Contents

Abstract iii

Resumen v

Resum vii

Acknowledgements ix

Abbreviations and Acronyms xvii

1 Introduction 1
- 1.1 Background and Motivation . . . 1
- 1.2 Objectives of the Thesis . . . 3
- 1.3 Organization of Thesis . . . 3

2 State of the Art 5
- 2.1 Sound Source Localization . . . 5
  - 2.1.1 Signal Model . . . 5
  - 2.1.2 Time Difference Of Arrival (TDOA) . . . 6
  - 2.1.3 Generalized Cross Correlation . . . 7
  - 2.1.4 Steered Response Power . . . 8
  - 2.1.5 SRP-PHAT Variants . . . 11
- 2.2 Automatic Speech Recognition . . . 12
  - 2.2.1 The Speech Recognition Problem . . . 12
  - 2.2.2 Hidden Markov Models . . . 14
  - 2.2.3 Speech Recognition Evaluations . . . 15
- 2.3 Source Separation and Enhancement . . . 16
  - 2.3.1 Mixing Models . . . 17
  - 2.3.2 Source Separation Tasks and Approaches . . . 20
  - 2.3.3 Underdetermined Source Separation . . . 21
  - 2.3.4 Time-Frequency Masking Limitations . . . 25

3 Paper Contributions and Discussion 27
- 3.1 Modified SRP-PHAT Functional . . . 28
  - 3.1.1 Abstract . . . 28
  - 3.1.2 Contributions . . . 28
- 3.2 A SRP Iterative Method for Source Localization . . . 28
  - 3.2.1 Abstract . . . 28
  - 3.2.2 Contributions . . . 29
- 3.3 A Real-Time SSL and Enhancement System . . . 29
  - 3.3.1 Abstract . . . 29
  - 3.3.2 Contributions . . . 29
- 3.4 Real-Time Speaker Localization and Detection System . . . 30
  - 3.4.1 Abstract . . . 30
  - 3.4.2 Contributions . . . 30
- 3.5 ASR in Cocktail-Party Situations . . . 30

3.5.1 Abstract . . . 30
3.5.2 Contributions . . . 31
3.6 Influence of Sound Separation Methods in ASR . . . 31
3.6.1 Abstract . . . 31
3.6.2 Contributions . . . 31

4 A Modified SRP-PHAT Functional for Robust Real-Time Sound Source Localization with Scalable Spatial Sampling 33
4.1 Introduction . . . 33
4.2 The SRP-PHAT Algorithm . . . 34
4.3 The Inter-Microphone Time Delay Function . . . 35
4.4 Proposed Approach . . . 37
4.4.1 Computation of Accumulation Limits . . . 37
4.4.2 Computational Cost . . . 38
4.5 Experiments . . . 38
4.6 Conclusion . . . 39
4.7 References . . . 40

5 A Steered Response Power Iterative Method for High-Accuracy Acoustic Source Localization 43
5.1 Introduction . . . 43
5.2 Modified SRP Algorithm . . . 44
5.3 Proposed Approach . . . 45
5.3.1 Mean-Based Functional . . . 45
5.3.2 Iterative Sub-Volume Decomposition . . . 45
5.3.3 Computational Cost . . . 46
5.4 Experiments . . . 47
5.4.1 Influence of Initial Resolution . . . 48
5.4.2 Algorithm Comparison . . . 49
5.4.3 Real Setup . . . 50
5.4.4 Discussion . . . 50
5.5 Conclusion . . . 50
5.6 References . . . 50

6 A Real-Time Sound Source Localization and Enhancement System Using Distributed Microphones 51
6.1 Introduction . . . 51
6.2 SRP-PHAT Sound Source Localization . . . 53
6.2.1 Modified SRP-PHAT . . . 53
6.3 Enhancement . . . 54
6.3.1 Delay and Sum Beamformer . . . 54
6.3.2 SRP-PHAT Binary Masking . . . 55
6.3.3 Estimation of Localization Error . . . 55
6.4 Experiments . . . 57
6.4.1 Results . . . 57
6.5 Conclusion . . . 58
6.6 References . . . 58

7 Real-Time Speaker Localization and Detection System for Camera Steering in Multiparticipant Videoconferencing Environments **61**
7.1 Introduction . . . 61
7.2 SRP-Based Source Localization . . . 62
7.2.1 Modified SRP-PHAT Functional . . . 63
7.3 Speaker Detection . . . 63
7.3.1 Distribution of Location Estimates . . . 63
7.3.2 Speech/Non-Speech Discrimination . . . 64
7.3.3 Camera Steering . . . 65
7.4 Experiments . . . 65
7.4.1 Results . . . 66
7.5 Conclusion . . . 67
7.6 References . . . 67

8 Automatic Speech Recognition in Cocktail-Party Situations: A Specific training for Separated Speech **69**
8.1 Introduction . . . 69
8.2 Robust Speech Recognition . . . 71
8.2.1 Speech Recognition in Cocktail-Party Situations . . . 71
8.3 Source Separation in Real Environments . . . 72
8.3.1 Speech Separation Based on Interclass Variance Maximization . . . 73
8.4 ASR Training for Reverberant and Simultaneous Speech . . . 73
8.4.1 Recognizer Specifications and Baseline System . . . 74
8.4.2 Training Data-sets . . . 74
8.5 Experiments . . . 76
8.5.1 Experiment 1: Single Speech Recognition . . . 76
8.5.2 Experiment 2: Speech Recognition With Interfering Speech . . . 76
8.5.3 Experiment 3: Simultaneous Speech Recognition . . . 78
8.6 Discussion . . . 78
8.7 Conclusion . . . 78
8.8 References . . . 79

9 Evaluating the Influence of Source Separation Methods in Robust Automatic Speech Recognition with a Specific Cocktail-Party Training **81**
9.1 Introduction . . . 81
9.2 Speech Recognition in Cocktail-Party Situations . . . 83
9.3 Sound Source Separation . . . 84
9.3.1 Multi-Level Thresholding Separation . . . 85
9.3.2 Separation with Full-Rank Spatial Convariance Models . . . 85
9.4 Experiments . . . 85
9.4.1 Experiment 1: Test Data-Set Using Multi-Level Thresholding Separation 86
9.4.2 Experiment 2: Test Data-Set Using Separation with Full-Rank Spatial Covariance Models . . . 86
9.5 Conclusion . . . 87
9.6 References . . . 87

10 Conclusions and Future Research **89**
10.1 Summary and Conclusions 89
10.2 Further Work 91

Bibliography **92**