

TK7882
S65
D44
2000

Discrete-Time Processing of Speech Signals



John R. Deller, Jr.
Michigan State University

John H. L. Hansen
University of Colorado at Boulder

John G. Proakis
Northeastern University

IEEE Signal Processing Society, *Sponsor*



The Institute of Electrical and Electronics Engineers, Inc., New York



A JOHN WILEY & SONS, INC., PUBLICATION
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

Contents

Preface to the IEEE Edition	xvii
Preface	xix
Acronyms and Abbreviations	xxiii

I Signal Processing Background

1 Propaedeutic	3
1.0 Preamble	3
1.0.1 The Purpose of Chapter 1	3
1.0.2 Please Read This Note on Notation	4
1.0.3 For People Who Never Read Chapter 1 (and Those Who Do)	5
1.1 Review of DSP Concepts and Notation	6
1.1.1 "Normalized Time and Frequency"	6
1.1.2 Singularity Signals	9
1.1.3 Energy and Power Signals	9
1.1.4 Transforms and a Few Related Concepts	10
1.1.5 Windows and Frames	16
1.1.6 Discrete-Time Systems	20
1.1.7 Minimum, Maximum, and Mixed-Phase Signals and Systems	24
1.2 Review of Probability and Stochastic Processes	29
1.2.1 Probability Spaces	30
1.2.2 Random Variables	33
1.2.3 Random Processes	42
1.2.4 Vector-Valued Random Processes	52
1.3 Topics in Statistical Pattern Recognition	55
1.3.1 Distance Measures	56
1.3.2 The Euclidean Metric and "Prewhitening" of Features	58

1.3.3	Maximum Likelihood Classification	63
1.3.4	Feature Selection and Probabilistic Separability Measures	66
1.3.5	Clustering Algorithms	70
1.4	Information and Entropy	73
1.4.1	Definitions	73
1.4.2	Random Sources	77
1.4.3	Entropy Concepts in Pattern Recognition	78
1.5	Phasors and Steady-State Solutions	79
1.6	Onward to Speech Processing	81
1.7	Problems	85
	Appendices: Supplemental Bibliography	90
1.A	Example Textbooks on Digital Signal Processing	90
1.B	Example Textbooks on Stochastic Processes	90
1.C	Example Textbooks on Statistical Pattern Recognition	91
1.D	Example Textbooks on Information Theory	91
1.E	Other Resources on Speech Processing	92
1.E.1	Textbooks	92
1.E.2	Edited Paper Collections	92
1.E.3	Journals	92
1.E.4	Conference Proceedings	93
1.F	Example Textbooks on Speech and Hearing Sciences	93
1.G	Other Resources on Artificial Neural Networks	94
1.G.1	Textbooks and Monographs	94
1.G.2	Journals	94
1.G.3	Conference Proceedings	95

II Speech Production and Modeling

2 Fundamentals of Speech Science

99

2.0	Preamble	99
2.1	Speech Communication	100
2.2	Anatomy and Physiology of the Speech Production System	101
2.2.1	Anatomy	101

2.2.2	The Role of the Vocal Tract and Some Elementary Acoustical Analysis	104
2.2.3	Excitation of the Speech System and the Physiology of Voicing	110
2.3	Phonemics and Phonetics	115
2.3.1	Phonemes Versus Phones	115
2.3.2	Phonemic and Phonetic Transcription	116
2.3.3	Phonemic and Phonetic Classification	117
2.3.4	Prosodic Features and Coarticulation	137
2.4	Conclusions	146
2.5	Problems	146

3 Modeling Speech Production

151

3.0	Preamble	151
3.1	Acoustic Theory of Speech Production	151
3.1.1	History	151
3.1.2	Sound Propagation	156
3.1.3	Source Excitation Model	159
3.1.4	Vocal-Tract Modeling	166
3.1.5	Models for Nasals and Fricatives	186
3.2	Discrete-Time Modeling	187
3.2.1	General Discrete-Time Speech Model	187
3.2.2	A Discrete-Time Filter Model for Speech Production	192
3.2.3	Other Speech Models	197
3.3	Conclusions	200
3.4	Problems	201
3.A	Single Lossless Tube Analysis	203
3.A.1	Open and Closed Terminations	203
3.A.2	Impedance Analysis, T-Network, and Two-Port Network	206
3.B	Two-Tube Lossless Model of the Vocal Tract	211
3.C	Fast Discrete-Time Transfer Function Calculation	217

III Analysis Techniques

4 Short-Term Processing of Speech

225

4.1	Introduction	225
-----	--------------	-----

4.2	Short-Term Measures from Long-Term Concepts	226
4.2.1	Motivation	226
4.2.2	"Frames" of Speech	227
4.2.3	Approach 1 to the Derivation of a Short-Term Feature and Its Two Computational Forms	227
4.2.4	Approach 2 to the Derivation of a Short-Term Feature and Its Two Computational Forms	231
4.2.5	On the Role of " $1/N$ " and Related Issues	234
4.3	Example Short-Term Features and Applications	236
4.3.1	Short-Term Estimates of Autocorrelation	236
4.3.2	Average Magnitude Difference Function	244
4.3.3	Zero Crossing Measure	245
4.3.4	Short-Term Power and Energy Measures	246
4.3.5	Short-Term Fourier Analysis	251
4.4	Conclusions	262
4.5	Problems	263

5 Linear Prediction Analysis

266

5.0	Preamble	266
5.1	Long-Term LP Analysis by System Identification	267
5.1.1	The All-Pole Model	267
5.1.2	Identification of the Model	270
5.2	How Good Is the LP Model?	280
5.2.1	The "Ideal" and "Almost Ideal" Cases	280
5.2.2	"Nonideal" Cases	281
5.2.3	Summary and Further Discussion	287
5.3	Short-Term LP Analysis	290
5.3.1	Autocorrelation Method	290
5.3.2	Covariance Method	292
5.3.3	Solution Methods	296
5.3.4	Gain Computation	325
5.3.5	A Distance Measure for LP Coefficients	327
5.3.6	Preemphasis of the Speech Waveform	329
5.4	Alternative Representations of the LP Coefficients	331
5.4.1	The Line Spectrum Pair	331
5.4.2	Cepstral Parameters	333
5.5	Applications of LP in Speech Analysis	333
5.5.1	Pitch Estimation	333
5.5.2	Formant Estimation and Glottal Waveform Deconvolution	336

5.6	Conclusions	342
5.7	Problems	343
5.A	Proof of Theorem 5.1	348
5.B	The Orthogonality Principle	350

6 Cepstral Analysis

352

6.1	Introduction	352
6.2	"Real" Cepstrum	355
6.2.1	Long-Term Real Cepstrum	355
6.2.2	Short-Term Real Cepstrum	364
6.2.3	Example Applications of the stRC to Speech Analysis and Recognition	366
6.2.4	Other Forms and Variations on the stRC Parameters	380
6.3	Complex Cepstrum	386
6.3.1	Long-Term Complex Cepstrum	386
6.3.2	Short-Term Complex Cepstrum	393
6.3.3	Example Application of the stCC to Speech Analysis	394
6.3.4	Variations on the Complex Cepstrum	397
6.4	A Critical Analysis of the Cepstrum and Conclusions	397
6.5	Problems	401

IV Coding, Enhancement and Quality Assessment

7 Speech Coding and Synthesis

409

7.1	Introduction	410
7.2	Optimum Scalar and Vector Quantization	410
7.2.1	Scalar Quantization	411
7.2.2	Vector Quantization	425
7.3	Waveform Coding	434
7.3.1	Introduction	434
7.3.2	Time Domain Waveform Coding	435
7.3.3	Frequency Domain Waveform Coding	451
7.3.4	Vector Waveform Quantization	457
7.4	Vocoders	459
7.4.1	The Channel Vocoder	460
7.4.2	The Phase Vocoder	462
7.4.3	The Cepstral (Homomorphic) Vocoder	462

7.4.4	Formant Vocoders	469
7.4.5	Linear Predictive Coding	471
7.4.6	Vector Quantization of Model Parameters	485
7.5	Measuring the Quality of Speech Compression Techniques	488
7.6	Conclusions	489
7.7	Problems	490
7.A	Quadrature Mirror Filters	494

8 Speech Enhancement

501

8.1	Introduction	501
8.2	Classification of Speech Enhancement Methods	504
8.3	Short-Term Spectral Amplitude Techniques	506
8.3.1	Introduction	506
8.3.2	Spectral Subtraction	506
8.3.3	Summary of Short-Term Spectral Magnitude Methods	516
8.4	Speech Modeling and Wiener Filtering	517
8.4.1	Introduction	517
8.4.2	Iterative Wiener Filtering	517
8.4.3	Speech Enhancement and All-Pole Modeling	521
8.4.4	Sequential Estimation via EM Theory	524
8.4.5	Constrained Iterative Enhancement	525
8.4.6	Further Refinements to Iterative Enhancement	527
8.4.7	Summary of Speech Modeling and Wiener Filtering	528
8.5	Adaptive Noise Canceling	528
8.5.1	Introduction	528
8.5.2	ANC Formalities and the LMS Algorithm	530
8.5.3	Applications of ANC	534
8.5.4	Summary of ANC Methods	541
8.6	Systems Based on Fundamental Frequency Tracking	541
8.6.1	Introduction	541
8.6.2	Single-Channel ANC	542
8.6.3	Adaptive Comb Filtering	545
8.6.4	Harmonic Selection	549
8.6.5	Summary of Systems Based on Fundamental Frequency Tracking	551

8.7	Performance Evaluation	552
8.7.1	Introduction	552
8.7.2	Enhancement and Perceptual Aspects of Speech	552
8.7.3	Speech Enhancement Algorithm Performance	554
8.8	Conclusions	556
8.9	Problems	557
8.A	The INTEL System	561
8.B	Addressing Cross-Talk in Dual-Channel ANC	565

9 Speech Quality Assessment 568

9.1	Introduction	568
9.1.1	The Need for Quality Assessment	568
9.1.2	Quality Versus Intelligibility	570
9.2	Subjective Quality Measures	570
9.2.1	Intelligibility Tests	572
9.2.2	Quality Tests	575
9.3	Objective Quality Measures	580
9.3.1	Articulation Index	582
9.3.2	Signal-to-Noise Ratio	584
9.3.3	Itakura Measure	587
9.3.4	Other Measures Based on LP Analysis	588
9.3.5	Weighted-Spectral Slope Measures	589
9.3.6	Global Objective Measures	590
9.3.7	Example Applications	591
9.4	Objective Versus Subjective Measures	593
9.5	Problems	595

V Recognition

10 The Speech Recognition Problem 601

10.1	Introduction	601
10.1.1	The Dream and the Reality	601
10.1.2	Discovering Our Ignorance	604
10.1.3	Circumventing Our Ignorance	605
10.2	The "Dimensions of Difficulty"	606
10.2.1	Speaker-Dependent Versus Speaker-Independent Recognition	607
10.2.2	Vocabulary Size	607

10.2.3	Isolated-Word Versus Continuous-Speech Recognition	608	
10.2.4	Linguistic Constraints	614	
10.2.5	Acoustic Ambiguity and Confusability	619	
10.2.6	Environmental Noise	620	
10.3	Related Problems and Approaches	620	
10.3.1	Knowledge Engineering	620	
10.3.2	Speaker Recognition and Verification	621	
10.4	Conclusions	621	
10.5	Problems	621	
11	Dynamic Time Warping		623
11.1	Introduction	623	
11.2	Dynamic Programming	624	
11.3	Dynamic Time Warping Applied to IWR	634	
11.3.1	DTW Problem and Its Solution Using DP	634	
11.3.2	DTW Search Constraints	638	
11.3.3	Typical DTW Algorithm: Memory and Computational Requirements	649	
11.4	DTW Applied to CSR	651	
11.4.1	Introduction	651	
11.4.2	Level Building	652	
11.4.3	The One-Stage Algorithm	660	
11.4.4	A Grammar-Driven Connected-Word Recognition System	669	
11.4.5	Pruning and Beam Search	670	
11.4.6	Summary of Resource Requirements for DTW Algorithms	671	
11.5	Training Issues in DTW Algorithms	672	
11.6	Conclusions	674	
11.7	Problems	674	
12	The Hidden Markov Model		677
12.1	Introduction	677	
12.2	Theoretical Developments	679	
12.2.1	Generalities	679	
12.2.2	The Discrete Observation HMM	684	
12.2.3	The Continuous Observation HMM	705	
12.2.4	Inclusion of State Duration Probabilities in the Discrete Observation HMM	709	
12.2.5	Scaling the Forward-Backward Algorithm	715	

12.2.6	Training with Multiple Observation Sequences	718
12.2.7	Alternative Optimization Criteria in the Training of HMMs	720
12.2.8	A Distance Measure for HMMs	722
12.3	Practical Issues	723
12.3.1	Acoustic Observations	723
12.3.2	Model Structure and Size	724
12.3.3	Training with Insufficient Data	728
12.3.4	Acoustic Units Modeled by HMMs	730
12.4	First View of Recognition Systems Based on HMMs	734
12.4.1	Introduction	734
12.4.2	IWR Without Syntax	735
12.4.3	CSR by the Connected-Word Strategy Without Syntax	738
12.4.4	Preliminary Comments on Language Modeling Using HMMs	740
12.5	Problems	740

13 Language Modeling

745

13.1	Introduction	745
13.2	Formal Tools for Linguistic Processing	746
13.2.1	Formal Languages	746
13.2.2	Perplexity of a Language	749
13.2.3	Bottom-Up Versus Top-Down Parsing	751
13.3	HMMs, Finite State Automata, and Regular Grammars	754
13.4	A "Bottom-Up" Parsing Example	759
13.5	Principles of "Top-Down" Recognizers	764
13.5.1	Focus on the Linguistic Decoder	764
13.5.2	Focus on the Acoustic Decoder	770
13.5.3	Adding Levels to the Linguistic Decoder	772
13.5.4	Training the Continuous-Speech Recognizer	775
13.6	Other Language Models	779
13.6.1	<i>N</i> -Gram Statistical Models	779
13.6.2	Other Formal Grammars	785
13.7	IWR As "CSR"	789
13.8	Standard Databases for Speech-Recognition Research	790
13.9	A Survey of Language-Model-Based Systems	791

13.10 Conclusions	801
-------------------	-----

13.11 Problems	801
----------------	-----

14 The Artificial Neural Network	805
---	------------

14.1 Introduction	805
-------------------	-----

14.2 The Artificial Neuron	808
----------------------------	-----

14.3 Network Principles and Paradigms	813
---------------------------------------	-----

14.3.1 Introduction	813
---------------------	-----

14.3.2 Layered Networks: Formalities and Definitions	815
--	-----

14.3.3 The Multilayer Perceptron	819
----------------------------------	-----

14.3.4 Learning Vector Quantizer	834
----------------------------------	-----

14.4 Applications of ANNs in Speech Recognition	837
---	-----

14.4.1 Presegmented Speech Material	837
-------------------------------------	-----

14.4.2 Recognizing Dynamic Speech	839
-----------------------------------	-----

14.4.3 ANNs and Conventional Approaches	841
---	-----

14.4.4 Language Modeling Using ANNs	845
-------------------------------------	-----

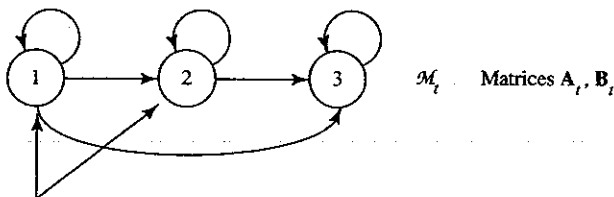
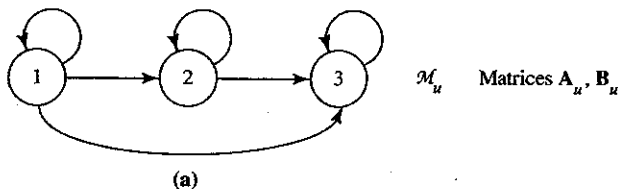
14.4.5 Integration of ANNs into the Survey Systems of Section 13.9	845
--	-----

14.5 Conclusions	846
------------------	-----

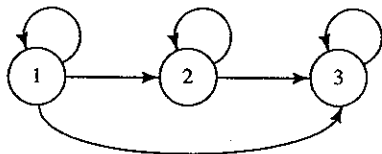
14.6 Problems	847
---------------	-----

Index	899
--------------	------------

Sample Illustration



Observation probabilities "tied" $b(k|1) = b(k|2) \forall k$.



Interpolated model

$$A = \epsilon_t A_t + (1 - \epsilon_t) A_u$$

$$B = \epsilon_t B_t + (1 - \epsilon_t) B_u$$

(c)

FIGURE 12.15. (a) A three-state HMM trained in the conventional (e.g., F-B algorithm) manner. (b) The "same" three-state HMM with tied states. (c) An "interpolated" model derived from the models of (a) and (b).