

Advances in

Digital Speech Transmission



Editors

Rainer Martin

Ulrich Heute

Christiane Antweiler

 WILEY

Contents

List of Contributors	xxi
Preface	xxvii
1 Introduction	1
<i>Rainer Martin, Ulrich Heute, Christiane Antweiler</i>	
1 Speech Quality Assessment	7
2 Speech-Transmission Quality: Aspects and Assessment for Wideband vs. Narrowband Signals	9
<i>Ulrich Heute</i>	
2.1 Introduction	9
2.2 Speech Signals	10
2.3 Telephone-Band Speech Signals	11
2.3.1 Narrowband Speech Intelligibility	12
2.3.2 Narrowband Speech-Sound Quality	12
2.4 Wideband Speech Signals	14
2.4.1 Wideband-Speech Intelligibility and Sound Quality	15
2.4.2 Wideband Speech Transmission and Processing	18
2.5 Speech-Quality Assessment	25
2.5.1 Auditory Quality Determination	25
2.5.2 Instrumental Quality Determination	27
2.6 Wideband Speech-Quality Assessment	30
2.6.1 Integral Quality Determination	30
2.6.2 Attribute-Oriented Quality Determination	34

2.6.3 Combined Direct and Attribute-Based Total Quality Determination	43
2.7 Concluding Remarks	43
Bibliography	44
3 Parametric Quality Assessment of Narrowband Speech in Mobile Communication Systems	51
<i>Marc Werner</i>	
3.1 Introduction	51
3.1.1 Subjective Listening Tests and Classes of Objective Measures .	52
3.1.2 Overview of Objective Speech Quality Measures	55
3.1.3 Development of Parametric Models	57
3.2 Simulations of GSM and UMTS Speech Transmissions	58
3.2.1 Simulation Environment	58
3.2.2 GSM Transmission Parameters	62
3.2.3 UMTS Transmission Parameters	64
3.3 Speech Quality Measures based on Transmission Parameters	65
3.3.1 Correlation Analysis	65
3.3.2 Parametric Speech Quality Measures	68
3.4 Discussion and Conclusions	73
Bibliography	73
II Adaptive Algorithms in Acoustic Signal Processing	77
4 Kalman Filtering in Acoustic Echo Control: A Smooth Ride on a Rocky Road	79
<i>Gerald Enzner</i>	
4.1 Introduction	79
4.1.1 Adaptive Filter Structures for Acoustic Echo Control	81
4.1.2 Control of Adaptive Filters	82
4.1.3 Open Problems / Organization of this Chapter	84

4.2	A Comprehensive Theory of Acoustic Echo Control	85
4.2.1	Stochastic Modeling of the Echo Path	85
4.2.2	Minimum Mean-Square Error (MMSE) Solution	87
4.2.3	MMSE Processor in the Gaussian Case	89
4.3	The Kalman Filter for Conditional Mean and Covariance Estimation .	90
4.3.1	Linear Echo Path Model in DFT-Matrix Form	91
4.3.2	Markov Model of the Time-Varying Echo Path	93
4.3.3	Exact Kalman Filter for the Conditional Mean and Covariance	94
4.3.4	Diagonalization of the Kalman Filter	96
4.3.5	Unification of Adaptive Filtering and Adaptation Control . . .	98
4.4	AEC Performance of the Frequency-Domain Adaptive Kalman Filter .	100
4.5	Discussion and Conclusions	102
	Bibliography	103

5 Noise Reduction - Statistical Analysis and Control of Musical Noise 107

Colin Breithaupt, Rainer Martin

5.1	Introduction	107
5.2	Speech Enhancement in the DFT Domain	109
5.2.1	Optimal Speech Estimators	109
5.3	Measurement and Assessment of Unnatural Fluctuations	115
5.3.1	Filter Analysis via Approximated Filter Input-Output Characteristics	116
5.3.2	Outlier Statistics	118
5.4	Avoidance of Processing Artifacts	120
5.5	Control of Spectral Fluctuations in the Cepstral Domain	123
5.6	Discussion and Conclusions	128
5.7	Acknowledgements	129
5.8	Appendix	129
5.8.1	Mean <i>a priori</i> SNR for different filter types and low SNR . . .	129
5.8.2	Approximation of the <i>decision-directed</i> approach for low SNR .	131
	Bibliography	131

Nilesh Madhu, Rainer Martin

6.1	Introduction	135
6.2	Signal Model	136
6.2.1	Continuous Time Model	136
6.2.2	Discrete Time Representation	137
6.2.3	Formulation in the Frequency Domain	138
6.2.4	Simplified Model for Localization	138
6.3	Localization Approach Taxonomy	141
6.4	Indirect Localization Approaches	141
6.4.1	Generalized Cross-Correlation (GCC)	142
6.4.2	Adaptive Eigenvalue Decomposition (AED)	144
6.4.3	Information Theoretic Approach to TDOA Estimation	146
6.4.4	Extension to Multiple Microphone Pairs	147
6.5	Direct Localization Approaches	148
6.5.1	Steered Response Power Beamforming	149
6.5.2	Minimum Mean Square (MMSE) Approach	150
6.5.3	Practical aspects	151
6.5.4	Subspace Based Approaches	151
6.5.5	Maximum Likelihood Estimation (MLE)	154
6.6	Evaluation of Localization Algorithms	156
6.6.1	Performance of the Indirect Methods	158
6.6.2	Performance of the Direct Methods	160
6.6.3	The Two-Source Case	165
6.7	Conclusions	166
	Bibliography	166

7

Multi-Channel System Identification with Perfect Sequences
– Theory and Applications –

171

Christiane Antweiler

7.1	Introduction	171
7.2	System Identification with Perfect Sequences	174
7.2.1	Geometric Interpretation of the NLMS Algorithm	175
7.2.2	Optimal Excitation of the NLMS Algorithm	176
7.2.3	Influence of Environmental Noise, Stepsize, and Period	179
7.2.4	Odd-Perfect Sequences	181
7.2.5	Tracking of Time-Variant Systems	183
7.2.6	Complexity	184
7.3	Multi-Channel System Identification	185
7.3.1	The Dual-Channel Case	185
7.3.2	Simulation Results	188
7.3.3	Generalization to the Multi-Channel Case	190
7.4	Applications	191
7.4.1	Simulation of Time-Variant RIRs for Stereophonic Echo Control	191
7.4.2	Acoustic Tube Endoscopy	194
7.5	Discussion and Conclusions	195
	Bibliography	195

III

Speech Coding for Heterogeneous Networks

199

8

Embedded Speech Coding: From G.711 to G.729.1

201

Bernd Geiser, Stéphane Ragot, Hervé Taddei

8.1	Introduction	201
8.2	Theory and Tools of Embedded Speech Coding	203
8.2.1	Basic Principles	203
8.2.2	Approximation Theory	205
8.2.3	Hierarchical Vector Quantization Methods	208

8.3	Embedded Speech Coding Methods	212
8.3.1	Embedded DPCM and ADPCM	212
8.3.2	Embedded CELP	213
8.3.3	Embedded Extensions of CELP Coders	216
8.3.4	Embedded Parameter Quantization	218
8.4	Standardized Embedded Speech Coders	219
8.4.1	ITU-T G.711 PCM Codec	219
8.4.2	ITU-T G.727 and G.722 ADPCM Codecs	220
8.4.3	MPEG-4 Scalable Speech Coding	220
8.4.4	Embedded Wideband Coding for VoIP: ITU-T G.729.1	223
8.5	Network Aspects of Embedded Speech Coding	232
8.5.1	Implementation and Utilization of Scalability	232
8.5.2	Unequal Error Protection and Encryption	236
8.6	Conclusions and Perspectives	237
	Bibliography	238

9 Backwards Compatible Wideband Telephony 249

Peter Jax

9.1	Introduction	249
9.2	From Narrowband Telephony to Wideband Telephony	250
9.3	Stand-Alone Bandwidth Extension	254
9.3.1	Estimation of the Wideband Spectral Envelope	255
9.3.2	Extension of the Excitation Signal	256
9.3.3	Performance and State-of-the-Art	257
9.4	Embedded Wideband Coding Using Bandwidth Extension Techniques	257
9.4.1	Transmission of BWE Information	258
9.4.2	Examples of Embedded Wideband Speech Codecs	260
9.4.3	Audio Coding	262
9.5	Combination of Bandwidth Extension with Watermarking	262
9.5.1	Digital Watermarking of Speech Signals	263
9.5.2	Transmission of BWE Information via Watermarking	265

9.5.3	Challenges and Status	266
9.6	Advanced Transmission of Highband Parameters	267
9.6.1	Coding with Side Information	268
9.6.2	Error Concealment with Side Information	270
9.7	Conclusions	274
	Bibliography	274

IV Joint Source-Channel Coding 279

10 Parameter Models and Estimators in Soft Decision Source Decoding 281

Tim Fingscheidt

10.1	Introduction	281
10.2	Overview to Soft Decision Source Decoding	283
10.2.1	Source Encoding	283
10.2.2	Equivalent Channel	284
10.2.3	Hard Decision and Soft Decision Source Decoding	285
10.3	The Markovian Parameter Model	287
10.3.1	Description of <i>A Priori</i> Knowledge	287
10.3.2	Quantification of Utilizable Residual Redundancy	288
10.3.3	Choice of the Model Order	289
10.4	Basic Extrapolative Estimators	290
10.4.1	Introduction and Simulation Settings	290
10.4.2	Estimators	291
10.4.3	Simulation Results	294
10.5	Joint Extrapolative Estimation of Two Different Parameters	298
10.5.1	Estimators	298
10.5.2	Simulation Results	299
10.6	Extrapolative Estimation with Repeated Parameter Transmission	301
10.6.1	Estimators	301
10.6.2	Simulation Results	303
10.7	Interpolative Estimation of a Parameter	304

10.7.1 Estimators	304
10.7.2 Simulation Results	306
10.8 Discussion and Conclusions	307
Bibliography	307
11 Optimal MMSE Estimation for Vector Sources with Spatially and Temporally Correlated Elements	311
<i>Stefan Heinen, Marc Adrat</i>	
11.1 Introduction	311
11.2 Source Model	312
11.3 Transmission Channel	316
11.4 Optimal MMSE Parameter Estimator	316
11.5 Near-Optimal MMSE Parameter Estimator	320
11.6 Illustrative Comparison	323
11.7 Simulation Results	325
11.8 Conclusions	327
Bibliography	327
12 Source Optimized Channel Codes & Source Controlled Channel Decoding	329
<i>Stefan Heinen, Thomas Hindelang</i>	
12.1 Introduction	329
12.2 The Transmission System Used as Reference	330
12.3 Source Optimized Channel Coding (SOCC)	332
12.3.1 Definition	333
12.3.2 Decoding of <i>Source Optimized Channel Codes</i>	334
12.3.3 Design of Source Optimized Channel Codes	335
12.3.4 Numerical Aspects of SOCC Design	336
12.3.5 Bit Allocation between Source and Channel Coding	336
12.3.6 Relation to Channel Optimized Vector Quantization	338
12.4 Source Controlled Channel Decoding (SCCD)	341
12.4.1 Channel Coding and Decoding in SCCD	341

12.4.2	A Priori Knowledge in Channel Decoding	345
12.4.3	Channel Decoding Using Intra-Parameter Correlation	347
12.4.4	Channel Decoding Using Inter-Frame Correlation	349
12.4.5	Channel Decoding Using Intra-Parameter and Inter-Frame Correlation	350
12.4.6	Simulation Results	352
12.4.7	Exploiting A Priori Knowledge in Source and/or Channel Decoding	354
12.5	Comparison of SOCC versus SCCD	357
12.6	Conclusions	362
	Bibliography	363

13 Iterative Source-Channel Decoding & Turbo DeCodulation 365

Marc Adrat, Thorsten Clevorn, Laurent Schmalen

13.1	Introduction	365
13.2	The Key of the Turbo Principle: Extrinsic Information	366
13.2.1	Terms of Reliability Information	367
13.2.2	Extrinsic Information of Channel Decoding	368
13.2.3	Extrinsic Information of Source Decoding	371
13.2.4	Extrinsic Information of Demodulation	374
13.2.5	EXIT Charts	376
13.3	Iterative Source-Channel Decoding (ISCD)	379
13.3.1	Transmission System and Algorithm	379
13.3.2	Simulation Examples	382
13.3.3	Advancements and Optimizations	385
13.4	Turbo DeCodulation (TDeC)	387
13.4.1	Transmission System and Algorithm	388
13.4.2	Simulation Examples	390
13.4.3	Advancements and Optimizations	393
13.5	Conclusions	394
	Bibliography	395

V	Speech Processing in Hearing Instruments	399
14	Binaural Signal Processing in Hearing Aids	401
	<i>Volkmar Hamacher, Ulrich Kornagel, Thomas Lotter, Henning Puder</i>	
14.1	Introduction	401
14.1.1	Monaural Hearing Aids - State of the Art	402
14.1.2	Binaural Hearing Aids	404
14.1.3	Organization of this Chapter	405
14.2	Wireless System for Hearing Aids	405
14.2.1	Comparison of Wireless Systems	405
14.2.2	Functional Description of the Wireless System for Hearing Aids	406
14.2.3	Applications of the Wireless System for Hearing Aids	409
14.3	Binaural Classification Systems	410
14.3.1	Motivation and Basic Principle	410
14.3.2	Binaural Classification	412
14.4	Binaural Beamformer	415
14.4.1	Dual Channel Input-Output Beamformer Design	416
14.4.2	Multichannel Postfilter	419
14.4.3	Performance Evaluation	420
14.5	Blind Source Separation (BSS): An Application for a Binaural Directional Microphone Array in Hearing Aids	422
14.5.1	Application Scenario	422
14.5.2	Specific Hearing Aid Challenges and Solutions	423
14.5.3	Signal Separation with Hearing Aid Constraints	424
14.5.4	Output Signal Selection	425
14.5.5	Binaural Output Generation	426
14.5.6	Concluding Remarks	427
14.6	Conclusions	427
	Bibliography	428

15 Auditory-profile-based Physical Evaluation of Multi-microphone Noise Reduction Techniques in Hearing Instruments **431**

Koen Eneman, Arne Leijon, Simon Doclo, Ann Spriet, Marc Moonen, Jan Wouters

15.1 Introduction	431
15.2 Multi-microphone Noise Reduction in Hearing Instruments	434
15.2.1 Classical Solutions	434
15.2.2 Generalized Sidelobe Canceler	436
15.2.3 Adaptive Two-stage Beamforming Approach	438
15.2.4 Spatially Preprocessed Speech-distortion-weighted Multichannel Wiener Filtering	439
15.3 Auditory-profile-based Physical Evaluation	441
15.3.1 Simulation of Hearing-impaired Perception	442
15.3.2 Physical Evaluation Measures	444
15.4 Test Conditions	449
15.5 Simulation Results	450
15.6 Discussion	452
15.7 Conclusions	455
Bibliography	456

VI Speech Processing for Human–Machine Interfaces **459**

16 Automatic Speech Recognition in Adverse Acoustic Conditions **461**

Hans-Günter Hirsch

16.1 Introduction	461
16.2 Structure of Speech Recognition Systems	462
16.2.1 Mel Frequency Cepstral Analysis	463
16.2.2 Modeling Speech Units as HMMs	466
16.3 Acoustic Scenarios during Speech Input	468
16.3.1 Simulation of the Acoustic Environment	469
16.3.2 Recognition Results for Different Distortion Effects	473
16.4 Improving the Recognition Performance in Adverse Conditions	476

16.4.1	Adapting HMMs to Reverberation	477
16.4.2	Adaptation of Delta Parameters	482
16.4.3	Recognition Experiments on Hands-Free Speech Input	485
16.4.4	Combined Adaptation to All Distortion Effects	487
16.4.5	Recognition Experiments on Hands-Free Speech Input in Noisy Environments	490
16.5	Conclusions	493
	Bibliography	494

17 Speaker Classification for Next-Generation Voice-Dialog Systems 497

Felix Burkhardt, Florian Metze, Joachim Stegmann

17.1	Introduction	497
17.2	Speaker Classification	498
17.2.1	Overview	498
17.2.2	Feature Extraction	499
17.2.3	Classification Algorithms	501
17.2.4	Evaluation of Classifiers	503
17.3	Detection of Age and Gender	505
17.3.1	Background	505
17.3.2	Algorithms	506
17.3.3	Results	509
17.4	Detection of Anger	510
17.4.1	Background	510
17.4.2	Algorithm	513
17.4.3	Results	514
17.5	Applications in IVR Systems	517
17.5.1	Adaptive Voice-Dialogs	518
17.5.2	A Voice Portal Based on Age/Gender Detection	519
17.5.3	Customer Self-Service Based on Anger Detection	521
17.6	Discussion and Conclusion	523
	Bibliography	525

Index	529
Permissions List	541