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

Springer

Series Advisory Board

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

Author

Dr. Wolfgang Herbordt
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

Springer is a part of Springer Science+Business Media

springeronline.com

Printed in The Netherlands

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

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, *Wolfgang Herbordt*
June 2004

Contents