

Lecture Notes
in Control and Information Sciences 315

Editors: M. Thoma · M. Morari

Wolfgang Herbordt

Sound Capture for Human/Machine Interfaces

**Practical Aspects
of Microphone Array Signal Processing**

With 73 Figures

Series Advisory Board

F. Allgöwer · P. Fleming · P. Kokotovic · A.B. Kurzhanski ·
H. Kwakernaak · A. Rantzer · J.N. Tsitsiklis

Author

Dr. Wolfgang Herboldt
ATR – Advanced Telecommunications Research Institute International
Spoken Language Translation Research Laboratories
2-2, Hikaridai, Seiko-cho, Soraku-gun
Kyoto 619-0288
Japan

ISSN 0170-8643

ISBN 3-540-23954-5 **Springer Berlin Heidelberg New York**

Library of Congress Control Number: 2005920066

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in other ways, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable to prosecution under German Copyright Law.

Springer is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2005
Printed in The Netherlands

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Data conversion by author.
Final processing by PTP-Berlin Protago- \TeX -Production GmbH, Germany
Cover-Design: design & production GmbH, Heidelberg
Printed on acid-free paper 89/3141/Yu - 5 4 3 2 1 0

Preface

For convenient human/machine interaction, acoustic front-ends are required which allow seamless and hands-free audio communication. For maximum speech intelligibility and optimum speech recognition performance, interference, noise, reverberation, and acoustic echoes of loudspeakers should be suppressed. Microphone array signal processing is advantageous to single-channel speech enhancement since the spatial dimension can be exploited in addition to the temporal dimension.

In this work, joint adaptive beamforming and acoustic echo cancellation with microphone arrays is studied with a focus on the challenges of practical systems. Particularly, the following aspects are efficiently solved, leading to a real-time system, which was successfully used in the real world:

- High suppression of both strongly time-varying interferers, as, e.g., competing speakers, and slowly time-varying diffuse noise, as, e.g., car noise in passenger cabins of cars,
- Efficient cancellation of acoustic echoes of multi-channel reproduction systems even in strongly time-varying acoustic conditions with high background noise levels and with limited computational resources,
- High output signal quality with limited array apertures and limited numbers of microphones because of product design constraints,
- Robustness against reverberation with respect to the desired signal, moving desired sources, or array imperfections such as position errors or gain and phase mismatch of the microphones.

Detailed theoretical analysis and experimental studies illustrate the performance of the system. Audio examples can be found on the web page <http://www.wolfgangherbordt.de/micarraybook/>. Special focus is put on the reproducibility of the results by giving detailed descriptions of the proposed algorithms and of the parameter settings.

The intended audience of this book is both specialists and readers with general knowledge of statistical and adaptive signal processing. For any question or comment, please don't hesitate to contact the author!

Acknowledgements

I would especially like to thank my supervisor, Prof. Walter Kellermann of the Friedrich-Alexander University in Erlangen, Germany, for the unique opportunity to unify scientific and private interests in his research group in a very fruitful atmosphere with many productive discussions.

Since the beginning, this research was funded by several grants from Intel Corp., which made this work possible and which particularly led to the practical aspects work. I am especially thankful to David Graumann of Intel Corp., Hillsboro, OR, and Jia Ying of the Intel China Research Center, Beijing, China, for continuously supporting and promoting this work within and outside of Intel. I would also like to thank all the other people working with Intel who made my stays in China and in the United States unforgettable experiences.

I would like to thank Prof. Rainer Martin of the Ruhr-University in Bochum, Germany, Prof. Heinrich Niemann of the Friedrich-Alexander University in Erlangen, and Darren Ward of the Imperial College in London, UK, for their interest in my work, for reviewing this thesis, and for finding the time to participate in the defense of this thesis.

I am very thankful to everybody working in the Telecommunications Laboratory in Erlangen who made my stay here so enjoyable. I especially would like to thank my 'office mate' Lutz Trautmann for his friendship, his advises, and for being so considerate with his experiments with new algorithms for the simulation of musical instruments. I also would like to thank Herbert Buchner for many fruitful discussions about adaptive filter theory and for the great collaboration. Further, I would like to thank Ursula Arnold for her invaluable administrative support, and Rüdiger Nägel and Manfred Lindner for the construction of microphone array hardware.

I would like to thank all the people that I know through my numerous business trips for their helpful discussions and for the great moments together. Especially, I would like to thank Henning Puder of Siemens Audiologische Technik in Erlangen for proof-reading this thesis and Satoshi Nakamura of

VIII Acknowledgements

ATR in Kyoto, Japan, for giving me the possibility to continue this research with a focus on automatic speech recognition in the ATR labs.

Finally, I am very thankful to my family and to my friends for their continuous encouragement, for their understanding, and for the infinite number of relaxed moments during the last years.

Erlangen,
June 2004

Wolfgang Herbordt

Contents

1	Introduction	1
2	Space-Time Signals	5
2.1	Propagating Wave Fields	6
2.2	Spatio-Temporal Random Fields.....	12
2.2.1	Statistical Description of Space-Time Signals	13
2.2.2	Spatio-temporal and Spatio-spectral Correlation Matrices	15
2.3	Summary	23
3	Optimum Linear Filtering	25
3.1	Generic Multiple-Input Multiple-Output (MIMO) Optimum Filtering	27
3.1.1	Structure of a MIMO Optimum Filter.....	27
3.1.2	Least-Squares Error (LSE) Optimization	28
3.1.3	Minimum Mean-Squared Error (MMSE) Optimization in the DTFT Domain	33
3.2	Applications of MIMO Optimum Filters.....	35
3.2.1	System Identification	35
3.2.2	Inverse Modeling	36
3.2.3	Interference Cancellation	38
3.3	Discussion	38
4	Optimum Beamforming for Wideband Non-stationary Signals	41
4.1	Space-Time Signal Model.....	43
4.1.1	Desired Signal	43
4.1.2	Interference.....	45
4.1.3	Sensor Noise	46
4.1.4	Sensor Signals	47
4.2	Space-Time Filtering with Sensor Arrays	48

4.2.1	Concept of Beamforming	48
4.2.2	Beamformer Response and Interference-Independent Performance Measures	51
4.2.3	Interference-Dependent Performance Measures	56
4.2.4	Spatial Aliasing and Sensor Placement	60
4.3	Data-Independent Beamformer Design	63
4.4	Optimum Data-Dependent Beamformer Designs	66
4.4.1	LSE/MMSE Design	67
4.4.2	Linearly-Constrained Least-Squares Error (LCLSE) and Linearly-Constrained Minimum Variance (LCMV) Design	77
4.4.3	Eigenvector Beamformers	91
4.4.4	Suppression of Correlated Interference	93
4.5	Discussion	94
5	A Practical Audio Acquisition System Using a Robust GSC (RGSC)	99
5.1	Spatio-temporal Constraints	100
5.2	RGSC as an LCLSE Beamformer with Spatio-temporal Constraints	101
5.2.1	Quiescent Weight Vector	102
5.2.2	Blocking Matrix	102
5.2.3	Interference Canceller	106
5.3	RGSC in the DTFT Domain	107
5.4	RGSC Viewed from Inverse Modeling and from System Identification	110
5.4.1	Blocking Matrix	110
5.4.2	Interference Canceller	113
5.5	Experimental Results for Stationary Acoustic Conditions	115
5.5.1	Performance Measures in the Context of the RGSC	116
5.5.2	Experimental Setup	117
5.5.3	Interference Rejection of the RGSC	118
5.5.4	Cancellation of the Desired Signal by the Blocking Matrix	122
5.6	Strategy for Determining Optimum RGSC Filters	127
5.6.1	Determination of the Optimum Filters for the Blocking Matrix	127
5.6.2	Determination of the Optimum Filters for the Interference Canceller	129
5.7	Relation to Alternative GSC Realizations	130
5.8	Discussion	131

6	Beamforming Combined with Multi-channel Acoustic Echo Cancellation	133
6.1	Multi-channel Acoustic Echo Cancellation	134
6.1.1	Problem Statement	134
6.1.2	Challenges	135
6.2	Combination of Beamforming and Acoustic Echo Cancellation	137
6.2.1	‘AEC First’	138
6.2.2	‘Beamformer First’	141
6.3	Integration of Acoustic Echo Cancellation into the GSC	143
6.3.1	AEC After the Quiescent Weight Vector (GSAEC)	145
6.3.2	AEC Combined with the Interference Canceller (GEIC)	152
6.4	Discussion	158
7	Efficient Real-Time Realization of an Acoustic Human/Machine Front-End	163
7.1	Multi-channel Block-Adaptive Filtering in the Discrete Fourier Transform (DFT) Domain	164
7.1.1	Optimization Criterion	165
7.1.2	Adaptive Algorithm	167
7.2	RGSC Combined with Multi-channel Acoustic Echo Cancellation in the DFT Domain	170
7.2.1	RGSC in the DFT Domain	170
7.2.2	Combination with the AEC	177
7.2.3	Real-Time Algorithm and Computational Complexity	181
7.3	Experimental Results	185
7.3.1	Comparison of the DFT-Bin-Wise Adaptation with a Full-Band Adaptation	185
7.3.2	Comparison of the RGSC with a GSC Using a Fixed Blocking Matrix	188
7.3.3	Comparison of the Proposed Adaptation Control with an ‘Ideal’ Adaptation Control	191
7.3.4	Application of the RGSC as a Front-End for an Automatic Speech Recognizer (ASR)	193
7.4	Discussion	202
8	Summary and Conclusions	205
A	Estimation of Signal-to-Interference-Plus-Noise Ratios (SINRs) Exploiting Non-stationarity	209
A.1	Biased Estimation of the SINR Using Spatial Information	210
A.1.1	Principle	210
A.1.2	Biased SINR Estimation in the DTFT Domain	211
A.1.3	Illustration	212
A.2	Unbiased SINR Estimation Using Spatial Coherence Functions	213
A.3	Algorithm of the Unbiased SINR Estimation	215

XII Contents

A.3.1	Estimation of the PSDs in the DFT Domain	215
A.3.2	Double-Talk Detection	216
A.3.3	Unbiased Estimation of the SINR	217
A.3.4	Robustness Improvement	218
A.3.5	Summary of the Algorithm and Experimental Results ..	219
B	Experimental Setups and Acoustic Environments	225
B.1	Passenger Cabin of a Car	226
B.2	Multimedia Room	227
C	Notations	229
C.1	Conventions	229
C.2	Abbreviations and Acronyms	229
C.3	Mathematical Symbols	230
	References	251
	Index	267