

Contents

List of Abbreviations	XXXI
1 Introduction to Speech Processing	
<i>J. Benesty, M. M. Sondhi, Y. Huang</i>	1
1.1 A Brief History of Speech Processing	1
1.2 Applications of Speech Processing	2
1.3 Organization of the Handbook	4
References	4
 Part A Production, Perception, and Modeling of Speech	
2 Physiological Processes of Speech Production	
<i>K. Honda</i>	7
2.1 Overview of Speech Apparatus	7
2.2 Voice Production Mechanisms	8
2.3 Articulatory Mechanisms	14
2.4 Summary	24
References	25
 3 Nonlinear Cochlear Signal Processing and Masking in Speech Perception	
<i>J. B. Allen</i>	27
3.1 Basics	27
3.2 The Nonlinear Cochlea	35
3.3 Neural Masking	45
3.4 Discussion and Summary	55
References	56
 4 Perception of Speech and Sound	
<i>B. Kollmeier, T. Brand, B. Meyer</i>	61
4.1 Basic Psychoacoustic Quantities	62
4.2 Acoustical Information Required for Speech Perception	70
4.3 Speech Feature Perception	74
References	81
 5 Speech Quality Assessment	
<i>V. Grancharov, W. B. Kleijn</i>	83
5.1 Degradation Factors Affecting Speech Quality	84
5.2 Subjective Tests	85
5.3 Objective Measures	90
5.4 Conclusions	95
References	96

Part B Signal Processing for Speech

6 Wiener and Adaptive Filters

<i>J. Benesty, Y. Huang, J. Chen</i>	103
6.1 Overview.....	103
6.2 Signal Models.....	104
6.3 Derivation of the Wiener Filter.....	106
6.4 Impulse Response Tail Effect.....	107
6.5 Condition Number.....	108
6.6 Adaptive Algorithms.....	110
6.7 MIMO Wiener Filter.....	116
6.8 Conclusions.....	119
References	120

7 Linear Prediction

<i>J. Benesty, J. Chen, Y. Huang</i>	121
7.1 Fundamentals.....	121
7.2 Forward Linear Prediction.....	122
7.3 Backward Linear Prediction.....	123
7.4 Levinson–Durbin Algorithm.....	124
7.5 Lattice Predictor.....	126
7.6 Spectral Representation.....	127
7.7 Linear Interpolation.....	128
7.8 Line Spectrum Pair Representation.....	129
7.9 Multichannel Linear Prediction.....	130
7.10 Conclusions.....	133
References	133

8 The Kalman Filter

<i>S. Gannot, A. Yeredor</i>	135
8.1 Derivation of the Kalman Filter.....	136
8.2 Examples: Estimation of Parametric Stochastic Process from Noisy Observations.....	141
8.3 Extensions of the Kalman Filter.....	144
8.4 The Application of the Kalman Filter to Speech Processing.....	149
8.5 Summary.....	157
References	157

9 Homomorphic Systems and Cepstrum Analysis of Speech

<i>R. W. Schafer</i>	161
9.1 Definitions.....	161
9.2 Z-Transform Analysis.....	164
9.3 Discrete-Time Model for Speech Production.....	165
9.4 The Cepstrum of Speech.....	166
9.5 Relation to LPC.....	169
9.6 Application to Pitch Detection.....	171

9.7	Applications to Analysis/Synthesis Coding	172
9.8	Applications to Speech Pattern Recognition	176
9.9	Summary	180
	References	180
10	Pitch and Voicing Determination of Speech with an Extension Toward Music Signals	
	<i>W. J. Hess</i>	181
10.1	Pitch in Time-Variant Quasiperiodic Acoustic Signals	182
10.2	Short-Term Analysis PDAs	185
10.3	Selected Time-Domain Methods	192
10.4	A Short Look into Voicing Determination	195
10.5	Evaluation and Postprocessing	197
10.6	Applications in Speech and Music	201
10.7	Some New Challenges and Developments	203
10.8	Concluding Remarks	207
	References	208
11	Formant Estimation and Tracking	
	<i>D. O'Shaughnessy</i>	213
11.1	Historical	213
11.2	Vocal Tract Resonances	215
11.3	Speech Production	216
11.4	Acoustics of the Vocal Tract	218
11.5	Short-Time Speech Analysis	221
11.6	Formant Estimation	223
11.7	Summary	226
	References	226
12	The STFT, Sinusoidal Models, and Speech Modification	
	<i>M. M. Goodwin</i>	229
12.1	The Short-Time Fourier Transform	230
12.2	Sinusoidal Models	242
12.3	Speech Modification	253
	References	256
13	Adaptive Blind Multichannel Identification	
	<i>Y. Huang, J. Benesty, J. Chen</i>	259
13.1	Overview	259
13.2	Signal Model and Problem Formulation	260
13.3	Identifiability and Principle	261
13.4	Constrained Time-Domain Multichannel LMS and Newton Algorithms	262
13.5	Unconstrained Multichannel LMS Algorithm with Optimal Step-Size Control	266
13.6	Frequency-Domain Blind Multichannel Identification Algorithms	268
13.7	Adaptive Multichannel Exponentiated Gradient Algorithm	276

13.8 Summary	279
References	279

Part C Speech Coding

14 Principles of Speech Coding

<i>W. B. Kleijn</i>	283
14.1 The Objective of Speech Coding	283
14.2 Speech Coder Attributes	284
14.3 A Universal Coder for Speech	286
14.4 Coding with Autoregressive Models	293
14.5 Distortion Measures and Coding Architecture	296
14.6 Summary	302
References	303

15 Voice over IP: Speech Transmission over Packet Networks

<i>J. Skoglund, E. Kozica, J. Linden, R. Hagen, W. B. Kleijn</i>	307
15.1 Voice Communication	307
15.2 Properties of the Network	308
15.3 Outline of a VoIP System	313
15.4 Robust Encoding	317
15.5 Packet Loss Concealment	326
15.6 Conclusion	327
References	328

16 Low-Bit-Rate Speech Coding

<i>A. V. McCree</i>	331
16.1 Speech Coding	331
16.2 Fundamentals: Parametric Modeling of Speech Signals	332
16.3 Flexible Parametric Models	337
16.4 Efficient Quantization of Model Parameters	344
16.5 Low-Rate Speech Coding Standards	345
16.6 Summary	347
References	347

17 Analysis-by-Synthesis Speech Coding

<i>J.-H. Chen, J. Thyssen</i>	351
17.1 Overview	352
17.2 Basic Concepts of Analysis-by-Synthesis Coding	353
17.3 Overview of Prominent Analysis-by-Synthesis Speech Coders	357
17.4 Multipulse Linear Predictive Coding (MPLPC)	360
17.5 Regular-Pulse Excitation with Long-Term Prediction (RPE-LTP)	362
17.6 The Original Code Excited Linear Prediction (CELP) Coder	363
17.7 US Federal Standard FS1016 CELP	367
17.8 Vector Sum Excited Linear Prediction (VSELP)	368
17.9 Low-Delay CELP (LD-CELP)	370

17.10 Pitch Synchronous Innovation CELP (PSI-CELP)	371
17.11 Algebraic CELP (ACELP)	371
17.12 Conjugate Structure CELP (CS-CELP) and CS-ACELP	377
17.13 Relaxed CELP (RCELP) – Generalized Analysis by Synthesis	378
17.14 eX-CELP	381
17.15 iLBC	382
17.16 TSNFC	383
17.17 Embedded CELP	386
17.18 Summary of Analysis-by-Synthesis Speech Coders	388
17.19 Conclusion	390
References	390

18 Perceptual Audio Coding of Speech Signals

<i>J. Herre, M. Lutzky</i>	393
18.1 History of Audio Coding	393
18.2 Fundamentals of Perceptual Audio Coding	394
18.3 Some Successful Standardized Audio Coders	396
18.4 Perceptual Audio Coding for Real-Time Communication	398
18.5 Hybrid/Crossover Coders	403
18.6 Summary	409
References	409

Part D Text-to-Speech Synthesis

19 Basic Principles of Speech Synthesis

<i>J. Schroeter</i>	413
19.1 The Basic Components of a TTS System	413
19.2 Speech Representations and Signal Processing for Concatenative Synthesis	421
19.3 Speech Signal Transformation Principles	423
19.4 Speech Synthesis Evaluation	425
19.5 Conclusions	426
References	426

20 Rule-Based Speech Synthesis

<i>R. Carlson, B. Granström</i>	429
20.1 Background	429
20.2 Terminal Analog	429
20.3 Controlling the Synthesizer	432
20.4 Special Applications of Rule-Based Parametric Synthesis	434
20.5 Concluding Remarks	434
References	434

21 Corpus-Based Speech Synthesis

<i>T. Dutoit</i>	437
21.1 Basics	437

21.2	Concatenative Synthesis with a Fixed Inventory	438
21.3	Unit-Selection-Based Synthesis	447
21.4	Statistical Parametric Synthesis	450
21.5	Conclusion	453
	References	453
22	Linguistic Processing for Speech Synthesis	
	<i>R. Sproat</i>	457
22.1	Why Linguistic Processing is Hard	457
22.2	Fundamentals: Writing Systems and the Graphical Representation of Language	457
22.3	Problems to be Solved and Methods to Solve Them	458
22.4	Architectures for Multilingual Linguistic Processing	465
22.5	Document-Level Processing	465
22.6	Future Prospects	466
	References	467
23	Prosodic Processing	
	<i>J. van Santen, T. Mishra, E. Klabbers</i>	471
23.1	Overview	471
23.2	Historical Overview	475
23.3	Fundamental Challenges	476
23.4	A Survey of Current Approaches	477
23.5	Future Approaches	484
23.6	Conclusions	485
	References	485
24	Voice Transformation	
	<i>Y. Stylianou</i>	489
24.1	Background	489
24.2	Source-Filter Theory and Harmonic Models	490
24.3	Definitions	492
24.4	Source Modifications	494
24.5	Filter Modifications	498
24.6	Conversion Functions	499
24.7	Voice Conversion	500
24.8	Quality Issues in Voice Transformations	501
24.9	Summary	502
	References	502
25	Expressive/Affective Speech Synthesis	
	<i>N. Campbell</i>	505
25.1	Overview	505
25.2	Characteristics of Affective Speech	506
25.3	The Communicative Functionality of Speech	508
25.4	Approaches to Synthesizing Expressive Speech	510
25.5	Modeling Human Speech	512

25.6 Conclusion	515
References	515

Part E Speech Recognition

26 Historical Perspective of the Field of ASR/NLU	
<i>L. Rabiner, B.-H. Juang</i>	521
26.1 ASR Methodologies	521
26.2 Important Milestones in Speech Recognition History	523
26.3 Generation 1 – The Early History of Speech Recognition	524
26.4 Generation 2 – The First Working Systems for Speech Recognition....	524
26.5 Generation 3 – The Pattern Recognition Approach to Speech Recognition	525
26.6 Generation 4 – The Era of the Statistical Model.....	530
26.7 Generation 5 – The Future	534
26.8 Summary	534
References	535
27 HMMs and Related Speech Recognition Technologies	
<i>S. Young</i>	539
27.1 Basic Framework	539
27.2 Architecture of an HMM-Based Recognizer.....	540
27.3 HMM-Based Acoustic Modeling	547
27.4 Normalization	550
27.5 Adaptation.....	551
27.6 Multipass Recognition Architectures	554
27.7 Conclusions.....	554
References	555
28 Speech Recognition with Weighted Finite-State Transducers	
<i>M. Mohri, F. Pereira, M. Riley</i>	559
28.1 Definitions	559
28.2 Overview.....	560
28.3 Algorithms	567
28.4 Applications to Speech Recognition	574
28.5 Conclusion	582
References	582
29 A Machine Learning Framework for Spoken-Dialog Classification	
<i>C. Cortes, P. Haffner, M. Mohri</i>	585
29.1 Motivation	585
29.2 Introduction to Kernel Methods	586
29.3 Rational Kernels.....	587
29.4 Algorithms	589
29.5 Experiments.....	591
29.6 Theoretical Results for Rational Kernels	593

29.7	Conclusion	594
	References	595
30	Towards Superhuman Speech Recognition	
	<i>M. Picheny, D. Nahamoo</i>	597
30.1	Current Status	597
30.2	A Multidomain Conversational Test Set	598
30.3	Listening Experiments	599
30.4	Recognition Experiments	601
30.5	Speculation	607
	References	614
31	Natural Language Understanding	
	<i>S. Roukos</i>	617
31.1	Overview of NLU Applications	618
31.2	Natural Language Parsing	620
31.3	Practical Implementation	623
31.4	Speech Mining	623
31.5	Conclusion	625
	References	626
32	Transcription and Distillation of Spontaneous Speech	
	<i>S. Furui, T. Kawahara</i>	627
32.1	Background	627
32.2	Overview of Research Activities on Spontaneous Speech	628
32.3	Analysis for Spontaneous Speech Recognition	632
32.4	Approaches to Spontaneous Speech Recognition	635
32.5	Metadata and Structure Extraction of Spontaneous Speech	640
32.6	Speech Summarization	644
32.7	Conclusions	647
	References	647
33	Environmental Robustness	
	<i>J. Droppo, A. Acero</i>	653
33.1	Noise Robust Speech Recognition	653
33.2	Model Retraining and Adaptation	656
33.3	Feature Transformation and Normalization	657
33.4	A Model of the Environment	664
33.5	Structured Model Adaptation	667
33.6	Structured Feature Enhancement	671
33.7	Unifying Model and Feature Techniques	675
33.8	Conclusion	677
	References	677
34	The Business of Speech Technologies	
	<i>J. Wilpon, M. E. Gilbert, J. Cohen</i>	681
34.1	Introduction	682

34.2	Network-Based Speech Services	686
34.3	Device-Based Speech Applications	692
34.4	Vision/Predictions of Future Services – Fueling the Trends	697
34.5	Conclusion	701
	References	702

35 Spoken Dialogue Systems

	<i>V. Zue, S. Seneff</i>	705
35.1	Technology Components and System Development	707
35.2	Development Issues	712
35.3	Historical Perspectives	714
35.4	New Directions	715
35.5	Concluding Remarks	718
	References	718

Part F Speaker Recognition

36 Overview of Speaker Recognition

	<i>A. E. Rosenberg, F. Bimbot, S. Parthasarathy</i>	725
36.1	Speaker Recognition	725
36.2	Measuring Speaker Features	729
36.3	Constructing Speaker Models	731
36.4	Adaptation	735
36.5	Decision and Performance	735
36.6	Selected Applications for Automatic Speaker Recognition	737
36.7	Summary	739
	References	739

37 Text-Dependent Speaker Recognition

	<i>M. Hébert</i>	743
37.1	Brief Overview	743
37.2	Text-Dependent Challenges	747
37.3	Selected Results	750
37.4	Concluding Remarks	760
	References	760

38 Text-Independent Speaker Recognition

	<i>D. A. Reynolds, W. M. Campbell</i>	763
38.1	Introduction	763
38.2	Likelihood Ratio Detector	764
38.3	Features	766
38.4	Classifiers	767
38.5	Performance Assessment	776
38.6	Summary	778
	References	779

Part G Language Recognition

39 Principles of Spoken Language Recognition

<i>C.-H. Lee</i>	785
39.1 Spoken Language	785
39.2 Language Recognition Principles	786
39.3 Phone Recognition Followed by Language Modeling (PRLM)	788
39.4 Vector-Space Characterization (VSC)	789
39.5 Spoken Language Verification	790
39.6 Discriminative Classifier Design	791
39.7 Summary	793
References	793

40 Spoken Language Characterization

<i>M. P. Harper, M. Maxwell</i>	797
40.1 Language versus Dialect	798
40.2 Spoken Language Collections	800
40.3 Spoken Language Characteristics	800
40.4 Human Language Identification	804
40.5 Text as a Source of Information on Spoken Languages	806
40.6 Summary	807
References	807

41 Automatic Language Recognition Via Spectral and Token Based Approaches

<i>D. A. Reynolds, W. M. Campbell, W. Shen, E. Singer</i>	811
41.1 Automatic Language Recognition	811
41.2 Spectral Based Methods	812
41.3 Token-Based Methods	815
41.4 System Fusion	818
41.5 Performance Assessment	820
41.6 Summary	823
References	823

42 Vector-Based Spoken Language Classification

<i>H. Li, B. Ma, C.-H. Lee</i>	825
42.1 Vector Space Characterization	826
42.2 Unit Selection and Modeling	827
42.3 Front-End: Voice Tokenization and Spoken Document Vectorization	830
42.4 Back-End: Vector-Based Classifier Design	831
42.5 Language Classification Experiments and Discussion	835
42.6 Summary	838
References	839

Part H Speech Enhancement

43 Fundamentals of Noise Reduction

<i>J. Chen, J. Benesty, Y. Huang, E. J. Diethorn</i>	843
43.1 Noise	843
43.2 Signal Model and Problem Formulation	845
43.3 Evaluation of Noise Reduction	846
43.4 Noise Reduction via Filtering Techniques	847
43.5 Noise Reduction via Spectral Restoration	857
43.6 Speech-Model-Based Noise Reduction	863
43.7 Summary	868
References	869

44 Spectral Enhancement Methods

<i>I. Cohen, S. Gannot</i>	873
44.1 Spectral Enhancement	874
44.2 Problem Formulation	875
44.3 Statistical Models	876
44.4 Signal Estimation	879
44.5 Signal Presence Probability Estimation	881
44.6 A Priori SNR Estimation	882
44.7 Noise Spectrum Estimation	888
44.8 Summary of a Spectral Enhancement Algorithm	891
44.9 Selection of Spectral Enhancement Algorithms	896
44.10 Conclusions	898
References	899

45 Adaptive Echo Cancellation for Voice Signals

<i>M. M. Sondhi</i>	903
45.1 Network Echoes	904
45.2 Single-Channel Acoustic Echo Cancellation	915
45.3 Multichannel Acoustic Echo Cancellation	921
45.4 Summary	925
References	926

46 Dereverberation

<i>Y. Huang, J. Benesty, J. Chen</i>	929
46.1 Background and Overview	929
46.2 Signal Model and Problem Formulation	931
46.3 Source Model-Based Speech Dereverberation	932
46.4 Separation of Speech and Reverberation via Homomorphic Transformation	936
46.5 Channel Inversion and Equalization	937
46.6 Summary	941
References	942

47 Adaptive Beamforming and Postfiltering	
<i>S. Gannot, I. Cohen</i>	945
47.1 Problem Formulation.....	947
47.2 Adaptive Beamforming.....	948
47.3 Fixed Beamformer and Blocking Matrix.....	953
47.4 Identification of the Acoustical Transfer Function.....	955
47.5 Robustness and Distortion Weighting.....	960
47.6 Multichannel Postfiltering.....	962
47.7 Performance Analysis.....	967
47.8 Experimental Results.....	972
47.9 Summary.....	972
47.A Appendix: Derivation of the Expected Noise Reduction for a Coherent Noise Field.....	973
47.B Appendix: Equivalence Between Maximum SNR and LCMV Beamformers.....	974
References	975
48 Feedback Control in Hearing Aids	
<i>A. Spriet, S. Doclo, M. Moonen, J. Wouters</i>	979
48.1 Problem Statement.....	980
48.2 Standard Adaptive Feedback Canceller.....	982
48.3 Feedback Cancellation Based on Prior Knowledge of the Acoustic Feedback Path.....	986
48.4 Feedback Cancellation Based on Closed-Loop System Identification.....	990
48.5 Comparison.....	995
48.6 Conclusions.....	997
References	997
49 Active Noise Control	
<i>S. M. Kuo, D. R. Morgan</i>	1001
49.1 Broadband Feedforward Active Noise Control.....	1002
49.2 Narrowband Feedforward Active Noise Control.....	1006
49.3 Feedback Active Noise Control.....	1010
49.4 Multichannel ANC.....	1011
49.5 Summary.....	1015
References	1015

Part I Multichannel Speech Processing

50 Microphone Arrays	
<i>G. W. Elko, J. Meyer</i>	1021
50.1 Microphone Array Beamforming.....	1021
50.2 Constant-Beamwidth Microphone Array System.....	1029
50.3 Constrained Optimization of the Directional Gain.....	1030
50.4 Differential Microphone Arrays.....	1031
50.5 Eigenbeamforming Arrays.....	1034

50.6	Adaptive Array Systems	1037
50.7	Conclusions	1040
	References	1040
51	Time Delay Estimation and Source Localization	
	<i>Y. Huang, J. Benesty, J. Chen</i>	1043
51.1	Technology Taxonomy	1043
51.2	Time Delay Estimation	1044
51.3	Source Localization	1054
51.4	Summary	1061
	References	1062
52	Convolutional Blind Source Separation Methods	
	<i>M. S. Pedersen, J. Larsen, U. Kjems, L. C. Parra</i>	1065
52.1	The Mixing Model	1066
52.2	The Separation Model	1068
52.3	Identification	1071
52.4	Separation Principle	1071
52.5	Time Versus Frequency Domain	1076
52.6	The Permutation Ambiguity	1078
52.7	Results	1084
52.8	Conclusion	1084
	References	1084
53	Sound Field Reproduction	
	<i>R. Rabenstein, S. Spors</i>	1095
53.1	Sound Field Synthesis	1095
53.2	Mathematical Representation of Sound Fields	1096
53.3	Stereophony	1100
53.4	Vector-Based Amplitude Panning	1103
53.5	Ambisonics	1104
53.6	Wave Field Synthesis	1109
	References	1113
	Acknowledgements	1115
	About the Authors	1117
	Detailed Contents	1133
	Subject Index	1161

Springer Handbook of Speech Processing

Benesty, J.; Sondhi, M.M.; Huang, Y. (Eds.)

2008, XXXVI, 1176 p. 456 illus.,

ISBN: 978-3-540-49127-9