
Contents

Preface to the Second Editionxvii
Preface to the First Editionxix
Publisher’s Notexxi
In Memoriam xxiii

Chapter 1 Introduction 1

- 1.1 Understanding the Enemy: Noise 2
 - 1.1.1 Noise Sources 2
 - 1.1.2 Noise and Speech Levels in Various Environments..... 3
- 1.2 Classes of Speech Enhancement Algorithms 6
- 1.3 Book Organization 7
- References 9

PART I *Fundamentals*

Chapter 2 Discrete-Time Signal Processing and Short-Time Fourier Analysis 13

- 2.1 Discrete-Time Signals 13
- 2.2 Linear Time-Invariant Discrete-Time Systems 16
 - 2.2.1 Difference Equations..... 16
 - 2.2.2 Linear Convolution..... 17
- 2.3 z -Transform 18
 - 2.3.1 Properties 18
 - 2.3.2 z -Domain Transfer Function..... 19
- 2.4 Discrete-Time Fourier Transform..... 20
 - 2.4.1 DTFT Properties 23
 - 2.4.2 Discrete Fourier Transform 24
 - 2.4.3 Windowing 26
- 2.5 Short-Time Fourier Transform 31
 - 2.5.1 Definition..... 31
 - 2.5.2 Interpretations of the STFT..... 33
 - 2.5.3 Sampling the STFT in Time and Frequency..... 34
 - 2.5.4 Short-Time Synthesis of Speech..... 35
 - 2.5.4.1 Filterbank Summation for Short-Time Synthesis of Speech 36
 - 2.5.4.2 Overlap-and-Add Method for Short-Time Synthesis 38

2.6 Spectrographic Analysis of Speech Signals 40
 2.7 Summary 42
 References 42

Chapter 3 Speech Production and Perception..... 45

3.1 Speech Signal 45
 3.2 Speech Production Process..... 45
 3.2.1 Lungs 46
 3.2.2 Larynx and Vocal Folds 47
 3.2.3 Vocal Tract 51
 3.3 Engineering Model of Speech Production 54
 3.4 Classes of Speech Sounds 55
 3.5 Acoustic Cues in Speech Perception 57
 3.5.1 Vowels and Diphthongs 57
 3.5.2 Semivowels 59
 3.5.3 Nasals 60
 3.5.4 Stops 60
 3.5.5 Fricatives 63
 3.6 Summary 65
 References 65

Chapter 4 Noise Compensation by Human Listeners 67

4.1 Intelligibility of Speech in Multiple-Talker Conditions..... 68
 4.1.1 Effect of Masker’s Spectral/Temporal Characteristics and Number of Talkers: Monaural Hearing 68
 4.1.2 Effect of Source Spatial Location: Binaural Hearing..... 70
 4.2 Acoustic Properties of Speech Contributing to Robustness..... 76
 4.2.1 Shape of the Speech Spectrum..... 76
 4.2.2 Spectral Peaks 78
 4.2.3 Periodicity 80
 4.2.4 Rapid Spectral Changes Signaling Consonants 81
 4.3 Perceptual Strategies for Listening in Noise 82
 4.3.1 Auditory Streaming..... 83
 4.3.2 Listening in the Gaps and Glimpsing..... 84
 4.3.3 Use of F0 Differences..... 84
 4.3.4 Use of Linguistic Knowledge 86
 4.3.5 Use of Spatial and Visual Cues 86
 4.4 Summary 87
 References 87

PART II Algorithms

Chapter 5	Spectral-Subtractive Algorithms.....	93
5.1	Basic Principles of Spectral Subtraction	93
5.2	Geometric View of Spectral Subtraction.....	97
5.2.1	Upper Bounds on the Difference between the Noisy and Clean Signals' Phases	98
5.2.2	Alternate Spectral-Subtractive Rules and Theoretical Limits	100
5.3	Shortcomings of the Spectral Subtraction Method	105
5.4	Spectral Subtraction Using Oversubtraction	107
5.5	Nonlinear Spectral Subtraction	113
5.6	Multiband Spectral Subtraction.....	115
5.7	Minimum Mean Square Error Spectral Subtraction Algorithm	120
5.8	Extended Spectral Subtraction	122
5.9	Spectral Subtraction Using Adaptive Gain Averaging	124
5.10	Selective Spectral Subtraction.....	127
5.11	Spectral Subtraction Based on Perceptual Properties	129
5.12	Performance of Spectral Subtraction Algorithms	130
5.13	Summary	132
	References	133
Chapter 6	Wiener Filtering	137
6.1	Introduction to Wiener Filter Theory	137
6.2	Wiener Filters in the Time Domain.....	138
6.3	Wiener Filters in the Frequency Domain	140
6.4	Wiener Filters and Linear Prediction	142
6.5	Wiener Filters for Noise Reduction	144
6.5.1	Square-Root Wiener Filter	151
6.5.2	Parametric Wiener Filters	152
6.6	Iterative Wiener Filtering	156
6.6.1	Mathematical Speech Production Model	157
6.6.2	Statistical Parameter Estimation of the All-Pole Model in Noise	158
6.7	Imposing Constraints on Iterative Wiener Filtering	165
6.7.1	Across-Time Spectral Constraints.....	165
6.7.2	Across-Iterations Constraints	168
6.8	Constrained Iterative Wiener Filtering.....	170
6.9	Constrained Wiener Filtering	172
6.9.1	Mathematical Definitions of Speech and Noise Distortions	172
6.9.2	Limiting the Noise Distortion Level	176

6.10	Estimating the Wiener Gain Function.....	179
6.11	Incorporating Psychoacoustic Constraints in Wiener Filtering	183
6.11.1	Shaping the Noise Distortion in the Frequency Domain	184
6.11.2	Using Masking Thresholds as Constraints.....	186
6.12	Codebook-Driven Wiener Filtering.....	190
6.13	Audible Noise Suppression Algorithm	194
6.14	Summary	199
	References	200
Chapter 7	Statistical-Model-Based Methods	203
7.1	Maximum-Likelihood Estimators.....	203
7.2	Bayesian Estimators	208
7.3	MMSE Estimator.....	209
7.3.1	MMSE Magnitude Estimator	211
7.3.2	MMSE Complex Exponential Estimator.....	215
7.3.3	Estimating the <i>A Priori</i> SNR	217
7.3.3.1	Maximum-Likelihood Method.....	217
7.3.3.2	Decision-Directed Approach	218
7.4	Improvements to the Decision-Directed Approach.....	219
7.4.1	Reducing the Bias.....	219
7.4.2	Improving the Adaptation Speed	221
7.5	Implementation and Evaluation of the MMSE Estimator	224
7.6	Elimination of Musical Noise.....	225
7.7	Log-MMSE Estimator	227
7.8	MMSE Estimation of the <i>p</i> th-Power Spectrum.....	230
7.9	MMSE Estimators Based on Non-Gaussian Distributions ...	234
7.10	Maximum <i>A Posteriori</i> (Map) Estimators	238
7.11	General Bayesian Estimators.....	241
7.12	Perceptually Motivated Bayesian Estimators	243
7.12.1	Psychoacoustically Motivated Distortion Measure	243
7.12.2	Weighted Euclidean Distortion Measure	243
7.12.3	Itakura–Saito Measure	248
7.12.4	Cosh Measure.....	249
7.12.5	Weighted Likelihood Ratio	251
7.12.6	Modified IS Distortion Measure	252
7.13	Incorporating Speech Absence Probability in Speech Enhancement	254
7.13.1	Incorporating Speech-Presence Uncertainty in Maximum-Likelihood Estimators.....	255
7.13.2	Incorporating Speech-Presence Uncertainty in MMSE Estimators.....	257
7.13.3	Incorporating Speech-Presence Uncertainty in Log-MMSE Estimators	261

7.13.4	Implementation Issues regarding <i>A Priori</i> SNR Estimation	263
7.14	Methods for Estimating the <i>A Priori</i> Probability of Speech Absence.....	264
7.15	Summary	269
	References	270
Chapter 8	Subspace Algorithms	273
8.1	Introduction	273
8.1.1	Definitions	273
8.1.2	Projections	274
8.1.3	Low-Rank Modeling	280
8.2	Using SVD for Noise Reduction: Theory.....	282
8.2.1	SVD Analysis of “Noisy” Matrices.....	282
8.2.2	Least-Squares and Minimum-Variance Estimates of the Signal Matrix	285
8.3	SVD-Based Algorithms: White Noise	287
8.3.1	SVD Synthesis of Speech.....	287
8.3.2	Determining the Effective Rank	293
8.3.3	Noise Reduction Algorithm.....	296
8.4	SVD-Based Algorithms: Colored Noise	297
8.5	SVD-Based Methods: A Unified View	301
8.6	EVD-Based Methods: White Noise	301
8.6.1	Eigenvalue Analysis of “Noisy” Matrices.....	302
8.6.2	Subspace Methods Based on Linear Estimators	306
8.6.2.1	Linear Minimum Mean-Square Estimator (LMMSE).....	307
8.6.2.2	Time-Domain-Constrained Estimator.....	309
8.6.2.3	Spectral-Domain-Constrained Estimator... ..	312
8.6.3	Implementation.....	318
8.6.3.1	Covariance Estimation.....	318
8.6.3.2	Estimating the Lagrange Multiplier.....	320
8.6.3.3	Estimating the Signal Subspace Dimension.....	322
8.7	EVD-Based Methods: Colored Noise.....	325
8.7.1	Pwhitening Approach.....	325
8.7.2	Signal/Noise KLT-Based Method	329
8.7.3	Adaptive KLT Approach.....	332
8.7.4	Subspace Approach with Embedded Pwhitening	335
8.7.4.1	Time-Domain-Constrained Estimator.....	335
8.7.4.2	Spectrum-Domain-Constrained Estimator.....	337
8.7.4.3	Implementation	339
8.7.4.4	Relationship between Subspace Estimators and Pwhitening.....	341

8.8	EVD-Based Methods: A Unified View	346
8.9	Perceptually Motivated Subspace Algorithms	346
8.9.1	Fourier to Eigen-Domain Relationship	347
8.9.2	Incorporating Psychoacoustic Model Constraints....	350
8.9.3	Incorporating Auditory Masking Constraints.....	353
8.10	Subspace-Tracking Algorithms	355
8.10.1	Block Algorithms	356
8.10.2	Recursive Algorithms.....	361
8.10.2.1	Modified Eigenvalue Problem Algorithms.....	362
8.10.2.2	Adaptive Algorithms	363
8.10.3	Using Subspace-Tracking Algorithms in Speech Enhancement	370
8.11	Summary	371
	References	372
Chapter 9	Noise-Estimation Algorithms	377
9.1	Voice Activity Detection vs. Noise Estimation	377
9.2	Introduction to Noise-Estimation Algorithms.....	379
9.3	Minimal-Tracking Algorithms	381
9.3.1	Minimum Statistics Noise Estimation	381
9.3.1.1	Principles	381
9.3.1.2	Derivation of the Bias Factor.....	384
9.3.1.3	Derivation of Optimal Time- and Frequency-Dependent Smoothing Factor.....	389
9.3.1.4	Searching for the Minimum	391
9.3.1.5	Minimum Statistics Algorithm.....	392
9.3.2	Continuous Spectral Minimum Tracking	394
9.4	Time-Recursive Averaging Algorithms for Noise Estimation.....	396
9.4.1	SNR-Dependent Recursive Averaging	397
9.4.2	Weighted Spectral Averaging.....	400
9.4.3	Recursive-Averaging Algorithms Based on Signal-Presence Uncertainty	405
9.4.3.1	Likelihood Ratio Approach	406
9.4.3.2	Minima-Controlled Recursive-Averaging (MCRA) Algorithms.....	409
9.5	Histogram-Based Techniques.....	420
9.6	Other Noise-Estimation Algorithms	427
9.7	Objective Comparison of Noise-Estimation Algorithms	429
9.8	Summary	432
	References	433

PART III Evaluation

Chapter 10	Evaluating Performance of Speech Enhancement Algorithms.....	439
10.1	Quality vs. Intelligibility	439
10.2	Evaluating Intelligibility of Processed Speech.....	440
10.2.1	Nonsense Syllable Tests	441
10.2.2	Word Tests	445
10.2.2.1	Phonetically Balanced Word Tests	445
10.2.2.2	Rhyming Word Tests	446
10.2.3	Sentence Tests	449
10.2.4	Measuring Speech Intelligibility.....	450
10.2.4.1	Speech Reception Threshold	451
10.2.4.2	Using Statistical Tests to Assess Significant Differences: Recommended Practice	453
10.3	Evaluating Quality of Processed Speech.....	459
10.3.1	Relative Preference Methods.....	460
10.3.2	Absolute Category Rating Methods	462
10.3.2.1	Mean Opinion Scores	463
10.3.2.2	Diagnostic Acceptability Measure	465
10.3.2.3	ITU-T P.835 Standard	468
10.4	Evaluating Reliability of Quality Judgments: Recommended Practice	471
10.4.1	Intra-Rater Reliability Measures	471
10.4.2	Inter-Rater Reliability Measures	473
10.5	Summary	475
	References	475
Chapter 11	Objective Quality and Intelligibility Measures.....	479
11.1	Objective Quality Measures	479
11.1.1	Segmental Signal-to-Noise Ratio Measures.....	479
11.1.2	Spectral Distance Measures Based on LPC.....	482
11.1.3	Perceptually Motivated Measures	484
11.1.3.1	Weighted Spectral Slope Distance Measure	485
11.1.3.2	Bark Distortion Measures.....	486
11.1.3.3	Perceptual Evaluation of Speech Quality Measure	491
11.1.4	Wideband PESQ.....	502
11.1.5	Composite Measures	503
11.1.6	Non-Intrusive Objective Quality Measures.....	505
11.1.7	Figures of Merit of Objective Measures.....	506

11.2	Evaluation of Objective Quality Measures.....	508
11.2.1	Correlations of Objective Measures with Quality and Signal/Background Distortions	512
11.2.2	Summary of Findings.....	514
11.3	Quality Measures: Summary of Findings and Future Directions	514
11.4	Speech Intelligibility Measures.....	516
11.4.1	Articulation Index	517
11.4.1.1	Theory.....	517
11.4.1.2	Transfer Function.....	521
11.4.1.3	Derivation of Band-Importance Functions	522
11.4.1.4	Articulation Index Measure.....	524
11.4.1.5	Speech Intelligibility Index (ANSI 1997 Standard)	526
11.4.1.6	Limitations of AI.....	529
11.4.2	Speech Transmission Index.....	537
11.4.2.1	Theory.....	538
11.4.2.2	Speech Modulation Spectrum	543
11.4.2.3	Computation of the STI Measure	546
11.4.2.4	Normalized-Covariance Measure.....	549
11.4.2.5	Limitations of STI Measures.....	553
11.4.3	Coherence-Based Index.....	557
11.4.3.1	Theory.....	557
11.4.3.2	Implementation of MSC	560
11.4.3.3	Computation of the CSII Measure.....	561
11.4.3.4	Computation of the Three-Level CSII.....	562
11.5	Evaluation of Intelligibility Measures	564
11.5.1	Correlation of Intelligibility Measures with Human Intelligibility Scores	565
11.5.2	Summary of Findings.....	567
11.6	Intelligibility Measures: Summary of Findings and Future Directions	568
11.7	Summary	571
	References	571

Chapter 12 Comparison of Speech Enhancement Algorithms 579

12.1	NOIZEUS: A Noisy Speech Corpus for Quality Evaluation of Speech Enhancement Algorithms.....	579
12.2	Comparison of Enhancement Algorithms: Speech Quality	581
12.2.1	Quality Evaluation: Procedure	582
12.2.2	Subjective Quality Evaluation: Results	583
12.2.3	Within-Class Algorithm Comparisons.....	583
12.2.4	Across-Class Algorithm Comparisons.....	589

12.2.5	Comparisons in Reference to Noisy Speech	593
12.2.6	Contribution of Speech and Noise Distortion to Judgment of Overall Quality	597
12.2.7	Summary of Findings.....	598
12.3	Comparison of Enhancement Algorithms: Speech Intelligibility	598
12.3.1	Listening Tests: Procedure	599
12.3.2	Intelligibility Evaluation: Results.....	600
12.3.3	Intelligibility Comparison among Algorithms.....	602
12.3.4	Intelligibility Comparison against Noisy Speech.....	603
12.3.5	Summary of Findings.....	604
12.5	Summary	605
	References	605

PART IV Future Steps

Chapter 13	Algorithms That Can Improve Speech Intelligibility	609
13.1	Reasons for the Absence of Intelligibility Improvement with Existing Noise-Reduction Algorithms	609
13.1.1	Influence of Speech Distortions	610
13.1.2	Lack of Effective SNR Increase.....	612
13.2	Algorithms Based on Channel Selection: A Different Paradigm for Noise Reduction	613
13.3	Channel-Selection Criteria	619
13.3.1	SNR Criterion.....	620
13.3.2	SNR _{ESI} Selection Criterion.....	623
13.3.3	Other Selection Criteria.....	628
13.3.4	Channel-Selection Criteria for Reverberation.....	629
13.3.5	Universal Selection Criterion for All Types of Background Interferers	631
13.4	Intelligibility Evaluation of Channel-Selection-Based Algorithms: Ideal Conditions	632
13.4.1	Broadband Noise Conditions.....	633
13.4.2	Competing-Talker Conditions	634
13.4.3	Reverberant Conditions	639
13.5	Implementation of Channel-Selection-Based Algorithms in Realistic Conditions	639
13.5.1	Simple, But Ineffective, Algorithms for Binary Mask Estimation.....	640
13.5.2	Effective Channel-Selection Algorithms Based on Binary Classifiers	641
13.5.3	Adapting to New Noise Environments.....	645

- 13.6 Evaluating Binary Mask Estimation Algorithms 646
- 13.7 Channel Selection and Auditory Scene Analysis 649
- 13.8 Summary 652
- References 652

- Appendix A: Special Functions and Integrals 657**
- Appendix B: Derivation of the MMSE Estimator 661**
- Appendix C: MATLAB® Code and Speech/Noise Databases 665**
- Index 677**