Time-Domain Beamforming and Blind Source Separation

Speech Input in the Car Environment

Lecture Notes in Electrical Engineering

Volume 3

Time-Domain Beamforming and Blind Source Separation
Julien Bourgeois, and Wolfgang Minker
ISBN 978-0-387-68835-0, 2007

Digital Noise Monitoring of Defect Origin
Telman Aliev
ISBN 978-0-387-71753-1, 2007

Multi-Carrier Spread Spectrum 2007
Simon Plass, Armin Dammann, Stefan Kaiser, and K. Fazel
ISBN 978-1-4020-6128-8, 2007

Julien Bourgeois • Wolfgang Minker

Time-Domain Beamforming and Blind Source Separation

Speech Input in the Car Environment

Springer

Julien Bourgeois
Universität Ulm
Institut für Informationstechnik
Albert-Einstein-Allee 43
89081 Ulm
Germany

Wolfgang Minker
Universität Ulm
Institut für Informationstechnik
Albert-Einstein-Allee 43
89081 Ulm
Germany

ISBN 978-0-387-68835-0 e-ISBN 978-0-387-68836-7
DOI 10.1007/978-0-387-68836-7
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2007932057

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The development of computer and telecommunication technologies led to a revolution in the way that people work and communicate with each other. One of the results is that large amount of information will increasingly be held in a form that is natural for users, as speech in natural language. In the presented work, we investigate the speech signal capture problem, which includes the separation of multiple interfering speakers using microphone arrays.

Adaptive beamforming is a classical approach which has been developed since the seventies. However it requires a double-talk detector (DTD) that interrupts the adaptation when the target is active, since otherwise target cancelation occurs. The fact that several speakers may be active simultaneously makes this detection difficult, and if additional background noise occurs, even less reliable. Our proposed approaches address this separation problem using continuous, uninterrupted adaptive algorithms. The advantage seems twofold: Firstly, the algorithm development is much simpler since no detection mechanism needs to be designed and no threshold is to be tuned. Secondly, the performance may be improved due to the adaptation during periods of double-talk.

In the first part of the book, we investigate a modification of the widely used NLMS algorithm, termed Implicit LMS (ILMS), which implicitly includes an adaptation control and does not require any threshold. Experimental evaluations reveal that ILMS mitigates the target signal cancelation substantially with the distributed microphone array. However, in the more difficult case of the compact microphone array, this algorithm does not sufficiently reduce the target signal cancelation. In this case, more sophisticated blind source separation techniques (BSS) seem necessary.

The second part is dedicated to blind separation techniques, much more recent than classical adaptive beamforming (the first results with real acoustic mixings appearing in the nineties). Our objective was to evaluate the performance of blind separation techniques relative to that of more mature beamforming approaches. In addition, we wanted to combine the advantages of beamforming, notably its performance and robustness, with those of blind source separation which does not require activity detection. Parra's frequency-domain block-diagonalization algorithm served as a benchmark in the field of blind source separation. However, we realized that this algorithm could not be flexibly combined with beamformers because it fails to cope with certain "acausal" type of source mixing. We therefore focused on the time-domain approach by Buchner. This approach has been extended to be applied in case there are more microphones than sources. At a moderate computational cost, the proposed Partial BSS scheme flexibly exploits all microphone signals and provides multiple interferer references. Moreover, we derive self-closed update rules that emerge as very robust relative to other algorithms in an experimental comparison. An emphasis is also placed on the theoretical study of BSS, evidencing the role of the causality of the mixing system.

In the last part of the book, we combine both, the beamforming and BSS approaches. While the input of geometrical prior information may increase the start-up performance, we show that the performance gain after the initial convergence is limited. The use of an adaptive interference canceler as a postprocessor leads to a higher interference suppression and also to a higher cancelation of the desired signal. However, we will see that the cancelation of the desired signal may be kept moderate by adequately combining BSS with the ILMS algorithm *and* geometrical prior information.

The presented book results from a cooperation between DaimlerChrysler and the University of Ulm. The industrial partner provided the privileged application field of the car environment and we applied two different, plausible experimental settings using compact and distributed microphone arrays. However, the proposed methods are quite general and should be easily portable to other environments and to different applications.

Acknowledgments

This project would not have been possible without the support of many people.

The authors would like to express their deepest gratitude to Jürgen Freudenberger from the University of Applied Sciences in Konstanz, who read the draft since the early revisions, was always open for discussions and provided numerous advices.

Klaus Linhard from DaimlerChrysler Research and Technology substantially supported the work by his profound technical experience and assistance.

Walter Kellermann from the University of Erlangen took the time to read the manuscript. His valuable comments and constructive criticism helped to make sense of the confusion.

Thanks to Guillaume Lathoud for the many stimulating discussions and his very sound feedback.

Thanks to Jason Ward and Caitlin Womersley from Springer for their assistance during the publishing process.

The authors acknowledge the financial support provided through the HOARSE project from the European Community's Human Potential Programme.

Contents

1 Introduction 1
1.1 Existing Approaches: A Brief Overview 2
1.2 Scope and Objective of the Book 3
1.3 Outline of the Book 4

2 Source Separation as a Multichannel Linear Filtering Problem 7
2.1 The Mixing Channels 7
2.2 The Separation Filters 10
2.2.1 Single Output Systems 10
2.2.2 Multiple Output Systems 13
2.2.3 The Spatial Response 14
2.2.4 Particular Cases 17
2.3 Spatial Filtering vs. Spectral Filtering 18
2.3.1 Minimum Filter Length for Spatial Separation 20
2.3.2 Particular Cases 21
2.4 Performance Measures 22
2.4.1 Compact Microphone Array 23
2.4.2 Distributed Microphone Array 23
2.4.3 Start-Up Performance and Performance after Initial Convergence 24
2.5 Summary and Conclusion 24

3 Linearly Constrained Minimum Variance Beamforming 27
3.1 LCMV Beamforming 27
3.2 From LCMV to Generalized Sidelobe Canceler (GSC) 30
3.3 Constraints for Compact and Distributed Setups 31
3.3.1 Constraint for Compact Microphone Array 31
3.3.2 Constraint for Distributed Microphone Array 33
3.4 The Target Signal Cancelation Problem 34
3.4.1 The Energy-Inversion Effect 35

3.4.2 Countermeasures: On the Necessity of a Double-Talk Detector ... 36
3.5 Summary and Conclusion ... 37

4 Implicit Adaptation Control for Beamforming ... 39
4.1 Adaptive Interference Canceler ... 39
4.2 Implicit Adaptation Control ... 42
4.3 Analysis of the ILMS Algorithm ... 43
4.3.1 Linking ILMS to the NLMS with Optimal Step-Size ... 43
4.3.2 ILMS Transient Behavior and Stability ... 46
4.4 Robustness Improvement ... 51
4.5 Experiments ... 52
4.5.1 Experiments with the Four-Element Compact Array Mounted in the Rear-View Mirror ... 55
4.5.2 Experiment with the Two-Element Distributed Array Mounted on the Car Ceiling ... 59
4.6 Summary and Conclusion ... 61

5 Second-Order Statistics Blind Source Separation ... 63
5.1 Problem and Notations ... 65
5.1.1 From a Scalar to a Convolutive Mixture Model ... 65
5.1.2 Separation Ambiguities ... 68
5.2 Nonstationarity and Source Separation ... 69
5.2.1 The Insufficiency of Decorrelation ... 69
5.2.2 Nonstationarity-Based Cost Function ... 70
5.3 Gradient-Based Minimization ... 73
5.3.1 Standard Gradient ... 73
5.3.2 Natural Gradient ... 74
5.4 Natural Gradient Algorithm for NonSquare Systems ... 75
5.5 Summary and Conclusion ... 79

6 Implementation Issues in Blind Source Separation ... 81
6.1 Natural Gradient in the Convolutive Case ... 81
6.1.1 Gradient in the Sylvester Subspace ... 82
6.1.2 From Matrices to z-Transforms ... 84
6.1.3 Self-Closed and Non-Self-Closed Natural Gradients ... 87
6.1.4 From z-Transforms Back to the Time Domain ... 89
6.1.5 Application to NG-SOS-BSS ... 91
6.1.6 Discussion: Which Natural Gradient is Best? ... 93
6.2 Online Adaptation ... 97
6.2.1 Block-Wise Batch Adaptation ... 97
6.2.2 Sample-Wise Adaptation ... 98
6.3 Experimental Results ... 100
6.3.1 Experiments with the Four-Element Compact Array Mounted in the Rear-View Mirror ... 101

6.3.2 Experiments with the Two-Element Distributed Array Mounted on the Car Ceiling ... 106
6.3.3 Comparison with Other BSS Algorithms in the Frequency Domain ... 108
6.4 Summary and Conclusion ... 111

7 On the Convergence and Stability in Second-Order Statistics BSS ... 113
7.1 Global Convergence ... 113
7.1.1 Difficulty of a Global Convergence Analysis ... 114
7.1.2 Convergence Analysis for a Simplified Algorithm ... 116
7.2 Local Stability ... 121
7.3 Summary and Conclusion ... 123

8 Comparison of LCMV Beamforming and Second-Order Statistics BSS ... 125
8.1 Properties of the Cost Functions ... 126
8.1.1 Shape of the Cost Function ... 126
8.1.2 On the Estimation Variance ... 128
8.2 Complexity ... 133
8.2.1 NLMS Complexity ... 134
8.2.2 Complexity of NG-SOS-BSS Algorithms ... 135
8.2.3 Comparison of NLMS and NG-SOS-BSS Complexities ... 138
8.3 Links with the ILMS Algorithm ... 140
8.4 Experimental Comparison ... 141
8.5 Summary and Conclusion ... 145

9 Combining Second-Order Statistics BSS and LCMV Beamforming ... 147
9.1 Existing Combinations ... 148
9.2 BSS and Geometric Prior Information ... 149
9.2.1 Causality Information ... 150
9.2.2 Prior Information on the Source Direction of Arrival ... 151
9.2.3 Geometric Information at the Initialization ... 154
9.2.4 Geometric Information as a Soft Constraint ... 156
9.2.5 Geometric Information as a Preprocessing ... 160
9.3 Combining SOS-BSS and the Power Criterion ... 163
9.4 Combining SOS-BSS with Geometric Prior Information and the Power Criterion ... 165
9.5 Experimental Results on Automatic Speech Recognition ... 167
9.6 Summary and Conclusion ... 171

10 Summary and Conclusions ... 173

Experimental Setups . . . 179
A.1 The Four-Element Compact Array Mounted in the Rear-View Mirror . . . 179
A.2 The Two-Element Distributed Array Mounted on the Car Ceiling . . . 179
A.3 Acoustic Characteristics of the Car Cabin . . . 181
A.4 Illustration of the Difficulty in the Design of a Reliable DTD . . . 181

Far- and Free-Field Acoustic Propagation Model and Null Beamforming . . . 185
B.1 Far- and Free-Field Model . . . 185
B.2 Null Beamforming . . . 186

The RGSC According to Hoshuyama et al. . . . 189
C.1 RGSC for the Four-Element Compact Array Mounted in the Rear-View Mirror . . . 189
C.2 RGSC for the Two-Element Distributed Array Mounted on the Car Ceiling . . . 191
C.3 Experimental Comparison: GSC vs. RGSC . . . 192
C.3.1 Experiments with the Four-Element Compact Array Mounted in the Rear-View Mirror . . . 194
C.3.2 Experiments with the Two-Element Distributed Array Mounted on the Car Ceiling . . . 195
C.4 Conclusion . . . 196

Stability Analysis . . . 199
D.1 Mixing and Separation Models . . . 199
D.2 Linearization of the NG-SOS-BSS Updates . . . 200
D.3 Local Stability Conditions . . . 203

Notations . . . 207

References . . . 215

Index . . . 223