Contents

2.2 STFT Phase: Background and Some Remarks 33
  2.2.1 Short-Time Fourier Transform 33
  2.2.2 Fourier Analysis of Speech: STFT Amplitude and Phase 34
2.3 Phase Unwrapping 35
  2.3.1 Problem Definition 35
  2.3.2 Remarks on Phase Unwrapping 38
  2.3.3 Phase Unwrapping Solutions 38
    2.3.3.1 Detecting Discontinuities 39
    2.3.3.2 Numerical Integration (NI) 40
    2.3.3.3 Isolating Sharp Zeros 41
    2.3.3.4 Iterative Phase Unwrapping 41
    2.3.3.5 Polynomial Factorization (PF) 42
    2.3.3.6 Time Series Approach 42
    2.3.3.7 Composite Method 43
    2.3.3.8 Schur–Cohn and Nyquist Frequency 44
2.4 Useful Phase-Based Representations 44
  2.4.1 Group Delay Representations 45
  2.4.2 Instantaneous Frequency 48
  2.4.3 Baseband Phase Difference 49
  2.4.4 Harmonic Phase Decomposition 50
    2.4.4.1 Background on the Harmonic Model 50
    2.4.4.2 Phase Decomposition using the Harmonic Model 51
  2.4.5 Phasegram: Unwrapped Harmonic Phase 52
    2.4.5.1 Definitions and Background 52
    2.4.5.2 Circular Mean and Variance 52
  2.4.6 Relative Phase Shift 53
  2.4.7 Phase Distortion 54
2.5 Experiments 57
  2.5.1 Experiment 2.1: One-Dimensional Phase Unwrapping 57
    2.5.1.1 Clean Signal Scenario 57
    2.5.1.2 Noisy Signal Scenario 58
  2.5.2 Experiment 2.2: Comparative Study of Phase Unwrapping Methods 58
  2.5.3 Experiment 2.3: Comparative Study on Group Delay Spectra 59
  2.5.4 Experiment 2.4: Circular Statistics of the Harmonic Phase 60
  2.5.5 Experiment 2.5: Circular Statistics of the Spectral Phase 62
  2.5.6 Experiment 2.6: Comparative Study of Phase Representations 63
2.6 Summary 65
References 65

3 Phase Estimation Fundamentals 71
Josef Kulmer and Pejman Mowlaee
3.1 Chapter Organization 71
3.2 Phase Estimation Fundamentals 71
  3.2.1 Background and Fundamentals 71
  3.2.2 Key Examples: Phase Estimation Problem 72
    3.2.2.1 Example 1: Discrete-Time Sinusoid 72
    3.2.2.2 Example 2: Discrete-Time Sinusoid in Noise 76
# Contents

3.2.3 Phase Estimation 80
3.2.3.1 Maximum Likelihood Estimation 80
3.2.3.2 Maximum a Posteriori Estimation 83
3.3 Existing Solutions 84
3.3.1 Iterative Signal Reconstruction 84
3.3.1.2 Griffin–Lim Algorithm (GLA) 85
3.3.1.3 Extensions of the GLA 87
3.3.2 Phase Reconstruction Across Time 89
3.3.3 Phase Reconstruction Across Frequency 90
3.3.4 Phase Randomization 91
3.3.5 Geometry-Based Phase Estimation 93
3.3.6 Least Squares (LS) 95
3.3.7 Spectro-Temporal Smoothing of Unwrapped Phase 97
3.3.7.1 Signal Segmentation 97
3.3.7.2 Linear Phase Removal 98
3.3.7.3 Apply Smoothing Filter 98
3.3.7.4 Reconstruction of the Enhanced-Phase Signal 101
3.4 Experiments 101
3.4.1 Experiment 3.1: Monte Carlo Simulation Comparing ML and MAP 101
3.4.2 Experiment 3.2: Monte Carlo Simulation on Window Impact 103
3.4.3 Experiment 3.3: Phase Recovery Using the Griffin–Lim Algorithm 105
3.4.4 Experiment 3.4: Phase Estimation for Speech Enhancement: A Comparative Study 105
3.5 Summary 107

References 108

## Part II Applications 113

4 Phase Processing for Single-Channel Speech Enhancement 115

*Johannes Stahl and Pejman Mowlaee*

4.1 Introduction and Chapter Organization 115
4.2 Speech Enhancement in the STFT Domain: General Concepts 116
4.2.1 *A priori* SNR Estimation 116
4.2.1.1 Decision-Directed *a priori* SNR Estimation 117
4.2.1.2 Cepstro-Temporal Smoothing 118
4.2.2 Noise PSD Estimation 118
4.2.2.1 Minimum Statistics 119
4.3 Conventional Speech Enhancement 119
4.3.1 Statistical Model 119
4.3.2 Short-Time Spectral Amplitude Estimation 121
4.4 Phase-Sensitive Speech Enhancement 123
4.4.1 Phase Estimation for Signal Reconstruction 123
4.4.2 Spectral Amplitude Estimation Given the STFT Phase 124
4.4.3 Iterative Closed-Loop Phase-Aware Single-Channel Speech Enhancement 126
5 Phase Processing for Single-Channel Source Separation 143

Pejman Mowlaee and Florian Mayer

5.1 Chapter Organization 143
5.2 Why Single-Channel Source Separation? 143
5.2.1 Background 143
5.2.2 Problem Formulation 144
5.3 Conventional Single-Channel Source Separation 145
5.3.1 Source-Driven SCSS 146
5.3.1.1 Ideal Binary Mask 147
5.3.1.2 Ideal Ratio Mask 147
5.3.2 Model-Based SCSS 147
5.3.2.1 Deep Learning 149
5.3.2.2 Non-Negative Matrix Factorization 150
5.4 Phase Processing for Single-Channel Source Separation 152
5.4.1 Complex Matrix Factorization Methods 152
5.4.1.1 Complex Matrix Factorization 152
5.4.1.2 Complex Matrix Factorization with Intra-Source Additivity 154
5.4.2 Phase Importance for Signal Reconstruction 155
5.4.2.1 Multiple Input Spectrogram Inversion 155
5.4.2.2 Partial Phase Reconstruction 156
5.4.2.3 Informed Source Separation Using Iterative Reconstruction (ISSIR) 157
5.4.2.4 Sinusoidal-Based PPR 158
5.4.2.5 Spectrogram Consistency 159
5.4.2.6 Geometry-Based Phase Estimation 160
5.4.2.7 Phase Decomposition and Temporal Smoothing 162
5.4.2.8 Phase Reconstruction of Spectrograms with Linear Unwrapping 163
5.4.3 Phase-Aware Time–Frequency Masks 164
5.4.3.1 Phase-Insensitive Masks 164
5.4.3.2 Phase-Sensitive Mask 165
5.4.3.3 Complex Ratio Mask 165
5.4.3.4 Complex Mask 166
5.4.4 Phase Importance in Signal Interaction Models 166
5.5 Experiments 168
5.5.1 Experiment 5.1: Phase Estimation for Proof-of-Concept Signal Reconstruction 168
5.5.2 Experiment 5.2: Comparative Study of GLA-Based Phase Reconstruction Methods 168
5.5.2.1 Convergence Analysis 169
5.5.2.2 Quantized Scenario 169
5.5.3 Experiment 5.3: Phase-Aware Time–Frequency Mask 170
5.5.4 Experiment 5.4: Phase-Sensitive Interaction Functions 172
5.5.5 Experiment 5.5: Complex Matrix Factorization 172
5.6 Summary 174
References 174

6 Phase-Aware Speech Quality Estimation 179
Pejman Mowlaee
6.1 Chapter Organization 179
6.2 Introduction: Speech Quality Estimation 179
6.2.1 General Definition of Speech Quality 180
6.2.2 Speech Quality Estimators: Amplitude, Phase, or Both? 181
6.3 Conventional Instrumental Metrics for Speech Quality Estimation 182
6.3.1 Perceived Quality 182
6.3.2 Speech Intelligibility 184
6.4 Why Phase-Aware Metrics? 188
6.4.1 Phase and Speech Intelligibility 188
6.4.2 Phase and Perceived Quality 188
6.5 New Phase-Aware Metrics 189
6.5.1 Group Delay Deviation 189
6.5.2 Instantaneous Frequency Deviation 190
6.5.3 Unwrapped MSE 190
6.5.4 Phase Deviation 190
6.5.5 UnHPSNR and UnRMSE 191
6.6 Subjective Tests 191
6.6.1 CCR Test 192
6.6.2 MUSHRA Test 192
6.6.3 Statistical Analysis 193
6.6.4 Speech Intelligibility Test 194
6.6.5 Evaluation of Speech Quality Measures 196
6.7 Experiments 198
6.7.1 Experiment 6.1: Impact of Phase Modifications on Speech Quality 199
6.7.2 Experiment 6.2: Phase and Perceived Quality Estimation 201
6.7.3 Experiment 6.3: Phase and Speech Intelligibility Estimation 202
6.7.4 Experiment 6.4: Evaluating the Phase Estimation Accuracy 203
6.8 Summary 205
References 205

7 Conclusion and Future Outlook 210
Pejman Mowlaee
7.1 Chapter Organization 210
7.2 Renaissance of Phase-Aware Signal Processing: Decline and Rise 210