
Contents

1 Introduction	1
<i>Jacob Benesty, Shoji Makino, Jingdong Chen</i>	
1.1 Speech Enhancement	1
1.2 Challenges and Opportunities	3
1.3 Organization of the Book	4
1.4 Further Reading	7
References	8
2 Study of the Wiener Filter for Noise Reduction	9
<i>Jacob Benesty, Jingdong Chen, Yiteng (Arden) Huang, Simon Doclo</i>	
2.1 Introduction	9
2.2 Estimation of the Clean Speech Samples	11
2.3 Estimation of the Noise Samples	13
2.4 Important Relationships Between Noise Reduction and Speech Distortion	14
2.5 Particular Case: White Gaussian Noise	21
2.6 Better Ways to Manage Noise Reduction and Speech Distortion ...	23
2.6.1 A Suboptimal Filter	23
2.6.2 Noise Reduction Exploiting the Speech Model	26
2.6.3 Noise Reduction with Multiple Microphones	27
2.7 Simulation Experiments	29
2.8 Conclusions	37
References	38
3 Statistical Methods for the Enhancement of Noisy Speech	43
<i>Rainer Martin</i>	
3.1 Introduction	43
3.2 Spectral Analysis	44
3.3 The Wiener Filter and its Implementation	45
3.4 Estimation of Spectral Amplitudes	51
3.4.1 MMSE Estimation	51
3.4.2 Maximum Likelihood and MAP Estimation	53
3.5 MMSE Estimation Using Super-Gaussian Speech Models	54
3.6 Background Noise Power Estimation	58
3.6.1 Minimum Statistics Noise Power Estimation	58
3.7 The MELPe Speech Coder	60
3.8 Conclusions	62
References	63

4 Single- and Multi-Microphone Spectral Amplitude Estimation Using a Super-Gaussian Speech Model	67
<i>Thomas Lotter</i>	
4.1 Introduction	67
4.2 Single-Channel Statistical Filter	68
4.2.1 Statistical Model	70
4.2.2 Speech Estimators	78
4.3 Multichannel Statistical Filter	83
4.3.1 Joint Statistical Model	84
4.3.2 Multichannel MAP Spectral Amplitude Estimation	86
4.4 Experimental Results	88
4.5 Conclusions	92
References	92
5 From Volatility Modeling of Financial Time-Series to Stochastic Modeling and Enhancement of Speech Signals . .	97
<i>Israel Cohen</i>	
5.1 Introduction	97
5.2 Problem Formulation	99
5.3 Spectral Analysis	101
5.4 Statistical Model for Speech Signals	104
5.5 Model Estimation	105
5.6 Experimental Results	106
5.7 Conclusions	108
References	111
6 Single-Microphone Noise Suppression for 3G Handsets Based on Weighted Noise Estimation	115
<i>Akihiko Sugiyama, Masanori Kato, Masahiro Serizawa</i>	
6.1 Introduction	115
6.2 Conventional Noise Suppression Algorithm	117
6.2.1 MMSE-STSA	117
6.2.2 Problem in Noise Estimation	119
6.3 New Noise Suppression Algorithm	120
6.3.1 Weighted Noise Estimation	121
6.3.2 Spectral Gain Modification	123
6.3.3 Computational Requirements	123
6.4 Evaluation	124
6.4.1 Objective Evaluation for Noise Estimation	125
6.4.2 Subjective Evaluation	127
6.5 Conclusions	131
References	131

7 Signal Subspace Techniques for Speech Enhancement	135
<i>Firas Jabloun, Benoit Champagne</i>	
7.1 Introduction	135
7.2 Signal and Noise Models	137
7.3 Linear Signal Estimation	138
7.3.1 Least-Squares Estimator	139
7.3.2 The Linear Minimum Mean Squared Error Estimator	139
7.3.3 The Time-Domain Constrained Estimator	140
7.3.4 The Spectral-Domain Constrained Estimator	141
7.4 Handling Colored Noise	143
7.4.1 Prewhitening	143
7.4.2 The Generalized Eigenvalue Decomposition Method	144
7.4.3 The Rayleigh Quotient Method	145
7.5 A Filterbank Interpretation	146
7.5.1 The Frequency to Eigendomain Transformation	146
7.5.2 The Eigen Filterbank	146
7.6 Implementation Issues	148
7.6.1 Estimating the Covariance Matrix	149
7.6.2 Parameter Analysis	150
7.7 Fast Subspace Estimation Techniques	152
7.7.1 Fast Eigenvalue Decomposition Methods	153
7.7.2 Subspace Tracking Methods	153
7.7.3 The Frame Based EVD (FBEVD) Method	154
7.8 Some Recent Developments	155
7.8.1 Auditory Masking	155
7.8.2 Multi-Microphone Systems	156
7.8.3 Subband Processing	156
7.9 Conclusions	157
References	157
8 Speech Enhancement: Application of the Kalman Filter in the Estimate-Maximize (EM) Framework	161
<i>Sharon Gannot</i>	
8.1 Introduction	161
8.2 Signal Model	166
8.3 EM - Based Algorithm	168
8.3.1 State Estimation (E-Step)	169
8.3.2 Parameter Estimation (M-Step)	170
8.3.3 Reduced Complexity	171
8.3.4 Discussion	171
8.4 Parameter Estimation Using Higher-Order Statistics	172
8.5 Gradient-Based Sequential Algorithm	174
8.6 All-Kalman Speech and Parameter Estimation	175
8.6.1 Dual Scheme	176
8.6.2 Joint Scheme	178

8.7	Experimental Study	181
8.7.1	Experimental Setup	181
8.7.2	Verifying the Gaussian Assumption	182
8.7.3	Objective Evaluation	183
8.7.4	Subjective Evaluation	186
8.7.5	Comparison Between EM-Based Algorithms	188
8.7.6	Evaluation of the UKF	188
8.8	Conclusions	189
	References	195
9 Speech Distortion Weighted Multichannel Wiener		
	Filtering Techniques for Noise Reduction	199
	<i>Simon Doclo, Ann Spriet, Jan Wouters, Marc Moonen</i>	
9.1	Introduction	199
9.2	GSC and Spatially Pre-Processed SDW-MWF	201
9.2.1	Notation and General Structure	201
9.2.2	Generalized Sidelobe Canceller	204
9.2.3	Speech Distortion Weighted Multichannel Wiener Filter	205
9.3	Frequency-Domain Criterion for SDW-MWF	207
9.3.1	Frequency-Domain Notation	207
9.3.2	Normal Equations	208
9.3.3	Adaptive Algorithm	210
9.3.4	Practical Implementation	212
9.4	Approximations for Reducing the Complexity	213
9.4.1	Block-Diagonal Correlation Matrices	213
9.4.2	Diagonal Correlation Matrices	216
9.4.3	Unconstrained Algorithms	217
9.4.4	Summary	218
9.5	Experimental Results	219
9.5.1	Setup and Performance Measures	219
9.5.2	SNR Improvement and Robustness Against Microphone Mismatch	220
9.5.3	Tracking Performance	223
9.6	Conclusions	224
	References	225
10 Adaptive Microphone Arrays Employing Spatial Quadratic Soft Constraints and Spectral Shaping		
	<i>Sven Nordholm, Hai Quang Dam, Nedelko Grbić, Siow Yong Low</i>	
10.1	Introduction	229
10.2	Signal Modelling and Problem Formulation	231
10.2.1	Analysis and Synthesis Filterbanks	232
10.2.2	The Wiener Solution	233
10.2.3	The Space Constrained Source Covariance Information	234
10.3	Robust Soft Constrained Adaptive Microphone Array (RSCAMA)	235

10.3.1 Problem Formulation	235
10.3.2 A Recursive Algorithm for the RSCAMA	237
10.4 Noise Statistics Updated Adaptive Microphone Array (NSUAMA)	238
10.4.1 Problem Formulation	238
10.4.2 The Noise Covariance Detector	238
10.4.3 Estimation of Power Spectrum of SOI	240
10.4.4 The NSUAMA Algorithm	241
10.5 Evaluations	242
10.5.1 The Simulation Scenario	242
10.5.2 Results for RSCAMA and NSUAMA Beamformers	242
10.6 Conclusions	244
References	244
11 Single-Microphone Blind Dereverberation	247
<i>Tomohiro Nakatani, Masato Miyoshi, Keisuke Kinoshita</i>	
11.1 Introduction	247
11.2 Overview of Existing Approaches	249
11.2.1 Blind Inverse Filtering	249
11.2.2 Dereverberation Based on Speech Signal Features	250
11.3 Harmonicity of Speech Signals and Its Robust Estimation	251
11.3.1 Model of Speech Harmonicity	251
11.3.2 Adaptive Harmonic Filtering	252
11.3.3 Robust F_0 Estimation and Voicing Detection	253
11.4 Harmonicity Based Dereverberation – HERB	255
11.4.1 Basic Idea	255
11.4.2 Model of Reverberant Speech Signal	256
11.4.3 Dereverberation Filter	258
11.4.4 Interpretation of the Dereverberation Filter	258
11.5 Implementation of a Prototype System	260
11.5.1 Dereverberation Filter Calculation	261
11.5.2 Heuristics Improving Accuracy of F_0 Estimation and Voicing Decisions with Reverberation	261
11.6 Simulation Experiments	262
11.6.1 Task: Dereverberation of Word Utterances	262
11.6.2 Energy Decay Curves of Impulse Responses	262
11.6.3 Speaker Dependent Word Recognition Rate	263
11.7 Future Directions	265
11.7.1 Theoretical Extension of HERB	266
11.7.2 Accuracy Improvement of Speech Model	266
11.7.3 Reduction of Training Data Size	268
11.8 Conclusions	268
References	269

12 Separation and Dereverberation of Speech Signals with Multiple Microphones	271
<i>Yiteng (Arden) Huang, Jacob Benesty, Jingdong Chen</i>	
12.1 Introduction	271
12.2 Signal Model and Problem Formulation	274
12.3 Blind Identification of a SIMO System	276
12.4 Separating Reverberant Speech and Concurrent Interference	279
12.4.1 Example: Removing Interference Signals in a 2×3 MIMO Acoustic System	279
12.4.2 Generalization	281
12.5 Speech Dereverberation	284
12.5.1 Principle	284
12.5.2 The Least-Squares Implementation	286
12.6 Simulations	287
12.6.1 Performance Measures	287
12.6.2 Experimental Setup	288
12.6.3 Experimental Results	290
12.7 Conclusions	296
References	297
13 Frequency-Domain Blind Source Separation	299
<i>Hiroshi Sawada, Ryo Mukai, Shoko Araki, Shoji Makino</i>	
13.1 Introduction	299
13.2 BSS for Convolutional Mixtures	301
13.3 Overview of Frequency-Domain Approach	302
13.4 Complex-Valued ICA	304
13.5 Source Localization	306
13.5.1 Basic Theory for Nearfield Model	307
13.5.2 DOA Estimation with Farfield Model	308
13.6 Permutation Alignment	311
13.6.1 Localization Approach	312
13.6.2 Correlation Approach	312
13.6.3 Integrated Method	314
13.7 Scaling Alignment	315
13.8 Spectral Smoothing	317
13.8.1 Windowing	318
13.8.2 Minimizing Error by Adjusting Scaling Ambiguity	319
13.9 Experimental Results	320
13.9.1 Linear Array	320
13.9.2 Planar Array	322
13.10 Conclusions	324
References	324

14 Subband Based Blind Source Separation	329
<i>Shoko Araki, Shoji Makino</i>	
14.1 Introduction	329
14.2 BSS of Convolutional Mixtures	331
14.2.1 Model Description	331
14.2.2 Frequency-Domain BSS and Related Issue	332
14.3 Subband Based BSS	333
14.3.1 Configuration of Subband BSS	333
14.3.2 Time-Domain BSS Implementation for a Separation Stage...	336
14.3.3 Solving the Permutation and Scaling Problems	337
14.4 Basic Experiments for Subband BSS	339
14.4.1 Experimental Setup	339
14.4.2 Subband System	340
14.4.3 Conventional Frequency-Domain BSS	340
14.4.4 Conventional Fullband Time-Domain BSS	341
14.4.5 Results	341
14.4.6 Discussion	343
14.5 Frequency-Appropriate Processing for Further Improvement	344
14.5.1 Longer Separation Filters in Low Frequency Bands	345
14.5.2 Overlap-Blockshift in Low Frequency Bands	346
14.5.3 Discussion	347
14.6 Conclusions	349
References	350
15 Real-Time Blind Source Separation for Moving Speech Signals	353
<i>Ryo Mukai, Hiroshi Sawada, Shoko Araki, Shoji Makino</i>	
15.1 Introduction	353
15.2 ICA Based BSS of Convolutional Mixtures	355
15.2.1 Frequency-Domain ICA	355
15.2.2 Permutation and Scaling Problems	356
15.2.3 Low Delay Blockwise Batch Algorithm	357
15.3 Residual Crosstalk Cancellation	358
15.3.1 Straight and Crosstalk Components of BSS	358
15.3.2 Model of Residual Crosstalk Component Estimation	359
15.3.3 Adaptive Algorithm and Spectrum Estimation	360
15.4 Experiments and Discussions	362
15.4.1 Experimental Conditions	362
15.4.2 Performance for Fixed Sources	364
15.4.3 Moving Target and Moving Interference	365
15.4.4 Performance of Blockwise Batch Algorithm with Postprocessing	366
15.4.5 Performance of Online Algorithm	367
15.5 Conclusions	367
References	368

16 Separation of Speech by Computational Auditory Scene Analysis	371
<i>Guy J. Brown, DeLiang Wang</i>	
16.1 Introduction	371
16.2 Auditory Scene Analysis	372
16.3 Computational Auditory Scene Analysis	373
16.3.1 Peripheral Auditory Processing and Feature Extraction	375
16.3.2 Monaural Approaches	376
16.3.3 Binaural Approaches	382
16.3.4 Frameworks for Cue Integration	387
16.4 Integrating CASA with Speech Recognition	391
16.5 CASA Compared to ICA	394
16.6 Challenges for CASA	395
16.7 Conclusions	398
References	398
Index	403