of Continuous Speech into Phones	436
Word Spotting	438
10.7 Dynamic Time Warping (DTW)	439
DTW Details	440
Search Space Modification and Reduction (‡)	441
Difficulties with DTW	443
Applying DTW to Connected-Word Speech (‡)	446
Hybrid Segmentation-DTW Recognition (‡)	450
10.8 Search Reduction	451
Speaker Dependence	451
Clustering of Reference Templates	451
Exploiting Vocabulary Redundancies	453
Using Subword Speech Units as Templates (‡)	453
Reference Template Selection	455
Applying Vector Quantization to ASR	456
10.9 Networks for Speech Recognition	459
Search Design	460
Hidden Markov Models	460
Training Markov Models (‡)	463
Classification via HMMs (‡)	466
Comparing HMMs and DTW	466
Aiding ASR via Grammar Networks	468
10.10 Labeling Phones in the Cognitive Approach to ASR	470
Phonological Rules	473
10.11 Using Prosodics to Aid Recognition	474
10.12 Commercial Systems and Hardware	475
10.13 Summary	476

Problems	477
11 Speaker Recognition	479
11.1 Introduction	479
11.2 Verification vs. Recognition	480
11.3 Recognition Techniques	482
Template Matching and Alignment	482
Text Dependence	483
Statistical vs. Dynamic Features	484
Vector Quantization (‡)	486
Similarity and Distance Measures	489
Cepstral Analysis	491
Orthogonal LPC Parameters (‡)	491
11.4 Features that Distinguish Speakers	493
Measures of the Effectiveness of Features for Recognition	494
Techniques to Choose Features	495
Spectral Features	495
Prosodic Features	496
11.5 System Design	497
Data Collection	498
Sequential Decision Strategy	498
Multiple-Stage Recognition (‡)	499
Effects of Different Communication Channels (‡)	499
11.6 Speaker Recognition by Humans	500
11.7 Summary	502
Problems	503
REFERENCES	504
ANSWERS TO SELECTED PROBLEMS	550
INDEX	558