

DIGITAL SIGNAL PROCESSING

M. Brandstein · D. Ward (Eds.)

Microphone Arrays

Signal Processing Techniques
and Applications



Springer

RICHARD STERN

Springer

Berlin

Heidelberg

New York

Barcelona

Hongkong

London

Milan

Paris

Singapore

Tokyo

Engineering  **ONLINE LIBRARY**

<http://www.springer.de/engine/>

Series Editors

Prof. Dr.-Ing. ARILD LACROIX
Johann-Wolfgang-Goethe-Universität
Institut für angewandte Physik
Robert-Mayer-Str. 2-4
D-60325 Frankfurt

Prof. Dr.-Ing.
ANASTASIOS VENETSANOPOULOS
University of Toronto
Dept. of Electrical and Computer Engineering
10 King's College Road
M5S 3G4 Toronto, Ontario
Canada

Editors

Prof. MICHAEL BRANDSTEIN
Harvard University,
Div. of Eng. and Applied Sciences
33 Oxford Street
MA 02138 Cambridge
USA

e-mail: msb@hrl.harvard.edu

Dr. DARREN WARD
Imperial College, Dept. of Electrical Engineering
Exhibition Road
SW7 2AZ London
GB
e-mail: d.ward@ic.ac.uk

ISBN 3-540-41953-5 Springer-Verlag Berlin Heidelberg New York

Cip data applied for

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 for prosecution act under German Copyright Law.

Springer-Verlag is a part of Springer Science+Business Media
<http://www.springer.de>

● Springer-Verlag Berlin Heidelberg New York 2001
Printed in Germany

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: Camera-ready copy by authors

Cover-Design: de'blik, Berlin

SPIN: 11559276 62/3111 5 4 3 2 1 Printed on acid-free paper

Preface

The study and implementation of microphone arrays originated over 20 years ago. Thanks to the research and experimental developments pursued to the present day, the field has matured to the point that array-based technology now has immediate applicability to a number of current systems and a vast potential for the improvement of existing products and the creation of future devices.

In putting this book together, our goal was to provide, for the first time, a single complete reference on microphone arrays. We invited the top researchers in the field to contribute articles addressing their specific topic(s) of study. The reception we received from our colleagues was quite enthusiastic and very encouraging. There was the general consensus that a work of this kind was well overdue. The results provided in this collection cover the current state of the art in microphone array research, development, and technological application.

This text is organized into four sections which roughly follow the major areas of microphone array research today. Parts I and II are primarily theoretical in nature and emphasize the use of microphone arrays for speech enhancement and source localization, respectively. Part III presents a number of specific applications of array-based technology. Part IV addresses some open questions and explores the future of the field.

Part I concerns the problem of enhancing the speech signal acquired by an array of microphones. For a variety of applications, including human-computer interaction and hands-free telephony, the goal is to allow users to roam unfettered in diverse environments while still providing a high quality speech signal and robustness against background noise, interfering sources, and reverberation effects. The use of microphone arrays gives one the opportunity to exploit the fact that the source of the desired speech signal and the noise sources are physically separated in space. Conventional array processing techniques, typically developed for applications such as radar and sonar, were initially applied to the hands-free speech acquisition problem. However, the environment in which microphone arrays is used is significantly different from that of conventional array applications. Firstly, the desired speech signal has an extremely wide bandwidth relative to its center frequency, meaning that conventional narrowband techniques are not suitable. Secondly, there

is significant multipath interference caused by room reverberation. Finally, the speech source and noise signals may be located close to the array, meaning that the conventional far-field assumption is typically not valid. These differences (amongst others) have meant that new array techniques have had to be formulated for microphone array applications. Chapter 1 describes the design of an array whose spatial response does not change appreciably over a wide bandwidth. Such a design ensures that the spatial filtering performed by the array is uniform across the entire bandwidth of the speech signal. The main problem with many array designs is that a very large physical array is required to obtain reasonable spatial resolution, especially at low frequencies. This problem is addressed in Chapter 2, which reviews so-called superdirective arrays. These arrays are designed to achieve spatial directivity that is significantly higher than a standard delay-and-sum beamformer. Chapter 3 describes the use of a single-channel noise suppression filter on the output of a microphone array. The design of such a post-filter typically requires information about the correlation of the noise between different microphones. The spatial correlation functions for various directional microphones are investigated in Chapter 4, which also describes the use of these functions in adaptive noise cancellation applications. Chapter 5 reviews adaptive techniques for microphone arrays, focusing on algorithms that are robust and perform well in real environments. Chapter 6 presents optimal spatial filtering algorithms based on the generalized singular-value decomposition. These techniques require a large number of computations, so the chapter presents techniques to reduce the computational complexity and thereby permit real-time implementation. Chapter 7 advocates a new approach that combines explicit modeling of the speech signal (a technique which is well-known in single-channel speech enhancement applications) with the spatial filtering afforded by multi-channel array processing.

Part II is devoted to the source localization problem. The ability to locate and track one or more speech sources is an essential requirement of microphone array systems. For speech enhancement applications, an accurate fix on the primary talker, as well as knowledge of any interfering talkers or coherent noise sources, is necessary to effectively steer the array, enhancing a given source while simultaneously attenuating those deemed undesirable. Location data may be used as a guide for discriminating individual speakers in a multi-source scenario. With this information available, it would then be possible to automatically focus upon and follow a given source on an extended basis. Of particular interest lately, is the application of the speaker location estimates for aiming a camera or series of cameras in a video-conferencing system. In this regard, the automated localization information eliminates the need for a human or number of human camera operators. Several existing commercial products apply microphone-array technology in small-room environments to steer a robotic camera and frame active talkers. Chapter 8 summarizes the various approaches which have been explored to accurately locate an individ-

ual in a practical acoustic environment. The emphasis is on precision in the face of adverse conditions, with an appropriate method presented in detail. Chapter 9 extends the problem to the case of multiple active sources. While again considering realistic environments, the issue is complicated by the presence of several talkers. Chapter 10 further generalizes the source localization scenario to include knowledge derived from non-acoustic sensor modalities. In this case both audio and video signals are effectively combined to track the motion of a talker.

Part III of this text details some specific applications of microphone array technology available today. Microphone arrays have been deployed for a variety of practical applications thus far and their utility and presence in our daily lives is increasing rapidly. At one extreme are large aperture arrays with tens to hundreds of elements designed for large rooms, distant talkers, and adverse acoustic conditions. Examples include the two-dimensional, harmonic array installed in the main auditorium of Bell Laboratories, Murray Hill and the 512-element Huge Microphone Array (HMA) developed at Brown University. While these systems provide tremendous functionality in the environments for which they are intended, small arrays consisting of just a handful (usually 2 to 8) of microphones and encompassing only a few centimeters of space have become far more common and affordable. These systems are intended for sound capture in close-talking, low to moderate noise conditions (such as an individual dictating at a workstation or using a hands-free telephone in an automobile) and have exhibited a degree of effectiveness, especially when compared to their single microphone counterparts. The technology has developed to the point that microphone arrays are now available in off-the-shelf consumer electronic devices available for under \$150. Because of their growing popularity and feasibility we have chosen to focus primarily on the issues associated with small-aperture devices. Chapter 11 addresses the incorporation of multiple microphones into hearing aid devices. The ability of beamforming methods to reduce background noise and interference has been shown to dramatically improve the speech understanding of the hearing impaired and to increase their overall satisfaction with the device. Chapter 12 focuses on the case of a simple two-element array combined with postfiltering to achieve noise and echo reduction. The performance of this configuration is analyzed under realistic acoustic conditions and its utility is demonstrated for desktop conferencing and intercom applications. Chapter 13 is concerned with the problem of acoustic feedback inherent in full-duplex communications involving loudspeakers and microphones. Existing single-channel echo cancellation methods are integrated within a beamforming context to achieve enhanced echo suppression. These results are applied to single- and multi-channel conferencing scenarios. Chapter 14 explores the use of microphone arrays for sound capture in automobiles. The issues of noise, interference, and echo cancellation specifically within the car environment are addressed and a particularly effective approach is detailed. Chapter 15 discusses the applica-

VIII Preface

tion of microphone arrays to improve the performance of speech recognition systems in adverse conditions. Strategies for effectively coupling the acoustic signal enhancements afforded through beamforming with existing speech recognition techniques are presented. A specific adaptation of a recognizer to function with an array is presented. Finally, Chapter 16 presents an overview of the problem of separating blind mixtures of acoustic signals recorded at a microphone array. This represents a very new application for microphone arrays, and is a technique that is fundamentally different to the spatial filtering approaches detailed in earlier chapters.

In the final section of the book, Part IV presents expert summaries of current open problems in the field, as well as personal views of what the future of microphone array processing might hold. These summaries, presented in Chapters 17 and 18, describe both academically-oriented research problems, as well as industry-focused areas where microphone array research may be headed.

The individual chapters that we selected for the book were designed to be tutorial in nature with a specific emphasis on recent important results. We hope the result is a text that will be of utility to a large audience, from the student or practicing engineer just approaching the field to the advanced researcher with multi-channel signal processing experience.

Cambridge MA, USA
London, UK
January 2001

Michael Brandstein
Darren Ward

Contents

Part I. Speech Enhancement

1 Constant Directivity Beamforming

<i>Darren B. Ward, Rodney A. Kennedy, Robert C. Williamson</i>	3
1.1 Introduction	3
1.2 Problem Formulation	6
1.3 Theoretical Solution	7
1.3.1 Continuous sensor	7
1.3.2 Beam-shaping function	8
1.4 Practical Implementation	9
1.4.1 Dimension-reducing parameterization	9
1.4.2 Reference beam-shaping filter	11
1.4.3 Sensor placement	12
1.4.4 Summary of implementation	12
1.5 Examples	13
1.6 Conclusions	16
References	16

2 Superdirective Microphone Arrays

<i>Joerg Bitzer, K. Uwe Simmer</i>	19
2.1 Introduction	19
2.2 Evaluation of Beamformers	20
2.2.1 Array-Gain	21
2.2.2 Beampattern	22
2.2.3 Directivity	23
2.2.4 Front-to-Back Ratio	24
2.2.5 White Noise Gain	24
2.3 Design of Superdirective Beamformers	24
2.3.1 Delay-and-Sum Beamformer	26
2.3.2 Design for spherical isotropic noise	26
2.3.3 Design for Cylindrical Isotropic Noise	30
2.3.4 Design for an Optimal Front-to-Back Ratio	30
2.3.5 Design for Measured Noise Fields	32
2.4 Extensions and Details	33
2.4.1 Alternative Form	33

X Contents

2.4.2 Comparison with Gradient Microphones	35
2.5 Conclusion	36
References	37
3 Post-Filtering Techniques	
<i>K. Uwe Simmer, Joerg Bitzer, Claude Marro</i>	39
3.1 Introduction	39
3.2 Multi-channel Wiener Filtering in Subbands	41
3.2.1 Derivation of the Optimum Solution	41
3.2.2 Factorization of the Wiener Solution	42
3.2.3 Interpretation	45
3.3 Algorithms for Post-Filter Estimation	46
3.3.1 Analysis of Post-Filter Algorithms	47
3.3.2 Properties of Post-Filter Algorithms	49
3.3.3 A New Post-Filter Algorithm	50
3.4 Performance Evaluation	51
3.4.1 Simulation System	52
3.4.2 Objective Measures	52
3.4.3 Simulation Results	54
3.5 Conclusion	57
4 Spatial Coherence Functions for Differential Microphones in Isotropic Noise Fields	
<i>Gary W. Elko</i>	61
4.1 Introduction	61
4.2 Adaptive Noise Cancellation	61
4.3 Spherically Isotropic Coherence	65
4.4 Cylindrically Isotropic Fields	73
4.5 Conclusions	77
References	84
5 Robust Adaptive Beamforming	
<i>Osamu Hoshuyama, Akihiko Sugiyama</i>	87
5.1 Introduction	87
5.2 Adaptive Beamformers	88
5.3 Robustness Problem in the GJBF	90
5.4 Robust Adaptive Microphone Arrays — Solutions to Steering-Vector Errors	92
5.4.1 LAF-LAF Structure	92
5.4.2 CCAF-LAF Structure	94
5.4.3 CCAF-NCAF Structure	95
5.4.4 CCAF-NCAF Structure with an AMC	97
5.5 Software Evaluation of a Robust Adaptive Microphone Array	99
5.5.1 Simulated Anechoic Environment	99
5.5.2 Reverberant Environment	101

5.6 Hardware Evaluation of a Robust Adaptive Microphone Array	104
5.6.1 Implementation	104
5.6.2 Evaluation in a Real Environment	104
5.7 Conclusion	106
References	106

6 GSVD-Based Optimal Filtering for Multi-Microphone Speech Enhancement

<i>Simon Doclo, Marc Moonen</i>	111
6.1 Introduction	111
6.2 GSVD-Based Optimal Filtering Technique	113
6.2.1 Optimal Filter Theory	114
6.2.2 General Class of Estimators	116
6.2.3 Symmetry Properties for Time-Series Filtering	117
6.3 Performance of GSVD-Based Optimal Filtering	118
6.3.1 Simulation Environment	118
6.3.2 Spatial Directivity Pattern	119
6.3.3 Noise Reduction Performance	121
6.3.4 Robustness Issues	121
6.4 Complexity Reduction	122
6.4.1 Linear Algebra Techniques for Computing GSVD	122
6.4.2 Recursive and Approximate GSVD-Updating Algorithms	123
6.4.3 Downsampling Techniques	125
6.4.4 Simulations	125
6.4.5 Computational Complexity	126
6.5 Combination with ANC Postprocessing Stage	127
6.5.1 Creation of Speech and Noise References	127
6.5.2 Noise Reduction Performance of ANC Postprocessing Stage	128
6.5.3 Comparison with Standard Beamforming Techniques	129
6.6 Conclusion	129
References	130

7 Explicit Speech Modeling for Microphone Array Speech Acquisition

<i>Michael Brandstein, Scott Griebel</i>	133
7.1 Introduction	133
7.2 Model-Based Strategies	136
7.2.1 Example 1: A Frequency-Domain Model-Based Algorithm	137
7.2.2 Example 2: A Time-Domain Model-Based Algorithm	140
7.3 Conclusion	148
References	151

Part II. Source Localization

8 Robust Localization in Reverberant Rooms

Joseph H. DiBiase, Harvey F. Silverman, Michael S. Brandstein 157

8.1 Introduction 157

8.2 Source Localization Strategies 158

8.2.1 Steered-Beamformer-Based Locators 159

8.2.2 High-Resolution Spectral-Estimation-Based Locators 160

8.2.3 TDOA-Based Locators 161

8.3 A Robust Localization Algorithm 164

8.3.1 The Impulse Response Model 164

8.3.2 The GCC and PHAT Weighting Function 166

8.3.3 ML TDOA-Based Source Localization 167

8.3.4 SRP-Based Source Localization 169

8.3.5 The SRP-PHAT Algorithm 170

8.4 Experimental Comparison 172

References 178

9 Multi-Source Localization Strategies

Elio D. Di Claudio, Raffaele Parisi 181

9.1 Introduction 181

9.2 Background 184

9.2.1 Array Signal Model 184

9.2.2 Incoherent Approach 185

9.2.3 Coherent Signal Subspace Method (CSSM) 185

9.2.4 Wideband Weighted Subspace Fitting (WB-WSF) 186

9.3 The Issue of Coherent Multipath in Array Processing 187

9.4 Implementation Issues 188

9.5 Linear Prediction-ROOT-MUSIC TDOA Estimation 189

9.5.1 Signal Pre-Whitening 189

9.5.2 An Approximate Model for Multiple Sources in Reverberant
Environments 191

9.5.3 Robust TDOA Estimation via ROOT-MUSIC 192

9.5.4 Estimation of the Number of Relevant Reflections 194

9.5.5 Source Clustering 195

9.5.6 Experimental Results 196

References 198

10 Joint Audio-Video Signal Processing for Object Localization and Tracking

Norbert Strobel, Sascha Spors, Rudolf Rabenstein 203

10.1 Introduction 203

10.2 Recursive State Estimation 205

10.2.1 Linear Kalman Filter 206

10.2.2 Extended Kalman Filter due to a Measurement Nonlinearity 210

10.2.3 Decentralized Kalman Filter 212

10.3 Implementation 218

10.3.1 System description	218
10.3.2 Results	219
10.4 Discussion and Conclusions	221
References	222

Part III. Applications

11 Microphone-Array Hearing Aids	
<i>Julie E. Greenberg, Patrick M. Zurek</i>	229
11.1 Introduction	229
11.2 Implications for Design and Evaluation	230
11.2.1 Assumptions Regarding Sound Sources	230
11.2.2 Implementation Issues	231
11.2.3 Assessing Performance	232
11.3 Hearing Aids with Directional Microphones	233
11.4 Fixed-Beamforming Hearing Aids	234
11.5 Adaptive-Beamforming Hearing Aids	235
11.5.1 Generalized Sidelobe Canceler with Modifications	236
11.5.2 Scaled Projection Algorithm	242
11.5.3 Direction of Arrival Estimation	243
11.5.4 Other Adaptive Approaches and Devices	243
11.6 Physiologically-Motivated Algorithms	244
11.7 Beamformers with Binaural Outputs	245
11.8 Discussion	246
References	249
12 Small Microphone Arrays with Postfilters for Noise and Acoustic Echo Reduction	
<i>Rainer Martin</i>	255
12.1 Introduction	255
12.2 Coherence of Speech and Noise	257
12.2.1 The Magnitude Squared Coherence	257
12.2.2 The Reverberation Distance	258
12.2.3 Coherence of Noise and Speech in Reverberant Enclosures ..	259
12.3 Analysis of the Wiener Filter with Symmetric Input Signals	263
12.3.1 No Near End Speech	265
12.3.2 High Signal to Noise Ratio	265
12.4 A Noise Reduction Application	266
12.4.1 An Implementation Based on the NLMS Algorithm	266
12.4.2 Processing in the 800 – 3600 Hz Band	268
12.4.3 Processing in the 240 – 800 Hz Band	269
12.4.4 Evaluation	269
12.4.5 Alternative Implementations of the Coherence Based Postfilter	271
12.5 Combined Noise and Acoustic Echo Reduction	271

XIV Contents

12.5.1 Experimental Results	274
12.6 Conclusions	275
References	276

**13 Acoustic Echo Cancellation for Beamforming
Microphone Arrays**

<i>Walter L. Kellermann</i>	281
13.1 Introduction	281
13.2 Acoustic Echo Cancellation	282
13.2.1 Adaptation algorithms	284
13.2.2 AEC for multi-channel sound reproduction	287
13.2.3 AEC for multi-channel acquisition	287
13.3 Beamforming	288
13.3.1 General structure	288
13.3.2 Time-invariant beamforming	290
13.3.3 Time-varying beamforming	291
13.3.4 Computational complexity	292
13.4 Generic structures for combining AEC with beamforming	292
13.4.1 Motivation	292
13.4.2 Basic options	293
13.4.3 'AEC first'	293
13.4.4 'Beamforming first'	296
13.5 Integration of AEC into time-varying beamforming	297
13.5.1 Cascading time-invariant and time-varying beamforming	297
13.5.2 AEC with GSC-type beamforming structures	301
13.6 Combined AEC and beamforming for multi-channel recording and multi-channel reproduction	302
13.7 Conclusions	303
References	303

**14 Optimal and Adaptive Microphone Arrays for Speech In-
put in Automobiles**

<i>Sven Nordholm, Ingvar Claesson, Nedelko Grbić</i>	307
14.1 Introduction: Hands-Free Telephony in Cars	307
14.2 Optimum and Adaptive Beamforming	309
14.2.1 Common Signal Modeling	309
14.2.2 Constrained Minimum Variance Beamforming and the Gen- eralized Sidelobe Canceler	310
14.2.3 <i>In Situ</i> Calibrated Microphone Array (ICMA)	312
14.2.4 Time-Domain Minimum-Mean-Square-Error Solution	313
14.2.5 Frequency-Domain Minimum-Mean-Square-Error Solution ..	314
14.2.6 Optimal Near-Field Signal-to-Noise plus Interference Beam- former	316
14.3 Subband Implementation of the Microphone Array	317
14.3.1 Description of LS-Subband Beamforming	318

14.4 Multi-Resolution Time-Frequency Adaptive Beamforming	319
14.4.1 Memory Saving and Improvements	319
14.5 Evaluation and Examples	320
14.5.1 Car Environment	320
14.5.2 Microphone Configurations	321
14.5.3 Performance Measures	321
14.5.4 Spectral Performance Measures	322
14.5.5 Evaluation on car data	323
14.5.6 Evaluation Results	323
14.6 Summary and Conclusions	324
References	326
 15 Speech Recognition with Microphone Arrays	
<i>Maurizio Omologo, Marco Matassoni, Piergiorgio Svaizer</i>	331
15.1 Introduction	331
15.2 State of the Art	332
15.2.1 Automatic Speech Recognition	332
15.2.2 Robustness in ASR	336
15.2.3 Microphone Arrays and Related Processing for ASR	337
15.2.4 Distant-Talker Speech Recognition	339
15.3 A Microphone Array-Based ASR System	342
15.3.1 System Description	342
15.3.2 Speech Corpora and Task	345
15.3.3 Experiments and Results	346
15.4 Discussion and Future Trends	348
References	349
 16 Blind Separation of Acoustic Signals	
<i>Scott C. Douglas</i>	355
16.1 Introduction	355
16.1.1 The Cocktail Party Effect	355
16.1.2 Chapter Overview	356
16.2 Blind Signal Separation of Convolutional Mixtures	357
16.2.1 Problem Structure	357
16.2.2 Goal of Convolutional BSS	359
16.2.3 Relationship to Other Problems	360
16.3 Criteria for Blind Signal Separation	362
16.3.1 Overview of BSS Criteria	362
16.3.2 Density Modeling Criteria	362
16.3.3 Contrast Functions	364
16.3.4 Correlation-Based Criteria	366
16.4 Structures and Algorithms for Blind Signal Separation	367
16.4.1 Filter Structures	367
16.4.2 Density Matching BSS Using Natural Gradient Adaptation	368
16.4.3 Contrast-Based BSS Under Prewhitening Constraints	370

XVI Contents

16.4.4 Temporal Decorrelation BSS for Nonstationary Sources	372
16.5 Numerical Evaluations	373
16.6 Conclusions and Open Issues	375
References	378

Part IV. Open Problems and Future Directions

17 Future Directions for Microphone Arrays	
<i>Gary W. Elko</i>	383
17.1 Introduction	383
17.2 Hands-Free Communication	383
17.3 The "Future" of Microphone Array Processing	385
17.4 Conclusions	387
18 Future Directions in Microphone Array Processing	
<i>Dirk Van Compernelle</i>	389
18.1 Lessons From the Past	389
18.2 A Future Focused on Applications	391
18.2.1 Automotive	391
18.2.2 Desktop	392
18.2.3 Hearing Aids	393
18.2.4 Teleconferencing	393
18.2.5 Very Large Arrays	393
18.2.6 The Signal Subspace Approach - An Alternative to Spatial Filtering ?	393
18.3 Final Remarks	394
Index	395

List of Contributors

Joerg Bitzer
Houpert Digital Audio
Bremen, Germany

Michael S. Brandstein
Harvard University
Cambridge MA, USA

Ingvar Claesson
Blekinge Inst. of Technology
Ronneby, Sweden

Joseph H. DiBiase
Brown University
Providence RI, USA

Elio D. Di Claudio
University of Rome "La Sapienza"
Rome, Italy

Simon Doclo
Katholieke Universiteit Leuven
Leuven, Belgium

Scott C. Douglas
Southern Methodist University
Dallas TX, USA

Gary W. Elko
Agere Systems
Murray Hill NJ, USA

Nedelko Grbić
Blekinge Inst. of Technology
Ronneby, Sweden

Julie E. Greenberg
Massachusetts Inst. of Technology
Cambridge MA, USA

Scott M. Griebel
Harvard University
Cambridge MA, USA

Osamu Hoshuyama
NEC Media Research Labs
Kawasaki, Japan

Walter L. Kellermann
University Erlangen-Nuremberg
Erlangen, Germany

Rodney A. Kennedy
The Australian National University
Canberra, Australia

Claude Marro
France Télécom R&D
Lannion, France

Rainer Martin
Aachen University of Technology
Aachen, Germany

Marco Matassoni
Istituto per la Ricerca Scientifica e
Tecnologica
Povo, Italy

Marc Moonen
Katholieke Universiteit Leuven
Leuven, Belgium

XVIII List of Contributors

Sven Nordholm

Curtin University of Technology
Perth, Australia

Maurizio Omologo

Istituto per la Ricerca Scientifica e
Tecnologica
Povo, Italy

Raffaele Parisi

University of Rome "La Sapienza"
Rome, Italy

Rudolf Rabenstein

University Erlangen-Nuremberg
Erlangen, Germany

Harvey F. Silverman

Brown University
Providence RI, USA

K. Uwe Simmer

Aureca GmbH
Bremen, Germany

Sascha Spors

University Erlangen-Nuremberg
Erlangen, Germany

Norbert Strobel

Siemens Medical Solutions
Erlangen, Germany

Akihiko Sugiyama

NEC Media Research Labs
Kawasaki, Japan

Piergiorgio Svaizer

Istituto per la Ricerca Scientifica e
Tecnologica
Povo, Italy

Dirk Van Compernelle

Katholieke Universiteit Leuven
Leuven, Belgium

Darren B. Ward

Imperial College of Science,
Technology and Medicine
London, UK

Robert C. Williamson

The Australian National University
Canberra, Australia

Patrick M. Zurek

Sensimetrics Corporation
Somerville MA, USA

Part I

Speech Enhancement

1 Constant Directivity Beamforming

Darren B. Ward¹, Rodney A. Kennedy², and Robert C. Williamson²

¹ Imperial College of Science, Technology and Medicine, London, UK

² The Australian National University, Canberra, Australia

Abstract. Beamforming, or spatial filtering, is one of the simplest methods for discriminating between different signals based on the physical location of the sources. Because speech is a very wideband signal, covering some four octaves, traditional narrowband beamforming techniques are inappropriate for hands-free speech acquisition. One class of broadband beamformers, called constant directivity beamformers, aim to produce a constant spatial response over a broad frequency range. In this chapter we review such beamformers, and discuss implementation issues related to their use in microphone arrays.

1.1 Introduction

Beamforming is one of the simplest and most robust means of *spatial filtering*, i.e., discriminating between signals based on the physical locations of the signal sources [1]. In a typical microphone array environment, the desired speech signal originates from a talker's mouth, and is corrupted by interfering signals such as other talkers and room reverberation. Spatial filtering can be useful in such an environment, since the interfering sources generally originate from points in space separate from the desired talker's mouth. By exploiting the spatial dimension of the problem, microphone arrays attempt to obtain a high-quality speech signal without requiring the talker to speak directly into a close-talking microphone.

In most beamforming applications two assumptions simplify the analysis: (i) the signals incident on the array are narrowband (the *narrowband assumption*); and (ii) the signal sources are located far enough away from the array that the wavefronts impinging on the array can be modeled as plane waves (the *farfield assumption*). For many microphone array applications, the farfield assumption is valid. However, the narrowband assumption is never valid, and it is this aspect of the beamforming problem that we focus on in this chapter (see [2] for techniques that also lift the nearfield assumption).

To understand the inherent problem in using a narrowband array for broadband signals, consider a linear array with a fixed number of elements separated by a fixed inter-element distance. The important dimension in measuring array performance is its size in terms of operating wavelength. Thus for high frequency signals (having a small wavelength) a fixed array will appear large and the main beam will be narrow. However, for low frequencies

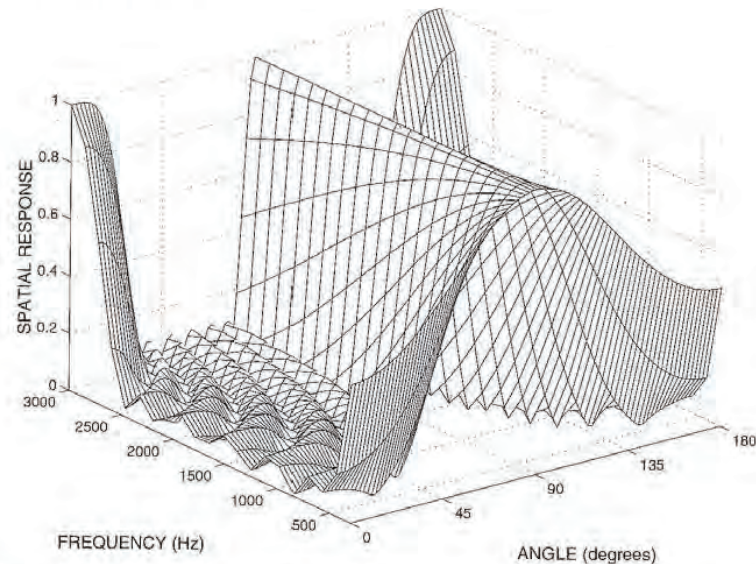


Fig. 1.1. Response of a narrowband array operated over a wide bandwidth.

(large wavelength) the same physical array appears small and the main beam will widen.

This is illustrated in Fig. 1.1 which shows the beampattern of an array designed for 1.5 kHz, but operated over a frequency range of 300 Hz to 3 kHz. If an interfering signal is present at, say, 60°, then ideally it should be attenuated completely by the array. However, because the beam is wider at low frequencies than at high frequencies, the interfering signal will be low-pass filtered rather than uniformly attenuated over its entire band. This "spectral tilt" results in a disturbing speech output if used for speech acquisition, and thus, such a narrowband array is unacceptable for speech applications. Another drawback of this narrowband design is that spatial aliasing is evident at high frequencies.¹

To overcome this problem, one must use a beamformer that is designed specifically for broadband applications. In this chapter we focus on a specific class of broadband beamformers, called *constant directivity beamformers* (CDB), designed such that the spatial response is the same over a wide frequency band. The response of a typical CDB is shown in Fig. 1.6 on page 15.

There have been several techniques proposed to design a CDB. Most techniques are based on the idea that at different frequencies, a different array should be used that has total size and inter-sensor spacing appropriate for that particular frequency. An example of this idea is the use of harmonically-

¹ Spatial aliasing comes about if a sensor spacing wider than half a wavelength is used. It is analogous to temporal aliasing in discrete-time signal processing.

nested subarrays, e.g., [3-5]. In this case, the array is composed of a set of nested equally-spaced arrays, with each subarray being designed as a narrowband array. The outputs of the various subarrays are then combined by appropriate bandpass filtering. The idea of harmonic nesting is to reduce the beampattern variation to that which occurs within a single octave. This approach can be improved by using a set of subarray filters to interpolate to frequencies between the subarray design frequencies [6].

A novel approach to CDB design was proposed by Smith in [7]. Noting that, for a given array, the beamwidth narrows at high frequencies, Smith's idea was to form several beams and to steer each individual beam in such a way that the width of the overall multi-beam was kept constant. Thus, as the individual beams narrow at higher frequencies, they are progressively "fanned" outwards in an attempt to keep the overall beamwidth constant. Unless a very large number of beams are formed, at high frequencies this fanning will result in notches in the main beam where the progressively narrower beams no longer overlap. This approach was applied to the design of microphone arrays in [8].

The first approach to CDB design that attempted to keep a constant beampattern over the entire spatial region (not just for the main beam) was presented by Doles and Benedict [9]. Using the asymptotic theory of unequally-spaced arrays [10,11], they derived relationships between beampattern characteristics and functional requirements on sensor spacings and weightings. This results in a filter-and-sum array, with the sensor filters creating a space-tapered array: at each frequency the non-zero filter responses identify a subarray having total length and spacing appropriate for that frequency. Although this design technique results in a beampattern that is frequency-invariant over a specified frequency band, it is not a general design technique, since it is based on a specific array geometry and beampattern shape. Other recent techniques for CDB design include [12] (based on a two-dimensional Fourier transform property [13] which exists for equally-spaced arrays) and [14] (based on a beam space implementation).

Prompted by the work of Doles and Benedict, we derived in [15] a very general design method for CDB's, suitable for three-dimensional array geometries. In this chapter we outline this technique, and discuss implementation issues specific to microphone array applications.

Time-domain versus frequency-domain beamforming

There are two general methods of beamforming for broadband signals: time-domain beamforming and frequency-domain beamforming. In time-domain beamforming an FIR filter is used on each sensor, and the filter outputs summed to form the beamformer output. For an array with M sensors, each feeding a L tap filter, there are ML free parameters. In frequency-domain beamforming the signal received by each sensor is separated into narrowband frequency bins (either through bandpass filtering or data segmentation

and discrete Fourier transform), and the data in each frequency bin is processed separately using narrowband techniques. For an array with M sensors, with L frequency bins within the band of interest, there are again ML free parameters. As with most beamformers, the method that we describe in this chapter can be formulated in either domain. A time-domain formulation has previously been given in [16], and hence, we restrict our attention to frequency-domain processing here.

1.2 Problem Formulation

Consider a linear array of $M = 2N + 1$ sensors located at $p_n, n = -N, \dots, N$. Assume that the data received at the n th sensor is separated into narrowband frequency bins, each of width Δf . Let the center frequency of the i th bin be f_i , and denote the frequencies within the bin as

$$F_i = [f_i - \Delta f/2, f_i + \Delta f/2).$$

The array data received in the i th bin at time k , is given by the M -vector:

$$\mathbf{x}_i(k) = \mathbf{a}(\theta, f_i) s_i(k) + \mathbf{v}_i(k).$$

The desired source signal is represented by $s_i(k)$, and the M -vector $\mathbf{v}_i(k)$ represents the interfering noise (consisting of reverberation and other unwanted noise sources). The array vector $\mathbf{a}(\theta, f)$ represents the propagation of the signal source to the array, and its n th element is given by

$$a_n(\theta, f) = e^{-j2\pi f c^{-1} p_n \cos \theta},$$

where c is the speed of wave propagation, and θ is the direction to the desired source (measured relative to the array axis). To simplify notation we will drop the explicit dependence on k in the sequel.

The beamformer output is formed by applying a weight vector to the received array data, giving

$$y_i = \mathbf{w}_i^H \mathbf{x}_i, \quad (1.1)$$

where ^{H} denotes Hermitian transpose, and $\mathbf{w}_i$ is the M -vector of array weights to apply to the i th frequency bin.²

The spatial response of the beamformer is given by

$$b(\theta, f) = \mathbf{w}_i^H \mathbf{a}(\theta, f), \quad f \in F_i, \quad (1.2)$$

which defines the transfer function between a source at location $\theta \in [-\pi, \pi)$ and the beamformer output. Also of interest is the *beam pattern*, defined as the squared magnitude of the spatial response.

² Note that it is a notational convention to use $\mathbf{w}^H$ rather than $\mathbf{w}^T$ [1].

The problem of designing a CDB can now be formulated as finding the array weights in each frequency bin such that the resulting spatial response remains constant over all frequency bins of interest.

One simple (but not very illuminating) approach to solving this problem is to perform a least-squares optimization in each frequency bin, i.e.,

$$\min_{w_i} \int_{2\pi} |b_{FI}(\theta) - w_i^H a(\theta, f_i)|^2 d\theta, \quad (1.3)$$

where $b_{FI}(\theta)$ is the desired frequency-invariant response. Thus, in each frequency bin there are M free parameters to optimize. Although this is a standard least-squares optimization problem and the required array weights are easily found, the solution provides very little insight into the problem. Specifically, there is no suggestion of any inherent structure in the CDB, and many important questions are left unanswered, such as how many sensors are required, and what range of frequencies can be used.

In an attempt to provide some insight into the problem of designing a CDB, we take an alternative theoretical approach in the following section, and then relate these theoretical results back to the problem of finding the required filter coefficients. As we will see, there is in fact a very strong implicit structure in the CDB, and exploiting this structure enables us to reduce the number of design parameters and find efficient implementations.

1.3 Theoretical Solution

It is well known that the important dimension in determining the array response is the physical array size, measured in wavelengths. Thus, to obtain the same beampattern at different frequencies requires that the array size remains constant in terms of wavelength. Specifically, consider a linear array with N elements located at $p_n, n = 1, \dots, N$, and assume the array weights are chosen to produce a desired beampattern $b(\theta)$ at a frequency f_1 . Then, at a frequency f_2 , the same beampattern $b(\theta)$ will be produced if the same array weights are used in an array with elements located at $p_n(f_1/f_2), n = 1, \dots, N$. In other words, the size of the array must scale directly with frequency to obtain the same beampattern.³ To obtain the same beampattern over a continuous range of frequencies would theoretically require a continuum of sensors.

1.3.1 Continuous sensor

Motivated by this interpretation, we consider the response of a theoretical continuous sensor. Assume that a signal $x(p, f)$ is received at a point p on

³ This is precisely the idea used in the harmonically-nested subarray technique.

the sensor at frequency f , and a weight $w(p, f)$ is applied to the sensor at this point and frequency. The output of the sensor is

$$y(f) = \int w(p, f) x(p, f) dp,$$

and the spatial response for a source at angle θ is

$$b(\theta, f) = \int w(p, f) e^{-j2\pi f c^{-1} p \cos \theta} dp. \quad (1.4)$$

We assume that the aperture has finite support in p , and thus, the integration has infinite limits.

Let $u = c^{-1} \cos \theta$. The response of the continuous sensor can now be written

$$b_u(u, f) = \int w(p, f) e^{-j2\pi f p u} dp.$$

Let the sensor weighting function be given by

$$w(p, f) = f B(pf), \quad (1.5)$$

where $B(\cdot)$ is an arbitrary, absolutely-integrable, finite-support function. Substitution gives

$$b_u(u, f) = \int f B(pf) e^{-j2\pi f p u} dp. \quad (1.6)$$

With the change of variable $\zeta = pf$, and noting that $d\zeta = f dp$, it is easily seen that the resulting spatial response is now independent of frequency, i.e.,

$$b_u(u, f) = \int B(\zeta) e^{-j2\pi \zeta u} d\zeta = b_{\text{FI}}(u). \quad (1.7)$$

This is an important result, since it states that if the weighting function is given by (1.5), then the resulting spatial response will be independent of frequency. In other words, (1.5) defines the weighting function for a CDB. It was shown in [15], that not only does (1.5) provide a sufficient condition, but it is in fact the necessary condition for a frequency-invariant spatial response.

1.3.2 Beam-shaping function

Equation (1.7) defines a Fourier transform relationship between $B(\cdot)$ and $b_{\text{FI}}(\cdot)$. To achieve some desired spatial response, the required function $B(\zeta)$ is thus easily found by taking the inverse Fourier transform of $b(u)$. We will refer to $B(\cdot)$ as the *beam-shaping* (BS) function, since it has a fundamental role in determining the spatial response.

Because of its symmetry with respect to space and frequency, the BS function can be interpreted as either a filter response at a certain point, i.e., $H_p(f) = B(pf)$, or equivalently, as an aperture weighting function at a certain frequency, i.e., $A_f(p) = B(pf)$.

We will assume that the BS function is Hermitian symmetric, i.e., $B(-\zeta) = B^*(\zeta)$. This implies that the resulting spatial response is real-valued.

1.4 Practical Implementation

Whilst we have shown theoretically that it is possible to produce a beam pattern that is exactly frequency-invariant using a continuous sensor, in practice we must attempt to approximate such a response using a finite array of discrete sensors. The problem of approximating a continuous aperture by a discrete array has been considered in [17]. One simple but effective technique is to approximate the integral in (1.6) using a Riemann sum—this is the approach we take here. In particular, we use trapezoidal integration to approximate the integral (1.6) by a summation of the form:

$$\hat{b}_{\text{FI}}(u) = \sum_{n=-N}^N f B(p_n f) e^{-j2\pi f p_n u} \Delta_n \quad (1.8)$$

where p_n is the location of the n th discrete sensor, and $\hat{b}_{\text{FI}}$ denotes an approximation of b_{FI} . We assume that the array is Hermitian symmetric about the origin, so that $B(-pf) = B(pf)^*$, and $p_{-n} = -p_n$. Although the technique is suitable for an arbitrary array geometry, a symmetric geometry simplifies implementation, and ensures that the position of the array phase center does not vary with frequency. The length of the n th subinterval is

$$\Delta_n = \frac{p_{n+1} - p_{n-1}}{2}, \quad (1.9)$$

which we refer to as the *spatial weighting term*.

Relating (1.8) to the response of a general array (1.2), we find that for a CDB the weight on the n th sensor in the i th frequency bin is

$$w_{i,n} = f_i \Delta_n B(p_n f_i), \quad (1.10)$$

where, recall, p_n is the location of the sensor, and f_i is the center frequency of the bin.

1.4.1 Dimension-reducing parameterization

Define the *reference beam-shaping filter response* as

$$H(f) = B(p_{\text{ref}} f), \quad (1.11)$$

where p_{ref} is some reference location (to be defined later). Also define the *beam-shaping filter response* of the n th sensor as

$$H_n(f) = B(p_n f), \quad n = -N, \dots, N.$$

It immediately follows that the BS filters satisfy the following dilation property:

$$H_n(f) = H(\gamma_n f), \quad (1.12)$$

where

$$\gamma_n = \frac{p_n}{p_{\text{ref}}}$$

is the dilation factor for the n th sensor. This is an extremely important property, since it shows that the filter responses on all sensors can be derived from the single filter response, $H(f)$, and enables the following efficient implementation of the CDB.

Let the reference BS filter response be given by its standard FIR filter representation:

$$H(f) = \sum_l h[l] e^{-j2\pi f / f_s l},$$

where f_s is the sampling frequency, and $h[l]$ is a L -vector of *beam-shaping coefficients*. From (1.12), the n th BS filter response is given by

$$\begin{aligned} H_n(f) &= \sum_l h[l] e^{-j2\pi f / f_s \gamma_n l} \\ &= \mathbf{h}^H \mathbf{d}_n(f), \end{aligned} \quad (1.13)$$

where $\mathbf{d}_n(f)$ is the L -dimensional *dilation vector* for the n th sensor. From (1.10), we see that the weight to use on the n th sensor in the i th bin is

$$w_{i,n} = \mathbf{h}^H \mathbf{t}_{i,n}, \quad (1.14)$$

where

$$\mathbf{t}_{i,n} = f_i \Delta_n \mathbf{d}_n(f_i) \quad (1.15)$$

is a L -dimensional *transformation vector*.

Equation (1.14) demonstrates the efficient parameterization afforded by this particular formulation of the CDB problem. Whereas the naive least-squares approach (1.3) requires an optimization of M parameters w_i in each frequency bin, we find that it is really only necessary to choose L frequency-independent BS parameters $\mathbf{h}$. Changing the beampattern shape only requires modification of these BS coefficients, and the implicit structure imposed by the transformation vectors ensures that the resulting response has constant directivity over the design band.

1.4.2 Reference beam-shaping filter

The underlying principle of the CDB is that the size and shape of the active array aperture should scale directly with frequency. This frequency scaling operation is performed by the BS filters. In deciding the coefficients of the reference BS filter, and the location of the reference point p_{ref} , we must consider this scaling property in more detail.

Let the chosen aperture size be Q wavelengths. Assuming the array is symmetric about the origin, this means that at any wavelength λ , sensors further from the origin than $Q\lambda/2$ should be inactive. In other words, the n th sensor should have a low-pass characteristic with a cutoff frequency of

$$f_n = \frac{Qc}{2|p_n|}. \quad (1.16)$$

From (1.13), note that $\gamma_n > 1$ results in compression in the frequency domain, whereas $\gamma_n < 1$ results in frequency expansion. Since the discrete-time frequency response $H(f)$ is periodic, it follows that frequency compression may cause aliasing; this is extremely undesirable. Aliasing can be avoided in one of two ways. First, choosing $p_{\text{ref}} = \max |p_n|$ ensures that $\gamma_n \leq 1, \forall n$, thus avoiding aliasing altogether—however, this requires additional constraints on the reference BS coefficients to impose the low-pass property (1.16). Alternatively, for sensors having $\gamma_n > 1$, the weights $w_{i,n}$ are set to zero for frequency bins $f_i > f_n$ —the reference BS weights are now potentially unconstrained. Of these two approaches, the second is preferable, since it removes any constraints on the BS coefficients. Moreover, the requirement that the sensor weights within certain bins are always zero does not complicate implementation.

Assume that the frequency response of the reference BS filter is non-zero for all frequencies up to $f_s/2$, the Nyquist frequency; this is the most general case of $H(f)$. From (1.16), it follows that a sensor with non-zero frequency response up to $f_s/2$ would be positioned at $|p_n| = Qc/f_s$. Thus, for the most general case of $H(f)$ the reference location is chosen as

$$p_{\text{ref}} = \frac{Qc}{f_s}. \quad (1.17)$$

The reference BS coefficients can be found by using the Fourier transform relationship defined by (1.7). Specifically, the BS function $B(\zeta)$ is found by taking the Fourier transform of the desired frequency-invariant spatial response $b_{\text{FI}}(u)$. Setting $f = \zeta/p_{\text{ref}}$, $B(\zeta)$ now defines the frequency response of the reference BS filter. The BS coefficient vector $\mathbf{h}$ is found using any standard FIR filter design technique. In practise, low-order implementations of the reference BS filter are generally to be preferred; this point is demonstrated in the following section.

1.4.3 Sensor placement

The most common geometry for array processing applications is typically an equally-spaced array, usually with a spacing of one half-wavelength at the highest frequency of operation. Although such a geometry is valid for a CDB, less sensors are required if a logarithmically spaced array is used. In choosing an appropriate sensor geometry, the most important consideration is to ensure that at any frequency spatial aliasing is avoided.

The idea is to start with an equally-spaced array that is used at the highest frequency, and then progressively add more sensors with wider spacings as frequency decreases (and the wavelength increases). At any frequency f , the total active aperture size should be Qc/f , and the largest spacing within the active array should be $c/(2f)$. These requirements are met (using the least number of sensors) with the following symmetric array geometry:

$$p_n = n \frac{c}{2f_U}, \quad 0 \leq n \leq \frac{Q}{2} \quad (1.18a)$$

$$p_{n+1} = \frac{Q}{Q-1} p_n, \quad n > \frac{Q}{2}, \quad p_n < \frac{(Q-1)c}{2f_L} \quad (1.18b)$$

$$p_{-n} = -p_n. \quad (1.18c)$$

Note that a harmonically-nested subarray geometry is only produced if $Q = 2$.

1.4.4 Summary of implementation

1. Choose a set of L reference BS coefficients, $\mathbf{h}$.
2. Position the sensors according to (1.18a)–(1.18c).
3. In the i th frequency bin, the weight on the n th sensor is

$$w_{i,n} = \mathbf{h}^H \mathbf{t}_{i,n},$$

where

$$\begin{aligned} \mathbf{t}_{i,n} &= \begin{cases} f_i \Delta_n \mathbf{d}_n(f_i), & f_i < f_n \\ 0, & \text{otherwise,} \end{cases} \\ f_n &= \frac{Qc}{2|p_n|} \\ \Delta_n &= \frac{p_{n+1} - p_{n-1}}{2} \\ \mathbf{d}_n(f_i) &= \left[e^{j2\pi f/f_s \gamma_n (L-1)/2}, \dots, e^{-j2\pi f/f_s \gamma_n (L-1)/2} \right] \\ \gamma_n &= \frac{|p_n|}{p_{\text{ref}}} \\ p_{\text{ref}} &= \frac{Qc}{f_s} \end{aligned}$$

1.5 Examples

We now show an example of the CDB design technique. The design was for a bandwidth of 300–3000 Hz (i.e., the same bandwidth as used in Fig. 1.1), with an aperture size of $Q = 4$ wavelengths. Using an FFT size of 128 resulted in 44 bins within the design band, with each bin having a width of 62.5 Hz. The sensors were positioned according to (1.18a)–(1.18c), resulting in the $M = 25$ sensor array geometry shown in Fig. 1.2. For frequencies of 1000 Hz and 2000 Hz, the active sensors are also indicated in this figure.

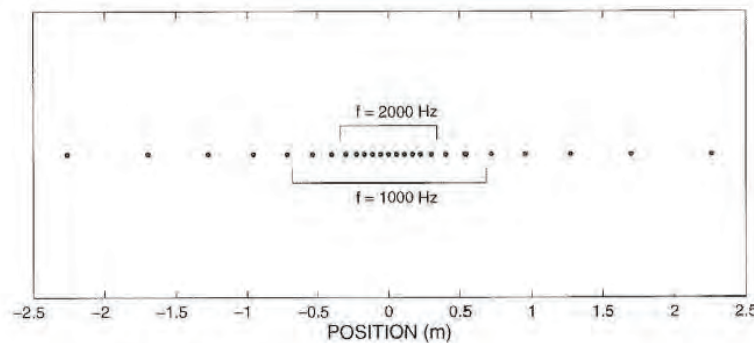


Fig. 1.2. Array geometry used for example CDB.

Assume we wish to design a standard sinc-like response (as produced by a uniformly weighted array). In this case it is known that the aperture function should be uniform. Thus, the BS function $B(\cdot)$ should ideally be a brick-wall low-pass filter. Assume we design the BS vector $\mathbf{h}$ to approximate an ideal low-pass filter using $L = 101$ filter coefficients. This results in the BS frequency responses shown in Fig. 1.3; for each sensor in the array, the weight required at each frequency is plotted. Note that these responses are all dilations of a single response, and that each has a low-pass characteristic.

Using these BS coefficients, the resulting spatial response of the CDB is shown in Fig. 1.4. Although the variation is not as great as for the narrow-band design in Fig. 1.1, the spatial response in Fig. 1.4 is far from frequency invariant. Why is this? The answer lies in the fact that the BS frequency response has a very sharp cutoff. Consider a single sensor. At low frequencies the sensor is always on. As frequency increases, there will come a point where the sensor will suddenly turn off, and at this frequency the aperture abruptly changes size. This abrupt change in the active aperture causes the alp-like appearance of the spatial response in Fig. 1.4.

Now, returning to the problem of designing the BS coefficients for the desired uniform spatial response, assume we design the BS vector $\mathbf{h}$ to ap-

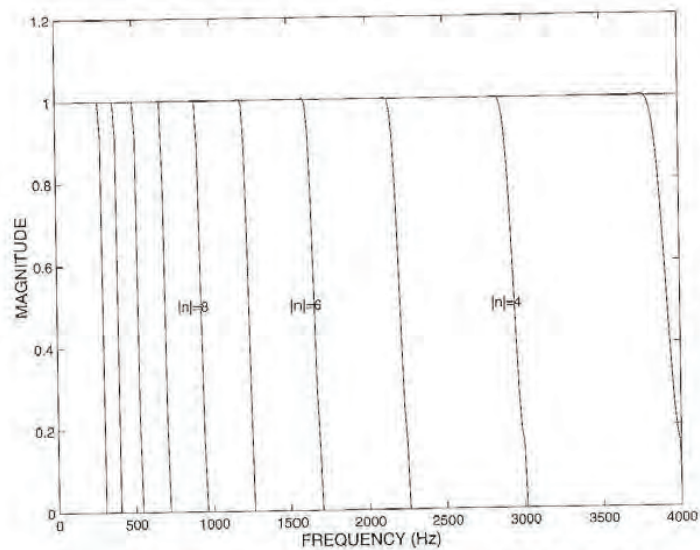


Fig. 1.3. Frequency responses of the weights on each sensor.

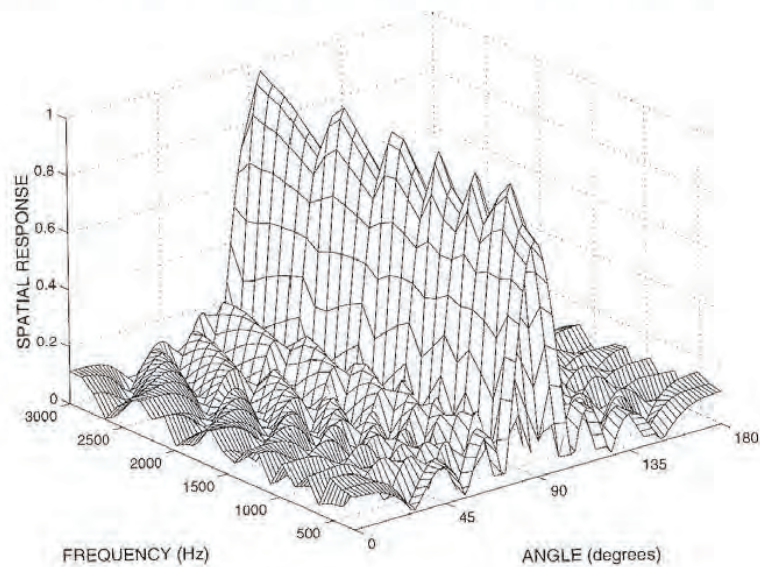


Fig. 1.4. Spatial response of example CDB.

proximate an ideal low-pass filter using only $L = 21$ filter coefficients. This results in the BS frequency responses shown in Fig. 1.5. In comparing this figure with Fig. 1.4, notice that the frequency responses exhibit a more gradual cutoff.

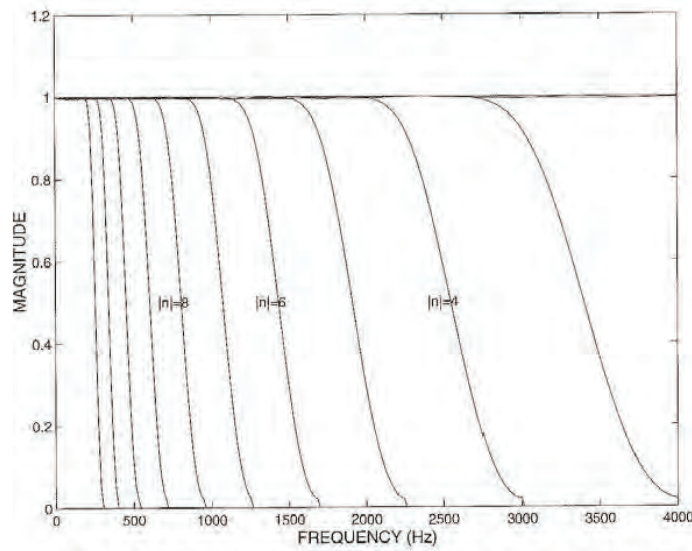


Fig. 1.5. Frequency responses of the weights on each sensor.

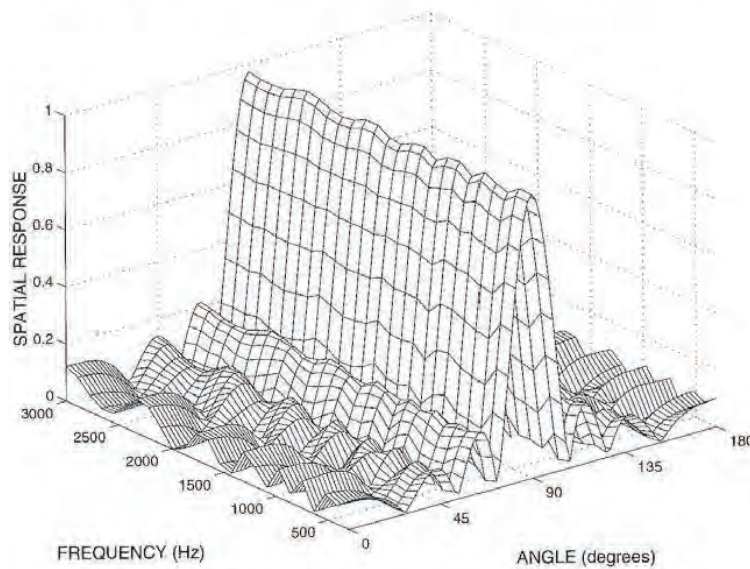


Fig. 1.6. Spatial response of example CDB.

Using these 21 BS coefficients, the resulting spatial response of the CDB is shown in Fig. 1.6. In this case the spatial response shows very little variation with frequency. This demonstrates that one should take careful consideration

of how well the underlying function can be approximated by the discrete array when choosing the required BS function.

1.6 Conclusions

Constant-directivity beamforming is a useful technique for spatial filtering in broadband signal environments in which the desired signal and the interference signals cover approximately the same bandwidth. In this chapter we have developed a technique for designing a CDB, and shown that there is an efficient parameterization and underlying structure exhibited by a CDB. The greatest drawback of a CDB in microphone array applications is that the size of the array is related to the lowest frequency of operation. Thus, producing an array that has a frequency-invariant spatial response down to, say, 300 Hz may require an array that is several meters long. In all but the largest rooms this is impractical. However, a constant spatial response can be readily achieved for mid and high frequencies (above say 1000 Hz) using an array with a total size of less than a meter. For the lower frequencies, other methods (such as the superdirective techniques described in the following chapter) are probably more appropriate.

References

1. B.D. Van Veen and K.M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4-24, Apr. 1988.
2. T.D. Abhayapala, R.A. Kennedy, and R.C. Williamson, "Nearfield broadband array design using a radially invariant modal expansion," *J. Acoust. Soc. Amer.*, vol. 107, no. 1, pp. 392-403, Jan. 2000.
3. J.L. Flanagan, D.A. Berkeley, G.W. Elko, J.E. West, and M.M. Sondhi, "Autodirective microphone systems," *Acustica*, vol. 73, pp. 58-71, 1991.
4. W. Kellermann, "A self-steering digital microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-91)*, 1991, vol. 5, pp. 3581-3584.
5. F. Khalil, J.P. Jullien, and A. Gilloire, "Microphone array for sound pickup in teleconference systems," *J. Audio Eng. Soc.*, vol. 42, no. 9, pp. 691-700, Sept. 1994.
6. J. Lardies, "Acoustic ring array with constant beamwidth over a very wide frequency range," *Acoust. Letters*, vol. 13, no. 5, pp. 77-81, 1989.
7. R. Smith, "Constant beamwidth receiving arrays for broad band sonar systems," *Acustica*, vol. 23, pp. 21-26, 1970.
8. M.M. Goodwin and G.W. Elko, "Constant beamwidth beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-93)*, 1993, vol. 1, pp. 169-172.
9. J.H. Doles III and F.D. Benedict, "Broad-band array design using the asymptotic theory of unequally spaced arrays," *IEEE Trans. Antennas Propagat.*, vol. 36, no. 1, pp. 27-33, Jan. 1988.

10. A. Ishimaru, "Theory of unequally-spaced arrays," *IRE Trans. Antennas Propagat.*, vol. AP-10, pp. 691-702, Nov. 1962.
11. A. Ishimaru and Y.S. Chen, "Thinning and broadbanding antenna arrays by unequal spacings," *IEEE Trans. Antennas Propagat.*, vol. AP-13, pp. 34-42, Jan. 1965.
12. T. Chou, "Frequency-independent beamformer with low response error," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-95)*, Detroit, USA, May 1995, pp. 2995-2998.
13. S. Haykin and J. Kesler, "Relation between the radiation pattern of an array and the two-dimensional discrete Fourier transform," *IEEE Trans. Antennas Propagat.*, vol. AP-23, no. 3, pp. 419-420, May 1975.
14. J.S. Marciano Jr. and T.B. Vu, "Reduced complexity beam space broadband frequency invariant beamforming," *Electronics Letters*, vol. 36, no. 7, pp. 682-683, Mar. 2000.
15. D.B. Ward, R.A. Kennedy, and R.C. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns," *J. Acoust. Soc. Amer.*, vol. 97, no. 2, pp. 1023-1034, Feb. 1995.
16. D.B. Ward, R.A. Kennedy, and R.C. Williamson, "FIR filter design for frequency-invariant beamformers," *IEEE Signal Processing Lett.*, vol. 3, no. 3, pp. 69-71, Mar. 1996.
17. C. Winter, "Using continuous apertures discretely," *IEEE Trans. Antennas Propagat.*, vol. AP-25, pp. 695-700, Sept. 1977.

2 Superdirective Microphone Arrays

Joerg Bitzer¹ and K. Uwe Simmer²

¹ Houpert Digital Audio, Bremen, Germany

² Aureca GmbH, Bremen, Germany

Abstract. This chapter gives an overview of so-called superdirective beamformers, which can be derived by applying the minimum variance distortionless response (MVDR) principle to theoretically well-defined noise fields, as for example the diffuse noise field. We show that all relevant performance measures for beamformer designs are functions of the coherence matrix of the noise field. Additionally, we present unconstrained and constrained MVDR-solutions using modified coherence functions. Solutions for different choices of the optimization criterion are given including a new solution to optimize the front-to-back ratio. Finally, we present a comparison of superdirective beamformers to gradient microphones and an alternative generalized sidelobe canceler (GSC) implementation of the superdirective beamformer.

2.1 Introduction

What is “super” about a superdirective microphone array? Compared to the standard delay-and-sum beamformer a superdirective array achieves a higher directivity. Therefore, “super”-directivity indicates that summing is not the optimal choice for combining sensor signals, if optimal directivity is desired. The term directivity describes the ability of a beamformer to suppress noise coming from all directions without affecting a desired signal from one principal direction.

A short historical overview in [6] shows that superdirectivity (or supergain) in connection with array processing was first mentioned in the first half of the last century. The solutions provided at that time were of academic interest only, since a lot of practical problems occurred which restricted the use of the theoretical work. The main reasons for failure were the self-noise and the gain and phase errors of the microphones. In order to overcome these problems a first constrained solution was published by Gilbert and Morgan in 1955 [15]. Early applications with slight modifications were seismic and sonar techniques [5]. It was not until the 90's that supergain was connected to microphone applications. Research in hearing aids highlighted the advantages of fixed beamformers over adaptive solutions [17]. Modern designs of superdirective beamformers include nearfield assumptions and the possibility to adapt the constraining to the actual problem.

This chapter is organized as follows: Section 2.2 introduces the measures to judge the different designs. In section 2.3 the optimal design will be derived

with respect to the given problems. Further extensions and special details are given in section 2.4. Concluding remarks close this chapter.

2.2 Evaluation of Beamformers

In order to get a better understanding of the features of the different designs of optimal beamformers, we first need to derive the measures to analyze their performance.

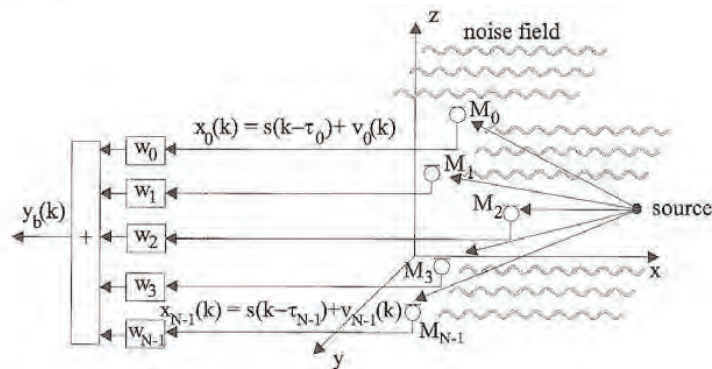


Fig. 2.1. Signal model consisting of noise field and desired source signal

The signal model is shown in Fig. 2.1. We assume that one sample of the discrete input sequence $x(k)$ at each sensor n consists of a delayed and attenuated version of the desired signal $a_n s(k - \tau_n)$ and a noise component $v_n(k)$ with arbitrary spatial statistics.

$$\begin{pmatrix} x_0(k) \\ x_1(k) \\ \vdots \\ x_{N-1}(k) \end{pmatrix} = \begin{pmatrix} a_0 s(k - \tau_0) \\ a_1 s(k - \tau_1) \\ \vdots \\ a_{N-1} s(k - \tau_{N-1}) \end{pmatrix} + \begin{pmatrix} v_0(k) \\ v_1(k) \\ \vdots \\ v_{N-1}(k) \end{pmatrix}$$

$$\mathbf{x}(k) = \mathbf{a} s(k - \tau) + \mathbf{v}(k) . \quad (2.1)$$

Since all relevant quantities and designs depend on the frequency, the following examinations are carried out in the frequency domain without any loss of generality. The Fourier-transform leads to

$$\mathbf{X}(e^{j\Omega}) = \mathbf{S}(e^{j\Omega})\mathbf{d} + \mathbf{V}(e^{j\Omega}) , \quad (2.2)$$

where $\mathbf{d}$ is the representation of the delays and the attenuation in the frequency domain which depends on the actual geometry of the array and the

direction of the source signal.

$$\mathbf{d}^T = [a_0 \exp(-j\Omega\tau_0), a_1 \exp(-j\Omega\tau_1), \dots, a_{N-1} \exp(-j\Omega\tau_{N-1})] . \quad (2.3)$$

Finally, the output signal

$$Y_b(e^{j\Omega}) = \sum_{n=0}^{N-1} W_n^*(e^{j\Omega}) X_n(e^{j\Omega}) = \mathbf{W}^H \mathbf{X} , \quad (2.4)$$

where $W_n(e^{j\Omega})$ denotes the frequency-domain coefficients of the beamformer of sensor n at the frequency Ω and the operator ^{H} denotes a conjugated transposition (Hermitian operator). The inverse Fourier-transform results in the discrete-time output signal $y_b(k)$.

2.2.1 Array-Gain

The array-gain (AG) is the measure which shows the improvement of the signal-to-noise ratio (SNR) between one sensor and the output of the whole array¹. Therefore,

$$G = \frac{SNR_{\text{Array}}}{SNR_{\text{Sensor}}} . \quad (2.5)$$

Assuming stationary signals, the SNR of one sensor is given by the ratio of the power spectral densities (PSD) of the signal Φ_{SS} and the average noise $\Phi_{V_a V_a}$.

The SNR at the output can be computed by deriving the PSD of the output signal

$$\Phi_{Y_b Y_b} = \mathbf{W}^H \Phi_{\mathbf{X} \mathbf{X}} \mathbf{W} , \quad (2.6)$$

where

$$\Phi_{\mathbf{X} \mathbf{X}} = \begin{pmatrix} \Phi_{X_0 X_0} & \Phi_{X_0 X_1} & \dots & \Phi_{X_0 X_{N-1}} \\ \Phi_{X_1 X_0} & \Phi_{X_1 X_1} & \dots & \Phi_{X_1 X_{N-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_{X_{N-1} X_0} & \Phi_{X_{N-1} X_1} & \dots & \Phi_{X_{N-1} X_{N-1}} \end{pmatrix} \quad (2.7)$$

is a power spectral density matrix of the array input signals. When the desired signal is present only, the output is

$$\Phi_{Y_b Y_b} \Big|_{\text{Signal}} = \Phi_{SS} |\mathbf{W}^H \mathbf{d}|^2 , \quad (2.8)$$

¹ The dependence on Ω is omitted for the sake of brevity and readability.

and for the noise-only case the output is

$$\left. \Phi_{Y_b Y_b} \right|_{\text{Noise}} = \Phi_{V_a V_a} W^H \Phi_{VV} W, \quad (2.9)$$

where Φ_{VV} is a normalized cross power spectral density matrix of the noise². Therefore,

$$G = \frac{|W^H d|^2}{W^H \Phi_{VV} W}. \quad (2.10)$$

Assuming a homogeneous noise field (2.10) can be expressed in terms of the coherence matrix

$$\Gamma_{VV} = \begin{pmatrix} 1 & \Gamma_{V_0 V_1} & \Gamma_{V_0 V_2} & \cdots & \Gamma_{V_0 V_{N-1}} \\ \Gamma_{V_1 V_0} & 1 & \Gamma_{V_1 V_2} & \cdots & \Gamma_{V_1 V_{N-1}} \\ \vdots & \vdots & \ddots & \vdots & \\ \Gamma_{V_{N-1} V_0} & \Gamma_{V_{N-1} V_1} & \Gamma_{V_{N-1} V_2} & \cdots & 1 \end{pmatrix}, \quad (2.11)$$

where

$$\Gamma_{V_n V_m}(e^{j\Omega}) = \frac{\Phi_{V_n V_m}(e^{j\Omega})}{\sqrt{\Phi_{V_n V_n}(e^{j\Omega}) \Phi_{V_m V_m}(e^{j\Omega})}} \quad (2.12)$$

is the coherence function [4].

Thus,

$$G = \frac{|W^H d|^2}{W^H \Gamma_{VV} W}. \quad (2.13)$$

This representation allows an easier examination of beamformers for different noise fields, since many theoretically defined noise fields can be expressed by their coherence function.

2.2.2 Beampattern

One way to evaluate beamformers is to compute the response of the array to a wavefront coming from a specific frequency and a specific angle, depending on azimuth φ and elevation θ in a spherical coordinate system. Computing this response over all angles and frequencies leads to the spatial-temporal transfer function

$$|H(\varphi, \theta)|^2 \Big|_{\text{dB}} = -10 \log_{10} \left(\frac{|W^H d|^2}{W^H \Gamma_{VV} \Big|_{\text{Wavefront}} W} \right) \quad (2.14)$$

² The normalization factor is set to force the trace of the matrix to equal N.

called the farfield beampattern, which is usually displayed on a logarithmic scale. It can be computed by using (2.13) and the knowledge of the coherence function of a single wavefront with frequency Ω and an angle of arrival φ, θ . Additionally, f_s denotes the sampling frequency, $c = 340$ m/s the speed of sound, and l_{nm} the distances between the sensors in the Cartesian coordinate system

$$\Gamma_{V_n V_m} \Big|_{\text{Wavefront}} = \exp(j\Omega \tau_{nm}), \quad (2.15)$$

where

$$\tau_{nm} = \frac{f_s}{c} (l_{x,nm} \sin(\theta) \cos(\varphi) + l_{y,nm} \sin(\theta) \sin(\varphi) + l_{z,nm} \cos(\theta)). \quad (2.16)$$

Since the beampattern depends on three variables, it is not possible to display it in a single plot. Fortunately, line arrays aligned to the z-axis have a rotational symmetry and, therefore, the beampattern is independent of φ . Examples of beampatterns for line arrays will be shown in section 2.3.

2.2.3 Directivity

A common quantity to evaluate beamformers is the directivity factor, or its logarithmic equivalent the directivity index (DI) which describes the ability of the array to suppress a diffuse noise field. Therefore, we can compute the directivity factor by using (2.13) and inserting the coherence function of a diffuse noise field:

$$\begin{aligned} \Gamma_{V_n V_m}(e^{j\Omega}) \Big|_{\text{Diffuse}} &= \frac{\sin(\Omega f_s l_{nm}/c)}{\Omega f_s l_{nm}/c} \\ &= \text{sinc} \left\{ \frac{\Omega f_s l_{nm}}{c} \right\} \end{aligned} \quad (2.17)$$

where $\text{sinc}(x) = \sin(x)/x$. Thus, the DI is

$$\text{DI}(e^{j\Omega}) = 10 \log_{10} \left(\frac{|W^H d|^2}{W^H \Gamma_{VV} \Big|_{\text{Diffuse}} W} \right). \quad (2.18)$$

Another formal definition uses the transfer function (2.14) and describes the ratio of the transfer function of the look-direction θ_0, φ_0 of the array to the spatial integration over all directions of incoming signals.

$$\text{DI}(e^{j\Omega}) = 10 \log_{10} \left(\frac{|H(e^{j\Omega}, \varphi_0, \theta_0)|^2}{\frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} |H(e^{j\Omega}, \varphi, \theta)|^2 \sin(\theta) d\varphi d\theta} \right) \quad (2.19)$$

2.2.4 Front-to-Back Ratio

In many applications no principal look-direction exists, as for example in video-conferences or the recording of orchestras. Therefore, the DI is not the best quantity to describe the behavior of the array. In such applications a front-to-back ratio (FBR) is a better choice, since in most cases all desired sources are in front of the array and all unwanted disturbances are behind the array [19], [11]. The formal description utilizes the beampattern again:

$$\text{FBR}(e^{j\Omega}) = \frac{\int_{\theta_0-\pi/2}^{\theta_0+\pi/2} \int_{\varphi_0-\pi/2}^{\varphi_0+\pi/2} |H(e^{j\Omega}, \varphi, \theta)|^2 \sin(\theta) d\varphi d\theta}{\int_{\theta_0+\pi/2}^{\theta_0+3\pi/2} \int_{\varphi_0+\pi/2}^{\varphi_0+3\pi/2} |H(e^{j\Omega}, \varphi, \theta)|^2 \sin(\theta) d\varphi d\theta} \quad (2.20)$$

2.2.5 White Noise Gain

This last quantity shows the ability of the array to suppress spatially uncorrelated noise, which can be caused by self-noise of the sensors. Inserting the coherence matrix for this noise field

$$\Gamma_{VV}|_{\text{uncorr}} = I \quad (2.21)$$

into (2.13) results in the white noise gain:

$$\text{WNG}(e^{j\Omega}) = \frac{|W^H d|^2}{W^H W} \quad (2.22)$$

On a logarithmic scale positive values represent an attenuation of uncorrelated noise, whereas negative values show an amplification.

2.3 Design of Superdirective Beamformers

In order to design optimal beamformers, we have to minimize the power of the output signal $y_b(k)$ of the array. The output PSD is given by (2.6) and is a function of the input signal and the coefficients we want to determine. In order to avoid the trivial solution $W_n = 0$, the minimization is constrained to give an undistorted signal response in the desired look direction, i.e.,

$$W^H d = 1 \quad (2.23)$$

Therefore, the following constrained minimization problem has to be solved:

$$\min_W W^H \Phi_{XX} W \quad \text{subject to} \quad W^H d = 1 \quad (2.24)$$

Since we are only interested in the optimal suppression of the noise, and we assume a perfect correspondence between the direction of the desired signal and the look-direction of the array, only the noise PSD-matrix Φ_{VV} is used.

The well-known solution for (2.24) is called the Minimum Variance Distortionless Response (MVDR) beamformer [6]. It is given by

$$W = \frac{\Phi_{VV}^{-1} d}{d^H \Phi_{VV}^{-1} d}, \quad (2.25)$$

and can be derived by using the Lagrange-multiplier [13] or gradient computation [20], [9]. Assuming a homogeneous noise field the solution is a function of the coherence matrix:

$$W = \frac{\Gamma_{VV}^{-1} d}{d^H \Gamma_{VV}^{-1} d}. \quad (2.26)$$

Equations (2.25) or (2.26) can be interpreted as a spatial decorrelation process followed by a matched filter for the desired signal. The normalization in the denominator leads to unity signal response for the look direction.

The design procedure reduces to the choice of theoretically well-defined noise-fields in order to get optimal designs for different applications. Furthermore, different models for the desired signal can be included, leading to farfield and nearfield designs.

Examples for desired signal models are:

- Standard farfield model for linear arrays with equidistant sensors:

$$d^T = [1, \exp(-j\Omega f_s c^{-1} l \cos(\theta_0)), \exp(-j\Omega f_s c^{-1} 2l \cos(\theta_0)), \dots, \exp(-j\Omega f_s c^{-1} (N-1)l \cos(\theta_0))] \quad (2.27)$$

where l is the inter-sensor spacing.

- Nearfield design, including attenuation of the desired signal [14], [22]

$$d^T = [a_0 \exp(-j\omega\tau_0), a_1 \exp(-j\omega\tau_1), \dots, a_{N-1} \exp(-j\omega\tau_{N-1})], \quad (2.28)$$

$$a_i = \frac{\|q - p_{ref}\|}{\|q - p_i\|}, \quad (2.29)$$

$$\tau_i = \frac{\|q - p_{ref}\| - \|q - p_i\|}{c}, \quad (2.30)$$

where $\|q - p_{ref}\|$ and $\|q - p_i\|$ denote the distance between the vector location of the source q and a reference sensor p_{ref} , or the sensor p_i , respectively.

More elaborate examples for exact nearfield designs can be found in [18], [23]

2.3.1 Delay-and-Sum Beamformer

Although this chapter is called superdirective microphone arrays the well-known Delay-and-Sum Beamformer (DSB) is included for comparison purposes. It is an 'optimal' beamformer for optimizing the WNG. We can derive the coefficients from (2.26) by inserting the coherence matrix for spatial uncorrelated noise $\Gamma = I$. Thus,

$$W = \frac{1}{N} d. \quad (2.31)$$

The WNG is optimal in this case and reaches N . All other standard shading schemes like the Dolph-Chebyshev window [10] worsen the performance subject to WNG.

2.3.2 Design for spherical isotropic noise

In order to optimize the directivity factor, which depends on the noise-field of a spherical isotropic noise field (diffuse), we have to solve (2.26) by using the coherence matrix of the diffuse noise field, given by (2.17). The resulting coefficients represent the classic superdirective beamformer (SDB)³.

Figure 2.2 shows the beampattern of a DSB and a superdirective beamformer, both using five linear equispaced microphones ($l = 5$ cm) in endfire steering direction ($\theta_0 = \pi$). The x-axis represents the incoming spatial angle ($[0 \dots 2\pi]$) and the y-axis represents the frequency of the signal in kHz. The sampling-frequency was set to 8 kHz to cover the telephone bandwidth. The grey-scaled image represents the attenuation of the incoming signals in dB.

The look-direction is unmodified at all frequencies due to the linear constraint. Additionally, an unmodified region at higher frequencies can be seen caused by spatial aliasing, since our choice of the parameter does not fulfill the spatial sampling theorem, which is given by

$$l < \frac{\lambda}{2}, \quad (2.32)$$

where λ denotes the wavelength. The upper sampling frequency should therefore be restricted to $f_s = 6.8$ kHz, or the distance should not exceed $l = 4.25$ cm. However, in order to show some effects we will keep these parameters in all experiments.

Furthermore, the DSB is unable to suppress low frequency noise sources coming from any direction. In contrast, the superdirective beamformer attenuates very well sources coming from directions other than the look-direction

³ In this chapter the term superdirective beamformer is used for the beamformer which optimizes the directivity factor, independent of the frequency or the ratio of the wavelength to the distance between the sensor elements. In the classic definition this is often restricted to the case where the wavelength is large with respect to the distance between sensors.

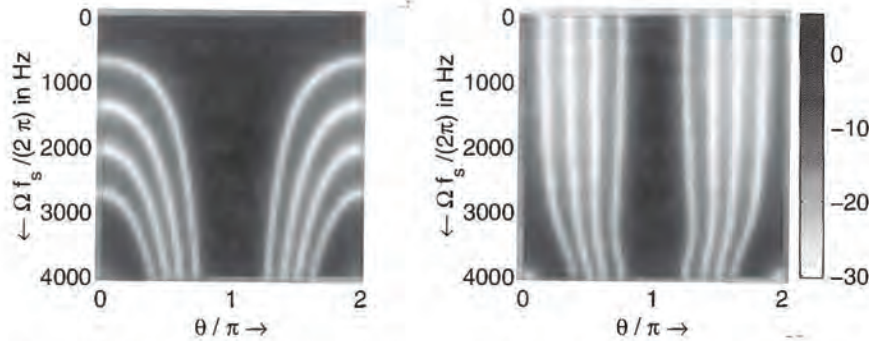


Fig. 2.2. Left: beam pattern of a delay-and-sum beamformer. Right: beam pattern of an optimal array for isotropic noise (superdirective beamformer). ($l = 5$ cm, $N = 5$, endfire steering direction)

over the whole frequency range. However, at higher frequencies the superdirective beamformer degrades to the DSB, since supergain can only be achieved if the signal wavelength is larger than two times the microphone distance.

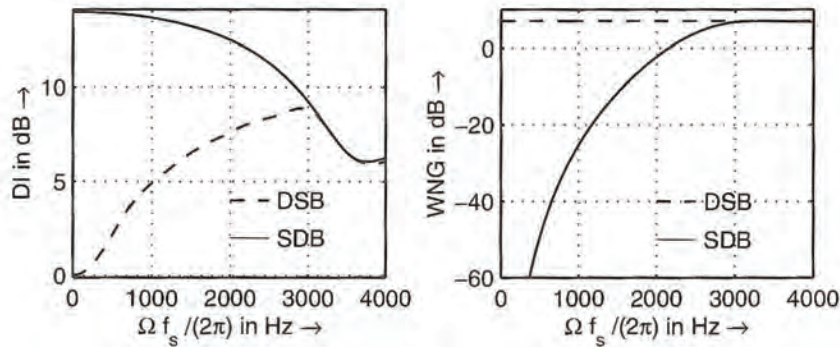


Fig. 2.3. Left: Directivity index (DI) for delay-and-sum beamformer and superdirective beamformer. Right: White noise gain (WNG) for delay-and-sum beamformer and superdirective beamformer. ($l = 5$ cm, $N = 5$, endfire steering direction)

Figure 2.3 shows the DI on the left side and the WNG on the right side for the same parameters as in the previous figure. The directivity index reaches zero at low frequencies for the DSB (as expected by analyzing the beam pattern) and N^2 for the superdirective beamformer. The proof for this limit in the endfire steering case can be found in [11]. At higher frequencies the directivity for both designs is nearly the same and it is given by N , since the $\text{sinc}\{\cdot\}$ function tends to zero, and the noise field is uncorrelated in this case.

If we now take a closer look at the WNG, we can see why this design is not suitable in real-world applications. Whereas the DSB suppresses uncorrelated noise equally at all frequencies, the SDB boosts uncorrelated noise at lower frequencies.

In order to give a deeper insight into how supergain works, we will compute the coefficients for an array of only two microphones. The distance is again 5 cm, and endfire steering is used.

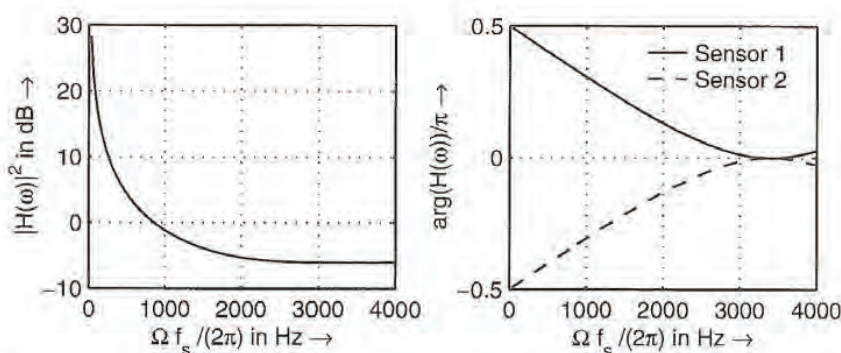


Fig. 2.4. Coefficients of a two channel SDB, left: Magnitude, right: Phase ($l = 5$ cm, $N = 2$, endfire steering direction)

In Fig. 2.4 the squared magnitude and the phase of the two coefficient vectors are shown. First of all, the coefficients are conjugate complex. Secondly, the filters force the phase between the noise components at each sensor to be π . Therefore, the correlated part of the noise will be compensated. Hence, the desired signal is also correlated, and therefore it is reduced as well. To fulfill the constraint of an undisturbed desired signal, the coefficients have to boost the input signals to compensate this behavior, which can be seen on the left side of Fig. 2.4. Therefore, uncorrelated noise will be amplified. At higher frequencies the correlation between the noise components vanishes and the beamformer degrades to the DSB. The magnitude of the coefficients reaches $1/2$.

In order to overcome the problem of self-noise amplification in superdirective designs, Gilbert and Morgan have proposed a method for solving (2.24) under a WNG constraint [15]. The method uses a small added scalar μ to the main diagonal of the normalized PSD or coherence matrix:

$$W_c = \frac{(\Gamma_{VV} + \mu I)^{-1} d}{d^H (\Gamma_{VV} + \mu I^{-1} d)} \quad (2.33)$$

We prefer a mathematically equivalent form, which preserves the interpretation as a coherence matrix with elements smaller than one. Instead of adding

the scalar to the main diagonal, we divide each non-diagonal element by $1 + \mu$. Therefore, μ can be interpreted as the ratio of the sensor noise σ^2 to the ambient noise power Φ_{VV} . For the diffuse noise field the non-diagonal elements are given by

$$\Gamma_{V_n V_m} = \frac{\text{sinc} \left\{ \frac{\Omega f_s l_{nm}}{c} \right\}}{1 + \frac{\sigma^2}{\Phi_{VV}}} \quad \forall \quad n \neq m. \quad (2.34)$$

The factor μ can vary from zero to infinity, which results in the unconstrained SDB or the DSB respectively. The WNG changes as a monotonic function between the two limits [15]. Typical values for μ are in the range between -10 dB to -30 dB. Unfortunately, there is no simple relation between μ and the resulting WNG. By using a frequency variant μ the WNG can be restricted at all frequencies, but not through direct computation.

There are two different iterative design schemes. The first one was published by Doerbecker [9]. It is a straightforward implementation of a trial-and-error strategy. Another iterative design method uses the scaled projection algorithm developed by Cox *et al.* for adaptive arrays [6]. Instead of the estimated PSD-matrix, the theoretically defined coherence or PSD-matrix is inserted in the scaled projection algorithm. This solution was presented in [17]. Both algorithms result in similar coefficients and can be implemented easily.

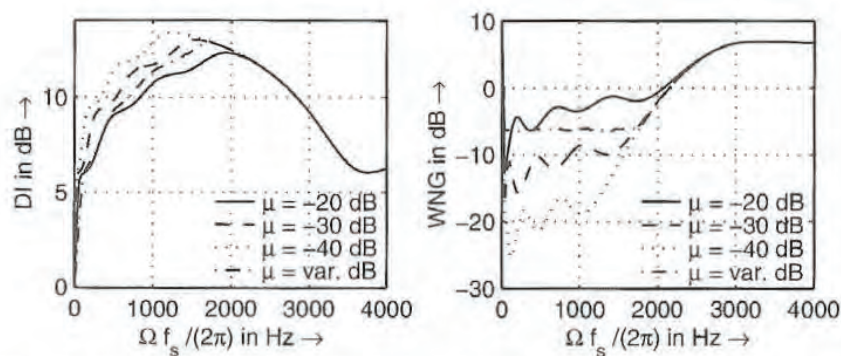


Fig. 2.5. Left: Directivity index (DI) for different constrained designs. Right: White noise gain (WNG) for different constrained designs. ($l = 5$ cm, $N = 5$, endfire steering direction)

Figure 2.5 depicts the effects for three fixed and one variable μ as constraining parameters. For the variable μ , the WNG constraint was set to -6 dB. The constrained design facilitates a good compromise between DI

and WNG. A careful design can optimize such arrays for a wide range of applications.

2.3.3 Design for Cylindrical Isotropic Noise

In some applications a spherical isotropic noise field is not the best choice or the best approximation of a given noise-field. Another well-defined noise-field can be used, if we reduce the three dimensions to two dimensions. We get a noise-field which is defined by infinite noise sources of a circle with an infinite radius. This kind of noise can arise if a lot of people speak in large rooms where the ceiling and the floor are damped well, or in the free-field (cocktail-party noise)⁴. The coherence between two sensors is given by [7]

$$\Gamma_{X_n X_m}(\omega) = J_0\left(\frac{\omega l_{nm}}{c}\right), \quad (2.35)$$

where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind. This leads to the solution of [8] as an improved design for speech enhancement for a hearing-aid application. In order to constrain the coefficients, a similar technique as in (2.34) has to be carried out.

In comparison to the design for a diffuse noise-field the differences are not large, but at lower frequencies a better suppression of noise sources behind the look direction can be observed. Elko [11] has shown that the directivity factor is less and its limit is $2N - 1$, in contrast to N^2 in the unconstrained case ($\mu = 0$). A design example will be given in the next section.

2.3.4 Design for an Optimal Front-to-Back Ratio

A last data-independent design tries to optimize the front-to-back ratio. In many applications the look direction of the desired signal cannot be predetermined, but in most cases the desired signal is in front of the array and all disturbances are at the rear, e.g. when recording an orchestra or in video-conferences.

Our suggestion for a different design strategy is not to use an isotropic noise field, but to restrict the assumed infinite noise sources to the back half of a circle or a sphere.

The resulting noise-field between two sensors separated by the distance l can be described by an integration over an infinite number of uncorrelated noise sources. The resulting function in the two-dimensional case is:

$$f(e^{j\Omega}, \theta_0) = \frac{1}{\pi} \int_{\theta_0 + \pi/2}^{\theta_0 + 3\pi/2} \exp(j\Omega f_s c^{-1} l \cos(\theta)) d\theta. \quad (2.36)$$

⁴ The origin of this cylindrical isotropic noise-field is the sonar application in shallow water.

Using numerical integration methods, inserting the resulting complex values in the coherence matrix and solving (2.26), results in a new design which suppresses noise sources from the rear very well.

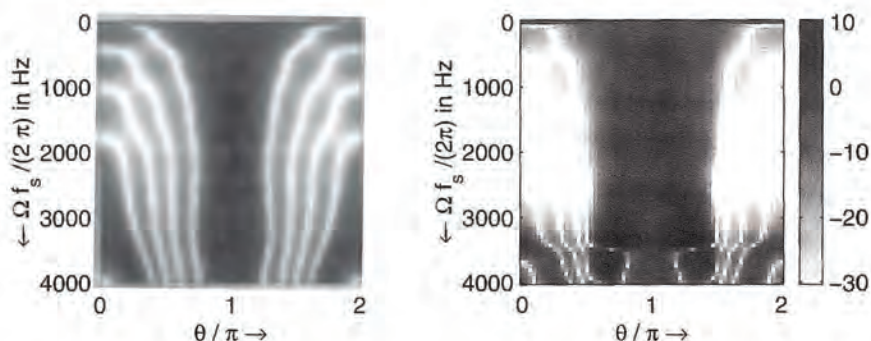


Fig. 2.6. Left: beampattern of a constrained superdirective beamformer. Right: beampattern of a constrained beamformer, designed with (2.36). ($l = 5$ cm, $N = 5$, $\mu = 0.01$, endfire steering direction)

Figure 2.6 shows beampatterns of two constrained beamformers ($\mu = 0.01$). The left side is computed with optimized coefficients for a diffuse noise-field, and the right side uses coefficients designed with the help of (2.36). At lower frequencies the constraining parameter is dominant and therefore, both designs do not perform well. From 300 Hz to 2800 Hz the new design suppresses all signals coming from the rear at the cost of a wider main lobe; this is sometimes an advantage, for example if the source is not exactly in endfire position.

At higher frequencies, especially if spatial aliasing occurs, the new design boosts signals coming from directions near the look direction, which can cause some unnatural coloring of the signal and the remaining noise. Therefore, special care has to be taken when choosing the parameters of the new design scheme.

In order to show the advantages of the new schemes, Fig. 2.7 depicts the DI and the FBR measure for the three different designs. At lower frequencies the small advantage of the cylindrical optimal design against the spherical design for the FBR can be seen, but the differences are very small over the whole frequency range. On the other hand, the behavior of the new design is completely different. Measuring the DI leads to much smaller values, but the FBR is very high, especially in the mid-frequency range.

Interestingly, we can transform between the optimal design for cylindrical isotropic noise and the new design by introducing a new variable which

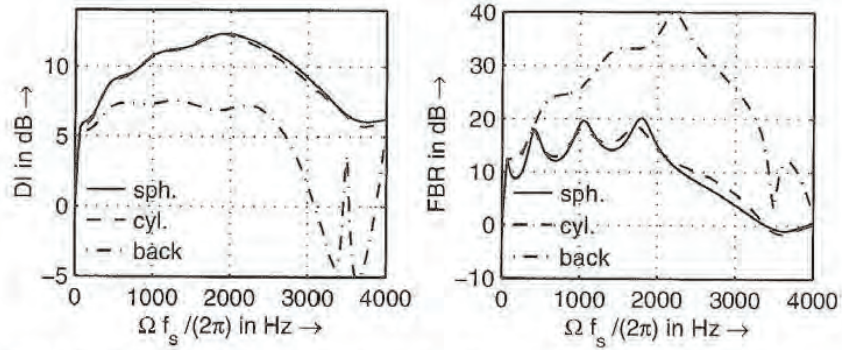


Fig. 2.7. Left: Directivity index (DI) for three optimal designs. Right: Front-to-back ratio (FBR) for three optimal designs. ($l = 5$ cm, $N = 5$, $\mu = 0.01$, endfire steering direction)

adjusts the limits of the integral, i.e.,

$$f(e^{j\Omega}, \theta_0, \delta) = \frac{1}{2(\pi - \delta)} \int_{\theta_0 + \delta}^{\theta_0 - \delta + 2\pi} \exp(j\Omega f_s c^{-1} l \cos(\theta)) d\theta \quad 0 \leq \delta \leq \pi \quad (2.37)$$

Setting $\delta = 0$ corresponds to the isotropic noise case, and $\delta = \pi/2$ results in (2.36).

2.3.5 Design for Measured Noise Fields

So far, only data-independent designs have been considered. If *a priori* knowledge is available, however, it should be used to improve the performance. For example, this information could be a prescribed direction ($\theta =$ angle) of an incoming noise source. Assuming the noise source is in the far field of the microphone array, the complex coherence function between two sensors is given by

$$\text{Re}\{\Gamma_{X_n X_m}(\omega)\} = \cos\left(\frac{\Omega f_s \cos(\theta) l_{nm}}{c}\right) \quad (2.38)$$

$$\text{Im}\{\Gamma_{X_n X_m}(\omega)\} = -\sin\left(\frac{\Omega f_s \cos(\theta) l_{nm}}{c}\right) \quad (2.39)$$

Inserting the complete coherence matrix in (2.26) forms a null in that direction over the whole frequency range. In order to restrict the WNG a constrained design is necessary.

Furthermore, if we assume stationarity we can measure the actual noise-field and solve the design equation which results in the MVDR solution. Adaptive algorithms like the constrained projection by Cox [6], or the original

algorithm by Frost [13], will converge exactly to the same solution under the assumption of stationary noise and an infinitely small step-size.

2.4 Extensions and Details

After describing the main form of the MVDR beamformer and typical data-independent designs, we will compare them to their analogue counterparts, the gradient microphones. Furthermore, an alternative implementation structure will be given which can reduce the computational complexity and open superdirective designs for future extensions.

2.4.1 Alternative Form

Assuming a time-aligned input signal, the optimal weights are defined differently, since the look-direction vector $\mathbf{d}$ is replaced by the column-vector

$$\mathbf{1} = \underbrace{[1, 1, \dots, 1]^T}_N$$

containing only ones, and the PSD-matrix or the coherence matrix contain the statistical information after time alignment (see Fig. 2.8). This gives

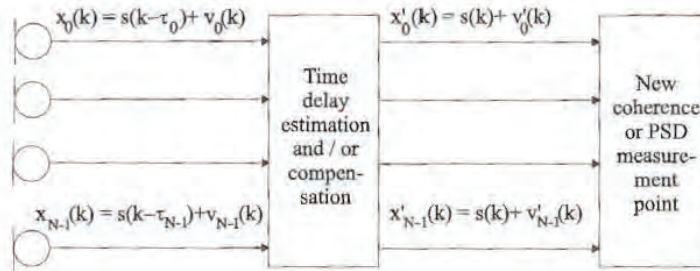


Fig. 2.8. Signal model after time delay compensation

$$W|_{ta} = \frac{\mathbf{1}^T (\mathbf{I}'_{VV} + \mu \mathbf{I})^{-1}}{\mathbf{1}^T (\mathbf{I}'_{VV} + \mu \mathbf{I})^{-1} \mathbf{1}}. \quad (2.40)$$

This solution of the constrained minimization problem can be decomposed into two orthogonal parts, following the ideas of Griffith and Jim [16]. One part represents the constraints only and the other part represents the unconstrained coefficients to minimize the output power of the noise.

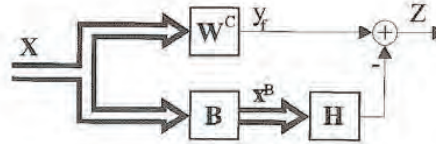


Fig. 2.9. Schematic description of the decomposition of the optimal weight vector into two orthogonal parts

The decomposed structure is depicted in Fig. 2.9. The multi-channel time-aligned input signal X is multiplied by W^C to fulfill the constraints. Furthermore, the input signal is projected onto the noise-only subspace⁵ by a blocking matrix B . The resulting vector X^B is multiplied by the optimal vector H and then subtracted from the output of the upper part of the structure to get the noise-reduced output signal Z . Several authors have shown the equivalence between this structure and the standard beamformer [16], [3], [12], if

$$W^C = \frac{1}{N} \mathbf{1},$$

which represents a delay-and-sum beamformer. Additionally, B has to fulfill the following properties:

- The size of the matrix is $(N - 1) \times N$
- The sum of all values in one row is zero
- The matrix has to be of rank $N-1$.

An example for $N = 4$ is given by

$$B = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \quad (2.41)$$

Another well-known example is the original Griffith-Jim matrix which subtracts two adjacent channels only:

$$B = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 1 & -1 \end{pmatrix}.$$

The last step to achieve a solution equivalent to (2.25) is the computation of the optimal filter H . A closer look at Fig. 2.9 shows that Y_f , X^B and Z describe exactly the problem of a multiple input noise canceler, described by

⁵ Which means that the desired signal is spatially filtered out (blocked).

Widrow and Stearns [24]. Therefore, this structure is called the generalized sidelobe canceler (GSC), if an adaptive implementation is used. The non-adaptive multi-channel Wiener solution of this problem can be found in [21]

$$H = \Phi_{X^B X^B}^{-1} \Phi_{X^B Y_f}, \quad (2.42)$$

where $\Phi_{X^B X^B}$ denotes the PSD-matrix of all signals after the matrix B , and $\Phi_{X^B Y_f}$ is the cross-PSD vector between the fixed beamformer output and the output signals X^B . Additionally, the coefficient vector can be computed as a function of the input PSD-matrix:

$$H = (B\Phi_{VV'}B^H)^{-1} B\Phi_{VV'}W^C. \quad (2.43)$$

If we now assume a homogeneous noise field, the PSD-matrix can be replaced by the coherence matrix of the delay-compensated noise field to compute the optimal coefficients:

$$H = (B\Gamma_{VV'}B^H)^{-1} B\Gamma_{VV'}W^C. \quad (2.44)$$

Therefore, all designs presented in section 2.3 can be implemented by using the GSC-structure. However, why should we do that? First of all, the new structure needs one filter less than the direct implementation. Using the first blocking matrix (2.41) further reduces the number of filters [1]. Secondly, a DSB output is available which can be used for future extensions. Thirdly, the new structure allows us to combine superdirective beamformers with adaptive post-filters for further noise reduction [2], and the new structure gives a deeper insight into MVDR-beamforming. For example, we can see that optimal beamforming is an averaging process combined with noise compensation.

2.4.2 Comparison with Gradient Microphones

Other devices with superdirectional characteristics are optimized gradient microphones [11]. In Fig. 2.10 a typical structure of a first order gradient microphone and its technical equivalent (composed of two omni-directional microphones) is shown.

The acoustic delay between the two open parts of the microphone can be realized by placing the diaphragm not exactly in the middle, or by using a material with a slower speed of sound.

The output of such systems is given by

$$E(\omega, \theta) = P_0 (1 - \exp(-j\omega[\tau + c^{-1}l \cos(\theta)])) , \quad (2.45)$$

where τ is the acoustic delay and P_0 denotes the amplitude of the source signal. If we now assume a small spacing with respect to the wavelength, an approximate solution can be derived:

$$E(\omega, \theta) \approx P_0 \omega (\tau + c^{-1}l \cos(\theta)) . \quad (2.46)$$

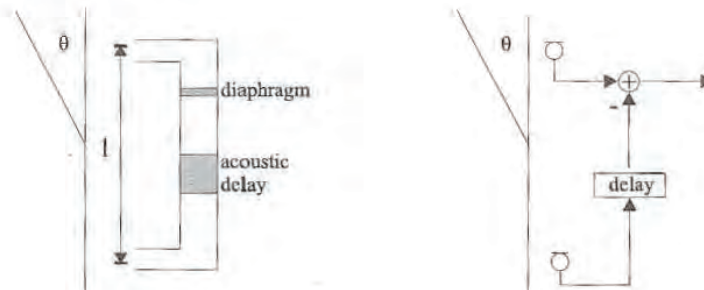


Fig. 2.10. Schematic description of a first order gradient microphone

A proper choice of τ leads to the different superdirective designs, called cardioid, supercardioid and hypercardioid. For example, the beampattern for a hypercardioid first order gradient microphone shows its zeros at $\approx \pm 109^\circ$. This type of microphone is designed to optimize the directivity factor and therefore, it represents the analogue equivalent of a two-sensor superdirective array. For a deeper insight and a complete review of higher order gradient microphones see [11].

At lower frequencies the two systems react more or less equally. The advantages of the analogue system are the smaller size of the device, and that no analogue-to-digital conversion is necessary. The advantages of the digital array technique are its flexibility, the easy scaling for many microphones, and the possible extensions with post-filters or other adaptive techniques.

At higher frequencies, if the assumption of small spacing is not valid anymore, the differences become visible. Through careful manufacturing these frequencies are much higher than the covered bandwidth. However, at some high frequencies the analogue microphone cancels the desired signal completely. On the other hand the array system reacts like a DSB at these frequencies, and no cancellation occurs.

2.5 Conclusion

Designing a so-called superdirective array or an optimal array for theoretically well-defined noise fields can be reduced to solving a single equation. Even nearfield assumptions and measured noise fields can be easily included. We have shown that the spatial characteristic, described by the coherence function, plays a key role in designing arrays. Most of the evaluation tools like the beampattern or the directivity index are directly connected to the coherence function. Beamformer designs with optimized directivity or higher front-to-back ratio also use the coherence.

One of the new aspects included in this chapter was a new noise model to improve the front-to-back ratio. Furthermore, we emphasized the close

relationship between superdirective arrays and adaptive beamformers and their well-known implementation as a generalized sidelobe canceler.

References

1. J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "An alternative implementation of the superdirective beamformer", in *Proc. IEEE Workshop Applicat. Signal Processing to Audio Acoust.*, pp. 7-10, New Paltz, NY, USA, Oct 1999.
2. J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "Multi-microphone noise reduction by post-filter and superdirective beamformer", in *Proc. Int. Workshop Acoust. Echo and Noise Control*, pp. 100-103, Pocono Manor, USA, Sep 1999.
3. K. M. Buckley, "Broad-band beamforming and the generalized sidelobe canceler", *IEEE Trans. Acoust. Speech Signal Processing*, vol. 34, pp. 1322-1323, Oct 1986.
4. G. C. Carter, *Coherence and Time Delay Estimation*, IEEE Press, 1993.
5. H. Cox, R. M. Zeskind, and T. Kooij, "Practical supergain", *IEEE Trans. Acoust. Speech Signal Processing*, vol. 34, pp. 393-398, Jun 1986.
6. H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming", *IEEE Trans. Acoust. Speech Signal Processing*, vol. 35, pp. 1365-1375, Oct 1987.
7. B. F. Cron and C.H. Sherman, "Spatial-correlation functions for various noise models", *J. Acoust. Soc. Amer.*, vol. 34, pp. 1732-1736, Nov 1962.
8. M. Doerbecker, "Speech enhancement using small microphone arrays with optimized directivity", in *Proc. Int. Workshop Acoust. Echo and Noise Control*, pp. 100-103, London, UK, Sep 1997.
9. M. Doerbecker, *Mehrkanalige Signalverarbeitung zur Verbesserung akustisch gestörter Sprachsignale am Beispiel elektronischer Hörhilfen*. PhD thesis, Dept. of Telecommunications, University of TH Aachen, Verlag der Augustinus Buchhandlung, Aachen, Germany, Aug 1998.
10. C. L. Dolph, "A current distribution for broadside arrays which optimizes the relationship between beamwidth and sidelobe level", *Proc. IRE*, pp. 335-348, Jun 1946.
11. G. W. Elko, "Superdirectional microphone arrays", in *Acoustic Signal Processing for Telecommunication*, S. L. Gay and J. Benesty, eds, ch. 10, pp. 181-235, Kluwer Academic Press, 2000.
12. M.H. Er and A. Cantoni, "Transformation of linearly constrained broadband processors to unconstrained partitioned form", *IEE Proc. Pt. H*, vol. 133, pp. 209-212, June 1986.
13. O. L. Frost, "An algorithm for linearly constrained adaptive array processing", *Proc. IEEE*, vol. 60, pp. 926-935, Aug 1972.
14. J.G. Ryan and R. A. Goubran, "Optimal nearfield response for microphone arrays", in *Proc. IEEE Workshop Applicat. Signal Processing to Audio Acoust.*, New Paltz, NY, USA, Oct 1997.
15. E. N. Gilbert and S. P. Morgan, "Optimum design of directive antenna arrays subject to random variations", *Bell Syst. Tech. J.*, pp. 637-663, May 1955.
16. L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming", *IEEE Trans. Antennas Propagat.*, vol. 30, pp. 27-34, 1982.

17. J. M. Kates and M. R. Weiss, "A comparison of hearing-aid array-processing techniques", *J. Acoust. Soc. Amer.*, vol. 99, pp. 3138-3148, May 1996.
18. R. A. Kennedy, T. Abhayapala, D. B. Ward, and R. C. Williamson, "Nearfield broadband frequency invariant beamforming", in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing (ICASSP-96)*, pp. 905-908, April 1996.
19. R. N. Marshall and W. R. Harry, "A new microphone providing uniform directivity over an extended frequency range", *J. Acoust. Soc. Amer.*, vol. 12, pp. 481-497, 1941.
20. R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*, John Wiley and Sons, New York, 1980.
21. S. Nordholm, I. Claesson, and P. Eriksson, "The broad-band Wiener solution for Griffith-Jim beamformers", *IEEE Trans. Signal Processing*, vol. 40, pp. 474-478, Feb 1992.
22. W. Tager, "Near field superdirectivity (NFSD)", in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing (ICASSP-98)*, Seattle, WA, USA, 1998.
23. D. B. Ward and G. W. Elko, "Mixed nearfield/farfield beamforming: A new technique for speech acquisition in a reverberant environment", in *Proc. IEEE Workshop Applicat. Signal Processing to Audio Acoust.*, New Paltz, NY, USA, Oct 1997.
24. B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Englewood Cliffs, 1985.

3 Post-Filtering Techniques

K. Uwe Simmer¹, Joerg Bitzer², and Claude Marro³

¹ Aureca GmbH, Bremen, Germany

² Houpert Digital Audio, Bremen, Germany

³ France Télécom R&D, Lannion, France

Abstract. In the context of microphone arrays, the term post-filtering denotes the post-processing of the array output by a single-channel noise suppression filter. A theoretical analysis shows that Wiener post-filtering of the output of an optimum distortionless beamformer provides a minimum mean squared error solution. We examine published methods for post-filter estimation and develop a new algorithm. A simulation system is presented to compare the performance of the discussed algorithms.

3.1 Introduction

What can be gained by additional post-filtering if the Minimum Variance Distortionless Response (MVDR) beamformer already provides the optimum solution for a given sound field?

Assuming that signal and noise are mutually uncorrelated the MVDR beamformer minimizes the noise power (or variance) subject to the constraint of a distortionless look direction response. The solution can be shown to be optimum in the Maximum Likelihood (ML) sense and produces the best possible Signal to Noise Ratio (SNR) for a narrowband input [1]. However, it does not maximize the SNR for a broadband input such as speech. Furthermore, the MVDR beamformer does not provide a broadband Minimum Mean Squared Error (MMSE) solution. The best possible linear filter in the MMSE sense is the multi-channel Wiener filter. As shown below the broadband multi-channel MMSE solution can be factorized into a MVDR beamformer followed by a single-channel Wiener post-filter. The multi-channel Wiener filter generally produces a higher output SNR than the MVDR filter. Therefore, additional post-filtering can significantly improve the SNR, which motivates this chapter.

The squared error minimized by the single-channel Wiener filter is the sum of residual noise and signal distortion components at the output of the filter. As a result, linear distortion of the desired signal cannot be avoided entirely if Wiener filtering is used. Additional Wiener filtering is advantageous in practice, however, because signal distortions can be masked by residual noise and a compromise between signal distortion and noise suppression can be found. Using MVDR beamforming alone often does not provide sufficient noise reduction due to its limited ability to reduce diffuse noise and reverberation.

The first concept of an electronic multi-microphone device to suppress diffuse reverberation was proposed by Danilenko in 1968 [2]. His research was motivated by Békésy's [3] observation that human listeners are able to suppress reverberation if sounds are presented binaurally. In Danilenko's reverberation suppressor a main microphone signal is multiplied by a broadband gain factor that is equal to the ratio of short-time cross-correlation and energy measurements. Two auxiliary microphones were used to measure correlation and energy. Danilenko already noted that such a system would also suppress incoherent acoustic noise. However, the proposed analog, electronic tube version of this system was not realized at that time. Another proposal in [2] was to evaluate squared sum and differences of two microphone signals, an idea that later was developed independently by Gierl and others in the context of digital multi-channel spectral subtraction algorithms [4], [5], [6], [7], [8].

According to Danilenko, his correlation-based concept was first realized during Blauert's stay at Bell Labs. In [9], Allen *et al.* presented a digital, two-microphone algorithm for dereverberation based on short-term Fourier-Transform and the overlap-add method. In 1984, Kaneda and Tohyama extended the application of the correlation based post-filters to noise reduction [10]. The first multi-microphone solution was published by Zelinski [11], [12]. Simmer and Wasiljeff showed that Zelinski's approach does not provide an optimum solution in the Wiener sense if the noise is spatially uncorrelated, and developed a slightly modified version [13]. A deeper analysis of the Zelinski and the Simmer post-filter can be found in [14], [15].

In the last decade, several new combinations and extensions of the post-filter approach were published. Le-Bouquin and Faucon used the coherence function as a post-filter [16], [17] and extended their system by a coherence subtraction method to overcome the problem of insufficient noise reduction at low frequencies [18], [19]. The problem of time delay estimation and further improvement of the estimation of the transfer function was independently addressed by Kuczynski *et al.* [20], [21] and Drews *et al.* [22], [23]. Fischer and Simmer gave a first solution by associating a post-filter and a generalized sidelobe canceler (GSC) to improve the noise reduction in case the noise field is dominated by coherent sources [24], [25]. Another system for the same task was introduced by Hussain *et al.* [26] and was based on switching between algorithms. The same strategy of switching between different algorithms, where the decision is based on the coherence between the sensors, can be found in [27], [28]. Furthermore, Mamhoudi and Drygajlo used the wavelet-transform in combination with different post-filters to improve the performance [29], [30]. Bitzer *et al.* [31], [32] proposed a solution with a super-directive array and McCowan *et al.* used a near-field super-directive approach [33].

Reading these papers we find that a theoretical basis for post-filtering seems to be missing. Therefore, an analysis based on optimum MMSE multi-channel filtering is presented in the following section.

3.2 Multi-channel Wiener Filtering in Subbands

We use matrix notation for a compact derivation. Signal vector $\mathbf{x}$ and weight vector $\mathbf{w}$ denote the multi-channel signal at the output of the N microphones and the multi-channel beamformer coefficients, respectively. We assume that the input signal vector $\mathbf{x}(k)$ is decomposed into M complex subband signals $\mathbf{x}(k, i)$ by means of an analysis filter-bank, where k is the discrete time index and i is the subband index. The optimum weight vector $\mathbf{w}_{opt}(k, i)$ for transforming the input signal vector $\mathbf{x}(k, i) = \mathbf{s}(k, i) + \mathbf{v}(k, i)$ corrupted by additive noise $\mathbf{v}(k, i)$ into the best possible MMSE approximation of the desired scalar signal $s(k, i)$ is referred to as multi-channel Wiener filter [34]. We assume that the relation between the desired scalar signal $s(k, i)$ and the signal vector $\mathbf{s}(k, i)$ is linear and that the N elements of the column vectors $\mathbf{s}(k, i)$ and $\mathbf{v}(k, i)$ are random processes. In the following, ^{T} denotes transposition, * denotes complex conjugation, ^{H} denotes Hermitian transposition, and $E[\cdot]$ denotes the statistical expectation operator.

3.2.1 Derivation of the Optimum Solution

The error in subband i for an arbitrary weight vector $\mathbf{w}(k, i)$ is defined as the difference of the filter output

$$y(k, i) = \mathbf{w}^H(k, i)\mathbf{x}(k, i) = \mathbf{w}^H(k, i)[\mathbf{s}(k, i) + \mathbf{v}(k, i)] \quad (3.1)$$

and the scalar desired signal $s(k, i)$, that is

$$e(k, i) = s(k, i) - \mathbf{w}^H(k, i)\mathbf{x}(k, i). \quad (3.2)$$

Using the definitions for the power of a complex signal

$$\phi_{xx}(k, i) = E[x(k, i)x(k, i)^*], \quad (3.3)$$

the cross-correlation vector

$$\phi_{xy}(k, i) = E[\mathbf{x}(k, i)y^*(k, i)], \quad (3.4)$$

and the correlation matrix

$$\Phi_{xx}(k, i) = E[\mathbf{x}(k, i)\mathbf{x}^H(k, i)], \quad (3.5)$$

the squared error at time k may be written as

$$\begin{aligned} \phi_{ee}(i) &= E[\{s(i) - \mathbf{w}^H(i)\mathbf{x}(i)\}\{s^*(i) - \mathbf{x}^H(i)\mathbf{w}(i)\}] \\ &= \phi_{ss}(i) - \mathbf{w}^H(i)\phi_{xs}(i) - \phi_{xs}^H(i)\mathbf{w}(i) + \mathbf{w}^H(i)\Phi_{xx}(i)\mathbf{w}(i), \end{aligned} \quad (3.6)$$

where the time index k has been omitted without loss of generality. The optimum solution minimizes the sum of all error powers $\phi_{ee}(i)$:

$$\sum_{i=0}^M [\phi_{ss}(i) - \mathbf{w}^H(i)\phi_{xs}(i) - \phi_{xs}^H(i)\mathbf{w}(i) + \mathbf{w}^H(i)\Phi_{xx}(i)\mathbf{w}(i)]. \quad (3.7)$$

Since the error power is necessarily real-valued and nonnegative for all subbands, the sum can be minimized for the weight vector $\mathbf{w}(i)$ by minimizing the error power $\phi_{ee}(i)$ for each subband. Therefore, the frequency index i may also be omitted without loss of generality.

The power ϕ_{ee} is a quadratic function of $\mathbf{w}$ and therefore has a single, global minimum. The optimum weight vector minimizing the squared error is obtained by setting the gradient of ϕ_{ee} with respect to $\mathbf{w}$ equal to the null vector [35]:

$$\nabla_{\mathbf{w}}(\phi_{ee}) = 2 \frac{\partial \phi_{ee}}{\partial \mathbf{w}^*} = -2\phi_{xs} + 2\Phi_{xx}\mathbf{w} = 0. \quad (3.8)$$

The resulting expression is the subband version of the multi-channel Wiener-Hopf equation in its most general form

$$\Phi_{xx}\mathbf{w}_{opt} = \phi_{xs}, \quad (3.9)$$

where Φ_{xx} is the correlation matrix of the noisy input vector and ϕ_{xs} is the cross-correlation vector between the noisy input vector and the desired scalar signal s . Assuming Φ_{xx} to be nonsingular, we may solve (3.9) for the optimum weight vector:

$$\mathbf{w}_{opt} = \Phi_{xx}^{-1} \phi_{xs}. \quad (3.10)$$

3.2.2 Factorization of the Wiener Solution

In our application, the received signal is assumed to consist of a single desired scalar signal that is transformed by the acoustic path $\mathbf{d}$ and additive noise:

$$\mathbf{x} = s\mathbf{d} + \mathbf{v}. \quad (3.11)$$

The noise vector $\mathbf{v}$ is given by

$$\mathbf{v} = [v_0, v_1, \dots, v_{N-1}]^T \quad (3.12)$$

where v_n is a complex noise signal in subband i at microphone n . The complex propagation vector is

$$\mathbf{d} = [d_0, d_1, \dots, d_{N-1}]^T \quad (3.13)$$

where d_n describes the acoustic path from the desired source to the microphone n for subband i . The propagation vector $\mathbf{d}$ may include time delays, near-field effects, and the transfer functions of enclosure and microphones. With the definitions (3.3), (3.4), (3.5) and assuming that signal and noise are uncorrelated, the cross-correlation vector may be reduced to

$$\phi_{xs} = \phi_{ss}\mathbf{d} \quad (3.14)$$

and the correlation matrix may be expressed as

$$\Phi_{xx} = \phi_{ss} \mathbf{d} \mathbf{d}^H + \Phi_{vv}. \quad (3.15)$$

Consequently, the optimum weight vector may be written as

$$\mathbf{w}_{opt} = \Phi_{xx}^{-1} \phi_{ss} \mathbf{d} = [\phi_{ss} \mathbf{d} \mathbf{d}^H + \Phi_{vv}]^{-1} \phi_{ss} \mathbf{d}. \quad (3.16)$$

The multi-channel Wiener filter can now be factorized into an array processor and a single channel post-filter by applying the Sherman-Morrison-Woodbury formula

$$[\mathbf{A}^{-1} + \mathbf{B} \mathbf{C}^{-1} \mathbf{B}^H]^{-1} \equiv \mathbf{A} - \mathbf{A} \mathbf{B} (\mathbf{C} + \mathbf{B}^H \mathbf{A} \mathbf{B})^{-1} \mathbf{B}^H \mathbf{A} \quad (3.17)$$

which is also known as the matrix inversion lemma [35]. Substituting

$$\mathbf{A} = \Phi_{vv}^{-1}, \quad \mathbf{B} = \sqrt{\phi_{ss}} \mathbf{d}, \quad \text{and} \quad \mathbf{C} = 1 \quad (3.18)$$

into (3.17), and taking into account that the Hermitian form $\mathbf{d}^H \Phi_{vv}^{-1} \mathbf{d}$ is scalar and real valued, the MMSE solution (3.16) can be transformed into

$$\begin{aligned} \mathbf{w}_{opt} &= \left[\Phi_{vv}^{-1} - \frac{\phi_{ss} \Phi_{vv}^{-1} \mathbf{d} \mathbf{d}^H \Phi_{vv}^{-1}}{1 + \phi_{ss} \mathbf{d}^H \Phi_{vv}^{-1} \mathbf{d}} \right] \phi_{ss} \mathbf{d} \\ &= \left[1 - \frac{\phi_{ss} \mathbf{d}^H \Phi_{vv}^{-1} \mathbf{d}}{1 + \phi_{ss} \mathbf{d}^H \Phi_{vv}^{-1} \mathbf{d}} \right] \phi_{ss} \Phi_{vv}^{-1} \mathbf{d} \\ &= \left[\frac{\phi_{ss}}{1 + \phi_{ss} \mathbf{d}^H \Phi_{vv}^{-1} \mathbf{d}} \right] \Phi_{vv}^{-1} \mathbf{d} \\ &= \left[\frac{\phi_{ss}}{\phi_{ss} + (\mathbf{d}^H \Phi_{vv}^{-1} \mathbf{d})^{-1}} \right] \frac{\Phi_{vv}^{-1} \mathbf{d}}{\mathbf{d}^H \Phi_{vv}^{-1} \mathbf{d}}. \end{aligned} \quad (3.19)$$

Equation (3.19) shows that the multi-channel Wiener filter (3.10) can be written as the product of the weight vector of the MVDR beamformer, (see Chapter 2) and a real-valued scalar factor. A similar result is used in [36] and [1] to show that the multi-channel Wiener and the MVDR solution yield the same SNR if the input is narrowband. In this case the MVDR beamformer is preferable since it is data independent (i.e. completely defined by the spatial configuration of signal and noise sources), whereas the Wiener solution is data dependent (ϕ_{ss} must be known or estimated) and is therefore much more difficult to handle. However, MVDR and Wiener solutions yield the same SNR only if the input consists of a single frequency. For the broadband case (which has already been discussed in [37]), the scalar factor becomes a subband or frequency domain post-filter that may significantly improve the SNR.

To show that the optimum post-filter is also a Wiener filter that operates on the single-channel output data, we evaluate the power of the desired signal at the output of the MVDR processor as

$$\phi_{s_o s_o} = \phi_{s s} \mathbf{w}_{\text{mvdr}}^H \mathbf{d} \mathbf{d}^H \mathbf{w}_{\text{mvdr}} = \phi_{s s} \left| \frac{\mathbf{d}^H \Phi_{vv}^{-1} \mathbf{d}}{\mathbf{d}^H \Phi_{vv}^{-1} \mathbf{d}} \right|^2 = \phi_{s s}. \quad (3.20)$$

This demonstrates the distortionless magnitude response. Furthermore, we determine the power of the output noise as

$$\phi_{v_o v_o} = \mathbf{w}_{\text{mvdr}}^H \Phi_{vv} \mathbf{w}_{\text{mvdr}} = \frac{\mathbf{d}^H \Phi_{vv}^{-1} \mathbf{d}}{(\mathbf{d}^H \Phi_{vv}^{-1} \mathbf{d})^2} = \frac{1}{\mathbf{d}^H \Phi_{vv}^{-1} \mathbf{d}}. \quad (3.21)$$

Substituting (3.20) and (3.21) into (3.19), we can finally factorize the optimum MMSE solution into the following expression:

$$\mathbf{w}_{\text{opt}} = \underbrace{\left[\frac{\phi_{s_o s_o}}{\phi_{s_o s_o} + \phi_{v_o v_o}} \right]}_{\text{Wiener post-filter}} \underbrace{\left[\frac{\Phi_{vv}^{-1} \mathbf{d}}{\mathbf{d}^H \Phi_{vv}^{-1} \mathbf{d}} \right]}_{\text{MVDR array}}. \quad (3.22)$$

Equation (3.22) includes the complex weight vector of the MVDR beamformer

$$\mathbf{w}_{\text{mvdr}}(k, i) = \frac{\Phi_{vv}^{-1}(k, i) \mathbf{d}(k, i)}{\mathbf{d}^H(k, i) \Phi_{vv}^{-1}(k, i) \mathbf{d}(k, i)}, \quad (3.23)$$

and the scalar, single channel Wiener post-filter that depends on the SNR at the output of the beamformer:

$$H_{\text{post}}(k, i) = \frac{\phi_{s_o s_o}(k, i)}{\phi_{s_o s_o}(k, i) + \phi_{v_o v_o}(k, i)} = \frac{SNR_{\text{out}}(k, i)}{1 + SNR_{\text{out}}(k, i)}. \quad (3.24)$$

The output signal $z(k, i)$ of the factorized MMSE filter is the product of the output signal $y(k, i)$ of the MVDR array:

$$y(k, i) = \mathbf{w}_{\text{mvdr}}^H(k, i) \mathbf{x}(k, i), \quad (3.25)$$

and the transfer function $H_{\text{post}}(k, i)$ of a single-channel post-filter:

$$z(k, i) = y(k, i) H_{\text{post}}(k, i). \quad (3.26)$$

The MVDR solution (3.23) maximizes the directivity index if Φ_{vv} equals the correlation matrix of the diffuse sound field. The resulting system may therefore be called 'superdirective array with Wiener post-filter' (although the term superdirectivity originated in the context of analog microphones). Since the definition (3.13) of the propagation vector does not include any far-field assumptions, (3.23) may also be used to design a near-field superdirective array.

3.2.3 Interpretation

Although the above results are clearly related to Wiener's work on optimum filtering [38], some basic assumptions were different. First of all, Wiener considered continuous time signals which leads to the Wiener-Hopf integral equation. The corresponding equation in matrix form (3.10) usually determines the filter coefficients for an optimum discrete time FIR filter of order N . In our case, the delay line is defined by the spatial arrangement of the acoustic sensor and the taps are realized by the N microphones. The array and the weight vector form a spatial filter. Wiener assumed that signal and noise are ergodic and stationary random processes and he used the Fourier-transform to find a solution for the optimum filter. This leads to a linear, time invariant filter. Such a filter is not appropriate for speech signals that may be modeled as short-time stationary processes only. The derivation used here is based on ensemble averages (expectations) and does not assume stationarity. In practice, however, only an approximate realization of such a filter is possible.

There are two main sources of errors: the analysis and synthesis filter-bank, and the procedures to estimate the time-varying signal and noise powers in the individual subbands. For the design of the filter-banks, a compromise between frequency and time resolution has to be made. High resolution in the frequency domain leads to poor resolution in the time domain and vice versa. Therefore, the highest possible frequency resolution that does not violate the short-term stationarity of speech should be chosen. Furthermore, the minimum error in the time-domain is only reached if the filters have non-overlapping frequency regions (see the discussion of subband methods in [39]). Since such filters are physically unrealizable, overlapping of subbands cannot be avoided. As a result, the suppression of a noise-only subband may affect adjacent subbands containing desired signal components. In the following, we will use windowing, Fast Fourier Transform (FFT) and the overlap-add method to implement the filter-bank. However, (3.22) is general enough to allow any complex or real valued filter-bank method. If overlap-add is used, circular convolution should be avoided by zero padding and by constraints imposed on the estimated transfer function.

In the derivation of the optimum filter, expectations are used to estimate the parameters. This is a theoretical construction since the ensemble averages cannot be computed in practice. An approximation proposed in [9] is the recursive Welch periodogram:

$$\hat{\phi}_{xy}(k, i) = \alpha \hat{\phi}_{xy}(k-1, i) + (1-\alpha)x(k, i)y^*(k, i), \quad (3.27)$$

where $\alpha = \exp(-D/[\tau_\alpha f_s])$ is defined by the decimation factor D of the filter-bank, the time-constant τ_α (ms), and the sampling frequency f_s (kHz). The time constant is again a compromise. If τ_α is low, artifacts may occur due to the variation of the transfer function estimate. On the other hand, if a high time constant τ_α is chosen, the assumption of short time stationarity is violated and the output speech signal may sound reverberant.

Unfortunately, the factorized result (3.22) does not give any indication of how the Wiener post-filter could be estimated. A possible solution, which we discuss in the next section, is based on the observation that the correlation between two microphone signals is low if the sound field is diffuse and the microphone distance is large enough.

3.3 Algorithms for Post-Filter Estimation

Figure 3.1 shows the block diagram of the studied algorithms. The microphone signals are time aligned and decomposed by a frequency subband transform (FT). The coefficients w_n represent the weight vector $\mathbf{w}$ of the beamformer and H represents the post-filter. The inverse subband transform (IFT) synthesizes the output signal. The coefficients f_n for post-filter estimation form a vector $\mathbf{f}$. Unless otherwise noted we assume that $\mathbf{f} = \mathbf{w}$. We begin

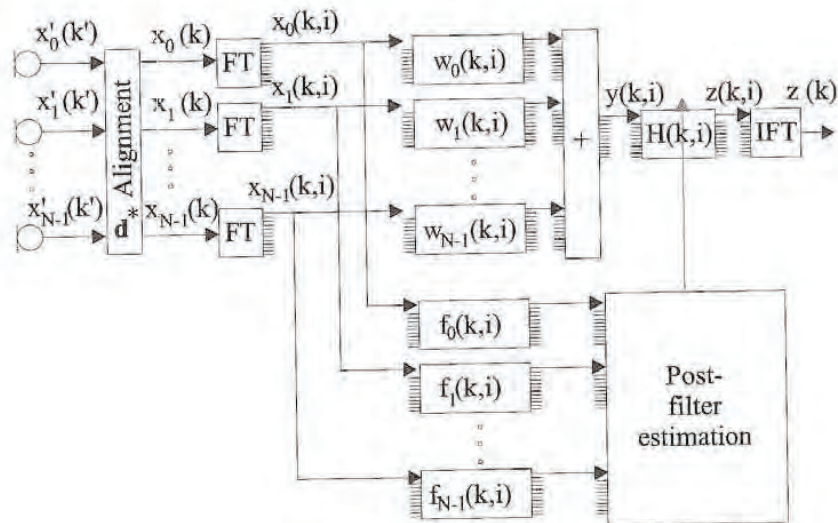


Fig. 3.1. General block diagram of the examined post-filters.

our analysis on multi-microphone post-filters by recalling some results on the performance of arrays from Chapter 2 since these results are needed later. We generally assume that the coefficients are normalized so that $\mathbf{w}^H \mathbf{1} \mathbf{1}^H \mathbf{w} = 1$ and $\mathbf{f}^H \mathbf{1} \mathbf{1}^H \mathbf{f} = 1$, where $\mathbf{1}$ is the N -vector of ones. Therefore, the array gain equals the noise reduction of the array. For convenience, we define a noise power attenuation factor that equals the inverse of the array gain:

$$A_F = \mathbf{w}^H \Gamma_{vv} \mathbf{w} = G^{-1}, \quad (3.28)$$

where the coherence matrix Γ_{vv} is the normalized noise correlation matrix $\Gamma_{vv} = \Phi_{vv}N/\text{trace}[\Phi_{vv}]$, and all quantities are assumed to be frequency dependent.

An examination of (3.28) shows that the noise attenuation of the array is the weighted sum of the complex coherence functions of all sensor pairs. Thus, all products appear in conjugate pairs $\Gamma_{mn} + \Gamma_{nm} = 2\text{Re}\{\Gamma_{nm}\}$. As a result, the noise reduction of the array is actually a function of the real part of the complex coherence between the sensors. The knowledge of the magnitude squared coherence is not sufficient.

The white noise gain is the array gain for spatially uncorrelated noise, where $\Gamma_{vv} = \mathbf{I}$. Thus, the attenuation factor for spatially white noise is

$$A_I = \mathbf{w}^H \mathbf{w} = \text{WNG}^{-1}. \quad (3.29)$$

The additional noise attenuation of the post-filter is given by

$$A_{\text{post}} = |H_{\text{post}}|^2. \quad (3.30)$$

The total noise attenuation of the combined system is the product of the attenuation of the array and the attenuation of the post-filter, or the respective sum in dB:

$$A_{\text{total}} \Big|_{\text{dB}} = 10 \log_{10}(A_I) + 10 \log_{10}(A_{\text{post}}). \quad (3.31)$$

3.3.1 Analysis of Post-Filter Algorithms

The first method for post-filter estimation we study is a generalized version of Zelinski's algorithms that was discussed by Marro *et al.* [15]. It covers several other algorithms as a special case.

$$H_{zm}(i) = \frac{\text{Re} \left\{ \sum_{n=0}^{N-2} \sum_{m=n+1}^{N-1} w_n(i) w_m^*(i) \Phi_{x_n x_m}(i) \right\} \sum_{n=0}^{N-1} |w_n(i)|^2}{\text{Re} \left\{ \sum_{n=0}^{N-2} \sum_{m=n+1}^{N-1} w_n(i) w_m^*(i) \right\} \sum_{n=0}^{N-1} |w_n(i)|^2 \Phi_{x_n x_n}(i)} \quad (3.32)$$

Equation (3.32) includes Danilenko's [2] idea to use the ratio of cross-correlation $\Phi_{x_n x_m}$ and power $\Phi_{x_n x_n}$ for suppressing incoherent noise, the complex sub-band approach of Allen *et al.* [9], Zelinski's proposal to average over all microphone pairs $m > n$ [11], and Marro's [40] extension to complex shading coefficients w_n . To write this algorithm in matrix notation, we note that

$$2\text{Re} \left\{ \sum_{n=0}^{N-2} \sum_{m=n+1}^{N-1} w_n w_m^* \Phi_{x_n x_m} \right\} = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} w_n w_m^* \Phi_{x_n x_m} - \sum_{n=0}^{N-1} w_n w_n^* \Phi_{x_n x_n}.$$

This is a Hermitian form of the shading coefficients w_n and the correlation matrix Φ_{xx} , minus the weighted sum of diagonal elements of Φ_{xx} . The algorithm (3.32) requires that the relative time-delay differences and gain ratios between the microphone signals have been compensated in advance so that $d = 1$. This leads to a modified noise correlation matrix Φ_{xx} (see Chapter 2). The transfer function of the post-filter (3.32) may now conveniently be written in matrix form as

$$H_{zm} = \frac{(\mathbf{w}^H \Phi_{xx} \mathbf{w} - \mathbf{w}^H \Phi_{xx}^D \mathbf{w}) \mathbf{w}^H \mathbf{w}}{(\mathbf{w}^H \mathbf{1} \mathbf{1}^H \mathbf{w} - \mathbf{w}^H \mathbf{w}) \mathbf{w}^H \Phi_{xx}^D \mathbf{w}}, \quad (3.33)$$

where Φ_{xx}^D is a diagonal matrix of the diagonal elements of Φ_{xx} . If the sound field is homogeneous, we have the same input power at each microphone, i.e. $\Phi_{xx}^D = \phi_{xx} \mathbf{I}$, and may write

$$H_{zm} = \frac{(\mathbf{w}^H \Phi_{xx} \mathbf{w} - \phi_{xx} \mathbf{w}^H \mathbf{w})}{\phi_{xx} (\mathbf{w}^H \mathbf{1} \mathbf{1}^H \mathbf{w} - \mathbf{w}^H \mathbf{w})}. \quad (3.34)$$

If signal and noise are uncorrelated we have $\Phi_{xx} = \Phi_{ss} + \Phi_{vv}$. Therefore,

$$H_{zm} = \frac{(\mathbf{w}^H \Phi_{ss} \mathbf{w} - \phi_{ss} \mathbf{w}^H \mathbf{w}) + (\mathbf{w}^H \Phi_{vv} \mathbf{w} - \phi_{vv} \mathbf{w}^H \mathbf{w})}{(\phi_{ss} + \phi_{vv}) (\mathbf{w}^H \mathbf{1} \mathbf{1}^H \mathbf{w} - \mathbf{w}^H \mathbf{w})}. \quad (3.35)$$

Assuming that the coefficients are normalized such that $\mathbf{w}^H \mathbf{1} \mathbf{1}^H \mathbf{w} = 1$, the desired signal is coherent, i.e., $\Phi_{ss} = \phi_{ss} \mathbf{1} \mathbf{1}^H$. With the noise correlation matrix being $\Phi_{vv} = \phi_{vv} \Gamma_{vv}$, where $\phi_{vv} = \text{trace}[\Phi_{vv}]/N$, we finally obtain

$$H_{zm} = \frac{\phi_{ss}}{\phi_{ss} + \phi_{vv}} + \frac{\phi_{vv} (\mathbf{w}^H \Gamma_{vv} \mathbf{w} - \mathbf{w}^H \mathbf{w})}{(\phi_{ss} + \phi_{vv}) (1 - \mathbf{w}^H \mathbf{w})}. \quad (3.36)$$

Although the designs of the MVDR array and the post-filter estimation algorithm do not seem to have much in common, the transfer function of the post-filter may be expressed as a function of the attenuation factors of the array by substituting (3.28) and (3.29) into (3.36):

$$H_{zm} = \frac{\phi_{ss}}{\phi_{ss} + \phi_{vv}} + \frac{\phi_{vv} (A_r - A_{\mathbf{I}})}{(\phi_{ss} + \phi_{vv}) (1 - A_{\mathbf{I}})}. \quad (3.37)$$

This is also true for the slightly modified version of Zelinski's algorithm [13]:

$$\bar{H}_{sm}(i) = \frac{\text{Re} \left\{ \sum_{n=0}^{N-2} \sum_{m=n+1}^{N-1} w_n(i) w_m^*(i) \Phi_{x_n x_m}(i) \right\}}{\text{Re} \left\{ \sum_{n=0}^{N-2} \sum_{m=n+1}^{N-1} w_n(i) w_m^*(i) \right\} \phi_{yy}(i)}, \quad (3.38)$$

where $\phi_{yy} = \phi_{ss} + \phi_{vv}A_I$ is the output power of the array. The modified post-filter can be expressed as

$$H_{sm} = \frac{\phi_{ss}}{\phi_{ss} + \phi_{vv}A_I} + \frac{\phi_{vv}A_I(A_I - A_1)}{(\phi_{ss} + \phi_{vv}A_I)(1 - A_1)} \quad (3.39)$$

These rather surprising results were first derived in [15]. They are used in the following section to discuss the properties of a large class of post-filtering algorithms.

3.3.2 Properties of Post-Filter Algorithms

First of all, we note that the shading coefficients w_n form a weight vector $\mathbf{w}$ that generally can be computed by using the design rule of the MVDR array. It is not necessary, however, to use the same design for array processor and post-filter (see Fig. 3.1). Both the MVDR weight vector and the array gain are functions of the noise correlation matrix. It should be noted that the correlation matrix that is used for the design may differ from the correlation matrix of the environment in which the array operates. Therefore, three different correlation matrices may be involved: a first one for the design of the array processor, a second one for the design of the post-filter, and a third one to determine the performance in the actual environment.

Analyzing (3.37) and (3.39) leads to the following conclusions:

- Optimum performance is only reached if $A_I = A_1$:
The difference of the two attenuation factors is zero only if the noise is spatially uncorrelated which was Danilenko's initial assumption in the design of his suppression system. In this case, (3.37) becomes a Wiener filter for the input signal of the beamformer. On the other hand, (3.39) becomes a Wiener filter for the beamformer output and therefore represents the MMSE solution for uncorrelated noise if the delay and sum beamformer is used. All other coefficient sets, including superdirective solutions, yield suboptimal performance. In a diffuse sound field, the noise is correlated at low frequencies which leads to poor performance for low frequency noise.
- Negative post-filter if $A_I < A_1$:
In a diffuse noise field, or if coherent sources are present, the difference of the attenuation factors ($A_I - A_1$) may cause a negative transfer-function. If negative parts of the transfer functions are set to zero, which is a common strategy, signal cancellation may occur.
- Infinite post-filter if $A_1 = 1$:
This is usually the case with superdirective designs which amplify uncorrelated noise at low frequencies.

To demonstrate the preceding results, we computed the theoretical performance of a four microphone end-fire array with 8 cm inter-microphone distance in a diffuse noise field ($\phi_{ss} = 0$). Figure 3.2 shows the attenuation

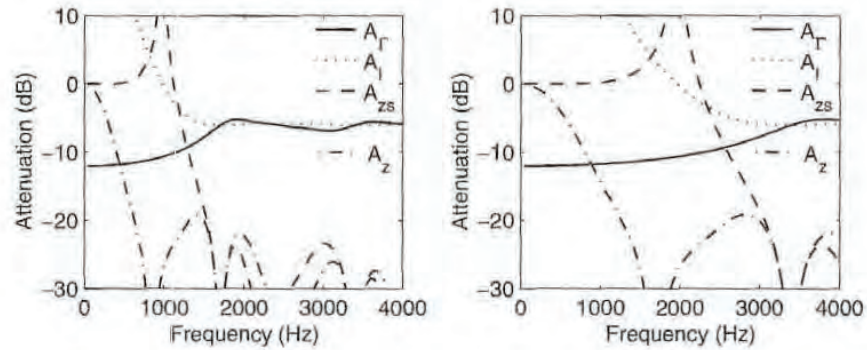


Fig. 3.2. Theoretical noise attenuation of an end-fire array for a diffuse noise field. Left: delay and sum beamformer coefficients. Right: superdirective coefficients.

factors A_r and A_l of the beamformer and the noise attenuation A_{post} of the post-filter (3.37). The left part depicts the attenuation for delay and sum beamformer coefficients ($\mathbf{f} = \mathbf{w} = 1/N$) and the right part depicts the attenuation for superdirective coefficients ($\mathbf{f} = \mathbf{w}_{\text{MVDR}}$).

The performance of the delay and sum beamformer and the respective post-filter is poor at low frequencies. At high frequencies the coherence of a diffuse noise field is approaching zero. Therefore, A_r is close to A_l and both post-filters perform nearly optimally.

The superdirective beamformer performs particularly well at low frequencies. The respective post-filter, however, does not benefit from using superdirective coefficients. The performance gets even worse at low frequencies and the transfer function is infinite at the frequency where A_l crosses 0 dB.

3.3.3 A New Post-Filter Algorithm

To derive an improved algorithm we note that in all cases the subtraction of the white noise attenuation A_l in (3.37) is causing the trouble. It reduces the performance for superdirective coefficients and is responsible for negative or infinite post-filters. Our straightforward approach for solving these problems is to replace the difference $A_r - A_l$ with A_r , since A_r is the parameter that is actually minimized by the design of the MVDR beamformer. Substituting $A_l = 0$ in (3.37) results in

$$H_{\text{apab}} = \frac{\phi_{ss}}{\phi_{ss} + \phi_{vv}} + \frac{\phi_{vv} A_r}{\phi_{ss} + \phi_{vv}} = \frac{\phi_{yy}}{\phi_{xx}}. \quad (3.40)$$

This new algorithm can be implemented easily by estimating the ratio of the output power ϕ_{yy} and the input power ϕ_{xx} of the beamformer for all sub-bands, where ϕ_{xx} is the power of the microphone closest to the desired source

or, alternatively, the average input power of the beamformer (see Fig. 3.3). This design is compatible with superdirective coefficients, is always positive, and provides good performance for low frequency noise. However, the new transfer function still approximates a Wiener filter for the input signal. It does not take into account that the noise has already been reduced by the MVDR beamformer. In order to correct this behavior, we may apply the following function to (3.40)

$$g(H, A) = \frac{H}{H + (1 - H) A} \quad (3.41)$$

This transforms the Wiener filter for the input to a Wiener filter for the output of the beamformer:

$$g\left(\frac{\phi_{ss}}{\phi_{ss} + \phi_{vv}}, A_f\right) = \frac{\phi_{ss}}{\phi_{ss} + \phi_{vv} A_f} \quad (3.42)$$

Since A_f is usually unknown, we may implement (3.40) directly and call this algorithm Adaptive Post-Filter for an Arbitrary Beamformer (APAB).

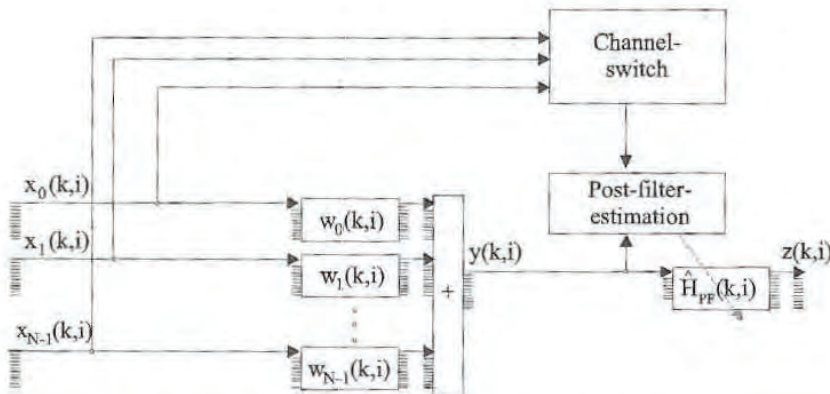


Fig. 3.3. Block diagram of the adaptive post-filter for an arbitrary beamformer (APAB).

3.4 Performance Evaluation

It is difficult to obtain reliable speech quality measures for the performance evaluation of noise reduction units. Subjective listening tests reach statistical significance only for a large number of trained listeners and are expensive and time-consuming. On the other hand, objective measures are often less

sensitive than the human auditory system to artifacts such as musical tones. Therefore, we did not rely exclusively on objective measures to optimize the noise reduction algorithms. Accompanying informal listening tests were conducted to validate the objective results.

3.4.1 Simulation System

Our simulation system consists of three parts: A signal generation module, the device or algorithm under test (DUT), and the evaluation unit. In a first step, clean speech $s(k)$ and a pure noise signal $v(k)$ are convolved with room impulse responses (RIR) that are computed using the image method of Allen and Berkley [41]. In Fig. 3.4, we show the room configuration used. Noise is added to the computed multi-channel signals to produce a given signal-to-noise ratio (SNR). The resulting noisy signal is fed into the DUT.

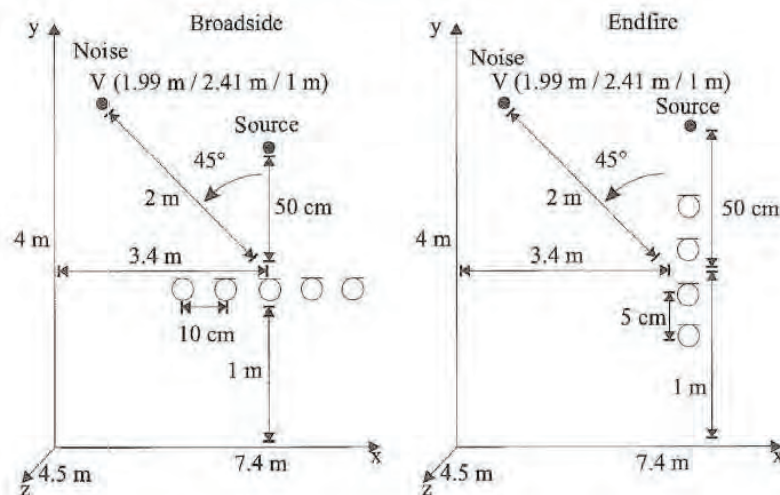


Fig. 3.4. Configuration of the simulated room.

The adaptive coefficients of the algorithm are copied to two slave algorithms which process speech or noise only. Thus, we have access to the processed speech signal $y_s(k)$, the processed noise signal $y_v(k)$, and a processed sum $y_{s+v}(k)$. Finally, these three output signals and the input signals are used in the evaluation unit to compute several speech quality measures. See Fig. 3.5 for a graphical description of the complete system.

3.4.2 Objective Measures

We are using three different quantities to obtain objective information about the tested algorithm. The first one is the segmental signal-to-noise ratio en-

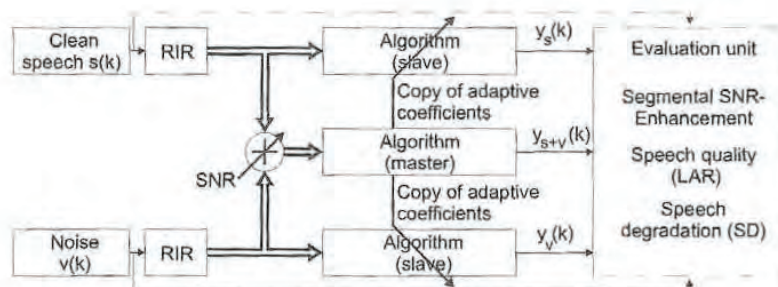


Fig. 3.5. Graphical description of the complete simulation system.

hancement (SNRE):

$$SNRE(l) = SNR_{in}(l) - SNR_{out}(l). \quad (3.43)$$

The segmental SNR is computed from consecutive samples with block-length $B = 256$ at a sampling frequency of 8 kHz:

$$SNR_{in}(l) = 10 \cdot \log_{10} \frac{\sum_{k=lB+1}^{(l+1)B} s^2(k)}{\sum_{k=lB+1}^{(l+1)B} v^2(k)}, \quad (3.44)$$

$$SNR_{out}(l) = 10 \cdot \log_{10} \frac{\sum_{k=lB+1}^{(l+1)B} y_s^2(k)}{\sum_{k=lB+1}^{(l+1)B} y_v^2(k)}. \quad (3.45)$$

The second objective measure is the log-area-ratio distance (LAR) which has been tested with good results in [42]. This quantity can be computed in three steps:

1. Estimate the PARTIAL CORrelation coefficients (PARCOR) of a block of samples. The block-size should be small enough to hold the assumption of stationarity but large enough to reduce bias and variance of the estimated values. A good choice is a block-size of 256 for a model order of $P = 12$. An algorithm for estimating PARCOR coefficients is the well-known Burg-algorithm [35].
2. Determine the area-coefficients by

$$a(p, l) = \frac{1 + k(p, l)}{1 - k(p, l)} \quad \forall \quad 1 \leq p \leq 12 \quad (3.46)$$

where $k(p, l)$ is the p th PARCOR coefficient of block l .

3. Compute the LAR of block l

$$LAR(l) = \sum_{p=1}^{12} 20 \log_{10} \left| \frac{g_s(p, l)}{g_{y_{s+v}}(p, l)} \right|. \quad (3.47)$$

The final quantity we use is a speech degradation measure, which can be defined by the LAR of the input and the output speech signals only

$$SD(l) = \sum_{p=1}^{12} 20 \log_{10} \left| \frac{g_s(p, l)}{g_{y_s}(p, l)} \right|. \quad (3.48)$$

It includes the room reverberation, the signal distortion caused by the tested algorithm, and the dereverberation features of the tested algorithm only. Finally, the average of all blocks containing speech is computed.

3.4.3 Simulation Results

The described simulation system was used to evaluate the performance of four different post-filter algorithms:

1. Zel88: The algorithm by Zelinski in the frequency-domain implementation [21].
2. Sim92: The algorithm by Simmer described in [13].
3. APAB: The adaptive post-filter for an arbitrary beamformer, described in section 3.3 with a constrained MVDR-beamformer designed for an isotropic noise field in three dimensions (superdirective beamformer). The constraining parameter is set to $\mu = 0.01$ (see Chapter 2).
4. APES: The adaptive post-filter extension for superdirective beamformers [32].

For comparison, we include the results of the case in which no algorithm is used (No NR).

The speech sample we used is the sentence "I am now speaking to you from a distance of 50 cm from the microphone" spoken by an adult male. The length of this file leads to 98 blocks containing speech. The noise file was white Gaussian noise used in order to give technical results which can be reproduced by other researchers. The input SNR was computed only for blocks containing speech by using the segmental SNR.

In the first experiment, the broadside array shown on the left side of Fig. 3.4 is examined. Figure 3.6 depicts the results for the SNRE. The left side shows the dependence on the input-SNR if the reverberation time is set to $\tau_{60} = 300$ ms. The right figure shows the results for SNR=5 dB as a function of the reverberation time. This provides information on the behavior of the algorithms for different spatial conditions. The noise-field is coherent for low reverberation time and approximately diffuse for high values.

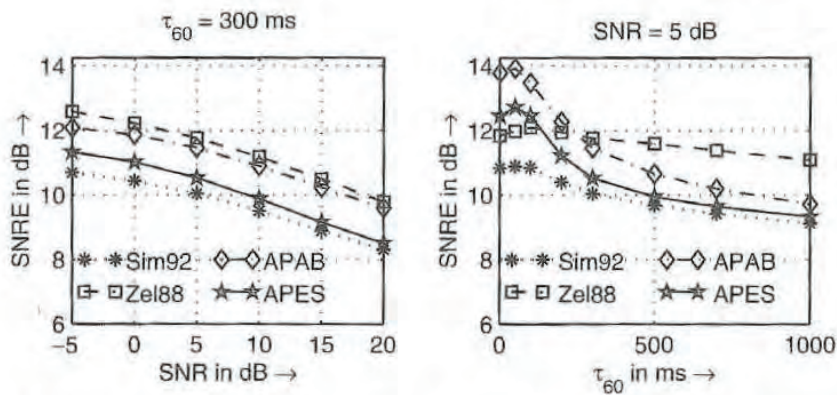


Fig. 3.6. Left: SNRE vs. input-SNR. Right: SNRE vs. reverberation time τ_{60} (Broadside).

Although not optimal the Zel88 algorithm performs quite well, especially for high reverberation times where it provides the best results of all tested algorithms (if only the SNRE is considered). At low reverberation times APAB and APES can benefit from the better suppression at low frequencies by using a superdirective beamformer instead of a standard delay and sum beamformer.

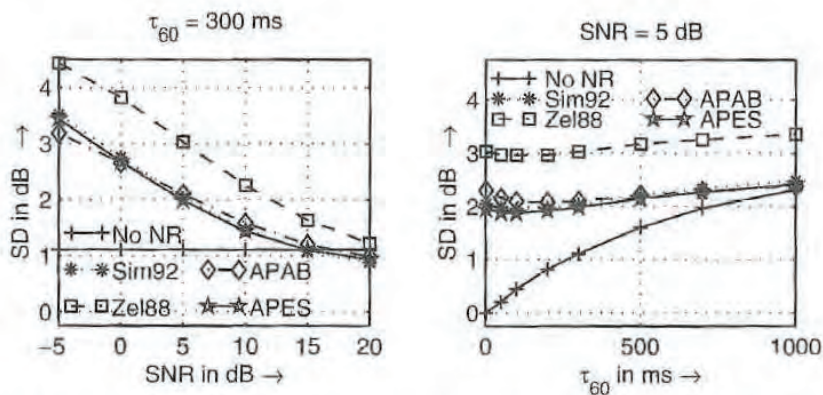


Fig. 3.7. Left: SD vs. input-SNR. Right: SD vs. reverberation time τ_{60} (Broadside).

If we take into account the next two measures shown in Fig. 3.7 and 3.8, which describe the performance in terms of speech quality, the results are different. All algorithms enhance the speech quality in comparison to the

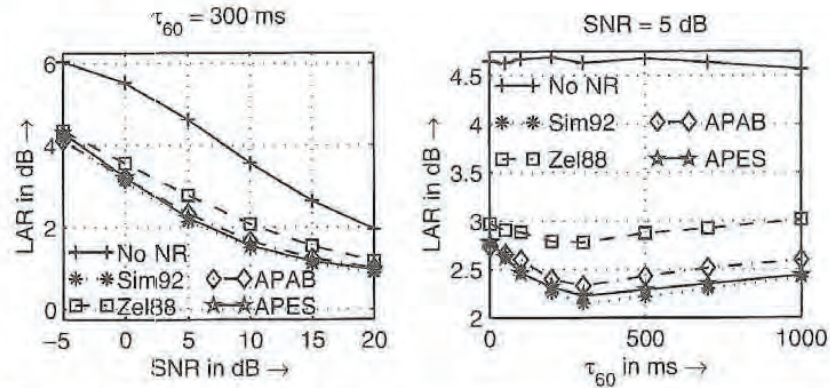


Fig. 3.8. Left: LAR vs. input-SNR. Right: LAR vs. reverberation time τ_{60} (Broad-side).

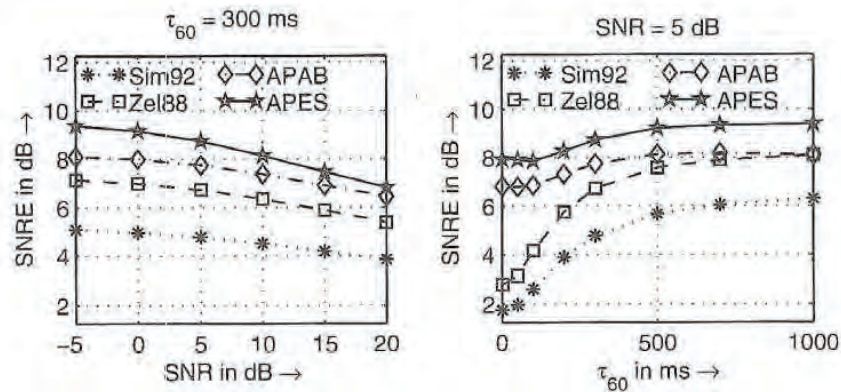


Fig. 3.9. Left: SNRE vs. input-SNR. Right: SNRE vs. reverberation time τ_{60} (End-fire).

unprocessed input signal¹. However, the algorithm with the highest SNRE does not produce the best LAR. A closer look at Fig. 3.7 explains this behavior. Since these figures show the speech degradation only, the non-processed signal is constant versus the SNR and reduces to zero if no reverberation is added to the speech signal. The algorithms cause signal distortion at low SNR and the algorithm with the highest performance in SNRE induces the largest distortion, whereas APAB and APES provide the best speech quality (LAR). At very good conditions (SNR > 15 dB), these algorithms are able to suppress reverberation without introducing speech degradation. The lack of artifacts was corroborated through informal listening tests.

¹ Smaller values indicate better quality.

In a second experiment (right side of Fig. 3.4), we changed the orientation of the array and the inter-microphone distance. Additionally, only four microphones were used to reduce the array size. In Fig. 3.9 the SNRE results of the simulation are shown. The performance of the Sim92 and Zel88 algorithms degrades drastically, since the inherent delay and sum beamformer does not perform well at low frequencies due to the small array size. On the other hand, APAB and APES perform well under all conditions. The SNRE for APES at high reverberation time is close to the result for the broadside-experiment although the number of microphones is reduced. Thus, we conclude that end-fire steering is preferable for this algorithm.

3.5 Conclusion

Wiener post-filtering of the output signal of an MVDR beamformer provides an optimum MMSE solution for signal enhancement. A large number of published algorithms for post-filter estimation are based on the assumption of spatially uncorrelated noise. This assumption leads to post-filtering algorithms with suboptimal performance in coherent and diffuse noise fields. In this chapter we presented a new algorithm which performs considerably better in correlated noise fields by using the gain of an arbitrary array. Small size end-fire arrays comprising an MVDR beamformer and optimized post-filters showed the best performance in our simulations.

References

1. R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*, John Wiley and Sons, New York, 1980.
2. L. Danilenko, *Binaurales Hören im nichtstationären diffusen Schallfeld*, PhD thesis, RWTH Aachen, Aachen, Germany, 1968.
3. G. von Békésy, *Experiments in Hearing*, McGraw-Hill, New York, 1960.
4. S. Gierl, *Geräuschreduktion bei Sprachübertragung mit Hilfe von Mikrofonarraysystemen*, PhD thesis, Universität Karlsruhe, Karlsruhe, Germany, 1990.
5. S. Gierl, "Noise reduction for speech input systems using an adaptive microphone-array", in *Int. Symp. Automotive Tech. and Automation (ISATA)*, Florence, Italy, May 1990, pp. 517-524.
6. H.-Y. Kim, F. Asano, Y. Suzuki, and T. Sone, "Speech enhancement based on short-time spectral amplitude estimation with two-channel beamformer", *IEICE Trans. Fundament.*, vol. E79-A, no. 12, pp. 2151-2158, Dec. 1996.
7. M. Dörbecker and S. Ernst, "Combination of two-channel spectral subtraction and adaptive Wiener post-filtering for noise reduction and dereverberation", in *Proc. EURASIP European Signal Proc. Conf. (EUSIPCO)*, Trieste, Italy, Sept. 1996.
8. K. Kroschel, A. Czyzewski, M. Ihle, and M. Kuropatwinski, "Adaptive noise cancellation of speech signals in a noise reduction system based on a microphone array", in *102nd Audio Eng. Soc. Conv., preprint 4450*, Munich, Germany, Mar. 1997.

9. J. B. Allen, D. A. Berkley, and J. Blauert, "Multimicrophone signal-processing technique to remove room reverberation from speech signals", *J. Acoust. Soc. Amer.*, vol. 62, no. 4, pp. 912-915, Oct. 1977.
10. Y. Kaneda and M. Tohyama, "Noise suppression signal processing using 2-point received signals", *Electron. Communicat. Japan*, vol. 67-A, no. 12, pp. 19-28, Apr. 1984.
11. R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms", in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*, New York, USA, Apr. 1988, pp. 2578-2581.
12. R. Zelinski, "Noise reduction based on microphone array with LMS adaptive post-filtering", *Electron. Lett.*, vol. 26, no. 24, pp. 2036-2037, Nov. 1990.
13. K. U. Simmer and A. Wasiljeff, "Adaptive microphone arrays for noise suppression in the frequency domain", in *Second Cost 229 Workshop Adapt. Alg. Communicat.*, Bordeaux, France, Oct. 1992, pp. 185-194.
14. Y. Mahieux and C. Marro, "Comparison of dereverberation techniques for videoconferencing applications", in *100th Audio Eng. Soc. Conv., preprint 4231*, Copenhagen, Denmark, May 1996.
15. C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering", *IEEE Trans. Speech and Audio Processing*, vol. 6, no. 3, pp. 240-259, May 1998.
16. R. Le Bouquin and G. Faucon, "On using the coherence function for noise reduction", in *Proc. EURASIP European Signal Proc. Conf. (EUSIPCO)*, Barcelona, Spain, Sept. 1990, pp. 1103-1106.
17. R. Le Bouquin and G. Faucon, "Study of a noise cancellation system based on the coherence function", in *Proc. EURASIP European Signal Proc. Conf. (EUSIPCO)*, Brussels, Belgium, Aug. 1992, pp. 1633-1636.
18. G. Faucon and R. Le Bouquin-Jeannes, "Optimization of speech enhancement techniques coping with uncorrelated and correlated noise", in *Proc. IEEE Int. Conf. on Communication Technology (ICCT-96)*, Beijing, China, May 1996, pp. 416-419.
19. R. Le Bouquin-Jeannes, A. A. Azirani, and G. Faucon, "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator", *IEEE Trans. Speech and Audio Processing*, vol. 5, no. 5, pp. 484-487, Sept. 1997.
20. P. Kuczynski, *Mehrkanal-Analyse von Sprachsignalen zur adaptiven Störunterdrückung*, PhD thesis, University of Bremen, Shaker Verlag, Aachen, Germany, Sept. 1995.
21. K. U. Simmer, P. Kuczynski, and A. Wasiljeff, "Time delay compensation for adaptive multichannel speech enhancement systems", in *Proc. Int. Symp. Signals, Syst. Electron. ISSSE-92*, Paris, France, Sept. 1992, pp. 660-663.
22. M. Drews and M. Streckfuß, "Multi channel speech enhancement using an adaptive post-filter with channel selection and auditory constraints", in *Proc. Int. Workshop Acoust. Echo and Noise Control*, London, UK, Sept. 1997, pp. 77-80.
23. M. Drews, *Mikrofonarrays und mehrkanalige Signalverarbeitung zur Verbesserung gestörter Sprache*, PhD thesis, Technische Universität Berlin, Berlin, Germany, 1999.

24. K. U. Simmer, S. Fischer, and A. Wasiljeff, "Suppression of coherent and incoherent noise using a microphone array", *Annals of Telecommunications*, vol. 49, no. 7/8, pp. 439-446, July 1994.
25. S. Fischer and K. U. Simmer, "Beamforming microphone arrays for speech acquisition in noisy environments", *Speech Commun.*, vol. 20, no. 3-4, pp. 215-227, Dec. 1996.
26. A. Hussain, D.R. Campbell, and T.J. Moir, "A new metric for selecting sub-band processing in adaptive speech enhancement systems", in *Proc. ESCA European Conf. Speech Communicat. Tech. (EUROSPEECH)*, Rhodes, Greece, Sept. 1997, pp. 1489-1492.
27. R. Atay, E. Mandridake, D. Bastard, and M. Najim, "Spatial coherence exploitation which yields non-stationary noise reduction in subband domain", in *Proc. EURASIP European Signal Proc. Conf. (EUSIPCO)*, Rhodes, Greece, Sept. 1998, pp. 1489-1492.
28. J. Gonzales-Rodríguez, J. L. Sanchez-Bote, and J. Ortega-Garcia, "Speech dereverberation and noise reduction with a combined microphone array approach", in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*, Istanbul, Turkey, Apr. 2000, pp. 1489-1492.
29. D. Mahmoudi and A. Drygajlo, "Combined Wiener and coherence filtering in wavelet domain for microphone array speech enhancement", in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*, Atlanta, USA, May 1998, pp. 1489-1492.
30. D. Mahmoudi and A. Drygajlo, "Wavelet transform based coherence function for multi-channel speech enhancement", in *Proc. EURASIP European Signal Proc. Conf. (EUSIPCO)*, Rhodes, Greece, Sept. 1998, pp. 1489-1492.
31. J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "An alternative implementation of the superdirective beamformer", in *Proc. IEEE Workshop Applicat. Signal Processing to Audio and Acoust.*, New Paltz, New York, Oct. 1999, pp. 7-10.
32. J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "Multi-microphone noise reduction by post-filter and superdirective beamformer", in *Proc. Int. Workshop Acoust. Echo and Noise Control*, Pocono Manor, USA, Sept. 1999, pp. 100-103.
33. I. A. McCowan, C. Marro, and L. Mauuary, "Robust speech recognition using near-field superdirective beamforming with post-filtering", in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*, Istanbul, Turkey, Apr. 2000.
34. J. P. Burg, "Three-dimensional filtering with an array of seismometers", *Geophysics*, vol. 29, no. 5, pp. 693-713, Oct. 1964.
35. S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 3rd edition, 1996.
36. L. W. Brooks and I. S. Reed, "Equivalence of the likelihood ratio processor, the maximum signal-to-noise ratio filter, and the Wiener filter", *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-8, no. 5, pp. 690-692, Sept. 1972.
37. D. J. Edelblute, J. M. Fisk, and G. L. Kinnison, "Criteria for optimum-signal-detection theory for arrays", *J. Acoust. Soc. Amer.*, vol. 41, no. 1, pp. 199-205, Jan. 1967.
38. N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*, Wiley, New York, 1949.
39. W. Kellermann, "Analysis and design of multirate systems for cancelling of acoustic echoes", in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*, Munich, Germany, Apr. 1988, pp. 2570-2573.

40. C. Marro, Y. Mahieux, and K. U. Simmer, "Performance of adaptive dereverberation techniques using directivity controlled arrays", in *Proc. EURASIP European Signal Proc. Conf. (EUSIPCO)*, Trieste, Italy, Sept. 1996, pp. 1127-1130.
41. J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics", *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943-950, Apr. 1979.
42. S. R. Quakenbusch, T. P. Barnwell, and B. A. Clemens, *Objective Measures of Speech Quality*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

5 Robust Adaptive Beamforming

Osamu Hoshuyama and Akihiko Sugiyama

NEC Media Research Labs, Kawasaki, Japan

Abstract. This chapter presents robust adaptive beamforming techniques designed specifically for microphone array applications. The basics of adaptive beamformers are first reviewed with the Griffiths-Jim beamformer (GJBF). Its robustness problems caused by steering vector errors are then discussed with some conventionally proposed robust beamformers. As better solutions to the conventional robust beamformers, GJBFs with an adaptive blocking matrix are presented in the form of a microphone array. Simulation results and real-time evaluation data show that a new robust adaptive microphone array achieves improved robustness against steering vector errors. Good sound quality of the output signal is also confirmed by a subjective evaluation.

5.1 Introduction

Beamforming is a technique which extracts the desired signal contaminated by interference based on directivity, i.e. spatial signal selectivity [1]–[5]. This extraction is performed by processing the signals obtained by multiple sensors such as microphones, antennas, and sonar transducers located at different positions in the space. The principle of beamforming has been known for a long time. Because of the vast amount of necessary signal processing, most research and development effort has been focused on geological investigations and sonar, which can afford a higher cost. With the advent of LSI technology, the required amount of signal processing has become relatively small. As a result, a variety of research projects where acoustic beamforming is applied to consumer-oriented applications, have been carried out [6].

Applications of beamforming include microphone arrays for speech enhancement. The goal of speech enhancement is to remove undesirable signals such as noise and reverberation. Among research areas in the field of speech enhancement are teleconferencing [7]–[8], hands-free telephones [9]–[11], hearing aids [12]–[21], speech recognition [22]–[23], intelligibility improvement [24]–[25], and acoustic measurement [26].

Beamforming can be considered as multidimensional signal processing in space and time. Ideal conditions assumed in most theoretical discussions are not always maintained. The target DOA (direction of arrival), which is assumed to be stable, does change with the movement of the speaker. The sensor gains, which are assumed uniform, exhibit significant distribution. As a result, the performance obtained by beamforming may not be as good as

is expected. Therefore, robustness against steering-vector errors caused by these array imperfections are becoming more and more important.

This chapter presents robust adaptive beamforming with the emphasis on microphone arrays as its application. In Section 2, the basics of adaptive beamformers are reviewed with the Griffiths-Jim beamformer (GJBF). Section 3 discusses robustness problems in the GJBF. Robust adaptive microphone arrays as solutions to the robustness problem are presented in Section 4. Finally in Section 5 evaluations of a robust adaptive microphone array are presented with simulation results and real-time evaluation data.

5.2 Adaptive Beamformers

A beamformer which adaptively forms its directivity pattern is called an adaptive beamformer. It simultaneously performs beam steering and null steering. In most acoustic beamformers, however, only null steering is performed with an assumption that the target DOA is known *a priori*. Due to adaptive processing, deep nulls can be developed even when errors in the propagation model exist. As a result, adaptive beamformers naturally exhibit higher interference suppression capability than its fixed counterpart. Among various adaptive beamformers, the Griffiths-Jim beamformer (GJBF) [27], or the generalized sidelobe canceler, is most widely known.

Figure 5.1 depicts the structure of the GJBF. It comprises a fixed beamformer (FBF), a multiple-input canceler (MC), and a blocking matrix (BM). The FBF is designed to form a beam in the look direction so that the target signal is passed and all other signals are attenuated. On the contrary, the BM forms a null in the look direction so that the target signal is suppressed and all other signals are passed through.

The simplest structure for the BM is a delay-and-subtract beamformer which was described in the previous section. Assuming a look direction perpendicular to the array surface, no delay element is necessary. Thus, a set of subtracters which take the difference between the signals at the adjacent microphones can be used as a BM. This structure is actually the one shown in Fig. 5.1. The BM was named after its function, which is to block the target signal.

The MC is composed of multiple adaptive filters each of which is driven by a BM output, $z_n(k)$ ($n=0, 1, \dots, N-2$). The BM outputs, $z_n(k)$, contain all the signal components except that in the look direction. Based on these signals, the adaptive filters generate replicas of components correlated with the interferences. All the replicas are subtracted from a delayed output signal, $b(k-L_1)$,¹ of the fixed beamformer which has an enhanced target signal component. As a result, in the subtracter output $y(k)$, the target signal is

¹ The L_1 -sample delay is introduced to compensate for the signal processing delay in the BM and the MC.

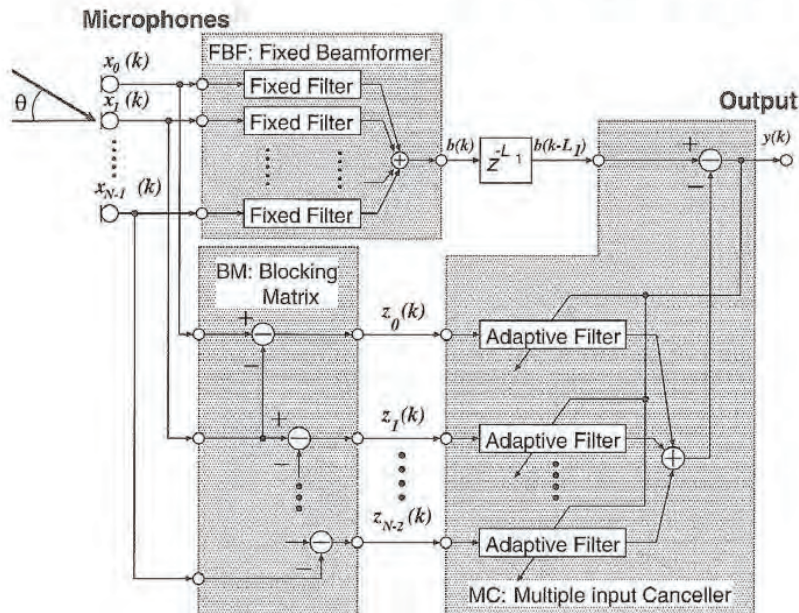


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceller (MC), and a blocking matrix (BM).

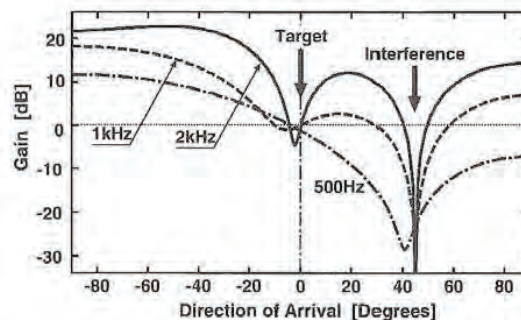


Fig. 5.2. Example directivity pattern of the Griffiths-Jim beamformer.

enhanced and undesirable signals such as ambient noise and interferences are suppressed.

The GJBF can be considered as an adaptive noise canceller with multiple reference signals, each of which is preprocessed by the BM. In an adaptive noise canceller, the auxiliary microphone is located close to the noise source to obtain a best possible noise reference. On the other hand, the BM in the

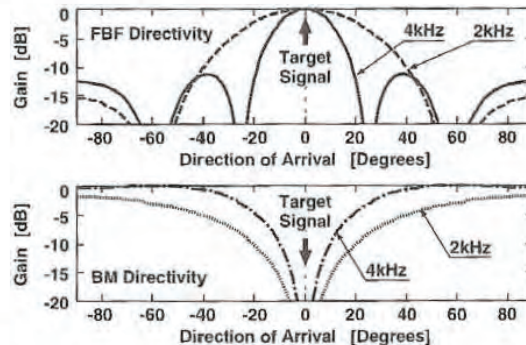


Fig. 5.3. Directivity pattern of a fixed beamformer (FBF) and a blocking matrix (BM).

GJBF extracts, with its directivity, the signal components correlated with the noise.

Figure 5.2 depicts an example directivity pattern obtained by the GJBF. In the direction of the target signal, almost constant gains close to 0 dB are obtained over a wide range of frequencies. On the contrary, in the direction of the interference, a deep null is formed. Although the directivity has frequency dependency, target signal extraction and interference suppression are simultaneously achieved.

With the same microphone array, adaptive beamformers generally achieve better interference suppression than fixed beamformers. This is because nulls are sharper than beams. The effect is demonstrated in Fig. 5.3, where directivity patterns of the FBF and the BM are illustrated. The null of the BM and the main lobe (beam) of the FBF are located in the target direction. It is also clear from the figure that they are orthogonal to each other. The BM in Fig. 5.1 has a simple delay-and-sum structure, however, a filter-and-sum beamformer [28,29] may also be employed.

5.3 Robustness Problem in the GJBF

The GJBF suffers from target-signal cancellation due to steering-vector errors, which is caused by an undesirable phase difference between $x_n(k)$ and $x_{n+1}(k)$ for the target. A phase error leads to target signal leakage into the BM output signal. As a result, blocking of the target becomes incomplete, which results in target signal cancellation at the microphone array output.

Steering-vector errors are inevitable because the propagation model does not always reflect the nonstationary physical environment. The steering vector is sensitive to errors in the microphone positions, those in the microphone characteristics, and those in the assumed target DOA (which is also known

as the look direction). For teleconferencing and hands-free communication in the car, the error in the assumed target DOA is the dominant factor.

A variety of techniques to reduce target-signal cancellation have been proposed mainly in the field of antennas and radars. The beamformers with these techniques are called robust beamformers. Typical approaches are reduction of the target-signal leakage in the BM outputs and restraint of coefficient growth in the MC. The former can be considered as a direct approach which reduces the target leakage in the BM output. The latter takes the form of an indirect approach. Even if there is target leakage in the BM output used as the MC input, the MC tries to minimize its influence.

Techniques to reduce target-signal leakage include:

- Target Tracking: The look direction is steered to the continuously estimated DOA [30]–[32]. Mistracking to interference may occur in the absence of a target signal.
- Multiple Constraints in BM: Multiple constraints are imposed on the BM so that signals from multiple DOAs are eliminated [33]. To compensate for the loss of the degrees of freedom for interference reduction with a large DOA error, additional microphones are needed.
- Constrained Gradient for Look-Direction Sensitivity: Gradient of the sensitivity at the look direction is constrained for a smaller variance of the sensitivity [34,35]. For a large error, loss in the degrees of freedom is inevitable.
- Improved Spatial Filter: A carefully designed spatial filter is used to eliminate the target signal [28]. Such a spatial filter also loses degrees of freedom.

Techniques that attempt to restrain excess coefficient growth include:

- Noise Injection: Artificially-generated noise is added to the error signal used to update the adaptive filters in the MC. This noise causes errors in the adaptive filter coefficients, preventing tap coefficients from growing excessively [36]. A higher noise level is needed to allow a larger look-direction error, resulting in less interference suppression.
- Norm Constraint: The coefficient norm of the adaptive filters in the MC is constrained by an inequality to suppress the growth of the tap coefficients [37]. In spite of its simplicity, interference reduction is degraded when the constraint is designed to allow a large error.
- Leaky Adaptive Algorithm: A leaky coefficient adaptation algorithm such as leaky LMS is used for the adaptive filters in the MC [28]. A large leakage is needed to allow a large look-direction error, leading to degraded interference-reduction.
- Adaptation Mode Control: Coefficient adaptation in the MC is controlled so that adaptation is carried out only when there is no target signal [38]. If there is no target signal when coefficients are adapted in the MC, the target leakage, if any, will have no effect on the performance of the beamformer.

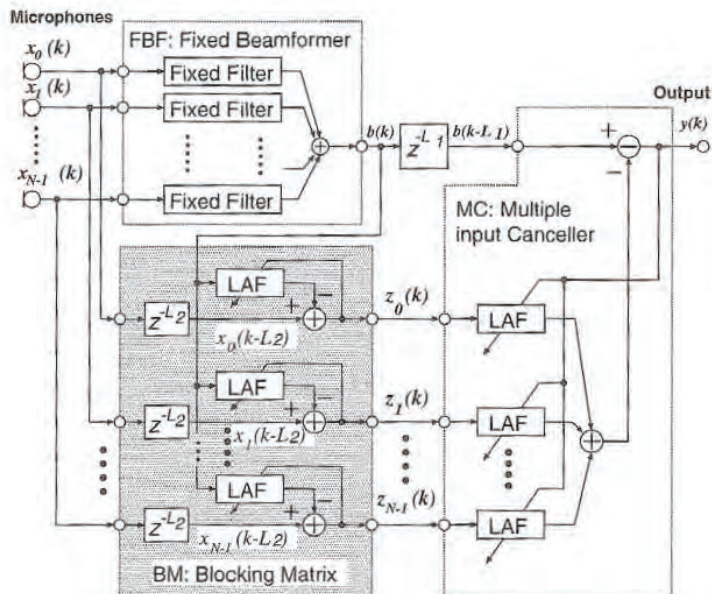


Fig. 5.4. GJBF with a LAF-LAF Structure.

These methods have been developed for a small look-direction error, typically less than 10 degrees. In the case of microphone arrays, the variance of the target DOA is typically much larger than in antennas and radar applications. No single conventional technique for robustness is sufficient for microphone arrays with a larger phase errors.

5.4 Robust Adaptive Microphone Arrays — Solutions to Steering-Vector Errors

5.4.1 LAF-LAF Structure

A target-tracking method with leaky adaptive filters (LAF) in the BM was proposed as a solution to target signal cancellation in [39]. It is combined with leaky adaptive filters in the MC [28], thereby called a LAF-LAF structure. Figure 5.4 depicts its block diagram. The leaky adaptive filters in the BM alleviate the influence of phase error, which results in the robustness. This structure can pick up a target signal with little distortion when the error between the actual and the assumed DOAs is not small. It does not need matrix products, and provides easy implementation. The n th output

$z_n(k)$ ($n = 0, 1, \dots, N-1$) of the BM can be obtained as follows:

$$z_n(k) = x_n(k - L_2) - \mathbf{h}_n^T(k) \mathbf{b}(k), \quad (5.1)$$

$$\mathbf{h}_n(k) \triangleq [h_{n,0}(k), h_{n,1}(k), \dots, h_{n,M_1-1}(k)]^T, \quad (5.2)$$

$$\mathbf{b}(k) \triangleq [b(k), b(k-1), \dots, b(k-M_1+1)]^T, \quad (5.3)$$

where $[\cdot]^T$ denotes vector transpose and $x_n(k)$ is the n th microphone signal. L_2 is the number of delay samples for causality, $\mathbf{h}_n(k)$ is the coefficient vector of the n th LAF, and $\mathbf{b}(k)$ is the signal vector consisting of delayed signals of $b(k)$ (which is the FBF output). Each LAF is assumed to have M_1 taps. The adaptation by the normalized LMS (NLMS) algorithm [40] is described as follows:

$$\mathbf{h}_n(k+1) = \mathbf{h}_n(k) - \delta \cdot \mathbf{h}_n(k) + \alpha \frac{z_n(k)}{\mathbf{b}(k)^T \mathbf{b}(k)} \mathbf{b}(k), \quad (5.4)$$

where α is the step size for the adaptation algorithm, and $\delta, 0 \leq \delta \leq 1$, is the leakage constant.

LAFs are also used in the MC for enhancing the robustness obtained in the BM. The LAFs prevent undesirable target-signal cancellation caused by the remaining correlation with the target signal in $z_n(k)$. Tap coefficient vectors $\mathbf{w}_n(k)$ of the MC have M_2 taps and are updated by an equation similar to (5.4), where $\mathbf{h}_n$, $\mathbf{b}$, and $z_n(k)$ are replaced with $\mathbf{w}_n$, $\mathbf{z}_n$, and $y(k)$, respectively. The leakage constant δ and the step size α are replaced with γ and β respectively, and may take different values from those in (5.4).

With the LAFs in the BM, the LAF-LAF structure adaptively controls the look direction, which is fixed in the GJBF. Due to robustness by the adaptive control of the look direction, the LAF-LAF structure does not lose degrees of freedom for interference reduction. Thus, no additional microphones are required compared to the conventional robust beamformers. Target signal leakage in the BM is sufficiently small to use a minimum leakage constant γ in the MC even for a large look-direction error. Such a value of γ leads to a higher interference-reduction performance in the MC. The output of the LAFs are summed and subtracted from an L_1 sample delayed version of the FBF output to generate the microphone array output $y(k)$.

The width of the allowable DOA for the target is determined by the leaky constants and the step sizes in both the BM and the MC. Generally, smaller values of these parameters make the allowable target DOA wider. The allowable DOA width for the target is not a simple function of the parameters, however, and is not easy to prescribe. It is reported [39] that the interference is attenuated by more than 18 dB when it is designed, through simulations, to allow 20 degree directional error. Tracking may not be sufficiently precise for a large tracking range. Thus, there is a trade-off between the degree of target-signal cancellation and the amount of interference suppression.

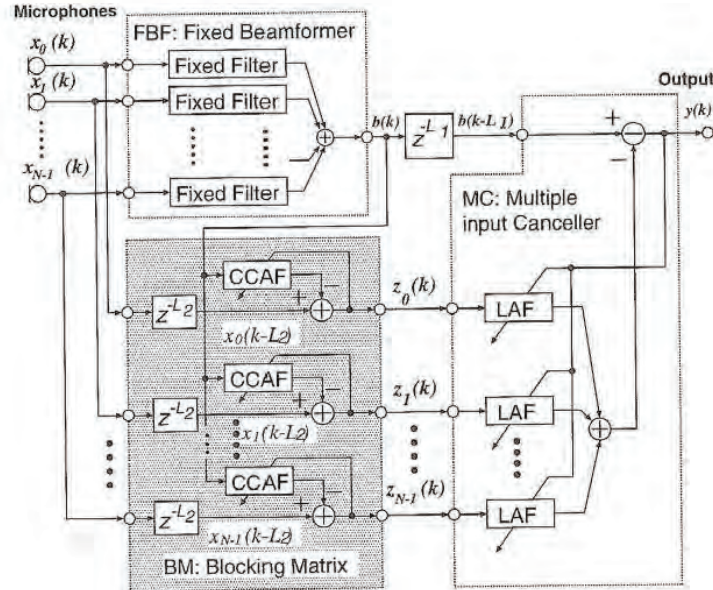


Fig. 5.5. GJBF with a CCAF-LAF Structure.

5.4.2 CCAF-LAF Structure

A more effective solution is to use coefficient-constrained adaptive filters (CCAFs) in the BM [41,42]. When combined with leaky adaptive filters in the MC as depicted in Fig. 5.5, the result is called a CCAF-LAF structure. CCAFs behave like adaptive noise cancelers. The input signal to each CCAF is the output of the FBF, and the output of the CCAF is subtracted from the delayed microphone signal. The CCAF coefficient vectors $\mathbf{h}_n(k)$ are adapted with constraints. Adaptation by the NLMS algorithm is described as follows:

$$\mathbf{h}'_n(k+1) = \mathbf{h}_n(k) + \alpha \frac{z_n(k)}{\mathbf{b}(k)^T \mathbf{b}(k)} \mathbf{b}(k), \quad (5.5)$$

$$\mathbf{h}_n(k+1) = \begin{cases} \phi_n, & \text{for } \mathbf{h}'_n(k+1) > \phi_n \\ \psi_n, & \text{for } \mathbf{h}'_n(k+1) < \psi_n \\ \mathbf{h}'_n(k+1), & \text{otherwise.} \end{cases} \quad (5.6)$$

$$\phi_n \triangleq [\phi_{n,0}, \phi_{n,1}, \dots, \phi_{n,M_1-1}]^T, \quad (5.7)$$

$$\psi_n \triangleq [\psi_{n,0}, \psi_{n,1}, \dots, \psi_{n,M_1-1}]^T, \quad (5.8)$$

where each CCAF is assumed to have M_1 taps and $\mathbf{h}'_n(k+1)$ is a temporal coefficient vector for limiting functions. ϕ_n and ψ_n are the upper and lower

bounds for coefficients. In the output signal $z_n(k)$, the components correlated with $\mathbf{b}(k)$ are cancelled by the CCAFs.

Each coefficient of the CCAFs is constrained based on the fact that filter coefficients for target-signal minimization vary significantly with the target DOA. An example of filter-coefficient variation is illustrated in Fig. 5.6. By the design of the constrained regions of the CCAF coefficients, the maximum allowable look-direction error can be specified. For example, when the CCAF coefficients are constrained in the hatched region in Fig. 5.6, up to 20° error in look direction could be allowed. Only the signal that arrives from a DOA in the limited DOA region is minimized at the outputs of the BM and remains at the output of the MC. If no interference exists in the region, which is common with microphone arrays, no mistracking occurs. For details on the design of upper and lower bounds, refer to [42].

Figure 5.7 illustrates a qualitative comparison between the LAF and the CCAF with respect to look-direction error and coefficient error from the optimum for signal blocking. Both the CCAF and the LAF give error characteristics approximating the ideal nonlinearity for target tracking. However, the coefficient error of the CCAF is a better approximation to the ideal nonlinearity than that of the LAF as shown by Fig. 5.7. The coefficient error of the CCAF becomes effective only when the look-direction error exceeds the threshold, otherwise it has no effect. On the other hand, the coefficient error of the LAF varies continuously with the look-direction error. Therefore, the CCAF leads to precise target tracking, which results in sharper spatial selectivity and less target-signal cancellation.

5.4.3 CCAF-NCAF Structure

It is possible to combine the BM with CCAFs [42] and the MC with norm-constrained adaptive filters (NCAFs) [37]. This is a CCAF-NCAF structure [43]. NCAFs subtract from $b(k - L_1)$ the components correlated with $z_n(k)$ ($n = 0, \dots, N - 1$). Let M_2 be the number of taps in each NCAF, and let $\mathbf{w}_n(k)$ and $\mathbf{z}_n(k)$ be the coefficient vector and the signal vector of the n th NCAF, respectively. The signal processing in the MC is described by

$$y(k) = b(k - L_1) - \sum_{n=0}^{N-1} \mathbf{w}_n^T(k) \mathbf{z}_n(k), \quad (5.9)$$

where

$$\mathbf{w}_n(k) \triangleq [w_{n,0}(k), w_{n,1}(k), \dots, w_{n,M_2-1}(k)]^T, \quad (5.10)$$

$$\mathbf{z}_n(k) \triangleq [z_n(k), z_n(k-1), \dots, z_n(k-M_2+1)]^T. \quad (5.11)$$