



uOttawa

L'Université canadienne  
Canada's university

**FACULTÉ DES ÉTUDES SUPÉRIEURES  
ET POSTDOCTORALES**



**uOttawa**

L'Université canadienne  
Canada's university

**FACULTY OF GRADUATE AND  
POSTDOCTORAL STUDIES**

**Christian Bouchard**

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

**Ph.D. (Electrical and Computer Engineering)**

GRADE / DEGREE

**School of Information Technology and Engineering**

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

**Beamforming with Microphone Arrays for Directional Sources**

TITRE DE LA THÈSE / TITLE OF THESIS

**Martin Bouchard**

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

**David Havelock**

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

**Miodrag Bolic**

**Hilmi Dijami**

**Rafik Goubran**

**Vijay Parsa (University of Western  
Ontario)**

**Gary W. Slater**

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

**BEAMFORMING WITH MICROPHONE ARRAYS  
FOR DIRECTIONAL SOURCES**

Christian Bouchard, M.Sc., B.Sc.A.

Thesis submitted to the  
Faculty of Graduate and Postdoctoral Studies  
In partial fulfillment of the requirements  
For the Ph.D. degree in Electrical and Computer Engineering

School of Information Technology and Engineering  
Faculty of Engineering  
University of Ottawa

© Christian Bouchard, Ottawa, Canada, 2010



Library and Archives  
Canada

Published Heritage  
Branch

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque et  
Archives Canada

Direction du  
Patrimoine de l'édition

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
ISBN: 978-0-494-65561-0  
*Our file* *Notre référence*  
ISBN: 978-0-494-65561-0

**NOTICE:**

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

**AVIS:**

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**



## Abstract

Beamforming may be used to improve signal-to-noise ratio, reduce reverberation, cancel interference, or estimate source location. It is done with an array of sensors using a model of the arriving wavefront. Conventionally, a plane wave or point source model is used and this can cause decreased array gain or even total breakdown of beamforming when the source is directional. The directivity of some real world acoustic sources is reviewed in this work. Examples are shown where using a point source model leads to large performance penalties, such as when the array subtends a significant angle about the source. The performance penalty of using a point source model to beamform on the human talker is estimated using phase information available in the literature.

An extension of beamforming applicable to directional sources with unknown directivity is presented. In the thesis, the proposed method is mostly intended for sound pickup of audio sources. The proposed beamforming method defines a set of “sub-beamformers”, each designed to respond to a different spatial mode of the source. The outputs of the individual sub-beamformers are combined in a weighted sum to give an overall output of better quality than that of a conventional (point source) beamformer. It is shown that with appropriate weighting, the optimum array gain can be achieved. In the case that the source happens to be a point source, the method reduces to conventional beamforming.

Methods are demonstrated to estimate the weighted sum, based on the observed data. Simulation and experimentally measured results for a simple directional source and an array of 16 microphones show that the proposed beamforming method provides an array gain of about 11 dB while conventional beamforming, with a value of -4 dB, does not provide any useful gain. Simulation results for simple sources in the presence of diffuse noise, diffuse noise and reverberation, or diffuse noise and an interferer are also presented. The computational complexity of the proposed beamforming method is also compared to that of conventional beamforming.

## **Acknowledgements**

I would like to thank my family for their support and encouragements.

My thanks to my thesis supervisors, Dr. Martin Bouchard and Dr. David Havelock, whose advices on the choice of the topic as well as the guidance during the research and the writing process were greatly appreciated. I am particularly in debt to Dr. Havelock for the numerous technical discussions.

I would also like to thank everybody at the Acoustics and Signal Processing Group of the National Research Council of Canada (NRC) where the research was conducted.

I would like to acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC) and the NRC through its Graduate Student Scholarship Supplement Program (GSSSP).

# Table of Contents

Abstract .....	ii
Acknowledgements .....	iii
Table of Contents .....	iv
List of Figures .....	xi
List of Tables.....	xviii
List of Symbols .....	xxi
 <i>Chapter 1</i>	
Introduction .....	1
1.1 Distributed array and directional sources .....	3
1.2 Proposed beamforming method .....	4
1.3 Some applications of microphone arrays and beamforming.....	5
1.3.1 Sound pickup (sound capture) .....	6
1.3.2 Source tracking .....	6
1.3.3 Spatial differentiation.....	7
1.4 Objective of the method and contributions of the thesis.....	7
1.5 Thesis outline .....	10
 <i>Chapter 2</i>	
Arrays and Beamforming.....	12
2.1 Introduction.....	12
2.2 Beamforming .....	12
2.2.1 Source model .....	14
2.2.2 Near-field, far-field, and conventional beamforming .....	15
2.3 Performance measures .....	16

2.3.1	Array gain and white-noise gain .....	16
2.3.2	Beampattern .....	17
2.4	Design method: picking beamforming weights .....	19
2.4.1	Data independent beamforming .....	19
2.4.2	Statistically optimum beamforming .....	20
2.4.3	Adaptive beamforming .....	24
2.5	Example of beamforming on a directional source .....	25
2.5.1	Beamforming on a point source .....	25
2.5.2	Beamforming on a quadrupole .....	27
2.6	Human talker directionality and why the point source model has been used .	29
2.6.1	Directivity magnitude .....	29
2.6.2	Impact of the phase .....	31
2.7	Chapter conclusion.....	33

### *Chapter 3*

Practical Examples of Directional Sources .....	34
3.1 Introduction.....	34
3.2 Musical instruments .....	35
3.3 Aeroacoustics.....	37
3.4 Chapter conclusion.....	39

### *Chapter 4*

An Extension of Beamforming Applicable to Directional Sources .....	40
4.1 Introduction and assumptions .....	40
4.2 Genesis of the method.....	42
4.3 Development of the method.....	44

4.3.1 Method with source decomposition into arbitrary modes set .....	49
4.4 Estimating mode strengths: gain-optimized (GO) estimator .....	50
4.5 Expression for the resulting array gain .....	52
4.6 Reduction of computational complexity .....	54
4.7 Difficulty of maximizing array gain subject to constraint on white-noise gain .....	55
4.8 Differences with other methods .....	56
4.8.1 Comparison with phase-mode spherical array .....	57
4.8.2 Comparison with near-field acoustical holography and Helmholtz equation least squares method .....	61
4.9 Chapter conclusion.....	61

## *Chapter 5*

Experimental and Simulation Results for the Proposed Beamforming Method .....	63
5.1 Introduction.....	63
5.2 Measurement setup and signal processing .....	63
5.3 White-noise gain for varying drive voltage .....	66
5.4 Beampattern along $x$ -axis.....	68
5.5 Chapter conclusion.....	69

## *Chapter 6*

Approaches to Estimate Mode Strengths .....	71
6.1 Introduction.....	71
6.2 Gain-optimized estimator (GO estimator) .....	72
6.2.1 Bias and variance of the GO estimator .....	72
6.3 Least squares (LS) estimators .....	73

6.3.1	Development .....	74
6.3.2	Difference between the exact source signal and the source model.....	77
6.3.3	Overall beamforming weight vector .....	78
6.3.4	Computation of the pseudoinverse and the source model .....	78
6.3.5	Details on estimated mode strengths.....	81
6.3.6	Least squares with source and interferer models (LSSI) .....	82
6.4	Proposed beamforming method to estimate the source model for MVDR beamforming .....	84
6.5	Estimation using many frames.....	84
6.6	Chapter conclusion.....	90

## *Chapter 7*

	Performance of the Proposed Beamforming Method in the Presence of Diffuse Noise, Reverberation, or Interfering Sources.....	91
7.1	Introduction.....	91
7.1.1	Organization.....	91
7.1.2	Array geometries.....	92
7.1.3	Synopsis of the results .....	93
7.2	Simulations: Diffuse noise.....	94
7.2.1	Simulation of diffuse noise .....	94
7.2.2	Simulation description .....	97
7.3	Simulation results: Diffuse noise .....	101
7.3.1	Linear array and dipole source as in the experimental measurements.....	101
7.3.2	Rectangular array and dipole source.....	122
7.3.3	Rectangular array and quadrupole source.....	127

7.3.4	Rectangular array and monopole source.....	128
7.3.5	Gain-optimized estimator with varying number of frames.....	130
7.3.6	Concluding remarks: Diffuse noise .....	136
7.4	Simulations: Diffuse noise + reverberation .....	137
7.4.1	Simulation of reverberation .....	137
7.4.2	Simulation description .....	139
7.5	Simulation results: Diffuse noise + reverberation .....	141
7.5.1	Rectangular array and dipole source.....	142
7.5.2	Rectangular array and monopole source.....	153
7.5.3	Concluding remarks: Diffuse noise + reverberation.....	161
7.6	Simulations: Diffuse noise + an interferer .....	161
7.6.1	Simulation of an interferer .....	161
7.6.2	Simulation description .....	162
7.7	Simulations results: Diffuse noise + an interferer.....	164
7.7.1	Rectangular array, dipole source, and monopole interference.....	164
7.7.2	Beamforming position error, rectangular array, and dipole source ..	188
7.7.3	Concluding remarks: Diffuse noise + an interferer.....	198
7.8	Chapter conclusion.....	199

## *Chapter 8*

Conclusion .....	200
8.1 Important points .....	200
8.2 Future work.....	206
8.2.1 Source directivity changes with frequency and time .....	206
8.2.2 Which optimization criterion to use.....	206

8.2.3	Estimators of the source mode strengths .....	207
8.2.4	Beamforming on real life directional sources .....	207

### *Appendix A*

Solution to the Wave Equation in the Spherical Coordinate System, Spherical Hankel Function, and Spherical Harmonics .....		209
A.1	Solution to the wave equation in the spherical coordinate system for an outgoing wave .....	209
A.2	Spherical Hankel function of the second kind .....	211
A.3	Spherical harmonics .....	214

### *Appendix B*

Theoretical Bias and Variance of the Gain-Optimized Estimator .....		219
B.1	Bias of the estimator .....	219
B.2	Variance of the estimator .....	221

### *Appendix C*

Less Promising Approaches to Mode Strength Estimation .....		223
C.1	Orthogonalization of spherical harmonic modes .....	223
C.2	Modes set with zero-bias estimator .....	225
C.3	MVDR to estimate the mode strengths .....	230

### *Appendix D*

Comparison of Computational Complexity Between the Proposed Beamforming Method with the GO Estimator and Conventional Beamforming .....		233
D.1	Comparison of computational complexity to obtain the beamformer output .....	233
D.2	Comparison of computational complexity to obtain the mode vectors .....	237



*Appendix E*

Email Exchange for Authorization to Reproduce Figures from IRCAM Technical Report [Rousseau 1996]..... 242

References..... 245

## List of Figures

### Chapter 2

Figure 2.1	Beamforming .....	13
Figure 2.2	Array geometry for arrays of (a) 1 microphone; (b) 5 microphones; (c) 51 microphones .....	25
Figure 2.3	Beamforming on a point source (a) Array geometry and source directivity; (b) White-noise gain for conventional beamforming and a varying number of microphones in the array .....	26
Figure 2.4	Beamforming on a quadrupole (a) Array geometry and source directivity; (b) White-noise gain for a varying number of microphones in the array.....	27
Figure 2.5	Beamforming on a quadrupole. White-noise gain for a varying number of microphones in the array, (—) quadrupole source model, (· · ·) proposed beamforming method with GO estimator, (—) conventional beamforming .....	29
Figure 2.6	Human talker directivity magnitude in dB, measured at 30 cm from the mouth in the horizontal plane, in three frequency bands .....	30

### Chapter 3

Figure 3.1	Measurements of Flute Directivity, with G fingering (a) Fundamental 410 Hz; (b) Overtone 820 Hz; (c) Overtone 1628 Hz .....	37
------------	---	----

### Chapter 4

Figure 4.1	Source with an arbitrary directivity and arbitrary discrete array .....	44
Figure 4.2	Proposed beamforming method schematic .....	49

### Chapter 5

Figure 5.1	Experiment setup .....	64
Figure 5.2	Array gain for varying drive voltage difference .....	67
Figure 5.3	Array response along the $x$ -axis for a dipole source .....	69

### Chapter 7

Figure 7.1	The two array geometries (with a dipole source) used for simulations: (a) Linear array; (b) Rectangular array .....	93
------------	---	----

Figure 7.2	Normalized cross-power spectral density as a function of continuous distance.....	97
------------	---	----

### Linear array, dipole source, and diffuse noise

Figure 7.3	Linear array and dipole source.....	102
Figure 7.4	Spread of the array gain for 10000 realizations. (a) GO estimator; (b) Exact mode strengths .....	104
Figure 7.5	Histogram of the array gain for 10000 realizations with the GO estimator for input SNR of (a) -50 dB; (b) -20 dB .....	104
Figure 7.6	Spread of the white-noise gain for 10000 realizations with the GO estimator.....	105
Figure 7.7	Spread of the array gain for 10000 realizations. (a) LS (SVD) estimator; (b) LS (Direct) estimator.....	106
Figure 7.8	Histogram of the array gain for 10000 realizations with the LS (SVD) estimator for input SNR of (a) -50 dB; (b) -20 dB.....	107
Figure 7.9	Spread of the white-noise gain for 10000 realizations with the LS (SVD) estimator .....	108
Figure 7.10	Spread of the array gain for 10000 realizations, microphone output vector as source model.....	109
Figure 7.11	Spread of the array gain for 10000 realizations. (a) GO estimator with MVDR; (b) Exact source model and MVDR .....	110
Figure 7.12	Array gain with MVDR and exact source signal as source model to which we add a random vector scaled by a factor .....	111
Figure 7.13	Spread of the array gain for 10000 realizations. GO estimator with (a) $N_{\text{order}} = 1$ ; (b) $N_{\text{order}} = 2$ ; (c) $N_{\text{order}} = 3$ ; (d) $N_{\text{order}} = 4$ .....	115
Figure 7.14	Spread of the array gain for 10000 realizations. (a) GO estimator (b) GO estimator and gain optimization with constraint on white-noise gain of 0 dB .....	117
Figure 7.15	Spread of the white-noise gain for 10000 realizations. (a) GO estimator (b) GO estimator and gain optimization with constraint on white-noise gain of 0 dB.....	118
Figure 7.16	Histogram of the array gain for the 10000 realizations with the GO estimator and gain optimization with a 0 dB constraint on the white-noise gain .....	119
Figure 7.17	Histogram of the white-noise gain for the 10000 realizations with the GO estimator and gain optimization with a 0 dB constraint on the white-noise gain.....	120

Figure 7.18	White-noise gain vs. array gain for the 10000 realizations with the GO estimator and gain optimization with a 0 dB constraint on white-noise gain .....	121
Rectangular array, dipole source, and diffuse noise		
Figure 7.19	Rectangular array and dipole source.....	122
Figure 7.20	With the GO estimator and 10000 realizations, spread of the (a) Array gain; (b) White-noise gain.....	124
Figure 7.21	Histogram of the array gain for the 10000 realization with the GO estimator for input SNR of (a) -50 dB; (b) -20 dB; (c) 0 dB; (d) 10 dB .....	125
Figure 7.22	Spread of the array gain for 10000 realizations. (a) LS (SVD); (b) microphone output vector as the source model.....	126
Rectangular array, quadrupole source, and diffuse noise		
Figure 7.23	Rectangular array and quadrupole source.....	127
Rectangular array, monopole source, and diffuse noise		
Figure 7.24	Rectangular array and monopole source.....	129
Figure 7.25	Spread of the array gain for 10000 realizations with (a) GO estimator; (b) LS (SVD) .....	130
Rectangular array, dipole source, and diffuse noise, GO estimator with a varying number of frames		
Figure 7.26	Spread of the array gain for 10000 realizations with the GO estimator and estimation done with (a) 1 frame; (b) 5 frames; (c) 15 frames .....	131
Figure 7.27	Standard deviation of the array gain for the 10000 realizations with the GO estimator and estimation with 1 frame, 5 frames, 15 frames.....	132
Figure 7.28	Histogram of the array gain for the 10000 realizations with the GO estimator for input SNR of -20 dB and estimation with (a) 1 frame; (b) 5 frames; (c) 15 frames .....	133
Figure 7.29	Histogram of the array gain for the 10000 realizations with the GO estimator for input SNR of -10 dB and estimation with (a) 1 frame; (b) 5 frames; (c) 15 frames .....	134
Figure 7.30	Histogram of the array gain for the 10000 realizations with the GO estimator for input SNR of 0 dB and estimation with (a) 1 frame; (b) 5 frames; (c) 15 frames .....	135

### Rectangular array, dipole source, diffuse noise, and reverberation

- Figure 7.31 Spread of the array gain for 10000 realizations with the GO estimator. (a) Without reverberation; (b) With reverberation..... 143
- Figure 7.32 Spread of the resulting gain in SRR for the 10000 realizations with the GO estimator ..... 144
- Figure 7.33 Histograms of the resulting gain in SRR for the 10000 realizations with the GO estimator for input SNR of (a) -50 dB; (b) -20 dB; (c) 0 dB; (d) 10 dB ..... 145
- Figure 7.34 Gain in SRR vs. array gain for the 10000 realizations with the GO estimator for input SNR of (a) -50 dB; (b) -20 dB; (c) 0 dB; (d) 10 dB. .... 146
- Figure 7.35 Spread of the array gain for 10000 realizations with the LS estimator. (a) Without reverberation; (b) With reverberation..... 147
- Figure 7.36 Spread of the resulting gain in SRR for 10000 realizations with the LS estimator ..... 148
- Figure 7.37 Spread of the array gain for 10000 realizations with the microphone output vector as the source model. (a) Without reverberation; (b) With reverberation. .... 149
- Figure 7.38 Spread of the resulting gain in SRR for 10000 realizations with the microphone output vector as the source model ..... 149
- Figure 7.39 Spread of the array gain for 10000 realizations with the GO estimator and MVDR. (a) Without reverberation; (b) With reverberation ..... 150
- Figure 7.40 Spread of the resulting gain in SRR for 10000 realizations with the GO estimator and MVDR ..... 151
- Figure 7.41 Spread of the array gain for 10000 realizations with the exact source signal and MVDR. (a) Without reverberation; (b) With reverberation ..... 152
- Figure 7.42 Spread of the resulting gain in SRR for 10000 realizations with the exact source signal and MVDR ..... 152

### Rectangular array, monopole source, diffuse noise, and reverberation

- Figure 7.43 Spread of the array gain for 10000 realizations with the GO estimator. (a) Without reverberation; (b) With reverberation..... 154
- Figure 7.44 Spread of the resulting gain in SRR for 10000 realizations with the GO estimator ..... 155
- Figure 7.45 Spread of the array gain for 10000 realizations with the LS estimator. (a) Without reverberation; (b) With reverberation..... 157

Figure 7.46	Spread of the resulting gain in SRR for 10000 realizations with the LS estimator.....	157
Figure 7.47	Spread of the array gain for 10000 realizations with the microphone output vector as the source model. (a) Without reverberation (b) With reverberation .....	158
Figure 7.48	Spread of the resulting gain in SRR for 10000 realizations with the microphone output vector as the source model .....	159
Figure 7.49	Spread of the array gain for 10000 realizations with the GO estimator and MVDR. (a) Without reverberation; (b) With reverberation .....	160
Figure 7.50	Spread of the resulting gain in SRR for 10000 realizations with the GO estimator with MVDR.....	160

#### Rectangular array, dipole source, diffuse noise, and monopole interferer

Figure 7.51	Rectangular array, dipole source, and monopole interferer.....	165
Figure 7.52	Spread of the array gain for 10000 realizations with the GO estimator for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB .....	169
Figure 7.53	Spread of the array gain for 10000 realizations with the LS estimator for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB .....	170
Figure 7.54	Spread of the array gain for 10000 realizations with LSSI for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB .....	171
Figure 7.55	Spread of the array gain for 10000 realizations with the microphone output vector as the source model for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB .....	172
Figure 7.56	Mean array gain over the 10000 realizations, for input SNR of (a) -50 dB; (b) -50 dB.GO estimator, LS estimator, LSSI, microphone output as the source model. ....	173
Figure 7.57	Spread of the gain in SIR for 10000 realizations with the GO estimator for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB .....	175
Figure 7.58	Spread of the gain in SIR for 10000 realizations with the LS estimator for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB .....	176
Figure 7.59	Spread of the gain in SIR for 10000 realizations with LSSI for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB .....	177
Figure 7.60	Spread of the gain in SIR for 10000 realizations with the microphone output vector as source model for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB .....	178

Figure 7.61	Mean gain in SIR over the 10000 realizations, for input SNR of (a) -50 dB; (b) 50 dB. GO estimator, LS estimator, LSSI, microphone output as the source model .....	179
Figure 7.62	Gain in SIR vs. array gain for the 10000 realizations with the GO estimator at an input SIR of 0 dB and for input SNR of (a) -50 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB .....	180
Figure 7.63	Spread of the array gain for 10000 realizations with the GO estimator and MVDR for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.....	182
Figure 7.64	Spread of the array gain for 10000 realizations with LSSI and MVDR for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB .....	183
Figure 7.65	Spread of the array gain for 10000 realizations with the exact source model and MVDR for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB .....	184
Figure 7.66	Spread of the gain in SIR for 10000 realizations with the GO estimator and MVDR for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.....	185
Figure 7.67	Spread of the gain in SIR for 10000 realizations with LSSI and MVDR for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB .....	186
Figure 7.68	Spread of the gain in SIR for 10000 realizations with the exact source model and MVDR for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.....	187

#### Rectangular array, dipole source, diffuse noise, and beamforming position error

Figure 7.69	100 Beamforming position errors (a) $x$ - $z$ plane; (b) $x$ - $y$ plane.....	189
Figure 7.70	Histogram of the magnitude of the 100 beamforming position errors .....	189
Figure 7.71	Spread of the array gain for 10000 realizations, the GO estimator (a) beamforming at source position; (b) with beamforming position errors .....	191
Figure 7.72	Spread of the array gain for 10000 realizations, LS estimation (a) beamforming at source position; (b) with beamforming position errors .....	192

#### Rectangular array, dipole source, diffuse noise, monopole interferer, and beamforming position error

Figure 7.73	100 interferer beamforming position errors for the LSSI estimator (a) $x$ - $z$ plane; (b) $x$ - $y$ plane.....	193
-------------	---	-----

Figure 7.74	Histogram of the magnitude of the 100 interferer beamforming position errors for the LSSI estimator.....	193
Figure 7.75	Spread of the array gain for 10000 realizations, the GO estimator, input SIR of 0 dB (a) beamforming at source position; (b) with beamforming position errors.....	195
Figure 7.76	Spread of the array gain for 10000 realizations, LSSI estimation, input SIR of 0 dB (a) beamforming at source position; (b) with beamforming position errors.....	196
Figure 7.77	Spread of the gain in SIR for 10000 realizations, the GO estimator, input SIR of 0 dB (a) beamforming at source position; (b) with beamforming position errors.....	197
Figure 7.78	Spread of the gain in SIR for 10000 realizations, LSSI estimation, input SIR of 0 dB (a) beamforming at source position; (b) with beamforming position errors.....	197

### *Appendix A*

Figure A.1	Spherical Hankel function of the second kind $h_n^{(2)}(x)$ (a) real part; (b) imag part; (—) $n = 0$ , ( $\cdot \cdot \cdot$ ) $n = 1$ , ( $- -$ ) $n = 2$ , ( $\cdot -$ ) $n = 3$ ....	213
Figure A.2	Spherical harmonic $Y_0^0(\theta, \phi)$ .....	217
Figure A.3	Spherical harmonic $Y_1^{-1}(\theta, \phi)$ . (a) real part; (b) imaginary part ....	217
Figure A.4	Spherical harmonic $Y_1^1(\theta, \phi)$ . (a) real part; (b) imaginary part.....	218



## List of Tables

### Chapter 2

Table 2.1	Phase of the sound pressure in degrees relative to the phase on the mouth axis, in the horizontal plane for a life-size mannequin (Values from Table I of [Flanagan 1960]) .....	31
Table 2.2	Estimation of the penalty for using the point source model phase to beamform with an array of 11 microphones on a human talker .....	32

### Chapter 4

Table 4.1	Comparison of the proposed beamforming method and phase-mode spherical array when maximizing the array gain .....	60
-----------	---	----

### Chapter 5

Table 5.1	White-noise gain (dB) for each sub-beamformer $y_{nm}(k)$ in response to the dipole in the simulated case .....	68
-----------	---	----

### Chapter 6

Table 6.1	Estimated mode strengths, frame 1 .....	87
Table 6.2	Estimated mode strengths, frame 2 .....	87
Table 6.3	Estimated mode strengths after phase alignment, frame 1 .....	88
Table 6.4	Estimated mode strengths after phase alignment, frame 2 .....	88
Table 6.5	Estimated mode strengths after averaging across frames .....	89

### Chapter 7

#### Linear array, dipole source, and diffuse noise

Table 7.1	Array gain and white-noise gain .....	102
Table 7.2	Array gain and white-noise gain, exact source model, without and with constraint of 0 dB on white-noise gain .....	116

#### Rectangular array, dipole source, and diffuse noise

Table 7.3	Array gain and white-noise gain .....	123
-----------	---------------------------------------	-----

#### Rectangular array, quadrupole source, and diffuse noise

Table 7.4	Array gain and white-noise gain .....	128
-----------	---------------------------------------	-----

Rectangular array, monopole source, and diffuse noise

Table 7.5	Array gain and white-noise gain .....	128
-----------	---------------------------------------	-----

Rectangular array, dipole source, diffuse noise, and reverberation

Table 7.6	Array gain, white-noise gain, and gain in SRR .....	142
Table 7.7	Array gain and gain in SRR at high input SNR.....	143
Table 7.8	Minimum input SNR for which the gain in SRR is 0 dB or above for all 10000 realizations .....	144

Rectangular array, monopole source, diffuse noise, and reverberation

Table 7.9	Array gain, white-noise gain, and gain in SRR .....	153
Table 7.10	Array gain and gain in SRR at high input SNR.....	155
Table 7.11	Minimum input SNR for which the gain in SRR is 0 dB or above for all 10000 realizations .....	156

Rectangular array, dipole source, diffuse noise, and monopole interferer

Table 7.12	Array gain, white-noise gain, and gain in SIR.....	165
Table 7.13	Approximate input SNR for which all realizations provide an array gain above or equal to 0 dB .....	168
Table 7.14	Approximate input SNR for which all realizations provide an array gain above or equal to 10 dB .....	168
Table 7.15	Approximate input SNR for which all realizations provide a gain in SIR above or equal to 0 dB .....	174
Table 7.16	Approximate input SNR for which all realizations provide a gain in SIR above or equal to 10 dB .....	174

Rectangular array, dipole source, diffuse noise, and beamforming position error

Table 7.17	Array gain and white-noise gain, beamforming with exact mode strength but with beamforming position error .....	190
Table 7.18	Array gain and white-noise gain, conventional beamforming but with beamforming position error .....	190

Rectangular array, dipole source, diffuse noise, monopole interferer, and beamforming position error

Table 7.19	Array gain, white-noise gain, and gain in SIR, beamforming with exact mode strength but position error .....	194
Table 7.20	Array gain, white-noise gain, and gain in SIR, conventional beamforming with position error .....	194

*Appendix A*

Table A.1	Spherical Hankel functions of the second kind of order zero to three .....	212
Table A.2	Selected values of the associated Legendre functions .....	215
Table A.3	Spherical harmonics of orders up to 3 .....	216

*Appendix D*

Table D.1	Comparison of computational complexity to obtain the beamformer output .....	236
Table D.2	Run-time comparison for the computation of 10000 beamformer outputs.....	237
Table D.3	Comparison of computational complexity to obtain the spherical Hankel functions of second kind .....	238
Table D.4	Comparison of computational complexity to obtain the spherical harmonic functions .....	239
Table D.5	Run-time comparison for the computation of the mode vectors necessary to beamforming 1000 times.....	240
Table D.6	Run-time comparison for the computation of the mode vectors and beamforming 1000 times .....	241

## List of Symbols

$B(f, \theta, \phi)$	Beamformer response, with plane wave source model.
$B(f, r, \theta, \phi)$	Beamformer response, with point source model.
$B(f, r, \theta, \phi, \text{source directivity and orientation})$	Beamformer response with specific source model, other than plane wave or point source models.
$c$	Speed of sound.
$C_m^m(k)$	Complex coefficient equal to $\mathbf{M}_m^H(k)\mathbf{R}_{ee}^{-1}(k)\mathbf{M}_m(k)$ .
$C_{n'm'}^{nm}(k)$	Complex coefficient equal to $\mathbf{Y}_n^{mH}(k)\mathbf{R}_{ee}^{-1}(k)\mathbf{Y}_{n'}^{m'}(k)$ .
$\mathbf{C}(f)$	Constraint matrix for Linearly Constrained Minimum Variance beamforming (LCMV).
$d_{l_1, l_2}$	Distance separating the $l_1^{\text{th}}$ and $l_2^{\text{th}}$ microphones.
$d_{nm}(k)$	Gain-optimized (GO) estimator of the source mode strength for the spherical harmonic mode $(n, m)$ .
$d_{nm}(k, fr)$	Gain-optimized (GO) estimator of the source mode strength for the spherical harmonic mode $(n, m)$ in the $fr^{\text{th}}$ FFT frame.
$d_{n_{\text{ref}}, m_{\text{ref}}}(k, fr)$	Gain-optimized (GO) estimator of the source mode strength for the reference spherical harmonic mode $(n_{\text{ref}}, m_{\text{ref}})$ in the $fr^{\text{th}}$ FFT frame.
$d_{\text{ZB}, m}(k)$	Estimated source mode strengths for the mode $\mathbf{M}_{\text{ZB}, m}(k)$ .
$\text{DENOM}(\bullet)$	Denominator operator.
$e_l(k)$	$l^{\text{th}}$ element of the noise vector $\mathbf{e}(k)$
$e_l(q)$	$q^{\text{th}}$ time sample of the noise component at the output of the $l^{\text{th}}$ microphone.
$e_l(t)$	Noise at the output of the $l^{\text{th}}$ microphone, in the time-domain.
$E(\bullet)$	Expectation operator.
$\mathbf{e}(f)$	Noise vector, frequency component of the noise at the output of the microphones.

$\mathbf{e}(k)$	Noise vector, frequency component of the noise at the output of the microphones.
$\mathbf{e}_w$	Noise vector whose elements are uncorrelated random complex numbers, with correlation matrix equal to the identity matrix.
$\mathbf{E}(k)$	Matrix whose columns are the eigenvectors of $\mathbf{R}_{ee}(k)$ .
$f$	Continuous-time frequency parameter, in Hertz.
$fr$	Frame index
$F_s$	Sampling frequency
$G(f)$	Array gain, for frequency $f$ .
$G(k)$	Array gain, for frequency bin $k$ .
$G_{\text{out, dB}}(k)$	Optimized white-noise gain for the proposed beamforming method and the GO estimator in Chapter 5.
$G_{\text{SIR}}(k)$	Gain in signal-to-interference ratio (SIR), for frequency bin $k$ .
$G_{\text{SRR}}(k)$	Gain in signal-to-reverberation ratio (SRR), for frequency bin $k$ .
$G_w(f)$	White-noise gain, for frequency $f$ .
$G_w(k)$	White-noise gain, for frequency bin $k$ .
$G_{w, \text{opt}}(f)$	Optimum white-noise gain, for frequency $f$ .
$h_n^{(2)}(\kappa r)$	Spherical Hankel function of the second kind of order $n$ .
$H$	As a superscript, indicates the Hermitian conjugate transpose.
$i$	Square root of -1.
$i_0$	Image order (number of walls encountered by a ray)
$\mathbf{I}$	Identity matrix.
$j_n(\kappa_k r)$	Spherical Bessel function of the first kind of order $n$ .
$J(k)$	Cost function with least squares, for frequency bin $k$ .
$k$	Discrete-time frequency bin index.
$l$	Microphone index
$L$	Number of discrete microphones.
$m$	Degree of a spherical harmonic.
$m_{\text{ref}}$	Degree of the spherical harmonic mode of reference when estimation is performed with multiple FFT frames.

$M$	Number of constraints with Linearly Constrained Minimum Variance beamforming (LCMV), or number of modes considered with the proposed beamforming method.
$\mathbf{M}_m(k)$	Mode vector, for frequency bin $k$ , for frequency bin $k$ .
$\mathbf{M}_{\text{ZB}}(k)$	Matrix containing the $M$ modes with “zero-bias” $\mathbf{M}_{\text{ZB},m}(k)$ .
$\mathbf{M}_{\text{ZB},m}(k)$	Mode vector with “zero bias” such that $\mathbf{M}_{\text{ZB}}^H(k)\mathbf{R}_{\text{ee}}^{-1}(k)\mathbf{M}_{\text{ZB}}(k) = \mathbf{I}_{M \times M}$ , for frequency bin $k$ .
$n$	Discrete-time sample index. or order of a spherical harmonic.
$n_{\text{ref}}$	Order of the spherical harmonic mode of reference when estimation is performed with multiple FFT frames.
$N$	Number of time samples used in an FFT frame.
$N_{\text{frames}}$	Number of FFT frames in the mode strengths estimation.
$N_{\text{order}}$	Order maximum of sub-beamformers included in the overall beamforming output.
$N_r(i_0)$	Number of images of order $i_0$ .
$N_{\text{ref}}$	Number of modes included in the phase alignment when computing the GO estimator with more than one frame.
$\text{NUM}(\bullet)$	Numerator operator.
$p_l(f)$	Frequency-domain output of the $l^{\text{th}}$ microphone for the continuous time case, forward Fourier transform of $p_l(t)$ .
$p_l(k)$	Frequency-domain output of the $l^{\text{th}}$ microphone for the discrete-time case, normalized Fast Fourier Transform of $p_l(n)$ .
$p_l(k, fr)$	Frequency-domain output of the $l^{\text{th}}$ microphone, $fr^{\text{th}}$ FFT frame, in the discrete-time case, normalized Fast Fourier Transform of $p_l(n, fr)$ .
$p_l(n)$	Time-domain output of the $l^{\text{th}}$ microphone for the discrete time case.
$p_l(n, fr)$	Time-domain output of the $l^{\text{th}}$ microphone, $fr^{\text{th}}$ FFT frame, for the discrete time case
$p_l(t)$	Time-domain output of the $l^{\text{th}}$ microphone for the continuous time case.
$P_n^m$	Legendre function of order $n$ and degree $m$ .

$P_{\text{MVDR},nm}(k)$	Estimated source mode strength of order $n$ and degree $m$ with MVDR.
$\mathbf{p}(f)$	Microphone output vector, for frequency $f$ .
$\mathbf{p}(k)$	Microphone output vector, for frequency bin $k$ .
$\mathbf{Q}(k)$	Matrix obtained from the QR factorization of $\mathbf{Y}(k)$ .
$\mathbf{Q}'(k)$	Matrix obtained from the QR factorization of $\mathbf{Y}'(k)$ .
$r$	Distance between a point and the origin of the spherical coordinate system, or rank of a matrix.
$r_d(f, r_m, \theta_m, \phi_m)$	Desired array response at the $m^{\text{th}}$ location at frequency $f$ , with Linearly Constrained Minimum Variance beamforming (LCMV).
$r_l$	Distance between the $l^{\text{th}}$ microphone and the origin of the spherical coordinate system.
$\mathbf{r}_d(f)$	Response vector at frequency $f$ , with Linearly Constrained Minimum Variance beamforming (LCMV).
$\mathbf{R}(k)$	Matrix obtained from the QR factorization of $\mathbf{Y}(k)$ .
$\mathbf{R}'(k)$	Matrix obtained from the QR factorization of $\mathbf{Y}'(k)$ .
$\mathbf{R}_{\text{ee}}(f)$	Normalized noise cross-power spectral density matrix, for frequency $f$ .
$\mathbf{R}_{\text{ee}}(k)$	Normalized noise cross-power spectral density matrix, for frequency bin $k$ .
$\mathbf{R}_{\text{pp}}(f)$	Microphone output cross-power spectral density matrix, for frequency $f$ .
$\mathbf{R}_{\text{pp}}(k)$	Microphone output cross-power spectral density matrix, for frequency bin $k$ .
$\mathbf{R}_{\text{ss}}(f)$	Normalized signal cross-power spectral density matrix, for frequency $f$ .
$\mathbf{R}_{\text{ss}}(k)$	Normalized signal cross-power spectral density matrix, for frequency bin $k$ .
$\hat{\mathbf{R}}_{\text{ss}}(k)$	Estimated normalized signal cross-power spectral density matrix, for frequency bin $k$ , in Chapter 5.

$\hat{\mathbf{R}}_{ss, w \text{ constr}}(k)$	Estimated normalized signal cross-power spectral density matrix, for frequency bin $k$ for array gain optimization with constraint on white-noise gain.
$\mathbf{R}_{(s+s_r)(s+s_r)}(k)$	Normalized signal plus reverberation cross-power spectral density matrix, for frequency bin $k$ .
$\mathbf{R}_{s,s_l}(k)$	Normalized interferer cross-power spectral density matrix, for frequency bin $k$ .
$\mathbf{R}_{s,s_r}(k)$	Normalized reverberation cross-power spectral density matrix, for frequency bin $k$ .
$s(t, r_l, \theta_l, \phi_l)$	Component of the source of interest at the output of the $l^{\text{th}}$ microphone, in the time-domain.
$s_m(k)$	Source mode strength for the frequency bin $k$ , contribution of the mode $\mathbf{M}_m(k)$ from the source.
$s_{nm}(f)$	Source mode strength at frequency $f$ , contribution of the spherical harmonic mode $(n, m)$ from the source.
$s_{nm}(k)$	Source mode strength for the frequency bin $k$ , contribution of the spherical harmonic mode $(n, m)$ from the source.
$s_{nm,l}(k)$	Mode strength for the frequency bin $k$ , contribution of the spherical harmonic mode $(n, m)$ from the interferer.
$S$	Total surface of the walls of the rectangular room.
$S_{l_1, l_2}(f)$	Cross-power spectral density function of diffuse noise between the $l_1^{\text{th}}$ and $l_2^{\text{th}}$ omnidirectional microphones, for frequency $f$ .
$S_{xx}(f)$	(Self) power-spectral density of the diffuse noise at the output of the microphones, for frequency $f$ .
$\text{SIR}_{\text{in}}(k)$	Signal-to-interference ratio (SIR) at the input of the array (averaged over the $L$ microphones), for frequency bin $k$ .
$\text{SNR}_{\text{in}}(f)$	Reference input signal-to-noise ratio (SNR) at the input of the array (averaged over the $L$ microphones), for frequency $f$ .
$\text{SNR}_{\text{out}}(f)$	Signal-to-noise ratio (SNR) at the output of the array, for frequency $f$ .
$\text{SRR}_{\text{in}}(k)$	Input signal-to-reverberation ratio (averaged over the $L$ microphones), for frequency bin $k$ .



$\mathbf{s}(f)$	Source signal vector, component of the source at the output of the microphones, for frequency $f$ .
$\mathbf{s}(k)$	Source signal vector, component of the source at the output of the microphones, for frequency bin $k$ .
$\mathbf{s}(f, r, \theta, \phi)$	Source signal vector for a source located at $(r, \theta, \phi)$ , component of the source at the output of the microphones, for frequency $f$ (beam pattern, Chapter 2).
$\mathbf{s}_e(k)$	Vector of difference (at the microphones) between the source model and the source signal, for frequency bin $k$ .
$\mathbf{s}_I(k)$	Component at the output of the microphones from the interferer, for frequency bin $k$ .
$\mathbf{s}_{\text{mo}}(f)$	Source model, used for beamforming, for frequency $f$ .
$\mathbf{s}_{\text{mo,GO}}(k)$	Source model used for beamforming built using the GO estimator and mode vectors, for frequency bin $k$ .
$\mathbf{s}_{\text{mo},m}(f, r_m, \theta_m, \phi_m)$	Source model for the $m^{\text{th}}$ location at frequency $f$ , with Linearly Constrained Minimum Variance beamforming (LCMV).
$\mathbf{s}_{\text{mo,LS}}(k)$	Source model used for beamforming built using least-squares estimation of the source mode strengths, for frequency bin $k$ .
$\mathbf{s}_{\text{mo,QR}}(k)$	Source model built from the estimation of mode strengths following the QR factorization of $\mathbf{Y}(k)$ .
$\mathbf{s}_{\text{mo,ZB}}(k)$	Source model with the “zero-bias” modes in $\mathbf{M}_{\text{ZB}}(k)$ .
$\mathbf{s}_{\text{mo,s}}(k)$	Source model, used for beamforming built using least-squares with modes for both source and interferer locations, for frequency bin $k$ .
$\mathbf{s}_{\text{ps}}(f, r, \theta, \phi)$	Source signal for a point source located at $(r, \theta, \phi)$ , component of the source at the output of the microphones, for frequency $f$ (beam pattern, Chapter 2).
$\mathbf{s}_{\text{pw}}(f, \theta, \phi)$	Source signal for a plane wave coming from direction $(\theta, \phi)$ , component of the source at the output of the microphones, for frequency $f$ (beam pattern, Chapter 2).
$\mathbf{s}_r(k)$	Component at the output of the microphones from the reverberation of the source, for frequency bin $k$ .

$\hat{\mathbf{s}}_{\text{str},I}(k)$	Estimated interferer mode strengths with least-squares with modes for both source and interferer locations, for frequency bin $k$ .
$\hat{\mathbf{s}}_{\text{str},\text{LS}}(k)$	Estimated source mode strengths with least-squares for the modes considered with the proposed beamforming method, for frequency bin $k$ .
$\hat{\mathbf{s}}_{\text{str},\text{LSSI}}(k)$	Estimated source and interferer mode strengths with least-squares for the modes considered with the proposed beamforming method, for frequency bin $k$ .
$\hat{\mathbf{s}}_{\text{str},\text{QR}}(k)$	Estimated mode strengths for the QR factorization of $\mathbf{Y}(k)$ .
$\hat{\mathbf{s}}_{\text{str},s}(k)$	Estimated source mode strengths with least-squares with modes for both source and interferer locations, for frequency bin $k$ .
$\mathbf{s}_{\text{str},\text{ZB}}(k)$	Exact source mode strengths when the source signal is expressed with the modes in $\mathbf{M}_{\text{ZB}}(k)$ , for frequency bin $k$ .
$\mathbf{s}'_{\text{str}}(k)$	Exact source mode strengths for the modes considered with the proposed beamforming method, for frequency bin $k$ .
$\mathbf{S}_{r \times r}$	Diagonal matrix of the nonzero singular values.
$t$	Continuous-time parameter, in seconds.
$T$	Reverberation time of the rectangular room.
$\mathbf{u}_m$	$m^{\text{th}}$ left singular vector.
$\mathbf{U}$	Matrix containing the left singular vectors (column vectors).
$V$	volume of the rectangular room
$\text{Var}(\bullet)$	Variance operator.
$\mathbf{V}$	Matrix containing the right singular vectors (column vectors).
$\mathbf{v}_m$	$m^{\text{th}}$ right singular vector.
$w_l(f)$	Beamforming weight for the $l^{\text{th}}$ microphone, for frequency $f$ .
$\mathbf{w}(f)$	Beamforming weight vector, for frequency $f$ .
$\mathbf{w}_{\text{conv}}(k)$	Beamforming weight vector for conventional beamforming (point source model.)
$\mathbf{w}_{\text{GO}}(k)$	Beamforming weight vector for the proposed beamforming method and the gain-optimized (GO) estimation of the source mode strengths, for frequency bin $k$ .

$\mathbf{w}_{\text{LS}}(k)$	Beamforming weight vector for the proposed beamforming method and least-squares estimation of the source mode strengths, for frequency bin $k$ .
$\mathbf{w}_m(k)$	Gain-optimized array weight vector for the sub-beamformer of order $m$ (mode $\mathbf{M}_m(k)$ ), for frequency bin $k$ .
$\mathbf{w}_{\text{MVDR,GO}}(k)$	Beamforming weight vector computed with the source model built with the gain-optimized (GO) estimator used in MVDR, for frequency bin $k$ .
$\mathbf{w}_{\text{MVDR,LS}}(k)$	Beamforming weight vector computed with the source model built with least-squares estimation used in MVDR, for frequency bin $k$ .
$\mathbf{w}_{\text{MVDR,}nm}(k)$	Beamforming weight vector for estimation of source mode strength of order $n$ and degree $m$ with MVDR.
$\mathbf{w}_{nm}(k)$	Gain-optimized array weight vector for the sub-beamformer of order $n$ and degree $m$ , for frequency bin $k$ .
$\mathbf{w}_{\text{opt}}(f)$	Beamforming weight vector that maximizes the array gain, for frequency $f$ .
$\mathbf{w}_{\text{opt}}(k)$	Beamforming weight vector that maximizes the array gain, for frequency bin $k$ .
$\mathbf{w}_{\text{opt, w constr}}(k)$	Beamforming weight vector that maximizes the array gain subject to a constraint on the white-noise gain, for frequency bin $k$ .
$\mathbf{w}_{\text{ps}}(f)$	Beamforming weight vector computed with a point source model, for frequency $f$ .
$\mathbf{w}_{\text{pw}}(f)$	Beamforming weight vector computed with plane wave model, for frequency $f$ .
$\mathbf{w}_{\text{w,opt}}(f)$	Beamforming weight vector that maximizes the white-noise gain, for frequency $f$ .
$x$	First coordinate in the Cartesian coordinate system.
$y$	Second coordinate in the Cartesian coordinate system.
$y(f)$	Beamformed array output, for frequency $f$ .
$y_m(k)$	Gain-optimized sub-beamformer output of order $m$ (mode $\mathbf{M}_m(k)$ ), for frequency bin $k$ .

$y_n(\mathbf{r})$	Spherical Bessel functions of the second kind of order $n$ .
$y_{nm}(k)$	Gain-optimized sub-beamformer output (or simply sub-beamformer) of order $n$ and degree $m$ , for frequency bin $k$ .
$y_{nm, \text{noise}}(k)$	Output of the sub-beamformer of order $n$ and degree $m$ for the noise, for frequency bin $k$ .
$y_{nm, \text{signal}}(k)$	Output of the sub-beamformer of order $n$ and degree $m$ for the source signal, for frequency bin $k$ .
$y_{\text{noise}}(k)$	Beamformed array output for the noise, for frequency bin $k$ .
$y_{\text{signal}}(k)$	Beamformed array output for the source signal, for frequency bin $k$ .
$\mathbf{y}_{\text{ZB}}(k)$	Column vector containing the $M$ sub-beamformer outputs for the modes in $\mathbf{M}_{\text{ZB}}(k)$ .
$Y_n^m(\theta, \phi)$	Spherical harmonic function of order $n$ and degree $m$ for the direction $(\theta, \phi)$ .
$\mathbf{Y}(k)$	Matrix containing the mode vectors considered with the proposed beamforming method for the <i>source model</i> , for frequency bin $k$ .
$\mathbf{Y}'(k)$	Matrix containing the mode vectors considered with the proposed beamforming method for the <i>source signal</i> , for frequency bin $k$ .
$\mathbf{Y}^+(k)$	Pseudoinverse of $\mathbf{Y}(k)$ .
$\mathbf{Y}_n^m(k)$	Mode vector of order $n$ and degree $m$ , for frequency bin $k$ .
$\bar{\mathbf{Y}}_n^m(k)$	Contains the $L$ values of spherical harmonic function for the phase-mode spherical array.
$\mathbf{Y}_{n,1}^m(k)$	Mode vector of order $n$ and degree $m$ for the interferer in the frequency bin $k$ .
$\mathbf{Y}_1(k)$	Matrix containing the mode vectors for the interferer, for least-squares with modes for both source and interferer locations, for frequency bin $k$ .
$\mathbf{Y}_s(k)$	Matrix containing the mode vectors for the source, for least-squares with modes for both source and interferer locations, for frequency bin $k$ .
$z$	Third coordinate in the Cartesian coordinate system.

$\alpha$	Wall surface absorption coefficient.
$\alpha(f)$	Scaling factor, for frequency $f$ .
$\delta^2$	Constraint on white-noise gain.
$\theta$	Angle between a point and the $z$ axis. (Second coordinate of the spherical coordinate system.)
$\theta_l$	Angle between the $l^{\text{th}}$ microphone and the $z$ axis. (Second coordinate of the spherical coordinate system.)
$\varepsilon$	Lagrange multiplier
$\varepsilon(k)$	Error vector containing the difference between the microphone output vector and the source model estimated with least squares, for frequency bin $k$ .
$\kappa$	Wave number
$\lambda$	Wavelength.
$\lambda_l$	$l^{\text{th}}$ eigenvalue of $\mathbf{R}_{ee}(k)$ .
$\Lambda(k)$	Diagonal matrix whose elements are the eigenvalues of $\mathbf{R}_{ee}(k)$ .
$\sqrt{\Lambda(k)}$	Diagonal matrix whose elements are the square root of the eigenvalues of $\mathbf{R}_{ee}(k)$ .
$\sigma_e^2(f)$	Average signal power spectral density over the $L$ microphones of the array, for frequency $f$ .
$\sigma_e^2(k)$	Average signal power spectral density over the $L$ microphones of the array, for frequency bin $k$ .
$\sigma_m$	$m^{\text{th}}$ singular value.
$\sigma_s^2(f)$	Average signal power spectral density over the $L$ microphones of the array, for frequency $f$ .
$\sigma_s^2(k)$	Average signal power spectral density over the $L$ microphones of the array, for frequency bin $k$ .
$\sigma_{s+s_r}^2(k)$	Average signal plus reverberation power spectral density over the $L$ microphones of the array, for frequency bin $k$ .
$\sigma_{s_i}^2(k)$	Average interferer power spectral density over the $L$ microphones of the array, for frequency bin $k$ .
$\sigma_{s_r}^2(k)$	Average reverberation power spectral density over the $L$ microphones of the array, for frequency bin $k$ .

$\Sigma$	Matrix containing the singular values of $\mathbf{Y}(k)$ .
$\Sigma^+$	Pseudoinverse of the matrix $\Sigma$ .
$\phi$	Angle between the projection of a point on the $x$ - $y$ plane, and the $x$ axis. (Third coordinate of the spherical coordinate system.)
$\phi_l$	Angle between the projection of the location of the $l^{\text{th}}$ microphone on the $x$ - $y$ plane, and the $x$ axis. (Third coordinate of the spherical coordinate system.)
$\nabla$	Gradient operator.
$\bar{\nabla}^2$	Laplacian operator.

---

# CHAPTER *1*

---

## Introduction

The term *beamforming* comes from the idea of forming narrow spatial beams to receive signals coming from certain directions while attenuating signals coming from directions outside the beams [Van Veen 1988; p.4]. Beamforming can be used for radiation or reception of energy. The present thesis discusses the reception of sound from sources using a microphone array, so that we are only concerned with receiving beamforming.

Beamforming microphone arrays achieve a directional or spatially-specific response by coherently summing signals from multiple sensors. It relies upon a model of the wavefront (*source model*) arriving at the array to calculate the time delay (or frequency-domain phase shift) and possibly also the amplitude weighting that must be applied to the output of each sensor so that the component from the source of interest may either be summed coherently or be maximized according to some criteria. Signals from other locations arrive at the array out of phase relative to the source model and are attenuated through destructive interference. The source model must provide accurate time delay or phase shift and, if it is used, amplitude, for beamforming to perform as expected.

The source of interest is located at some position, and it has some directivity and some orientation with respect to the array. Even if the position of this source is known,

---

the exact source directivity and orientation are generally unknown. Most often in the literature and in practice, the source is assumed to radiate as a point source, and a point source is used as the source model. A point source, or monopole, radiates in the same manner in all directions, with the sound pressure inversely proportional to the distance from the source. If the wavefront curvature is negligible, a plane wave model is used and the beamforming is said to be *far-field*.

From seminal references, here are the source models used in the description of three sophisticated beamforming methods. In the 1986 paper titled “Practical supergain”, Cox et al. studied the problem of maximizing the array gain (a measure of the improvement in signal-to-noise ratio) subject to an equality on the white-noise gain (used here as a measure of sensitivity to random errors). In the paper, it says “When there is a single coherent wavefront from direction  $\theta$  [...]” [Cox 1986]. Because the only qualification of the wavefront is a direction, this indicates that the source model is a plane wave. We discuss the problem of maximizing the array gain subject to an equality on the white-noise gain in Sub-Section 2.4.2.

The first sentence in the abstract of the paper “An algorithm for linearly constrained adaptive array processing” [Frost 1972] is “A constrained least mean-squares algorithm has been derived which is capable of adjusting an array of sensors in real time to respond to a signal coming from a desired direction while discriminating against noises coming from other directions”. Again, the signal is qualified by a direction, indicating a plane wave model. The linearly constrained minimum variance (LCMV) beamforming from this reference is also discussed in Sub-Section 2.4.2.

To develop the phase-mode spherical array, sometimes called “eigenbeamformer”, Meyer and Elko, in their 2002 conference proceeding “A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield” [Meyer 2002], used the fact that the sound pressure on a sphere from a plane wave can be decomposed into an infinite sum of spherical harmonics. In this method, beamforming is separated in two parts. In the first part, the received soundfield is spatially decomposed into spherical harmonics. In the second part, the beam is formed by summing spherical harmonics. In his book “Modal array signal processing: Principles and



---

Applications of acoustic wavefield decomposition” [Teutsch 2007], Teutsch did the development for the more general case of a point source, which we use in our Sub-Section 4.8.1 to compare our proposed beamforming method and the phase-mode spherical array.

Those three earlier methods are not restricted to plane wave or point source models. For the methods described by Cox and Frost, it is necessary to know the source model. In the case of the phase-mode spherical array, not only must the source model be known but the source model must also have a spherical harmonic decomposition on the spherical array.

If a source radiates differently than its source model, the three beamforming methods just mentioned would not perform as expected. The following citation from [Van Veen 1988] describes what happens then for statistically optimum beamforming methods such as those discussed in [Cox 1986] and [Frost 1972]: “If the required knowledge [i.e. source model] is inaccurate, the optimum beamformer will attenuate the desired signal as if it were interference. Cancellation of the desired signal is often significant, especially if the SNR of the desired signal is large”.

Now, here is an explanation of why the point source model has been extensively and successfully used to beamform with microphone arrays on human talkers. If the array subtends a small angle about the source, the array observes only a small portion of the source directivity pattern. As will be seen in this thesis, the human talker directivity pattern magnitude is fairly constant in the frontal hemisphere of the talker. Helping further, with arrays of small dimensions located in front of the talker, only a portion of the hemisphere is observed. In such a case, the point source is an acceptable model and does not lead to a much noticeable penalty, as can be seen in Section 2.6.

## **1.1 Distributed array and directional sources**

The point source may no longer be adequate as a source model for arrays covering a larger angular extent of a source, in particular with arrays surrounding a source (called *distributed array* in this thesis). This is particularly the case for sources that radiate very differently than point sources. Musical instruments and aeroacoustic

---

sources are examples of such sources and will be discussed in the thesis. Using the point source model to beamform on such sources may result in a decrease in the array gain or even in a total breakdown of beamforming. We present the decreased white-noise gain in the case of beamforming on a quadrupole source using a point source model in Section 2.5.

To avoid this, a more accurate source model is required, which would have the same directivity and orientation as the actual source. This source model could be used directly to perform the beamforming.

We found an example of beamforming on directional sources in two papers where the sources are aeroacoustic dipoles [Jordan 2002] [Liu 2008]. The authors were interested in finding the locations of the dipoles. They knew the source model (dipole) and its likely orientation. Unfortunately, as it was mentioned earlier, exact source directivity and orientation are unknown in most cases, so that such a more accurate source model is generally not available.

## 1.2 Proposed beamforming method

Because the source directivity and orientation of a source are unknown in most cases, this thesis introduces an extension of beamforming applicable to directional sources. While previous works started from a known source model, most often for a plane wave or a point source, no assumption is made on the source directivity in this thesis. The source directivity is treated as an unknown.

The general principle of the proposed beamforming method is to define a set of “sub-beamformers”, each beamforming on a different spatial mode of the source. The outputs of the individual sub-beamformers are combined in a weighted sum to give an overall output of better quality than that of a beamformer using the monopole (point source) model.

The thesis shows that, with the appropriate weighting, the sum of the sub-beamformers gives the same optimum array gain as beamforming directly with the exact source model. The weight to give each sub-beamformer is unknown and a few methods to perform the estimation, based on the observed data, are demonstrated in the thesis.

It is also possible to think of the proposed beamforming method as building a source model from spatial modes and the estimates of the strength of each mode composing the source. Beamforming weights are then computed with this source model.

Further comparing the proposed beamforming method with beamforming using a point source model:

- the proposed beamforming method can overcome deteriorated beamformer responses that occur when a null of the source directivity falls on the receiving array [Jordan 2002];
- in a stationary environment, once the sub-beamformer weights are evaluated, the computational complexity of the proposed beamforming method can be made the same as for a point source model beamformer by combining the sub-beamformers;
- the proposed beamforming method has a similar performance as a point source model beamformer when beamforming on point sources;
- the proposed beamforming method works with the same array geometry and hardware as a point source model beamformer.

The thesis demonstrates through simulations and experiments the advantages of the proposed beamforming method over beamforming with a point source when beamforming on a directional source.

### **1.3 Some applications of microphone arrays and beamforming**

This section presents three applications of employing microphone arrays and beamforming: sound pickup, source localization, and spatial differentiation, without going into much detail. An appropriate source model is required for proper beamforming operation. If the source model is different from how the actual source radiates, the behavior of the array response will be different than expected. The proposed beamforming method is mostly aimed at sound pickup.

---

### 1.3.1 Sound pickup (sound capture)

Positioning the microphone close to the source (referred to as “close-miking”) provides the best results for sound pickup because the direct path from the source easily dominates more distant interfering sources and reverberation. Unfortunately, this approach is obtrusive. People generally do not want to position a microphone by their mouth and a microphone mounted on or near devices such as musical instruments limits their operation and movement.

The use of a single microphone located at a significant distance from the source suffers from reverberation and interference. The reader may have experienced the poor sound quality when listening to someone talking on a speakerphone, compared to when they speak using a handset held close to their mouth. Although speakerphones typically utilize directional (cardioid) microphones, they usually have insufficient directivity to achieve good sound pickup. Cardioid microphones are a type of differential array and, with few exceptions [Teutsch 2001], they are not steerable.

Using an array, sound quality can be improved by spatially isolating the desired source from reverberation and interference.

### 1.3.2 Source tracking

An array can also be used to localize a source and then to track it. If the overall objective is sound pickup, this tracking information can then be used to focus the same array (or a different array) on the source. Localization methods are sometimes divided into three categories:

- (1) methods which try to find the beamforming location where the output power of the beamforming array is maximum,
- (2) methods adapted from the field of high-resolution spectral analysis (autoregressive modeling, minimum variance spectral estimation, and eigenanalysis-based techniques), and
- (3) methods employing time-difference of arrival (TDOA) estimates [DiBiase 2001].

---

The last category may be seen as different from beamforming as the outputs of the microphones are not all combined together. Instead, time-difference of arrival is estimated between pairs of microphones. To meet the time-difference estimate of a given microphone pair, the source must be located somewhere on one hyperbolic curve. The intersection of all the hyperbolic curves is then found in some optimal sense (in the sense of accuracy, robustness, and efficiency) to obtain the source location estimate [Huang 2001].

With the source localization estimated at some time interval, tracking can then be realized using an algorithm such as the well known Kalman filter [Kalman 1960].

### **1.3.3 Spatial differentiation**

While in the previous subsection the objective was to localize a physical source, here the objective is to characterize the overall sound field, i.e. to map from where the sound components are coming.

For example, measurement results for an auditorium, employing 98 measurements around a hard sphere and plane wave decomposition into spherical harmonics up to order 6, are presented in [Park 2005]. Measurement results for a small meeting room, a videoconferencing room, and a lecture theater, done with a spherical array of 32 microphones steered in 60 symmetric directions are presented in [Gover 2004].

## **1.4 Objective of the method and contributions of the thesis**

The main objective of the method presented in this thesis is to provide improved sound pickup of directional sources compared to beamforming using a point source model which can have deteriorated performances when beamforming on such sources.

To avoid being obtrusive, the microphones have to be at a certain distance from the source. If the microphones could be close to the source, “close-miking” would be a much simpler technique to capture the sound, and it is indeed a proven method in the audio recording industry.

The proposed beamforming method is aimed at sound pickup and not at characterizing the source, its directivity, or its vibration. To characterize sources, other methods exist, such as near-field acoustical holography (NAH) which requires the microphones to be close to the source to measure the evanescent acoustic field. This last point does not meet our requirement for the microphones being at a distance from the source.

No assumption is made on the source directivity in the proposed beamforming method. The source directivity is treated as an unknown while previous beamforming works have assumed a known source model, most often a plane wave or a point source.

Specifically, the contributions of the thesis are

- (1) Development of an extension to beamforming methods, applicable to directional sources whose directivity and orientation are unknown [Bouchard 2009].
  - No assumption is made on the source directivity (treated as unknown).
  - The array output is a weighted sum of “sub-beamformers”, each beamforming on a different spatial mode of the source.
  - With the appropriate weighting (mode strengths), the sum of the sub-beamformers gives the same optimum array gain as beamforming directly with the exact source model.
  - The proposed beamforming method provides a source model more complex and more appropriate than the point source model to beamform on directional sources.
- (2) Development of several estimators of the mode strengths based on the observed data.
  - The estimator called *Gain-optimized* (GO) estimator is introduced and is computationally simple. Its bias and variance in the presence of noise are evaluated in the thesis [Bouchard 2009]. A comparison of the computational complexity between the proposed beamforming method with the GO estimator and conventional beamforming is included.

- 
- Two estimators based on least-squares are also presented, one of which, in addition to allowing the estimation of the source model also provides the estimation of the model for an interfering source if its location is known.
  - Three estimators, which are deemed relatively unsatisfactory, are also documented (appendix C).
- (3) Comparison of beamforming with the proposed beamforming method and using the point source model is realized with simulations and measurements for a simple directional source [Bouchard 2009].
- (4) Simulations and analysis of performance for the proposed beamforming method and the different estimators, with comparison to conventional beamforming, when beamforming on simple sources in the presence of
- diffuse noise,
  - diffuse noise and reverberation,
  - diffuse noise and an interferer.

A further contribution is the estimation of the performance penalty caused by beamforming for a human talker using a point source model without taking into account the fact that the phase of the human talker differs from that of a point source (Sub-Section 2.6.2). This was realized using phase information of the human talker available in the literature.

There has been little previous work on beamforming on directional sources so this thesis represents only the beginning of the work on the subject. The proposed beamforming method reduces to conventional beamforming if a point source model is assumed. The analysis therefore examines how the proposed beamforming method performs when beamforming on directional sources and compares this to how other methods using the point source model perform when beamforming on the same directional sources.

---

## 1.5 Thesis outline

Beamforming is presented in Chapter 2. This includes a discussion of source model, near-field beamforming, far-field beamforming, and conventional beamforming. Two array performance measures, including array gain, are then presented. This is followed by the summary of three categories of array design methods. A simulation example of beamforming on a theoretical directional source is then presented to demonstrate the impact of using the point source model in such a case [Bouchard 2008]. Finally, the human talker directivity is discussed to explain why the point source model has been used so extensively for human talkers and why this would not be valid for a distributed array.

To clearly show that there are sources which radiate very differently than monopoles, Chapter 3 presents some examples of directional sources from the literature, discussing the directivity of some musical instruments and mentioning aeroacoustic sources [Bouchard 2008].

The proposed beamforming method is introduced in Chapter 4. It is shown that with the appropriate weighting, the sum of the sub-beamformers in the method gives the same optimum array gain as directly beamforming with the exact source model. The chapter also presents the gain-optimized (GO) estimator of the weight to give each sub-beamformer based on the observed data. It is then shown that if the source directivity pattern and source location with respect to the array do not change with time, it is possible to obtain the overall beamformer output without computing the output of each sub-beamformer. In this case, the method is computationally equivalent to beamforming directly with a source model. Comparisons of the proposed beamforming method with the phase-mode spherical array and with near-field acoustical holography are then presented. The methods are quite different even though spherical harmonic expansions are used in those two methods.

Chapter 5 compares simulations and measurements for a simple directional source when beamforming with both the proposed beamforming method and using the point source model. The content of this chapter is the subject of a technical paper published in the *Journal of the Acoustical Society of America* [Bouchard 2009].



Chapter 6 revisits the GO estimator and introduces two additional estimators which are based on least-squares. The use of the proposed beamforming method with MVDR and estimation using many FFT frames are also briefly discussed in this chapter.

Because the mathematical expression for the array gain resulting from the proposed beamforming method with an estimator is too complex and not mathematically tractable, performance is mostly investigated through simulations. Chapter 7 presents simulation results and analysis for a source in the presence of diffuse noise, diffuse noise and reverberation, and diffuse noise and an interferer. Chapter 8 then presents a final discussion.

Appendices include an introduction to spherical harmonics needed for the development of the proposed beamforming method, a mathematical computation of the bias and variance of the GO estimator, and three developments leading to less promising estimators. A comparison of the computational complexity between the proposed beamforming method with the GO estimator and conventional beamforming is also included.

---

# CHAPTER 2

---

## Arrays and Beamforming

### 2.1 Introduction

This chapter presents the concept of beamforming. This includes a discussion of source models, near-field beamforming, far-field beamforming, and conventional beamforming. Two array performance measures including array gain are presented. This is followed by the summary of three categories of array design methods. A simulation example of beamforming on a theoretical directional source is then presented to demonstrate the impact of using the point source model in such a case [Bouchard 2008]. Finally, human talker directivity is discussed to explain why the point source model is used extensively and why a new approach is needed for a distributed array.

### 2.2 Beamforming

Mathematically, the continuous-time complex exponential  $e^{i2\pi ft}$  is used here for the time dependence. Where no ambiguity exists, the continuous-time and discrete-time cases are distinguished by the use of  $t$  (in seconds) or  $n$  (sample index) as the time parameter, and the use of  $f$  (in Hertz) or  $k$  (frequency bin index) as the frequency parameter, respectively.

Figure 2.1 shows an array of  $L$  discrete sensors, connected to a signal processing block called *beamformer*. The beamformer is represented as a filter-and-sum operation in the time-domain, but it could also have been represented in the frequency-domain.

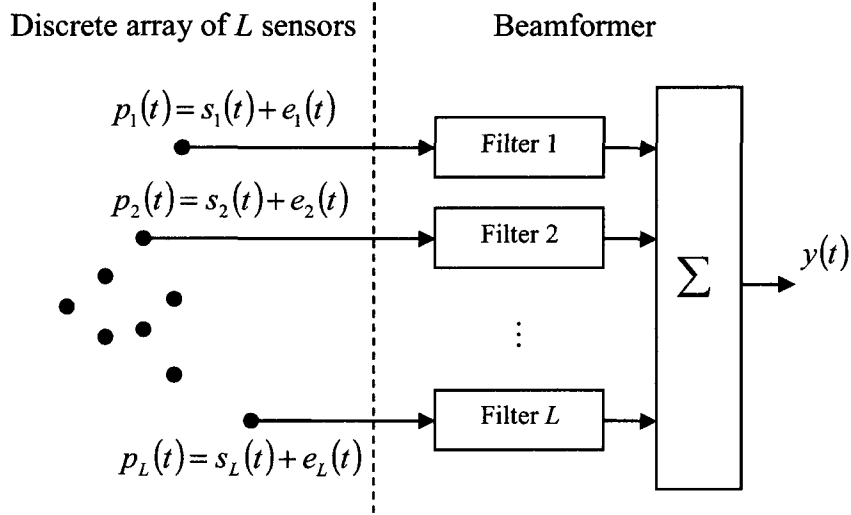


Figure 2.1. Beamforming.

Microphone  $l$  is physically located at the point  $(r_l, \theta_l, \phi_l)$  in a spherical coordinate system. In the time-domain, its output is  $p_l(t)$ . It is composed of a component  $s(t, r_l, \theta_l, \phi_l)$  from the source of interest, and an additive noise component  $e_l(t)$ . The noise can be composed of electrical noise, acoustical noise, signal from interfering sources, and components to account for sensor positioning errors. The noise is assumed to be uncorrelated with the source signal  $s(t, r_l, \theta_l, \phi_l)$ . We assume in the thesis that the microphones are omnidirectional.

The output of microphone  $l$  in the frequency-domain,  $p_l(f)$ , is obtained by applying the forward continuous-time Fourier transform to its time-domain output  $p_l(t)$ ,

$$p_l(f) = \int_{-\infty}^{\infty} p_l(t) e^{-i2\pi ft} dt, \quad (2.1)$$

provided, of course, that the integral exist. In the discrete-time case, a normalized Discrete Fourier Transform (DFT) (or normalized Fast Fourier Transform (FFT) in practice) can be used to obtain  $p_l(k)$  as

$$p_l(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} p_l(n) e^{-i2\pi kn/N}. \quad (2.2)$$

The reason for normalization by  $\sqrt{N}$ , where  $N$  is the number of time samples used in the transform, will become clear when discussing the proposed beamforming method and the gain optimized estimator.

In the frequency-domain, the beamformed output of an array of  $L$  microphones is defined as

$$y(f) = \sum_{l=1}^L w_l^*(f) p_l(f), \quad (2.3)$$

where the beamforming weight  $w_l(f)$  for the  $l^{\text{th}}$  microphone is generally a complex number. The superscript asterisk denotes complex conjugation. Using vector notation for each frequency, the beamforming weights corresponding to each array microphone are assembled in the weight vector  $\mathbf{w}(f)$ , and the microphone signals are assembled in a microphone output vector  $\mathbf{p}(f) = \mathbf{s}(f) + \mathbf{e}(f)$ . The vector  $\mathbf{s}(f)$  is called the *source signal* while  $\mathbf{e}(f)$  is the *noise vector*. The array output becomes

$$y(f) = \mathbf{w}^H(f) \mathbf{p}(f). \quad (2.4)$$

The superscript  $H$  indicates the Hermitian conjugate transpose.

### 2.2.1 Source model

As will be seen later, the beamforming weight vector  $\mathbf{w}(f)$  may be optimized in regard to some criteria, such as the signal-to-noise ratio (SNR) or directional response pattern. However to do that, it is necessary to have a model of the wavefront arriving at the array.

---

The source has some directivity, and it is located at some position and with some orientation with respect to the array. If there was no noise, for that source directivity, orientation, and position, the microphone outputs would be  $s_{\text{mo}}(f)$ . This is the source model vector or, simply, the *source model*. It accounts for the source directivity, orientation, location, and for the sound propagation.

The computation of the beamforming weight vector is done using the source model. Ideally, the source model and source signal would be identical. In practice, the exact source directivity and orientation are unknown and the source model used to compute the beamforming weights will differ from the actual source signal. The source model should be based on the actual source on which the array is beamforming.

### 2.2.2 Near-field, far-field, and conventional beamforming

In what is often called *near-field beamforming*, a point source (monopole) model is typically used [Ryan 2003]. If wavefront curvature is negligible, a plane wave model is used and the beamforming is said to be *far-field* [Ryan 1998].

The terms “near-field” and “far-field” can also refer to the source and are not to be confused with near-field and far-field beamforming. The near-field of the source is where evanescent sound components may exist while the far-field of the source is where “the angular field distribution is essentially independent of the distance” [Balanis 1997; p.33]. As an example, microphones surrounding a point source may be in the far-field of the source, however wavefront curvature cannot be neglected to use far-field beamforming.

In this thesis, the expression *conventional beamforming* will mean beamforming directly with the source model and the source model being a point source. This terminology makes no distinction between beamforming with a point source in the near-field or the far-field of an array; the latter case is just a point source located at infinity so that the wavefront is considered planar when it arrives on the array.

## 2.3 Performance measures

This section presents two array performance measures: array gain (and related white-noise gain) and the beampattern.

### 2.3.1 Array gain and white-noise gain

The *array gain* is a measure of the improvement in signal-to-noise ratio [Monzingo 1980] [Cox 1986]. It is defined as the ratio of the SNR at the output of the beamforming array to the SNR at the input (see below) of the array

$$G(f) = \frac{\text{SNR}_{\text{out}}(f)}{\text{SNR}_{\text{in}}(f)}. \quad (2.5)$$

The received source signal strength may be different at each microphone so a reference array input SNR is defined here, as it is in [Cox 1986], to be

$$\text{SNR}_{\text{in}}(f) = \frac{\sigma_s^2(f)}{\sigma_e^2(f)} \quad (2.6)$$

where,

$$\sigma_s^2(f) = \mathbf{s}^H(f)\mathbf{s}(f)/L \quad (2.7)$$

is the *average signal power spectral density* over the  $L$  microphones of the array, and

$$\sigma_e^2(f) = E\{\mathbf{e}^H(f)\mathbf{e}(f)\}/L \quad (2.8)$$

is the *average noise power spectral density* over the array.

The SNR at the output of the array is the output signal power density divided by the output noise power density

$$\text{SNR}_{\text{out}}(f) = \frac{|\mathbf{w}^H(f)\mathbf{s}(f)|^2}{E\{|\mathbf{w}^H(f)\mathbf{e}(f)|^2\}}. \quad (2.9)$$

Defining a *normalized signal cross-power spectral density matrix*

$$\mathbf{R}_{ss}(f) = \mathbf{s}(f)\mathbf{s}^H(f) / \sigma_s^2(f) \quad (2.10)$$

and a *normalized noise cross-power spectral density matrix*

$$\mathbf{R}_{ee}(f) = E\{\mathbf{e}(f)\mathbf{e}^H(f)\} / \sigma_e^2(f), \quad (2.11)$$

and using the definitions above in Eq. (2.5), the array gain at a specific frequency becomes

$$G(f) = \frac{\mathbf{w}^H(f)\mathbf{R}_{ss}(f)\mathbf{w}(f)}{\mathbf{w}^H(f)\mathbf{R}_{ee}(f)\mathbf{w}(f)}. \quad (2.12)$$

The beamforming weight vector  $\mathbf{w}(f)$  is based on the source model  $\mathbf{s}_{mo}(f)$  while the normalized signal cross-power spectral density matrix  $\mathbf{R}_{ss}(f)$  depends on the source signal  $\mathbf{s}(f)$ .

When the noise at the microphone outputs is additive white-noise uncorrelated from sensor to sensor, such as in spatially uncorrelated white-noise, and of equal power in each sensor,  $\mathbf{R}_{ee}(f)$  is equal to the identity matrix  $\mathbf{I}_L$ . (In this thesis, when referring to white-noise, it will always be assumed that the noise power is equal in each sensor.) In this case the array gain is called the *white-noise gain* and Eq. (2.12) becomes

$$G_w(f) = \frac{\mathbf{w}^H(f)\mathbf{R}_{ss}(f)\mathbf{w}(f)}{\mathbf{w}^H(f)\mathbf{w}(f)}. \quad (2.13)$$

This definition of white-noise gain, seen in [Cox 1986] [Ryan 2003] [Van Trees 2002; p.69], is not to be confused with others, for example the definition in [Van Veen 1988; p.12]: “[...] the output power due to unit variance white noise at the sensors,” which would equal  $\mathbf{w}^H(f)\mathbf{w}(f)$ . Only the former definition, as given in Eq. (2.13), will be used in this thesis.

### 2.3.2 Beampattern

The beampattern is a plot used to illustrate the spatial selectivity of a beamforming array. The beampattern represents the output magnitude of the array for

different source positions while the beamforming weight vector is kept constant. The plot is most often presented in dB, normalized so that the position where the array is steered has an output magnitude of 0 dB.

For conventional far-field beamforming (plane wave source model) the beampattern, at each frequency, is simply a function of arrival direction

$$B(f, \theta, \phi) = \left| \mathbf{w}_{pw}^H(f) \mathbf{s}_{pw}(f, \theta, \phi) \right|. \quad (2.14)$$

Again in Eq. (2.14), it is assumed that for any direction  $(\theta, \phi)$  the received  $\mathbf{s}_{pw}(f, \theta, \phi)$  signal originates from a source with a constant amplitude and phase. The beampattern can be represented as a single 3-D figure or as 2-D array response plots for specified planes (e.g. horizontal plane, vertical plane with  $\phi = 0$ , etc).

For near-field beamforming (point source model), the beampattern also depends upon the distance from the source to the receiving array

$$B(f, r, \theta, \phi) = \left| \mathbf{w}_{ps}^H(f) \mathbf{s}_{ps}(f, r, \theta, \phi) \right|. \quad (2.15)$$

For a plot presenting the array response with respect to the direction of arrival, it is necessary to specify the distance from the source to the array used for the computation.

For directional sources, the beampattern will depend also on the source directivity and orientation:

$$B(f, r, \theta, \phi, \text{source directivity and orientation}) = \left| \mathbf{w}^H(f) \mathbf{s}(f, r, \theta, \phi) \right|. \quad (2.16)$$

It should be noted that  $\mathbf{s}(f, r, \theta, \phi)$  implicitly depends upon source directivity and orientation. The actual application should dictate in function of what variables (e.g.  $r$ ,  $\theta$ ,  $\phi$ ,  $x$ ,  $y$ ,  $z$ , source directivity, orientation) to plot the array response.

One use of the beampattern is to evaluate the signal-to-interference ratio (SIR) at the output of the array. If all sources radiate as plane waves or all sources radiate as point sources, a single equation (Eq. (2.14) or Eq. (2.15) respectively) is used for the source and the interferer. The SIR can be gathered visually from a single plot by observing the difference in dB between the output magnitudes for the source position and the interferer



---

position. Or if the interferer position is unknown, the worst case scenario can still be found by observing at which position, other than the source, the maximum output magnitude is located.

Interpretation of the beampattern becomes more difficult for directional sources if the source and the interferer have different directivities or orientations. A different source signal vector  $\mathbf{s}(f, r, \theta, \phi)$  must then be used for the source and the interferer. Equation (2.16) has to be computed twice, once with the source signal vector for the source and once with the source signal vector for the interferer. Thus, the SIR cannot be obtained from a single plot. Furthermore, it is very difficult to evaluate the worst case scenario because it will depend, in addition to the interferer position, on the interferer directivity and orientation. This is why the beampattern is seldom used in the thesis.

## **2.4 Design method: picking beamforming weights**

The design of the beamformer is done by choosing appropriate beamforming weights. This section presents three categories of methods to choose those beamforming weights: data independent beamforming, statistically optimum beamforming, and adaptive beamforming. In all cases, it is necessary to have a source model. If the source model is different from how the actual source radiates, performances will not be as expected.

### **2.4.1 Data independent beamforming**

In data independent beamforming, the beamforming weights are computed based on what is known of the spatial position of the source of interest (and the position of interferers if some are present). This allows specifying a desired beampattern from which the required beamforming weight vector is found. The beamforming weights are not computed using the microphone output signals or the statistics of those signals, unlike the case of statistically optimum beamforming to be described later.

Many deterministic design techniques exist to find the beamforming weight vector [Van Trees 2002]. For example, for linear arrays with equal spacing between sensors and the plane wave model, techniques similar to spectral analysis of time series

(e.g. windowing, etc.) or FIR filters design (frequency sampling, optimum minimax error, etc.) can be used to find the beamforming weights. For array design, specifications are put on the beampattern to realize the spatial filtering while specifications are put on the frequency response for time signal filtering. For linear arrays with unequal spacing between sensors, the problem is more difficult but methods also exist.

In Chapter 4, to compare with the proposed beamforming method, the phase-mode spherical array (eigenbeamformer) will be developed as a statistically optimum beamformer. However, the phase-mode spherical array is most often presented as a data independent beamformer, see for example [Meyer 2002] where the design is to a desired beampattern, or [Rafaely 2004] where the design is for plane wave decomposition of the sound field.

The frequency invariant beamformer discussed in [Ward 1995] is also an example of data independent beamformer. In this case, proper positioning of the sensors and a proper choice of the filters at the output of those sensors allow the array to have a beampattern that is invariant over a broad frequency band.

## 2.4.2 Statistically optimum beamforming

In statistically optimum beamforming, the beamforming weights are computed based on the statistics of the microphone output signals. Two approaches will be discussed: 1) Maximization of signal-to-noise ratio and 2) Linearly Constrained Minimum Variance beamforming (LCMV). Other approaches include the multiple sidelobe canceller [Applebaum 1976], the use of a reference signal [Widrow 1967], and the Generalized Sidelobe Canceller [Griffiths 1982] which represents an alternative formulation of LCMV.

### *Maximization of signal-to-noise ratio*

In this approach, the array gain is maximized. Provided that  $\mathbf{R}_{ee}(f)$  is not singular, the array gain given by Eq.(2.12) is maximized with the following beamforming weight vector [Monzingo 1980] [Cox 1986]

$$\mathbf{w}_{\text{opt}}(f) = \alpha(f) \mathbf{R}_{ee}^{-1}(f) \mathbf{s}(f). \quad (2.17)$$

The scaling factor  $\alpha(f)$  is an arbitrary complex constant and it is set to unity in the remainder of the thesis. Using Eq. (2.17) in Eq. (2.12), the optimum array gain is

$$G_{\text{opt}}(f) = \frac{\mathbf{s}^H(f) \mathbf{R}_{\text{ee}}^{-1}(f) \mathbf{R}_{\text{ss}}(f) \mathbf{R}_{\text{ee}}^{-1}(f) \mathbf{s}(f)}{\mathbf{s}^H(f) \mathbf{R}_{\text{ee}}^{-1}(f) \mathbf{s}(f)}. \quad (2.18)$$

When the noise at the microphone outputs is additive white-noise uncorrelated from sensor to sensor, such as in spatially uncorrelated white-noise,  $\mathbf{R}_{\text{ee}}(f)$  is equal to the identity matrix  $\mathbf{I}_L$ . As described earlier, the array gain is then called white-noise gain and the optimum weight vector becomes equal to the source signal itself

$$\mathbf{w}_{\text{w,opt}}(f) = \mathbf{s}(f) \quad \text{when } \mathbf{R}_{\text{ee}} = \mathbf{I}_L. \quad (2.19)$$

The optimum white-noise gain is  $G_{\text{w,opt}}(f) = L$ , it equals the number of microphones.

Of course, the array designer does not have access to the exact source signal  $\mathbf{s}(f)$  and the source model  $\mathbf{s}_{\text{mo}}(f)$  is used instead to compute the beamforming weight vector. If the source model is different from the source signal, the resulting array gain will be less than the optimum value.

As a side note, the maximum array gain may not be achieved in practice even with the optimum beamforming weight vector found for the specific  $\mathbf{R}_{\text{ee}}(f)$  and the exact source signal as the source model. An optimum array gain requires a highly tuned system and can be very sensitive to random errors such as amplitude and phase errors in the microphone channels, or to slightly inexact microphone positions. Those errors are nearly uncorrelated from sensor to sensor and can be seen as adding uncorrelated white-noise to the output of each sensor. White-noise gain is thus a measure of robustness. With the optimum beamforming weight vector to maximize the array gain in the presence of the spatial noise characterized by  $\mathbf{R}_{\text{ee}}(f)$ , the white-noise gain can be very low. The beamformer is then very sensitive to errors. The solution is to find the beamforming weight vector that maximizes the array gain subject to an equality constraint on the white-noise gain [Cox 1987] [Cox 1986], mathematically, expressed as

$$\max_{\mathbf{w}(f)} G(f) \quad \text{subject to} \quad G_w(f) = \delta^2 \leq L. \quad (2.20)$$

The solution can be found using a Lagrange multiplier  $\varepsilon$  and minimizing  $1/G(f) + \varepsilon/G_w$ , that is

$$\min_{\mathbf{w}(f)} \frac{\mathbf{w}^H(f) \{ \mathbf{R}_{ee}(f) + \varepsilon \mathbf{I} \} \mathbf{w}(f)}{\mathbf{w}^H(f) \mathbf{R}_{ss}(f) \mathbf{w}(f)} \quad (2.21)$$

where  $\mathbf{I}$  is the identity matrix. The solution is [Cox 1986]

$$\mathbf{w}_{\text{opt, w constr}}(f) = \{ \mathbf{R}_{ee}(f) + \varepsilon \mathbf{I} \}^{-1} \mathbf{s}(f). \quad (2.22)$$

When  $\varepsilon$  is equal to zero, there is no constraint on the white-noise gain and the optimum array gain is obtained. When  $\varepsilon$  is equal to infinity, the optimum white-noise gain is obtained.  $\varepsilon$  provides a continuous monotonic control over the white-noise constraint, from no constraint ( $\varepsilon$  is equal to zero) to optimization for maximum white-noise gain ( $\varepsilon$  is equal to infinity).

Unfortunately, there is no simple relation between the desired white-noise gain  $\delta^2$  and the value of  $\varepsilon$ . An iterative process is required. A first value of  $\varepsilon$  is used to find the beamforming weight vector  $\mathbf{w}_{\text{opt, w constr}}(f)$  which is then used to compute the resulting white-noise gain. If the white-noise gain is lower than  $\delta^2$ ,  $\varepsilon$  is increased. If the white-noise gain is higher than  $\delta^2$ ,  $\varepsilon$  is decreased. A new beamforming weight vector  $\mathbf{w}_{\text{opt, w constr}}(f)$  is computed and the process repeated.

The resulting array gain will not be as high as the optimum array gain but will be less sensitive to the errors mentioned above.

### ***Linearly Constrained Minimum Variance beamforming (LCMV)***

With *Linearly Constrained Minimum Variance* (LCMV) beamforming [Frost 1972] [Van Veen 1988] [Van Trees 2002], the array response is specified for the position of the source of interest. The beamforming weights are then found to minimize the array output power subject to the constraint(s). This method preserves the desired signal while attenuating noise and interfering signals at positions other than the source of interest. If

there are known interferers where a null response is desired, additional constraints with an array response of zero can also be included.

The desired array response  $r_d(f, r_m, \theta_m, \phi_m)$  is specified at  $M$  points  $(r_m, \theta_m, \phi_m)$ . For example, the desired array response can be set to one for the position of the source and to zero for the position of known interferers. (Other types of constraints exist, for example derivative, eigenvector, or quiescent pattern based constraints [Van Trees 2002].)

A source model  $\mathbf{s}_{\text{mo},m}(f, r_m, \theta_m, \phi_m)$  must be specified for each of those locations. The source models are combined in a *constraint matrix*

$$\mathbf{C}(f) = [\mathbf{s}_{\text{mo},1}(f, r_1, \theta_1, \phi_1) \quad \mathbf{s}_{\text{mo},2}(f, r_2, \theta_2, \phi_2) \quad \dots \quad \mathbf{s}_{\text{mo},M}(f, r_M, \theta_M, \phi_M)] \quad (2.23)$$

and the desired responses in a *response vector*

$$\mathbf{r}_d(f) = [r_d(f, r_1, \theta_1, \phi_1) \quad r_d(f, r_2, \theta_2, \phi_2) \quad \dots \quad r_d(f, r_M, \theta_M, \phi_M)]^H. \quad (2.24)$$

For LCMV, the beamforming weight vector is then found such that

$$\min_{\mathbf{w}(f)} \mathbf{w}^H(f) \mathbf{R}_{\text{pp}}(f) \mathbf{w}(f) \quad \text{subject to} \quad \mathbf{C}^H(f) \mathbf{w}(f) = \mathbf{r}_d(f), \quad (2.25)$$

where  $\mathbf{R}_{\text{pp}}(f)$  is the *microphone output cross-power spectral density matrix*. Note that  $\mathbf{w}^H(f) \mathbf{R}_{\text{pp}}(f) \mathbf{w}(f)$  represents the overall (i.e. target plus noise) array output power at frequency  $f$ . The constraints are assumed to be linearly independent so  $\mathbf{C}(f)$  has rank  $M$ . The solution for the beamforming weight vector is [Van Veen 1988]

$$\mathbf{w}(f) = \mathbf{R}_{\text{pp}}^{-1}(f) \mathbf{C}(f) [\mathbf{C}^H(f) \mathbf{R}_{\text{pp}}^{-1}(f) \mathbf{C}(f)]^{-1} \mathbf{r}_d(f). \quad (2.26)$$

It is also possible to do the minimization with respect to the normalized noise cross-power spectral density matrix  $\mathbf{R}_{\text{ee}}(f)$ , in which case the beamforming weight vector is found such that

$$\min_{\mathbf{w}(f)} \mathbf{w}^H(f) \mathbf{R}_{\text{ee}}(f) \mathbf{w}(f) \quad \text{subject to} \quad \mathbf{C}^H(f) \mathbf{w}(f) = \mathbf{r}_d(f), \quad (2.27)$$

whose solution has the same form as Eq. (2.26) with  $\mathbf{R}_{ee}(f)$  replacing  $\mathbf{R}_{pp}(f)$ . In the majority of the literature, there is no distinction in name between the approach using  $\mathbf{R}_{ee}(f)$  and the one using  $\mathbf{R}_{pp}(f)$ . However, Van Trees calls Linear Constrained Minimum Variance (LCMV) the case using  $\mathbf{R}_{ee}(f)$  and Linear Constrained Minimum Power (LCMP) the case using  $\mathbf{R}_{pp}(f)$ .

The *Minimum Variance Distortionless Response* (MVDR) beamformer is a special case of LCMV where there is only one constraint for which the desired response is equal to one

$$\mathbf{s}_{mo}^H(f, r_1, \theta_1, \phi_1) \mathbf{w}(f) = 1. \quad (2.28)$$

Van Trees calls *Minimum Variance Distortionless Response* (MVDR) the case using  $\mathbf{R}_{ee}(f)$  to find the beamforming weight and *Minimum Power Distortionless Response* (MPDR) the case using  $\mathbf{R}_{pp}(f)$ .

### 2.4.3 Adaptive beamforming

The methods described in the statistically optimum beamforming section require knowledge of either the microphone output cross-power spectral density matrix  $\mathbf{R}_{pp}(f)$  or the noise cross-power spectral density matrix  $\mathbf{R}_{ee}(f)$ . These matrices are generally unknown and must be estimated. If signal statistics changes with time, the matrices will have to be estimated periodically. Then to find the beamforming weight vector, for example using Eqs. (2.17) or (2.26), the cross-power spectral density matrix must be inverted. With a large number of microphones, matrix inversions become computationally complex and so is computing directly the beamforming weight vector.

Adaptive methods usually apply iterative estimation to reduce computations, particularly avoiding large matrix inversions. It is possible to implement statistically optimum approaches using the Least Mean Square (LMS) or Recursive Least Squares (RLS) algorithms [Van Veen 1988] [Van Trees 2002].

## 2.5 Example of beamforming on a directional source

This section provides a simulation example to show the consequences of using the point source model to beamform on a directional source. This example was part of a presentation at the 156<sup>th</sup> Meeting of the Acoustical Society of America [Bouchard 2008].

### 2.5.1 Beamforming on a point source

The point source is located at the center of a circle of radius  $r$  equal to eight times the wavelength  $\lambda = c/f$  ( $c$  is the speed of sound), lying in the  $x$ - $y$  plane. There are 102 microphone positions equally spaced on the circle. For an array of one microphone, only the position at the extreme right of the source is used. To increase the number of microphones, microphones are added in pairs, with one on each side of the existing microphones. The adjacent element spacing does not change. There is always an odd number of microphones in the array, except when 102 microphones are included. Figure 2.2 shows examples for arrays of 1, 5, and 51 microphones.

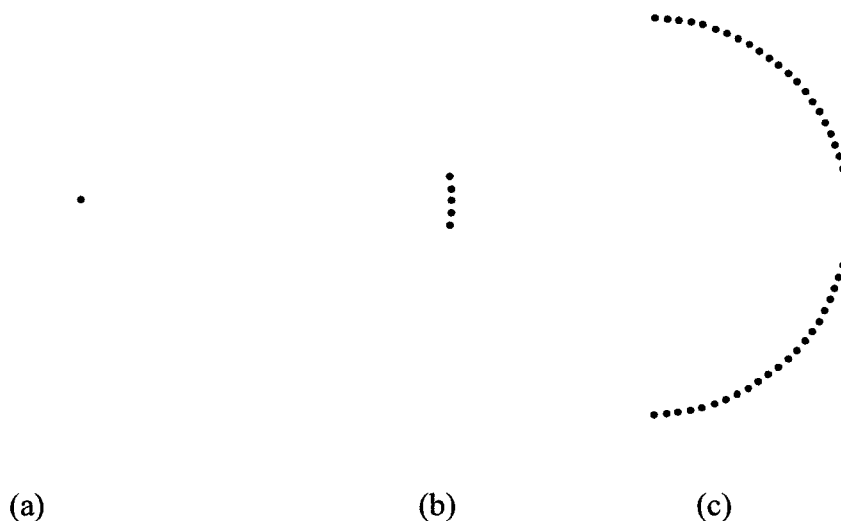


Figure 2.2. Array geometry for arrays of (a) 1 microphone; (b) 5 microphones; (c) 51 microphones.

Figure 2.3(a) shows the 102 microphone positions, with source directivity and location. The source is a point source whose location is shown with the dot at the center

and its directivity is shown as the continuous circle. The received signal at the microphones is

$$s_{\text{ps}}(f, r, \theta, \phi) = Y_0^0(\theta, \phi) h_0^{(2)}(\kappa r) = \frac{1}{\sqrt{4\pi}} i \frac{e^{-i\kappa r}}{\kappa r}, \quad (2.29)$$

where  $Y_0^0(\theta, \phi)$  and  $h_0^{(2)}(\kappa r)$  are the spherical harmonic function of order 0 and the Hankel function of the second kind of order 0, respectively (see Appendix A).  $\kappa = 2\pi f/c$  is the wave number.

Figure 2.3(b) shows the white-noise gain as a function of the number of microphones in the array using the beamforming weights  $w_{\text{ps}}(f, r, \theta, \phi) = s_{\text{ps}}(f, r, \theta, \phi)$ . The maximum white-noise gain should equal the number of microphones in the array. This is indeed the case on the figure. This indicates that the point source is the optimum source model when beamforming on a point source.

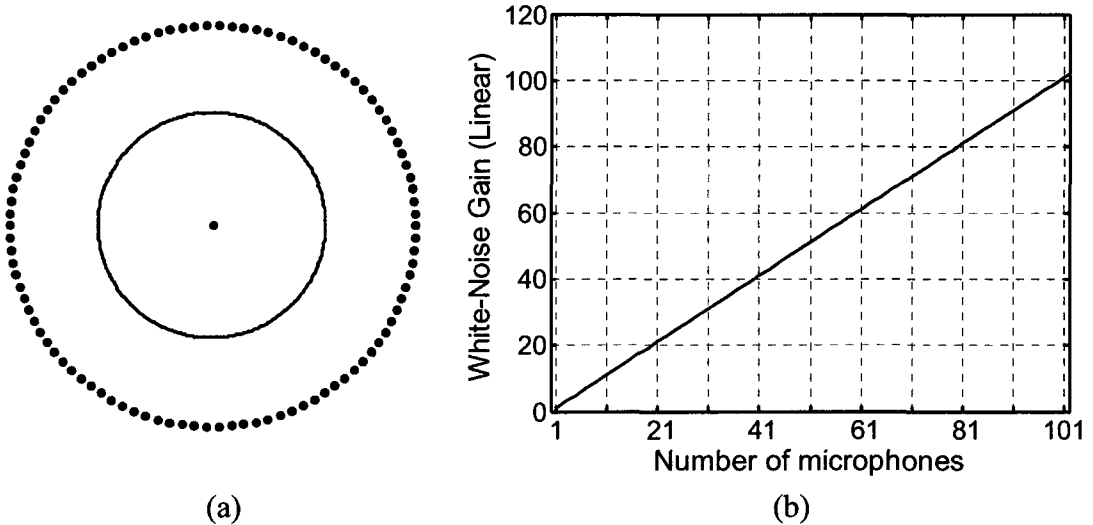


Figure 2.3. Beamforming on a point source (a) Array geometry and source directivity; (b) White-noise gain for conventional beamforming and a varying number of microphones in the array.



## 2.5.2 Beamforming on a quadrupole

Keeping the same scheme for the array, but now using a quadrupole source in the  $x$ - $y$  plane with directivity as shown in Fig. 2.4(a). The “+” and “-” signs indicate that adjacent lobes have opposite phases. The received signal at the microphones is

$$\begin{aligned} s_{\text{quad}}(f, r, \theta, \phi) &= [Y_2^{-2}(\theta, \phi) + Y_2^2(\theta, \phi)] h_2^{(2)}(\kappa r) \\ &= 6\sqrt{\frac{5}{96}} \sin^2 \theta \cos(2\phi) \left[ -i \frac{e^{-i\kappa r}}{\kappa r} - 3 \frac{e^{-i\kappa r}}{(\kappa r)^2} + 3i \frac{e^{-i\kappa r}}{(\kappa r)^3} \right], \end{aligned} \quad (2.30)$$

where  $Y_2^{-2}(\theta, \phi)$  and  $Y_2^2(\theta, \phi)$  are the spherical harmonic function of order 2, degrees -2 and 2, respectively.  $h_2^{(2)}(\kappa r)$  is the Hankel function of the second kind of order 2 (see Appendix A).

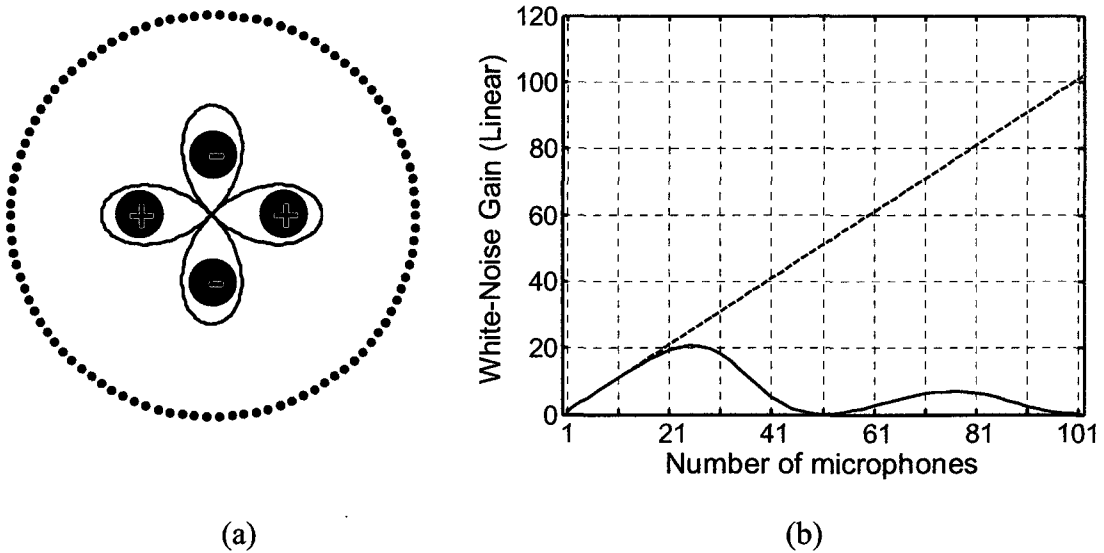


Figure 2.4. Beamforming on a quadrupole (a) Array geometry and source directivity; (b) White-noise gain for a varying number of microphones in the array, (---) quadrupole source model, (—) conventional beamforming.

The continuous curve in Fig. 2.4(b) is the white-noise gain as a function of the number of sensors for conventional beamforming, which uses the beamforming weight  $w_{\text{ps}}(f, r, \theta, \phi) = s_{\text{ps}}(f, r, \theta, \phi)$ . It can be seen that conventional beamforming no longer

gives the maximum white-noise gain. Of course, this being a simulation, we know the directivity and orientation of the source so we can easily use a quadrupole with the same directivity and orientation as the source model. The dashed line is the white-noise gain when this source model is used (the beamforming weight is  $w(f, r, \theta, \phi) = s_{\text{quad}}(f, r, \theta, \phi)$ ). In this case the optimum white-noise gain is obtained.

In practice, the source directivity and orientation would most likely be unknown so the exact source model could not be used. The proposed beamforming method could however be used. The dotted line on Fig. 2.5 is the white-noise gain for the proposed beamforming method with the GO estimator and  $N_{\text{order}} = 4$  in the absence of noise at the output of the microphones. The maximum white-noise gain penalty is approximately 0.9 dB, for an array of 39 microphones. For this number of microphones, the optimum white-noise gain is 15.9 dB. The penalty comes from the bias of the estimator. The proposed beamforming method is introduced in detail in Chapter 4.

Back to Fig. 2.4(b), we also observe that for an array of approximately 15 microphones or less, the white-noise gain is similar for both source models. The point source may be an adequate source model in this particular case. The validity of using a point source as the source model will depend on the array geometry, on the source directivity, orientation, and position as well as on the array output optimization criterion. To be more specific, if the array subtends a small angle about the source, the point source model may be acceptable, in particular if the array observes a small portion of a single lobe of the directivity pattern. This is clearly not the case in the example here when there are more than 20 microphones. This thesis focuses on such cases of arrays covering a large angular extent of the source.

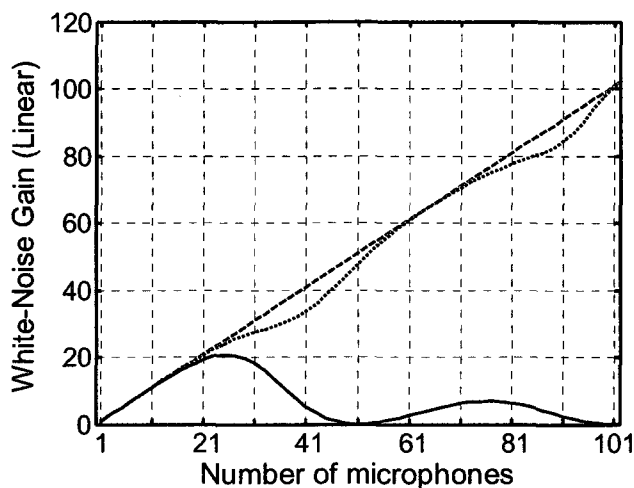


Figure 2.5. Beamforming on a quadrupole. White-noise gain for a varying number of microphones in the array, (---) quadrupole source model, ( $\cdot \cdot \cdot$ ) proposed beamforming method with GO estimator, (—) conventional beamforming.

## 2.6 Human talker directionality and why the point source model has been used

The directivity of the human talker is now discussed to explain why the point source model has been used so extensively. The directivity magnitude of the human talker is first examined. There are few measurements in the literature for directivity phase of the human talker, from what is available we try to estimate the impact of the phase on array performance.

### 2.6.1 Directivity magnitude

Measurements of the magnitude of the directivity for human talkers are provided in [Dunn 1939] [Chu 2002] [Warnock 2002]. As an example, Figure 2.6 shows the directivity magnitude of the human talker in the horizontal plane at the height of the mouth, for three frequency bands, plotted from Table I of [Dunn 1939]. These measurements were performed at 30 cm from the mouth. The plot corresponds to the talker facing the right of the figure, with his left shoulder toward the top and his back toward the left. It is observed that the magnitude is fairly constant in the hemisphere in

front of the talker. Furthermore, with arrays of small dimensions, only a portion of the hemisphere is observed. This explains why the point source model is often successfully used to beamform with microphone arrays on human talkers.

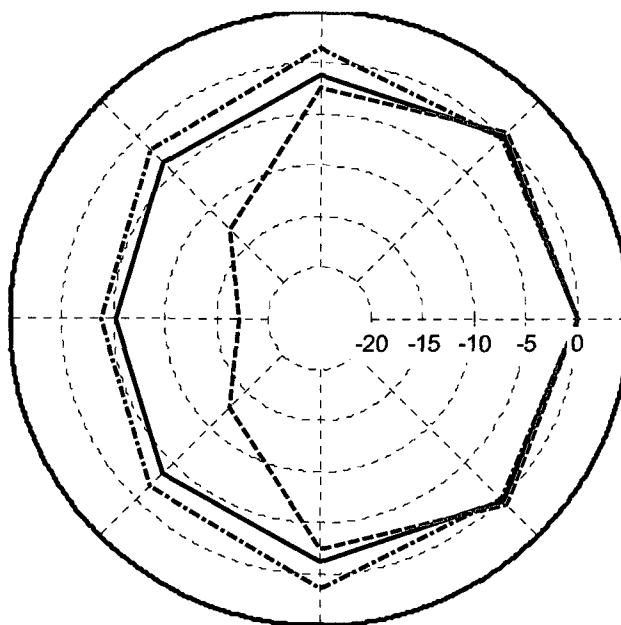


Figure 2.6. Human talker directivity magnitude in dB, measured at 30 cm from the mouth in the horizontal plane, in three frequency bands: (—) 250-500 Hz, (· -) 700-1000 Hz, (- -) 2800-4000 Hz (Plotted using selected data from Table I of [Dunn 1939]).

However, the figure shows a significant change in directivity from the hemisphere in front of the talker to the hemisphere at his back. In the context of an array surrounding the talker, this source directivity would impact the beamforming performances if the point source was used as the source model. Beamforming simulation results using curve fitting of the data from [Dunn 1939] to obtain the directivity magnitude used for both the source signal and the source model are presented in [Betlehem 2003]. Though they have used a directivity magnitude that changes with direction angle, they seem to have modeled the phase as that of a point source since [Dunn 1939] does not contain phase information. They simulated a 16-element circular array of 2.2 m of radius surrounding the source. Modest improvements to performance were obtained compared to using the point source model.

## 2.6.2 Impact of the phase

Few phase measurements have been published on the human talker. Flanagan presents measurements for a life-size mannequin that includes some phase information [Flanagan 1960]. The phase of the sound pressure at a distance of 30.48 cm (12 in.) from the mouth, relative to its value on the mouth axis, was measured at points in the vertical and horizontal planes in the frontal hemisphere. Phase values for frequencies of 1 kHz and 4 kHz in the horizontal plane, taken from Table I of [Flanagan 1960] are presented in Table 2.1. If the source would behave like a point source, the relative phase would be exactly zero at all points at a constant radius from the source.

Table 2.1. Phase of the sound pressure in degrees relative to the phase on the mouth axis, in the horizontal plane for a life-size mannequin (Values from Table I of [Flanagan 1960]).

Frequency	1 kHz	4 kHz
Angle(deg.)	Rel. phase (deg.)	
-90	-20	25
-80	-10	30
-60	-5	20
-40	0	20
-20	-5	10
0	0	0
20	-5	0
40	-5	10
60	-10	20
80	-10	30
90	-20	30

To estimate the impact of using a point source model to find the beamforming weights, values from Table 2.1 can be taken as the phase at the outputs of 11 microphones, all at the same distance from the source and forming an array covering 180 degrees of the human talker. A source signal magnitude of one at the 11 sensors is used so that only the phase behavior is different from that of a point source. (With the point source model, all the beamforming weights are equal to one with a phase of zero.) To maximize the white-noise gain, the optimum beamforming weights are found using Eq. (2.19), and are equal to the source signal. The white-noise gain is computed using Eq. (2.13). The maximum white-noise gain for any array of 11 sensors is  $10 \log(11) = 10.4 \text{ dB}$ . We find that the penalty for using the point source model to find the beamforming weights instead of the source signal itself is 2.2 dB at 1 kHz and 6.8 dB at 4 kHz, as summarized in Table 2.2.

Table 2.2. Estimation of the penalty for using the point source model phase to beamform with an array of 11 microphones on a human talker.

Frequency	1 kHz	4 kHz
Array	Penalty (dB)	
Covering 180 degrees	2.2	6.8
Covering 20 degrees	0.3	1.3

Values from Table 2.1 can also be used to show that if the array subtends a small angle about the talker, the impact of the phase difference with respect to the point source model has only a small effect on the white-noise gain. For example using 11 sensors again, with five sensors having the phase error for an angle position of 40 degrees (-5 deg. at 1 kHz, 10 deg. at 4 kHz) and six sensors having the phase error for an angle position of 60 degrees (10 deg. at 1 kHz, 20 deg. at 4 kHz), the penalty for using the point source model to find the beamforming weights is now 0.3 dB at 1 kHz and 1.3 dB at 4 kHz. Those results are also included in Table 2.2.

---

## 2.7 Chapter conclusion

This chapter presented the mathematics of beamforming, array performance measures, and methods to choose the beamforming weights.

A simulation example was discussed to show the impact of using the point source model on a directional source, demonstrating that it is necessary to have an appropriate source model to perform beamforming. When the array subtends a small angle about the source and the array observes a small portion of a single lobe of the source directivity pattern, the point source model may be adequate. It is no longer the case when the array observes many lobes as shown in the quadrupole example.

The directivity of the human talker was then presented to examine the validity of using the point source model when the array does not subtend a small angle about the source. It was shown that the point source model may not be valid when the array covers a larger region about the talker. There is little information available in the literature concerning the phase of the human talker. Still, with what is available, we have demonstrated that it would indeed impact beamforming if the point source model was used for a distributed array.

---

# CHAPTER

# 3

---

## Practical Examples of Directional Sources

### 3.1 Introduction

The previous chapter presented an example to show the impact of using conventional beamforming on an idealized directional source (quadrupole). It also discussed the directivity of the human talker from what is available in the literature. It showed that if the array subtends a small angle about the talker then conventional beamforming is acceptable but it raised doubts on the validity of conventional beamforming when the array does not subtend a small angle about the talker.

This chapter presents sources which are more directional than the human talker from two different fields: musical instruments and aeroacoustics. This thesis is mainly concerned with sound pickup, so musical instruments are of special interest. Aeroacoustics examples mentioned in this chapter are more applicable to localization than on sound pickup.

The information on directivity is taken from the available literature and, once again, little information is available on the phase. The microphones will usually be placed at a distance from the source, this is where we need to know the magnitude and phase of the source directivity to construct a signal model that will be used when beamforming.



## 3.2 Musical instruments

Most of the literature on musical instruments addresses the sound production mechanisms or the vibration of the instruments, rather than the propagation of the sound from the instrument to a distant point. This may be motivated by the interest in creating virtual musical instruments based on physical modeling of the actual instruments [Smith 2008] [Scavone 1998]. To realize such virtual instruments, efficient algorithms representing the physical models are implemented on dedicated digital signal processors or on general purpose computers. They can be played in real-time using, for example, a MIDI (Musical Instrument Digital Interface) controller without knowing how to play the real instruments. Or it may be driven by the desire to understand what physical characteristics differentiate a good instrument from a bad one [Bissinger 2008a] [Bissinger 2008b].

What follows is information on the directivity of string and wind instruments selected from the available literature to show examples of sources which do not behave as point sources.

Directivity measurements for violins, violas, cellos, and contrabasses are presented in [Meyer 1972], where it is stated “Each instrument group has a nondirectional characteristic in its lowest frequency range except for the contrabass.” It also says that at higher frequencies, all instruments studied are directional with directivity patterns getting more complex (increasing number of lobes) as the frequencies increase. The directivity figures presented for each instrument group is the average of 5 instruments, indicating that the directivity changes from one instrument to the next, even within a group. The measurements were made in an anechoic chamber with the instrument on a turntable. The sound was received with a single microphone positioned 1 m away from the violins and violas and 3.5 m from the cellos and contrabasses. Instead of bowing the strings, the instrument was excited by vibrating a needle pressed into the bridge while the strings were damped with two pieces of foam rubber.

Near-field acoustical holography (NAH) has also been used to study musical instruments. For example, in [Wang 2001] multiplanar NAH is used “[...] to characterize the radiated sound field when the violin is played at frequencies extending from 294 Hz

---

to 3 kHz, and more specifically, to localize the regions on the violin structure from which significant sound energy originates”. Not much detail is provided for the far-field directivity pattern other than saying that at low frequencies the far-field radiation is omnidirectional, and that around 880 Hz the far-field pattern is cardioid. It does not qualify the far-field directivity pattern at higher frequencies, but does specify that “the regions of energy sources become more localized and are distributed asymmetrically on the violin body”. This would indicate that the directivity pattern becomes more elaborate at higher frequencies as specified in [Meyer 1972].

Figure 3.1 reproduces measured directivity patterns from [Rousseau 1996] which presents simulations and measurements of radiation of lateral toneholes wind instruments. Those measured directivity patterns are for the fundamental frequency and two overtone frequencies of a flute with a G fingering. The directivity patterns correspond to a flute with the mouth and tone holes toward the top and the opening at the end of the flute toward the right in all figures. None of the directivity patterns resembles that of a point source, and the directivity patterns are different for different frequencies. The measurements were made with the instrument held by a support and not by a human player, which would modify the radiation pattern.

The conclusion is that the directivity changes with frequency and is different for each instrument group. In fact, it also changes within an instrument group. To beamform on a musical instrument, different source models would be necessary for each instrument and for each frequency. Furthermore, as the instrumentalist plays, the source model would have to be changed dynamically according to the note played.

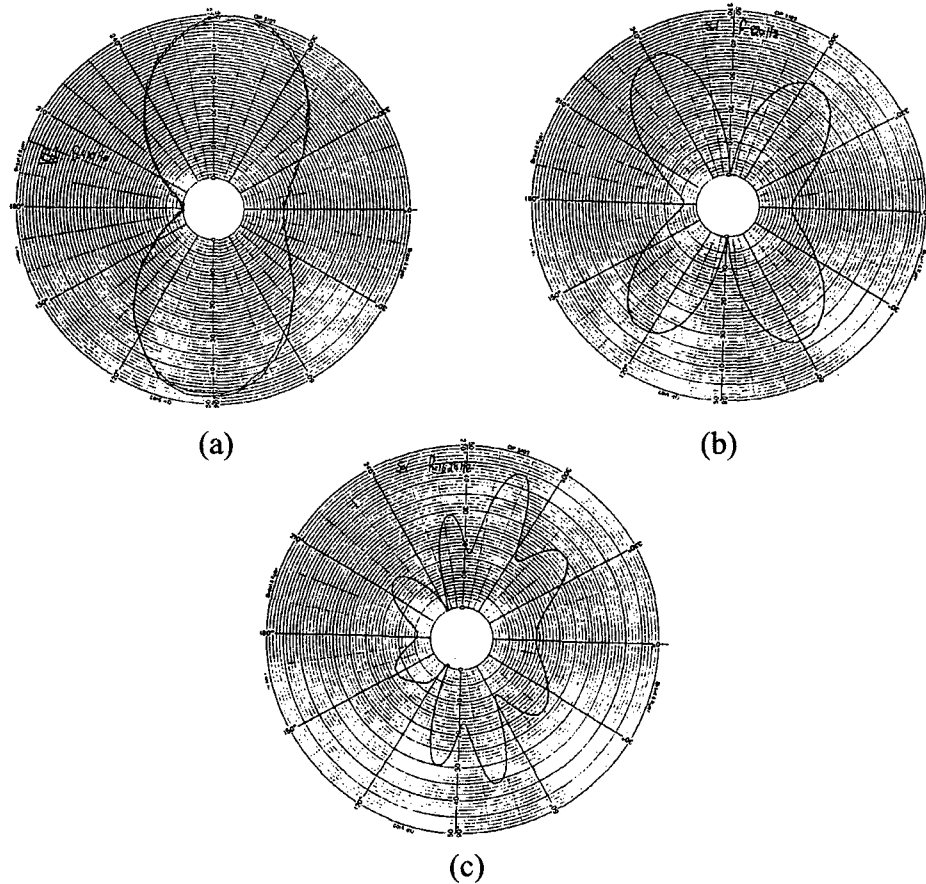


Figure 3.1. Measurements of Flute Directivity, with G fingering (a) Fundamental 410 Hz; (b) Overtone 820 Hz; (c) Overtone 1628 Hz. Reproduced from [Rousseau 1996] with permission.

### 3.3 Aeroacoustics

Here is how Goldstein introduces the topic: “Aeroacoustics is concerned with sound generated by aerodynamic forces or motions originating in a flow rather than by externally applied forces or motions of classical acoustics” [Goldstein 1976]. In aeroacoustics, sources which radiate differently than point sources are common, for example, sources that radiate as dipoles, quadrupoles, etc. The sound production mechanisms are beyond the scope of this thesis but this is discussed in details in [Powell 1990].

---

This section presents the application of beamforming to characterize the location and strength of noise sources on aircrafts, which is done either with real aircrafts flying at low altitude over an array or with scaled models in wind tunnels.

As mentioned above, in aeroacoustics the sources often do not behave as point sources. For example, in [Blake 2002; p.233] the sound produced at the trailing edge of an airfoil is described as “[...] radiated by a line of sources along the edge, with equal magnitudes to both sides of the airfoil, and the sound radiated to one side is 180 degrees out of phase with the sound radiated to the other,” that is, a line of dipoles located along the trailing edge, with the out-of-phase lobes perpendicular to the airfoil.

When using a point source model to realize the beamforming, restrictions are imposed on the physical setup in the hope that the array only observes a small portion of a single lobe of each multipole. This is described by Brooks and Humphreys for DAMAS, a deconvolution method that uses beamforming as a first step: “However, too large an array would present a measurement dilemma in aeroacoustics, in that the array would enclose differing directivity regions of a source being measured, thus nullifying the measured results (since uniform radiation of any source being measured is assumed over the array traditional processing).” [Brooks 2005]

In a paper describing measurement results of acoustic sources for the Airbus A340 obtained from actual flyovers, Piet et al. indicate that with the point source model, the source location accuracy is adequate even though some bias error may occur because of propagation effects, aircraft tracking, and the wings bending upwards during flight. He states however that it is much more difficult to estimate the source strength. Part of the possible solution is the “use of a multi-pole source model instead of the monopole one used here” [Piet 2002]. The authors do not go further, for example to propose methods to obtain such a model.

Papers have been published in recent years on the subject of beamforming on aeroacoustics dipole obtained by placing a small rod in an air flow [Jordan 2002][Liu 2008].

### **3.4 Chapter conclusion**

This chapter presented sources more directional than the human talker for musical instruments and aeroacoustics. From the simulation example of beamforming on a quadrupole discussed in Section 2.5.2, we would also expect deteriorated array performances when beamforming on the directional sources presented in this chapter if the point source model were used. The need for new techniques has been identified by researchers.

---

**CHAPTER****4**

---

**An Extension of Beamforming Applicable to Directional Sources****4.1 Introduction and assumptions**

It is obvious by now that it is necessary to have an appropriate source model to beamform on directional sources. The proper source model would have the same directivity and orientation as the actual source, and this model would be used directly to perform the beamforming. Unfortunately, as was said before, source directivity and orientation are unknown in most cases.

This chapter introduces the proposed new beamforming method. It is an extension of beamforming applicable to directional sources whose directivity is unknown. The general principle is to define a set of “sub-beamformers”, each beamforming on a different *spatial mode* of the source. The outputs of the individual sub-beamformers are combined in a weighted sum to give an overall output of better quality than that of a beamformer using the monopole model. The method is aimed at directional sources but it also works if the source is a point source. To go further, if the source is known to be a point source, the method is equivalent to conventional filter-and-sum beamforming.

Section 4.2 presents how the idea for the proposed beamforming method came to the author. Instead of continuing the demonstration this way, a more general presentation is then offered in Section 4.3. It shows that the sum of the sub-beamformers gives the same optimum array gain as beamforming directly with the exact source model.

A method to estimate the sub-beamformer weights based on the observed data is demonstrated in Section 4.4. It is called *gain-optimized estimator*. An expression for the resulting array gain is derived in Section 4.5.

Section 4.6 shows that if the source directivity pattern and source location with respect to the array do not change with time, it is possible to obtain the overall beamformer output without computing the output of each sub-beamformer. In this case, the method is computationally equivalent to beamforming directly with a source model.

It was shown in Chapter 2 how maximizing the array gain subject to a constraint on the white-noise gain could decrease the array gain sensitivity to random errors such as amplitude and phase errors in the microphone channels, or to slightly inexact microphone positions. Section 4.7 demonstrates how applying such a constraint could be done with the proposed beamforming method. The performance is investigated in Chapter 7 with a simulation.

Comparisons of the proposed beamforming method with the phase-mode spherical array and with near-field acoustical holography are presented in Section 4.8. The methods are quite different even though spherical harmonic expansions are used in all cases.

No assumptions are made on the source directivity. It is treated as an unknown. Six assumptions are used:

- The source location is known and does not change with time;
- The array geometry is known and does not change with time;
- Sound propagates as in free field between the source and each of the microphones;

- Microphones are omnidirectional. They have identical and ideal individual responses (i.e. no microphone mismatch);
- The noise component at the microphone outputs has an expected value of zero (this is not a strong assumption, as it is always the case for acoustic sources);
- The second order statistics of the noise field, as expressed by  $\mathbf{R}_{ee}(k)$ , are known and remain constant during a frame of measured samples (i.e. an “FFT frame”).

These assumptions are often made, for example in [Cox 1986] [Ryan 2003], even though they may not be explicitly stated.

## 4.2 Genesis of the method

The idea for the proposed beamforming method came by observing the form of the spherical Fourier transform while reading on the phase-mode spherical array in [Rafaely 2004]. The spherical Fourier transform of a function  $x(\theta, \phi)$ , square integrable on the unit sphere, is

$$x_{nm} = \int_0^{2\pi} \int_0^\pi x(\theta, \phi) Y_n^{m*}(\theta, \phi) \sin \theta d\theta d\phi. \quad (4.1)$$

The author observed that this equation is similar to the computation of the output  $x_{nm}$  of a continuous spherical surface array of radius equal to one, centered on the origin of the spherical coordinate system. The signal at point  $(\theta, \phi)$  on the continuous sensor surface is  $x(\theta, \phi)$  and the beamforming weight is  $Y_n^m(\theta, \phi)$ , which is the spherical harmonic function of order  $n$  and degree  $m$  (see appendix A). To avoid notation ambiguity later, Eq. (4.1) is re-written using  $y_{nm}$  to designate the array output,

$$y_{nm} = \int_0^{2\pi} \int_0^\pi x(\theta, \phi) Y_n^{m*}(\theta, \phi) \sin \theta d\theta d\phi. \quad (4.2)$$

The outgoing wave from a source of arbitrary directivity and frequency  $f$ , located at the origin of the coordinate system, can be decomposed [Williams 1999] into spherical harmonic modes  $(n', m')$  so that the signal on the sensor surface is



$$x(\theta, \phi) = x(f, r = 1, \theta, \phi) = \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} x_{n'm'}(f) h_{n'}^{(2)}(\kappa r) Y_{n'}^{m'}(\theta, \phi), \quad (4.3)$$

where  $h_{n'}^{(2)}(\kappa r)$  is the Hankel function of the second kind of order  $n'$  and  $\kappa = 2\pi f/c$  is the wave number.  $x_{n'm'}(f)$  (which is different from  $x_{nm}$ ) is generally a complex number, describing the strength of the spatial mode  $(n', m')$  in the source.

Substituting Eq. (4.3) in Eq. (4.2), the array output becomes

$$y_{nm} = \int_0^{2\pi} \int_0^{\pi} \left\{ \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} x_{n'm'}(f) h_{n'}^{(2)}(\kappa r) Y_{n'}^{m'}(\theta, \phi) \right\} Y_n^{m*}(\theta, \phi) \sin \theta d\theta d\phi. \quad (4.4)$$

Changing the order of summations and integrations, and keeping only terms depending on  $(\theta, \phi)$  within the integrals, leads to

$$y_{nm} = \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} x_{n'm'}(f) h_{n'}^{(2)}(\kappa r) \int_0^{2\pi} \int_0^{\pi} Y_{n'}^{m'}(\theta, \phi) Y_n^{m*}(\theta, \phi) \sin \theta d\theta d\phi. \quad (4.5)$$

Spherical harmonics have the property of being orthonormal:

$$\int_0^{2\pi} \int_0^{\pi} Y_{n'}^{m'}(\theta, \phi) Y_n^{m*}(\theta, \phi) \sin \theta d\theta d\phi = \begin{cases} 1 & n = n', m = m' \\ 0 & \text{otherwise} \end{cases}. \quad (4.6)$$

Using this property with Eq. (4.5), the array output is

$$y_{nm} = x_{nm}(f) h_n^{(2)}(\kappa r). \quad (4.7)$$

This equation shows that a continuous spherical array using the spherical harmonic function of order  $n$  and degree  $m$  to beamform on a source located at its center has for output only the component of the source with the same order and degree. If the source does not have a component of this order and degree, the corresponding output would be zero (except for noise).

The orthonormality property is not applicable if the source is not located at the center of the continuous spherical array, because the Hankel function of second kind varies with distance and it can not be taken out of the integrals.

To continue in the development of the proposed beamforming method, the next step would be to demonstrate how a properly weighted sum of beamformers of all orders and degrees  $(n, m)$  provides a response equivalent to that of beamforming directly with the exact source model while maximizing the array gain. Instead of continuing with the more abstract continuous spherical array case with the source located at its center, the next section begins the development of the proposed beamforming method again for the more practical case of a discrete array of arbitrary geometry. To emphasize, no requirement is set on the geometry of the array for the development of the proposed beamforming method.

### 4.3 Development of the method

Figure 4.1 illustrates schematically the directivity magnitude (the phase of the directivity is not shown) of an arbitrary source, at some frequency  $f$ , and a discrete sensor array with some arbitrary geometry. For notational convenience, and without loss of generality, the source is located at the center of the spherical coordinate system.

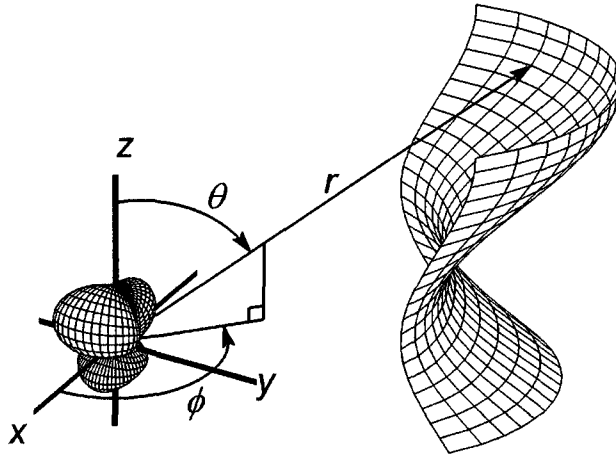


Figure 4.1. Source with an arbitrary directivity (magnitude shown, phase not shown) and arbitrary discrete array. (The microphones are located at the vertices.)

Using a decomposition [Williams 1999] into spherical harmonic modes  $(n, m)$  at frequency  $f$ , the outgoing wave from a source generates a signal at a point  $(r, \theta, \phi)$

$$s(f, r, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm}(f) h_n^{(2)}(\kappa r) Y_n^m(\theta, \phi), \quad (4.8)$$

where  $h_n^{(2)}(\kappa r)$  is the Hankel function of the second kind of order  $n$ ,  $\kappa = 2\pi f/c$  is the wave number, and  $Y_n^m(\theta, \phi)$  is the spherical harmonic function of order  $n$  and degree  $m$ . The factor  $s_{nm}(f)$  in the summation is a complex number equal to the contribution of the spherical harmonic mode  $(n, m)$  from the source at frequency  $f$ ; it is referred to here simply as the *mode strength*.

The directivity of the source, and so the mode strengths  $s_{nm}$ , are often unknown. This section demonstrates that if these mode strengths can be determined then a properly weighted linear sum of specially defined beamformers (called *sub-beamformers*) provides a response equivalent to that of beamforming with the exact source model. In this case, the optimum array gain is achieved.

For a discrete and time sampled array, the signal in the  $k^{\text{th}}$  frequency bin (obtained in practice by applying a Fast Fourier Transform (FFT) to a frame of time sampled data) for the  $l^{\text{th}}$  sensor is

$$s_l(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm}(k) h_n^{(2)}(\kappa_k r_l) Y_n^m(\theta_l, \phi_l), \quad (4.9)$$

where  $\kappa_k$  is the wave number at the center of frequency bin  $k$ . The source signal vector  $\mathbf{s}(k) = [s_1(k), \dots, s_L(k)]^T$  is formed from the signal component of all microphones to obtain

$$\mathbf{s}(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm}(k) \mathbf{Y}_n^m(k), \quad (4.10)$$

where, for notational convenience, we also define the *mode vector*

$$\mathbf{Y}_n^m(k) = [h_n^{(2)}(\kappa_k r_1) Y_n^m(\theta_1, \phi_1) \dots h_n^{(2)}(\kappa_k r_L) Y_n^m(\theta_L, \phi_L)]^T. \quad (4.11)$$

The reader is cautioned that the letter 'Y' is used both for the mode vector  $\mathbf{Y}_n^m(k)$  and for the spherical harmonic function  $Y_n^m(\theta_l, \phi_l)$ , albeit in different font properties, and that the mode vector contains both the spherical harmonic and the Hankel function of the second kind.

The weight vector to maximize the array gain given by Eq. (2.17) is

$$\mathbf{w}(k) = \mathbf{R}_{ee}^{-1}(k) \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm}(k) \mathbf{Y}_n^m(k), \quad (4.12)$$

or, rearranging,

$$\mathbf{w}(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^m(k). \quad (4.13)$$

Inserting this weight vector into Eq. (2.4), the optimized array beamformer output is

$$y(k) = \left\{ \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^m(k) \right\}^H \mathbf{p}(k). \quad (4.14)$$

Applying the Hermitian inside the curly bracket, using the fact that  $\{\mathbf{R}_{ee}^{-1}(k)\}^H = \mathbf{R}_{ee}^{-1}(k)$ , and rearranging once again, the array output becomes

$$y(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm}^*(k) \{ \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{p}(k) \}. \quad (4.15)$$

Finally, defining the term inside the curly bracket as

$$y_{nm}(k) = \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{p}(k), \quad (4.16)$$

the optimized array beamformer output is

$$y(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm}^*(k) y_{nm}(k). \quad (4.17)$$

Each factor  $y_{nm}(k)$  has the form of a beamformer (viz. Eq. (2.4)) with beamforming weight vector

$$\mathbf{w}_{nm}(k) = \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^m(k). \quad (4.18)$$

Considering the form of the gain-optimized array weight vector given in Eq. (2.17), the beamformer  $y_{nm}(k)$  is called a *gain-optimized sub-beamformer* of order  $n$  and degree  $m$ , or simply *sub-beamformer* of order  $n$  and degree  $m$ .

We will now demonstrate that when the sub-beamformer of order  $n$  and degree  $m$  is given the weight  $s_{nm}(k)$  in the double sum of Eq. (4.17), the optimum array gain is indeed achieved. The array gain is the ratio of the SNR at the output of the beamforming array to the SNR at the input. We will begin by finding the signal strength at the output of the array. Using Eq. (4.16), the signal component at the output of the sub-beamformer of order  $n$  and degree  $m$  is

$$y_{nm, \text{signal}}(k) = \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{s}(k). \quad (4.19)$$

After substituting this equation in Eq. (4.17), the signal component at the output of the array is

$$y_{\text{signal}}(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm}^*(k) \left\{ \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{s}(k) \right\}, \quad (4.20)$$

which, after moving the curly brackets becomes

$$y_{\text{signal}}(k) = \left\{ \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm}(k) \mathbf{Y}_n^m(k) \right\}^H \mathbf{R}_{ee}^{-1}(k) \mathbf{s}(k). \quad (4.21)$$

The term in the curly bracket is the spherical harmonic expansion of the source signal  $\mathbf{s}(k)$  as given by Eq. (4.10), so that the signal component at the output of the array finally is

$$y_{\text{signal}}(k) = \mathbf{s}^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{s}(k). \quad (4.22)$$

Using Eq. (4.16) again, the noise component at the output of the sub-beamformer of order  $n$  and degree  $m$  is

$$y_{nm, \text{noise}}(k) = \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k). \quad (4.23)$$

Following the same steps as above, the noise component at the output of the array is found to be

$$y_{\text{noise}}(k) = \mathbf{s}^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k). \quad (4.24)$$

The SNR at the output of the array can now be found with

$$\text{SNR}_{\text{out}}(k) = E \left\{ \frac{|y_{\text{signal}}(k)|^2}{|y_{\text{noise}}(k)|^2} \right\}, \quad (4.25)$$

where  $E$  is the expectation operator. The signal and noise components at the output of the array are substituted to obtain

$$\text{SNR}_{\text{out}}(k) = E \left\{ \frac{[\mathbf{s}^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{s}(k)][\mathbf{s}^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{s}(k)]^H}{[\mathbf{s}^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k)][\mathbf{s}^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k)]^H} \right\}, \quad (4.26)$$

which after manipulations and using Eqs. (2.10) and (2.11) becomes

$$\text{SNR}_{\text{out}}(k) = \frac{\sigma_s^2 \mathbf{s}^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{R}_{ss}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{s}(k)}{\sigma_e^2 \mathbf{s}^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{s}(k)}. \quad (4.27)$$

Knowing that the SNR at the input of the array is  $\sigma_s^2/\sigma_e^2$  and using Eq. (2.5), the array gain is

$$G(k) = \frac{\mathbf{s}^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{R}_{ss}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{s}(k)}{\mathbf{s}^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{s}(k)}. \quad (4.28)$$

This is exactly the optimum array gain given by Eq. (2.18) (with the frequency bin index  $k$  replacing the continuous frequency symbol  $f$ ) proving that if the proper weight is given to each sub-beamformer, the proposed beamforming method is equivalent to beamforming directly with the exact source model.

Since the gain-optimized sub-beamformers are readily computed from the signals at the microphone array, then the linear sum of sub-beamformers in Eq. (4.17) can be used to perform beamforming with optimal gain *if* the weights  $s_{nm}(k)$  can be found somehow. Figure 4.2 shows a schematic of the proposed beamforming method with the array data processed in parallel through a set of sub-beamformers, the outputs  $y_{nm}(k)$  multiplied by  $s_{nm}^*(k)$  and the products summed.

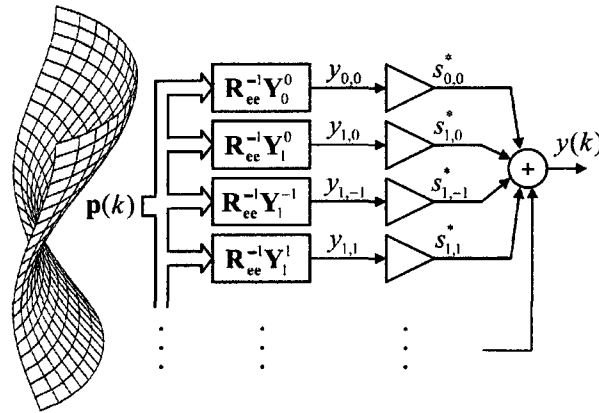


Figure 4.2. Proposed beamforming method schematic.

#### 4.3.1 Method with source decomposition into arbitrary modes set

It is noteworthy that the proposed beamforming method is not limited to using spherical harmonics. If the source signal can be decomposed into some other set of modes  $\mathbf{M}_m(k)$  as

$$\mathbf{s}(k) = \sum_{m=0}^M s_m(k) \mathbf{M}_m(k), \quad (4.29)$$

the optimum weight vector given by Eq. (2.17), after rearranging terms, becomes

$$\mathbf{w}(k) = \sum_{m=0}^M s_m(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{M}_m(k). \quad (4.30)$$

Inserting this weight vector into Eq. (2.4) and performing the same manipulations as were done when using spherical harmonics, the optimized array beamformer output is

$$y(k) = \sum_{m=0}^M s_m^*(k) \{ \mathbf{M}_m^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{p}(k) \}. \quad (4.31)$$

Defining a sub-beamformer (whose beamforming weight is  $\mathbf{w}_m(k) = \mathbf{R}_{ee}^{-1}(k) \mathbf{M}_m(k)$ )

$$y_m(k) = \mathbf{M}_m^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{p}(k), \quad (4.32)$$

the optimized array beamformer output finally is

$$y(k) = \sum_{m=0}^M s_m^*(k) y_m(k). \quad (4.33)$$

#### 4.4 Estimating mode strengths: gain-optimized (GO) estimator

Since the strength of each mode of the source is generally unknown, it is necessary to estimate  $s_{nm}(k)$  in Eq. (4.17). Here is the development to obtain a simple, low complexity estimator.

Replacing the microphone output vector  $\mathbf{p}(k)$  by its signal and noise components,  $\mathbf{s}(k) + \mathbf{e}(k)$ , for the sub-beamformer of order  $n$  and degree  $m$  in Eq. (4.16), and separating the terms we have

$$y_{nm}(k) = \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{s}(k) + \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k). \quad (4.34)$$

The source signal vector is then replaced by its spherical harmonic decomposition

$$\mathbf{s}(k) = \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_{n'}^{m'}(k) \quad (4.35)$$

to obtain

$$y_{nm}(k) = \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_{n'}^{m'}(k) + \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k). \quad (4.36)$$



Prime indices are used to distinguish the sub-beamformer mode from the modes making up the source. Bringing the summations to the front, the sub-beamformer output becomes

$$y_{nm}(k) = \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^{m'}(k) + \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k). \quad (4.37)$$

The mode  $(n, m)$  of the source, with the same order and degree as the sub-beamformer, is then taken out of the summation

$$y_{nm}(k) = s_{nm}(k) \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^m(k) + \sum_{\substack{n'=0 \\ (n',m') \neq (n,m)}}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^{m'}(k) + \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k) \quad (4.38)$$

Defining the coefficient

$$C_{n'm'}^{nm}(k) = \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^{m'}(k), \quad (4.39)$$

the sub-beamformer output becomes

$$y_{nm}(k) = C_{nm}^{nm}(k) s_{nm}(k) + \sum_{\substack{n'=0 \\ (n',m') \neq (n,m)}}^{\infty} \sum_{m'=-n'}^{n'} C_{n'm'}^{nm}(k) s_{n'm'}(k) + \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k). \quad (4.40)$$

Based on this equation, we define a possible estimator for the mode strength as

$$\begin{aligned} d_{nm}(k) &= \frac{1}{C_{nm}^{nm}(k)} y_{nm}(k) \\ &= \frac{1}{C_{nm}^{nm}(k)} \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{p}(k). \end{aligned} \quad (4.41)$$

Knowing the source position, the mode vector  $\mathbf{Y}_n^m(k)$  is computed with Eq. (4.11).  $C_{nm}^{nm}(k)$  is computed with Eq. (4.39) using  $\mathbf{Y}_n^m(k)$  and the inverse of the normalized noise cross-power spectral density matrix  $\mathbf{R}_{ee}(k)$ .  $\mathbf{p}(k)$  is the microphone output vector.  $\mathbf{p}(k)$  and  $\mathbf{Y}_n^m(k)$  are both column vectors, with the number of elements

equal to the number of microphones  $L$ .  $\mathbf{R}_{ee}(k)$  is a matrix  $L \times L$ , so the computation complexity to find its inverse will increase with the number of microphones in the array.

We call this estimator “gain-optimized estimator”, or “GO estimator” for short, because it uses  $\mathbf{w}_{nm}(k) = \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^m(k)$ , which is the beamforming weight vector that maximizes the array gain for the spherical harmonic mode of order  $n$  and degree  $m$ .

The bias and the variance of the GO estimator are given in Chapter 6 with their complete mathematical development provided in Appendix B. Other possible estimators are presented in Chapter 6.

## 4.5 Expression for the resulting array gain

In practice, each mode strength could be estimated with the GO estimator (either line of Eq. (4.41)). The array output would then be obtained by either summing sub-beamformer outputs

$$y(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n d_{nm}^*(k) y_{nm}(k) \quad (4.42)$$

or by first computing an overall beamforming weight vector (as will be seen in Section 4.6)

$$\mathbf{w}(k) = \mathbf{R}_{ee}^{-1}(k) \sum_{n=0}^{\infty} \sum_{m=-n}^n d_{nm}(k) \mathbf{Y}_n^m(k), \quad (4.43)$$

which would then be used to beamform directly with the microphone output vector

$$y(k) = \mathbf{w}(k)^H \mathbf{p}(k). \quad (4.44)$$

To evaluate the resulting array gain or white-noise gain, it is easier to work with the overall beamforming weight vector. The array gain is given by Eq. (2.12) and the white-noise gain is given by Eq. (2.13). Using the definitions of the normalized signal cross-power spectral density matrix Eq. (2.10) and the average signal power spectral density Eq. (2.7), the array gain and white-noise gain equations become respectively

$$G(k) = \frac{L}{\mathbf{s}^H(k)\mathbf{s}(k)} \frac{\mathbf{w}^H(k)\mathbf{s}(k)\mathbf{s}^H(k)\mathbf{w}(k)}{\mathbf{w}^H(k)\mathbf{R}_{ee}(k)\mathbf{w}(k)} \quad (4.45)$$

and

$$G_w(k) = \frac{L}{\mathbf{s}^H(k)\mathbf{s}(k)} \frac{\mathbf{w}^H(k)\mathbf{s}(k)\mathbf{s}^H(k)\mathbf{w}(k)}{\mathbf{w}^H(k)\mathbf{w}(k)}. \quad (4.46)$$

If the source radiates as a point source (or in a known way), it is easy to estimate either gain because the source signal  $\mathbf{s}(k)$  can be replaced in the equation. Unfortunately, we are assuming the source is unknown, thus  $\mathbf{s}(k)$  is also unknown and can not be replaced in the equation.

We can show the complexity of the resulting array gain (and white-noise gain) expression. To obtain the expression of the overall beamforming weight vector, the components composing the estimator are found (see Appendix B) to be

$$\begin{aligned} d_{nm}(k) = & s_{nm}(k) + \frac{1}{C_{nm}^{nm}(k)} \sum_{n'=0}^{\infty} \sum_{\substack{m'=-n' \\ (n',m') \neq (n,m)}}^{n'} C_{n'm'}^{nm}(k) s_{n'm'}(k) \\ & + \frac{1}{C_{nm}^{nm}(k)} \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k) \end{aligned} \quad (4.47)$$

Then substituting for Eq. (4.39), the estimator is

$$\begin{aligned} d_{nm}(k) = & s_{nm}(k) \\ & + \frac{1}{\mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^m(k)} \sum_{n'=0}^{\infty} \sum_{\substack{m'=-n' \\ (n',m') \neq (n,m)}}^{n'} s_{n'm'}(k) \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_{n'}^{m'}(k) \\ & + \frac{1}{\mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^m(k)} \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k) \end{aligned} \quad (4.48)$$

which then replaces  $s_{nm}(k)$  in the overall beamforming weight vector expression Eq. (4.12) to obtain (after moving  $\mathbf{R}_{ee}^{-1}(k)$  inside the summations)

$$\begin{aligned}
\mathbf{w}(k) = & \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^m(k) \\
& + \sum_{n=0}^{\infty} \sum_{m=-n}^n \left\{ \frac{1}{\mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^m(k)} \right. \\
& \quad \times \sum_{\substack{n'=0 \\ (n',m') \neq (n,m)}}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_{n'}^{m'}(k) \mathbf{Y}_n^m(k) \left. \right\} \\
& + \sum_{n=0}^{\infty} \sum_{m=-n}^n \left\{ \frac{1}{\mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^m(k)} \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k) \mathbf{Y}_n^m(k) \right\}.
\end{aligned} \tag{4.49}$$

An expression of the array gain or the white-noise gain is obtained by substituting the spherical harmonic decomposition of the source signal Eq. (4.10) and the beamforming weight vector  $\mathbf{w}(k)$  found above in Eqs. (4.45) or (4.46). We will not write the resulting expression because of its length. We will however mention that the expression is so complex that it is not insightful. The array gain (or white-noise gain) must be computed for specific cases of source directivity and location, and array geometry to evaluate the penalty with respect to the optimum value. Other estimators will have different equations.

## 4.6 Reduction of computational complexity

The beamforming weight vector  $\mathbf{w}_{nm}(k)$  and the source mode strength estimate  $d_{nm}(k)$  must be computed for each sub-beamformer. The number of such parameters to be estimated increases with the number  $N_{\text{order}}$  of spherical harmonic mode orders being considered to obtain the overall beamforming output

$$y(k) = \sum_{n=0}^{N_{\text{order}}} \sum_{m=-n}^n d_{nm}^*(k) y_{nm}(k). \tag{4.50}$$

In practice,  $N_{\text{order}}$  should be determined by the expected complexity of the source. A comparison of the computational complexity for the proposed beamforming method with the GO estimator and conventional beamforming is included in Appendix D.

If the source directivity pattern and source location with respect to the array do not change with time, once the source mode strengths are estimated it is possible to reduce the amount of computations necessary to obtain the overall beamformer. This is done by computing an overall beamforming weight vector. To find this vector, a source model is built using Eq. (4.10) with the estimate  $d_{nm}(k)$  replacing  $s_{nm}(k)$

$$\mathbf{s}_{\text{mo, GO}}(k) = \sum_{n=0}^{N_{\text{order}}} \sum_{m=-n}^n d_{nm}(k) \mathbf{Y}_n^m(k). \quad (4.51)$$

Substituting for the source model in Eq. (2.17), the overall beamforming weight vector is

$$\mathbf{w}(k) = \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{s}_{\text{mo, GO}}(k) \quad (4.52)$$

which is then used in Eq. (2.4) to obtain the overall beamformer output. This equation is shown again here for convenience

$$y(k) = \mathbf{w}^H(k) \mathbf{p}(k). \quad (4.53)$$

The overall beamforming output obtained by summing the sub-beamformers (Eq. (4.50)) is equivalent to the output (Eq. (4.53)) obtained using the source model. However, the second case maintains the same computational complexity as conventional beamforming since the computation of the individual sub-beamformer outputs is not required.

## 4.7 Difficulty of maximizing array gain subject to constraint on white-noise gain

The development of the proposed beamforming method was presented for the case of optimizing the array gain for a known noise field. The possible sensitivity of the array gain to random errors such as amplitude and phase errors in the microphone channels, or to slightly inexact microphone positions was discussed in Sub-Section 2.4.2. The solution was then presented as maximizing the array gain subject to an equality constraint on white-noise gain, for which the optimum beamforming weight vector was given by Eq. (2.22), repeated here for convenience

$$\mathbf{w}_{\text{opt, w constr}}(k) = \{\mathbf{R}_{\text{ee}}(k) + \varepsilon \mathbf{I}\}^{-1} \mathbf{s}(k). \quad (4.54)$$

To develop the proposed beamforming method in this case, the weight vector would become, after spherical harmonics expansion of the source signal

$$\mathbf{w}_{\text{opt, w constr}}(k) = \{\mathbf{R}_{\text{ee}}(k) + \varepsilon \mathbf{I}\}^{-1} \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm}(k) \mathbf{Y}_n^m(k). \quad (4.55)$$

Following a similar development as was done previously, the new estimator would be

$$d_{nm}(k) = \frac{1}{\mathbf{Y}_n^{mH}(k) \{\mathbf{R}_{\text{ee}}(k) + \varepsilon \mathbf{I}\}^{-1} \mathbf{Y}_n^m(k)} \mathbf{Y}_n^{mH}(k) \{\mathbf{R}_{\text{ee}}(k) + \varepsilon \mathbf{I}\}^{-1} \mathbf{p}(k), \quad (4.56)$$

and the overall beamforming weight vector would be

$$\mathbf{w}(k) = \{\mathbf{R}_{\text{ee}}(k) + \varepsilon \mathbf{I}\}^{-1} \sum_{n=0}^{\infty} \sum_{m=-n}^n d_{nm}(k) \mathbf{Y}_n^m(k). \quad (4.57)$$

An iterative process would then be required to find the value of  $\varepsilon$  providing the desired white-noise gain (see Section 2.4.2). Unfortunately, since the source directivity is unknown, so is the exact  $\mathbf{R}_{\text{ss}}(k)$ . It is not possible to compute the exact value for the resulting white-noise gain using Eq. (2.13). One possibility would be to go through the iterative process to find the value of  $\varepsilon$  while using the source model, Eq. (4.51), obtained with the GO estimator, Eq. (4.56), at each iteration to estimate the source signal cross-power spectral density matrix as  $\hat{\mathbf{R}}_{\text{ss, w constr}}(k) = \mathbf{s}_{\text{mo, GO}}(k) \mathbf{s}_{\text{mo, GO}}^H(k)$  and thus estimate the white-noise gain. The quality of the estimate would determine how close the actual white-noise gain is to the constraint. A simulation for this is performed at the end of Sub-Section 7.3.1.

## 4.8 Differences with other methods

This section first compares the proposed beamforming method with the phase-mode spherical array, whose development is presented in the context of optimizing the array gain. A few words are then said concerning near-field acoustical holography and the Helmholtz equation least squares method.

### 4.8.1 Comparison with phase-mode spherical array

The phase-mode spherical array [Meyer 2002] [Rafaely 2005] [Teutsch 2007] has some similarities in form to the proposed beamforming method. Both methods perform the beamforming in two steps. In the first step, both methods use spherical harmonic decomposition. The beamforming outputs of the first step are called eigenbeamformers for the phase-mode spherical array and are called sub-beamformers for the proposed beamforming method. For both methods, the output of the second step is a linear combination of the beamforming outputs obtained in the first step. Beyond this, the two methods are very different.

The eigenbeamformers mathematically decompose the signal received at the array using spherical harmonics, whereas the proposed beamforming method uses the fact that the source itself can be decomposed into spatial modes, each one propagating independently to the array. The eigenbeamformers use a spherical array to make the decomposition straight forward, whereas the proposed beamforming method applies to any array geometry. The phase-mode spherical array applies to known source models, specifically, plane waves or point sources, whereas the proposed beamforming method addresses more general and unknown sources.

To do the comparison, the mathematics of the phase-mode spherical array will be reviewed with the array gain maximization, similarly to what was done for the proposed beamforming method in Section 4.3. The development is done for a point source.

The array is made up of  $L$  microphones located on an open sphere of radius  $r$ , centered on the origin of the spherical coordinate system. In this coordinate system, the microphones are located at  $(r, \theta_l, \phi_l)$ . Because the received signal at the microphones will be decomposed into spherical harmonics it is simpler if the origin of the spherical coordinate system is at the center of the spherical array. For the proposed beamforming method, the spherical harmonic expansion was performed on the source so it was easier if the origin of the spherical coordinate system was at the source location.

The microphone outputs are combined in a vector  $\mathbf{p}(k)$ . Beginning from the array output

$$y(k) = \mathbf{w}^H(k) \mathbf{p}(k), \quad (4.58)$$

the beamforming weight vector  $\mathbf{w}(k)$  is replaced with the optimum solution to maximize the array gain

$$\mathbf{w}_{\text{opt}}(k) = \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{s}_{\text{mo}}(k), \quad (4.59)$$

where again,  $\mathbf{s}_{\text{mo}}(k)$  is the source model.

The point source is positioned at  $(r_0, \theta_0, \phi_0)$  and has a frequency located in frequency bin  $k$ . Decomposing the signal on the open sphere of radius  $r$  using spherical harmonic decomposition, the signal model for microphone  $l$  is [Teutsch 2007]

$$s_{\text{mo}}(k, r, \theta_l, \phi_l) = i \sum_{n=0}^{\infty} j_n(\kappa_k r) h_n^{(2)}(\kappa_k r_0) \sum_{m=-n}^n Y_n^{m*}(\theta_0, \phi_0) Y_n^m(\theta_l, \phi_l), \quad (4.60)$$

where  $j_n(\kappa_k r)$  is the spherical Bessel function of the first kind and  $h_n^{(2)}(\kappa_k r_0)$  is the spherical Hankel function of the second kind. This equation is valid for  $r_0 > r$ . The signal model vector is obtained by combining the signal model for the  $L$  sensors into a single signal model vector

$$\mathbf{s}_{\text{mo}}(k) = i \sum_{n=0}^{\infty} j_n(\kappa_k r) h_n^{(2)}(\kappa_k r_0) \sum_{m=-n}^n Y_n^{m*}(\theta_0, \phi_0) \bar{\mathbf{Y}}_n^m(k), \quad (4.61)$$

where  $\bar{\mathbf{Y}}_n^m(k)$  is a vector containing the  $L$  values  $Y_n^m(\theta_l, \phi_l)$ . The bar above the letter "Y" is to differentiate it from the mode vector definition used with the proposed beamforming method, Eq. (4.11), which also contains the spherical Hankel function of the second kind.

Substituting the signal model vector into the beamforming weight vector Eq. (4.59) which is then replaced into Eq. (4.58), the array output is

$$y(k) = \left\{ \mathbf{R}_{\text{ee}}^{-1}(k) i \sum_{n=0}^{\infty} j_n(\kappa_k r) h_n^{(2)}(\kappa_k r_0) \sum_{m=-n}^n Y_n^{m*}(\theta_0, \phi_0) \bar{\mathbf{Y}}_n^m(k) \right\}^H \mathbf{p}(k). \quad (4.62)$$



Applying the Hermitian inside the curly bracket, using the fact that  $\{\mathbf{R}_{ee}^{-1}(k)\}^H = \mathbf{R}_{ee}^{-1}(k)$ , and rearranging once again, the array output becomes

$$y(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n (-i) j_n(\kappa_k r) h_n^{(2)*}(\kappa_k r_0) Y_n^m(\theta_0, \phi_0) \{\bar{\mathbf{Y}}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{p}(k)\}. \quad (4.63)$$

The term inside the curly bracket is the *eigenbeamformer* of order  $n$  and degree  $m$

$$y_{nm}(k) = \bar{\mathbf{Y}}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{p}(k), \quad (4.64)$$

which has a beamforming weight vector  $\mathbf{R}_{ee}^{-1}(k) \bar{\mathbf{Y}}_n^m(k)$ .

Finally, the optimized array beamformer output is

$$\begin{aligned} y(k) &= \sum_{n=0}^{\infty} \sum_{m=-n}^n \{i j_n(\kappa_k r) h_n^{(2)}(\kappa_k r_0) Y_n^{m*}(\theta_0, \phi_0)\}^* y_{nm}(k) \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^n \{w_{nm}(k)\}^* y_{nm}(k) \end{aligned} \quad (4.65)$$

The phase-mode spherical array output is seen to be a linear sum of weighted eigenbeamformers. The eigenbeamformers are computed in a first step, and then summed in a second step.

Now, a comparison can be made between the phase-mode spherical array (Eqs. (4.64) and (4.65)) and the proposed beamforming method (Eqs. (4.16) and (4.17)) when maximizing the array gain. The number of modes used by the phase-mode spherical array is determined by the desired accuracy of approximation to beamforming directly with the point source model, this means at which value to limit  $n$  in the summation, whereas the number of modes used with the proposed beamforming method is determined by the source complexity. For example, with a point source the phase-mode spherical array requires an infinite number of modes to obtain the optimum array gain, whereas the proposed beamforming method reduces identically to the conventional beamformer.

As seen in Eq. (4.64), the beamforming weight vector of the eigenbeamformer  $(n, m)$  is  $\mathbf{R}_{ee}^{-1}(k) \bar{\mathbf{Y}}_n^m(k)$  with  $\bar{\mathbf{Y}}_n^m(k)$  only depending on the microphone positions on the

sphere. The weight vector of each sub-beamformer, Eq. (4.18), depends on the source position and orientation with respect to the array, the array geometry, and the frequency.

The weights used in the output stage of the phase-mode spherical array, Eq. (4.65), are given by  $w_{nm}(k) = i j_n(\kappa_k r) h_n^{(2)}(\kappa_k r_0) Y_n^{m*}(\theta_0, \phi_0)$  and depend on the position of the source, the array radius  $r$ , the frequency through  $\kappa_k$ , whereas the weights to obtain the overall output of the proposed beamforming method, Eq. (4.17), depend only on the source mode strengths. The comparison is summarized in Table 4.1.

Table 4.1. Comparison of the proposed beamforming method and phase-mode spherical array when maximizing the array gain.

	<b>Proposed beamforming method</b>	<b>Phase-mode spherical array</b>
<b>Array geometry</b>	Arbitrary	Spherical
<b>Source directivity</b>	Arbitrary	Point source
<b>Source model</b>	Estimated	Point source
<b>Why is processing in two steps possible</b>	Source itself is decomposed into spherical harmonics	Array is spherical: incoming signal on the array is decomposed into spherical harmonics
<b>Number of first-step beamformers needed</b>	The same number as modes in the source	An infinity
<b>First-step beamformer weight vector depends on</b>	Source and sensor positions, frequency, $\mathbf{R}_{ee}^{-1}(k)$	Sensor position, $\mathbf{R}_{ee}^{-1}(k)$
<b>Second-step output weights depend on</b>	Contribution of each mode to the source at each frequency	Source position, sphere radius, frequency

The proposed beamforming method addresses a problem with significantly more degrees of freedom (source directivity and orientation) than beamforming with the point source model, and this is reflected in the need to estimate more parameters for the same order of harmonic decomposition.

---

#### 4.8.2 Comparison with near-field acoustical holography and Helmholtz equation least squares method

The proposed beamforming method may also seem similar to near-field acoustical holography (NAH) or to the Helmholtz equation least squares method as those also use source decomposition into spatial modes.

However, the proposed beamforming method is developed to address sound pickup at a distance from the source. Unlike near-field acoustical holography [Williams 1999] or other techniques [Cho 2005] [Wu 2000], the method does not address reconstruction of the acoustic pressure field near the source or evaluation of the surface velocity of the source. Furthermore, unlike NAH, the near-field criterion for beamforming is wavefront curvature rather than the presence of evanescent components. To realize sound pickup with the proposed beamforming method, it is not required to measure the evanescent components so the microphones do not have to be close to the source.

### 4.9 Chapter conclusion

The end of Chapter 2 had discussed the consequences of using the point source model to beamform with an array covering a large angle about either a quadrupole or a human talker. Chapter 3 presented directivity information for real world directional sources.

Now Chapter 4 introduced the proposed alternative method for beamforming on directional sources. It showed that with the appropriate weighting, the sum of the sub-beamformers gives the same optimum array gain as beamforming directly with the exact source model. An estimator of the weight to give each sub-beamformer based on the observed data was presented. An expression for the resulting array gain was developed though it is so complex that it is not insightful.

It was shown that if the source directivity pattern and source location with respect to the array do not change with time, it is possible to obtain the overall beamformer output without computing the output of each sub-beamformer, thus maintaining the same computational complexity as conventional beamforming.

Comparisons of the proposed beamforming method with the phase-mode spherical array and with near-field acoustical holography were also presented.

---

# CHAPTER 5

---

## Experimental and Simulation Results for the Proposed Beamforming Method

### 5.1 Introduction

This chapter compares simulations and measurements for a simple directional source when beamforming with both the proposed beamforming method (with the GO estimator) and using the point source model. The content of this chapter is from a technical paper the author and his supervisors published in the Journal of the Acoustical Society of America [Bouchard 2009].

### 5.2 Measurement setup and signal processing

The setup of the demonstration is shown schematically in Fig. 5.1. The linear array has 16 uniformly spaced  $\frac{1}{4}$ -inch (6 mm) electret microphones with 8 cm spacing. The array is positioned on the floor with the microphone centers approximately 9 mm above the floor. The source is constructed with two 1-inch (26 mm) PVC pipes, approximately 3 m long, each connected to a compression driver. The pipes are approximately parallel with their open ends in contact. The centers of the opening are  $h=39.7$  cm directly above the microphone array and in line with the axis of the array.

A 1 kHz test signal is sent to each compression driver. The input signals to the drivers are always 180 degrees out of phase with respect to each other and their amplitudes are independently adjustable. The acoustic signal measured at the microphone array includes the 1 kHz test tone, room reverberation, noise, and distortion. Identical drivers, pipes, and couplers were used to minimize differences between the two source channels so that a dipole source is approximated when the drive voltages are equal. Since reverberation (image sources) is inevitable in a non-anechoic environment, the microphones are positioned as close as possible to the floor (which is vinyl over wood frame) so that the dominant (nearest) image source is symmetrically positioned about the array axis and will not affect the apparent source directivity. To reduce the effects of other source images, various absorbing panels and plush furniture are positioned about the room, as distant as possible from the source and array. Due to the hard floor and ceiling, the reverberation time in the vicinity of the array is about 700 ms.

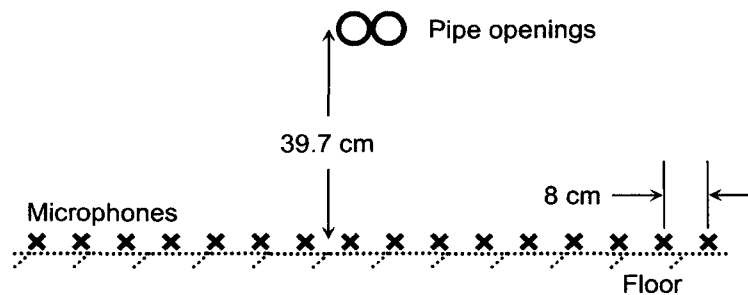


Figure 5.1. Experiment setup.

For analysis and simulation, a coordinate system is defined with the origin between the centers of the pipe openings, the  $xy$ -plane parallel to the floor, the  $x$ -axis parallel to the microphone array, and the  $z$ -axis passing through the array center, so that the array center is located at  $(0,0,-h)$ . The center of the pipe openings are located at  $(\pm dx/2, 0, 0)$ , where  $dx$  is the pipe diameter. The source due to each of the pipes is modeled as a point source at the effective position  $(\pm dx/2, -dl, 0)$ , which is a distance  $dl = 9$  mm in front of the pipe end [Ando 1969/1970]. In the simulations, this small displacement of the source is modeled but reverberation and noise are ignored.

Digital signals are acquired from the microphone array using a 16-channel analog-to-digital converter with 16-bit resolution and a sampling rate of 48 kHz. Processing is done offline using a 4096-point FFT with Hamming windowing on each microphone signal to obtain the Fourier coefficient for the single FFT bin containing the 1 kHz test tone. The Fourier coefficients from the 16 microphones are assembled into the microphone signal vector  $\mathbf{p}(k)$ .

As an indication of the signal-to-noise ratio, the FFT bin containing the 1 kHz tone for each of the 16 channels was observed to have a power spectral density at least 40 dB higher when the signal was present than when no signal was present.

In this demonstration of the proposed beamforming method, only sub-beamformers of order  $N_{\text{order}} = 4$  or less are considered. This number was chosen because it is more than adequate for a dipole and provides low computational complexity. Typically, some knowledge of the source characteristics would be used to guide the selection of a value for  $N_{\text{order}}$ . The contribution to each mode  $d_{nm}(k)$  is estimated using Eq. (4.41), and the overall beamforming weight vector is then obtained as

$$\mathbf{w}(k) = \sum_{n=0}^4 \sum_{m=-n}^n d_{nm}(k) \mathbf{Y}_n^m(k). \quad (5.1)$$

The optimized white-noise gain for the beamformer is computed as

$$G_{\text{out,dB}}(k) = 10 \log \left( \frac{\mathbf{w}^H(k) \hat{\mathbf{R}}_{\text{ss}}(k) \mathbf{w}(k)}{\mathbf{w}^H(k) \mathbf{w}(k)} \right). \quad (5.2)$$

This differs slightly from the white-noise gain in Eq. (2.13) because in our measurements the matrix  $\mathbf{R}_{\text{ss}}(k)$  is unknown and is estimated from a single microphone signal vector  $\mathbf{p}(k)$ , which includes some noise. It is estimated as

$$\hat{\mathbf{R}}_{\text{ss}}(k) = \frac{L}{\mathbf{p}^H(k) \mathbf{p}(k)} \cdot \mathbf{p}(k) \mathbf{p}^H(k). \quad (5.3)$$

### 5.3 White-noise gain for varying drive voltage

If the drive voltage to either one of the drivers is reduced to zero, the overall source will change from a dipole to a monopole. Figure 5.2 shows the white-noise gain according to Eq. (5.2) for the measurement and simulation data, for the proposed beamforming method and beamforming with a point source model, when the drive voltages are changed. The pressure magnitude and phase at the end of each tube are not available so white-noise gain is plotted as a function of drive voltage difference. Note that for the proposed beamforming method, Eq. (5.1) is evaluated at each drive voltage whereas the point source model for conventional beamforming is kept constant for all drive voltages.

The optimum white-noise gain is constant; it depends only on the number of sensors and is independent of the source. For an array of 16 sensors, the optimum white-noise gain is  $10 \log(16)$ , which is equal to 12.04 dB. As expected [Jordan 2002] [Bouchard 2006], beamforming with a point source model deteriorates and does not perform properly when the drive voltages are equal and the source is a dipole. In which case, Figure 5.2 shows a gain slightly below -4 dB for the measurement. On the other hand, for both the measurement and simulation, the proposed beamforming method maintains a gain of more than 11 dB.



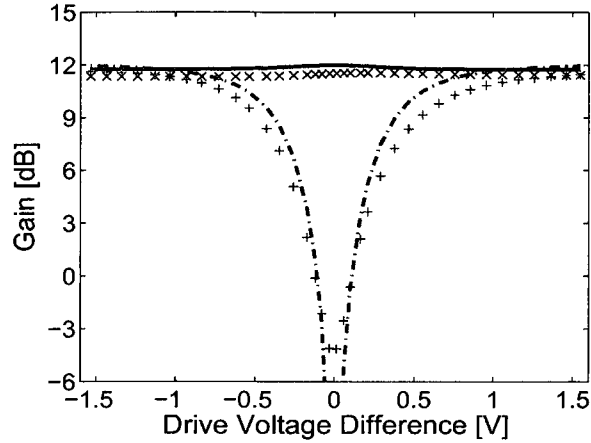


Figure 5.2. Array gain for varying drive voltage difference  $V_{\text{driver \#2}} - V_{\text{driver \#1}}$ . Driver #1 voltage is held at a fixed value of 1.534 V while driver #2 voltage is varied from 0 to 1.548 V, then driver #2 voltage is held at a fixed value of 1.548 V while driver #1 voltage is reduced to 0. Simulated (—) and measured (×) results for the proposed beamforming method, and simulated (· -) and measured (+) results for beamforming with a point source model.

At the source, only modes  $(n,m)=(1,-1)$  and  $(n,m)=(1,1)$  are necessary to generate a dipole with the orientation described for the setup. Table 5.1 presents the white-noise gain of each sub-beamformer in response to the dipole in the simulated case, computed using  $\mathbf{w}(k) = \mathbf{Y}_n^m(k)$  in Eq. (5.2). It shows that numerous sub-beamformers have similar (but slightly lower) white-noise gain than the two required modes. This is because the linear array has a limited view of the dipole and so the source looks somewhat similar to other modes. (This would not be the case for a dense array completely surrounding the source.)

As the overall source approaches a monopole (extreme left or right of the figure), conventional point source beamforming achieves slightly more gain than the proposed beamforming method.

Table 5.1. White-noise gain (dB) for each sub-beamformer  $y_{nm}(k)$  in response to the dipole in the simulated case.

n	0	1	2	3	4
m	White-noise gain (dB)				
-4					-13.34
-3				11.37	11.61
-2			-20.59	-19.57	-18.35
-1		12.04	11.90	10.83	6.73
0	-17.78	-17.90	-19.04	-22.46	-26.33
1		12.02	11.88	10.80	6.67
2			-10.72	-10.05	-9.38
3				11.34	11.57
4					-8.43

## 5.4 Beampattern along x-axis

A beampattern plot is often used to illustrate the spatial selectivity of a beamforming array using the conventional point source model. For conventional far-field beamforming (plane wave source model) the beampattern is simply a function of arrival direction while for near-field beamforming (point source model) it also depends upon the distance from the source to the receiving array. These beampatterns can be easily displayed as one- or two- dimensional plots but spatial selectivity is much more difficult to depict for the proposed beamforming method because it is also dependent upon the source directivity.

To provide some indication of spatial selectivity for the proposed beamforming method, Figure 5.3 illustrates how the response varies with the source position when the overall source approximates a dipole. It presents results for source locations at fixed height above the array and along the length of the array. (This is actually done by shifting the array.) The beamforming weight vector for the proposed beamforming method is computed using Eqs. (4.41) and (5.1), with the source over the center of the array. The

beamforming weight vector is then held constant for all other positions. Similarly, the beamforming weight vector for the conventional beamformer is based on a point source over the array center and is held fixed.

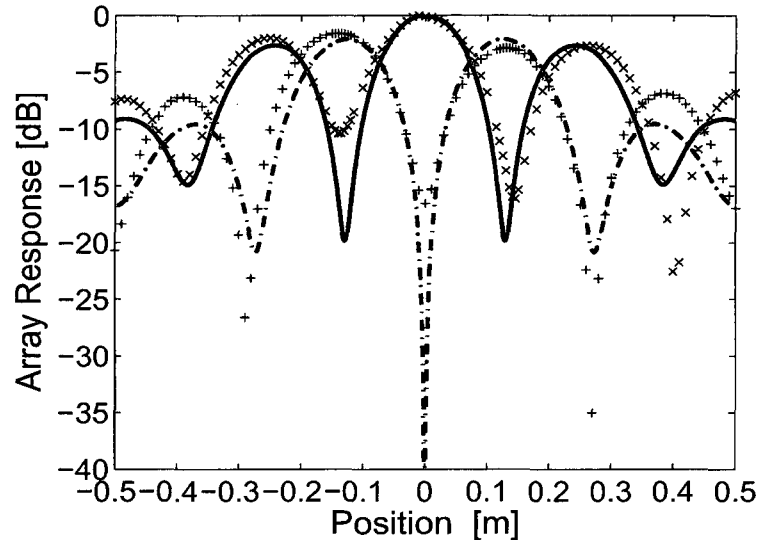


Figure 5.3. Array response along the  $x$ -axis for a dipole source. Simulated (—) and measured ( $\times$ ) results for the proposed beamforming method, and simulated (· -) and measured (+) results for beamforming with a point source model.

There is a good match between measurements and simulations results in Fig. 5.3. The side lobe levels obtained with the proposed beamforming method and with the conventional beamformer using the point source model are not greatly different but the proposed beamforming method overcomes the failure to sense the dipole source at the proper position.

## 5.5 Chapter conclusion

This chapter compared simulations and measurements for a simple directional source when beamforming with both the proposed beamforming method and using the point source model. Even in a non-anechoic environment, there was a good match between measurements and simulations.

For a measured dipole, the proposed beamforming method had an array white-noise gain of 11 dB (compared to a theoretical optimum array gain of 12 dB) while

conventional beamforming, with a value of -4 dB, does not provide any useful white-noise gain (With the array and source setup, the theoretical white-noise gain for conventional beamforming is  $-\infty$  dB).

The array response was simulated and measured along the  $x$ -axis. The proposed beamforming method overcame the failure to sense the dipole source at the proper position which conventional beamforming experiences.

---

# CHAPTER 6

---

## Approaches to Estimate Mode Strengths

### 6.1 Introduction

The proposed beamforming method for directional sources was introduced in Chapter 4, the development being in Section 4.3. It was shown that the optimum array gain can be obtained if the exact source mode strengths are used as the weights for the sub-beamformer outputs in a weighted linear sum. The question then becomes how to estimate the source mode strengths from the microphone outputs?

The gain-optimized (GO) estimator was provided in Section 4.4. It was used with the proposed beamforming method for the experiments and simulations in Chapter 5. We come back to the GO estimator in Section 6.2. Then estimators using least squares are introduced in Section 6.3. A short explanation of how the proposed beamforming method could be used with MVDR is presented in Section 6.4. Finally, estimation using many FFT frames is discussed in Section 6.5.

Section 4.5 showed that it is difficult to mathematically or analytically evaluate the performance of the method because of the complexity of the array gain equation. For this reason, the performance of the different estimators will be investigated using simulations in Chapter 7.

For completeness sake, other approaches that led to relatively unsatisfactory estimators are documented in Appendix C.

For the development of this chapter, we still use the assumptions given in Section 4.1. Again, no assumption is made on the source directivity. It is treated as an unknown.

## 6.2 Gain-optimized estimator (GO estimator)

In Section 4.4, the gain optimized estimator was given as

$$\begin{aligned} d_{nm}(k) &= \frac{1}{C_{nm}^{nm}(k)} y_{nm}(k) \\ &= \frac{1}{C_{nm}^{nm}(k)} \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{p}(k). \end{aligned} \quad (6.1)$$

Knowing the source position, the mode vector  $\mathbf{Y}_n^m(k)$  is computed with Eq. (4.11).  $C_{nm}^{nm}(k)$  is computed with Eq. (4.39) using  $\mathbf{Y}_n^m(k)$  and the inverse of the normalized noise cross-power spectral density matrix  $\mathbf{R}_{ee}(k)$ .

It was said in Section 4.6 that including the sub-beamformers up to order  $N_{\text{order}}$  to compute the overall beamformer output, while using the estimator above to estimate the mode strengths, was equivalent to using the source model

$$\mathbf{s}_{\text{mo,GO}}(k) = \sum_{n=0}^{N_{\text{order}}} \sum_{m=-n}^n d_{nm}(k) \mathbf{Y}_n^m(k). \quad (6.2)$$

### 6.2.1 Bias and variance of the GO estimator

The theoretical computation of the bias and variance of the GO estimator is provided in Appendix B. The results are repeated here. For the computation, we assumed that the noise field, and hence  $\mathbf{R}_{ee}(k)$ , is known. In practice however,  $\mathbf{R}_{ee}(k)$  would have to be estimated. In this case, the bias and variance would need to take into account the characteristics of the estimator of  $\mathbf{R}_{ee}(k)$ . We also assumed that  $\mathbf{R}_{ee}(k)$  is constant.

The bias of the estimator is

$$E\left\{d_{nm}(k)\right\} - s_{nm}(k) = \frac{1}{C_{nm}^{nm}(k)} \sum_{n'=0}^{\infty} \sum_{\substack{m'=-n' \\ (n',m') \neq (n,m)}}^{n'} C_{n'm'}^{nm}(k) s_{n'm'}(k). \quad (6.3)$$

This shows that the bias of the estimator for mode  $(n, m)$  is a weighted sum of the strengths  $s_{n'm'}(k)$  of the modes other than  $(n, m)$  present in the source. It also depends on the coefficients  $C_{n'm'}^{nm}(k)$  given by Eq. (4.39). Notice that the bias is independent of the average noise power over the array or at which order  $N_{\text{order}}$  we stop including modes.

The variance of the estimator is

$$\begin{aligned} \text{Var}\left\{d_{nm}(k)\right\} &= \frac{\sigma_e^2(k)}{C_{nm}^{nm}(k)} \\ &= \frac{\sigma_e^2(k)}{\mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{Y}_n^m(k)}. \end{aligned} \quad (6.4)$$

This shows that the variance of the estimator is proportional to the average noise power over the array.

Note that, as shown in Eq. (2.2), a normalized FFT is used to obtain the frequency-domain microphone outputs so that the variance of the estimator is independent of the FFT length. If the regular FFT had been used instead,  $\sigma_e^2(k)$  would be proportional to the FFT length  $N$  while  $\mathbf{R}_{ee}(k)$  would be independent of  $N$ , leading to an estimator variance depending on the FFT length.

Again, it is difficult to evaluate directly the beamforming performance of the proposed beamforming method by observing the equations of the bias and variance of the estimator. Performance of the proposed beamforming method with the GO estimator is investigated in Chapter 7.

### 6.3 Least squares (LS) estimators

In this approach, the least squares method is used to find the source mode strengths so as to minimize the sum of the square of the errors between the microphone outputs and a source model.

### 6.3.1 Development

The development begins by using the spherical harmonic decomposition for the received signal component vector (the source signal vector) Eq. (4.35) so that the microphone output vector is written as

$$\mathbf{p}(k) = \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_{n'}^{m'}(k) + \mathbf{e}(k). \quad (6.5)$$

As it was the case in Chapter 4, prime indices distinguish the modes making up the source from the modes used for beamforming.

Considering only sub-beamformers of order  $N_{\text{order}}$  or less in the proposed beamforming method, the double summation expressing the received signal is split in two, leading to

$$\mathbf{p}(k) = \sum_{n'=0}^{N_{\text{order}}} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_{n'}^{m'}(k) + \sum_{n'=N_{\text{order}}+1}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_{n'}^{m'}(k) + \mathbf{e}(k). \quad (6.6)$$

Matrix notation is then used to replace the first double summation as

$$\mathbf{p}(k) = \mathbf{Y}'(k) \mathbf{s}'_{\text{str}}(k) + \sum_{n'=N_{\text{order}}+1}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_{n'}^{m'}(k) + \mathbf{e}(k), \quad (6.7)$$

where

$$\mathbf{Y}'(k) = [\mathbf{Y}_{0'}^{0'}(k), \mathbf{Y}_{1'}^{-1'}(k), \mathbf{Y}_{1'}^{0'}(k), \mathbf{Y}_{1'}^{-1'}(k), \dots, \mathbf{Y}_{N_{\text{order}}'}^{N_{\text{order}}'}(k)] \quad (6.8)$$

is a matrix whose columns are the mode vectors for all spherical harmonic modes considered with the proposed beamforming method. The prime indicates that this matrix contains the exact (and unknown) mode vectors for the exact source and microphone positions to distinguish with the mode vectors used for beamforming, which may contain errors.  $\mathbf{Y}'(k)$  is a matrix with  $L$  rows and  $M = (N_{\text{order}} + 1)^2$  columns. The column vector

$$\mathbf{s}'_{\text{str}}(k) = [s_{0',0'}(k), s_{1',-1'}(k), s_{1',0'}(k), s_{1',1'}(k), \dots, s_{N_{\text{order}}',N_{\text{order}}'}(k)]^T \quad (6.9)$$



contains the exact source mode strengths for the modes considered with the proposed beamforming method.

The source model is defined as

$$\mathbf{s}_{\text{mo, LS}}(k) = \mathbf{Y}(k)\hat{\mathbf{s}}_{\text{str, LS}}(k). \quad (6.10)$$

Note the absence of primes and the “hat” over the vector containing the estimated mode strengths. The subscript “LS” signifies that the computation is done using the least-squares method.  $\mathbf{Y}(k)$  is computed from the microphone positions and the knowledge (or the estimation) of the source position. The task is then to estimate  $\hat{\mathbf{s}}_{\text{str, LS}}(k)$  such as to minimize the error between  $\mathbf{s}_{\text{mo, LS}}(k)$  and  $\mathbf{Y}'(k)\mathbf{s}'_{\text{str}}(k)$  (or even better, to minimize the error between  $\mathbf{s}_{\text{mo, LS}}(k)$  and the source signal  $\mathbf{s}(k)$ ).

Since  $\mathbf{Y}'(k)\mathbf{s}'_{\text{str}}(k)$  (or  $\mathbf{s}(k)$ ) is unavailable, an estimation is performed from the received microphone output vector, which is now written as

$$\begin{aligned} \mathbf{p}(k) &= \mathbf{s}_{\text{mo, LS}}(k) + \boldsymbol{\varepsilon}(k) \\ &= \mathbf{Y}(k)\hat{\mathbf{s}}_{\text{str, LS}}(k) + \boldsymbol{\varepsilon}(k), \end{aligned} \quad (6.11)$$

where  $\boldsymbol{\varepsilon}(k)$  is an error vector which regroups the noise vector, the components from the modes of orders above  $N_{\text{order}}$  not included in the model (the double sum remaining in Eq. (6.7)) as well as errors caused by the difference between the exact mode vectors in  $\mathbf{Y}'(k)$  and the mode vectors in  $\mathbf{Y}(k)$  used for beamforming.

Rearranging the terms, the error vector is

$$\boldsymbol{\varepsilon}(k) = \mathbf{p}(k) - \mathbf{Y}(k)\hat{\mathbf{s}}_{\text{str, LS}}(k), \quad (6.12)$$

from which a cost function is built as

$$\begin{aligned} J(k) &= \|\boldsymbol{\varepsilon}(k)\|^2 = \boldsymbol{\varepsilon}^H(k)\boldsymbol{\varepsilon}(k) \\ &= (\mathbf{p}(k) - \mathbf{Y}(k)\hat{\mathbf{s}}_{\text{str, LS}}(k))^H (\mathbf{p}(k) - \mathbf{Y}(k)\hat{\mathbf{s}}_{\text{str, LS}}(k)) \\ &= (\mathbf{p}^H(k) - \hat{\mathbf{s}}_{\text{str, LS}}^H(k)\mathbf{Y}^H(k))(\mathbf{p}(k) - \mathbf{Y}(k)\hat{\mathbf{s}}_{\text{str, LS}}(k)) \end{aligned} \quad (6.13)$$

which is equal to

$$J(k) = \mathbf{p}^H(k)\mathbf{p}(k) - \mathbf{p}^H(k)\mathbf{Y}(k)\hat{\mathbf{s}}_{\text{str,LS}}(k) - \hat{\mathbf{s}}_{\text{str,LS}}^H(k)\mathbf{Y}^H(k)\mathbf{p}(k) + \hat{\mathbf{s}}_{\text{str,LS}}^H(k)\mathbf{Y}^H(k)\mathbf{Y}(k)\hat{\mathbf{s}}_{\text{str,LS}}(k) \quad (6.14)$$

The minimum of the cost function is found by setting its gradient, with respect to  $\hat{\mathbf{s}}_{\text{str,LS}}^*(k)$ , equal to zero. To find the gradient of the different complex terms, Table A.2 of [Therrien 1992] was used, leading to

$$\nabla_{\hat{\mathbf{s}}_{\text{str,LS}}^*(k)} J(k) = 0 - 0 - \mathbf{Y}^H(k)\mathbf{p}(k) + \mathbf{Y}^H(k)\mathbf{Y}(k)\hat{\mathbf{s}}_{\text{str,LS}}(k) = 0, \quad (6.15)$$

or, after reorganizing terms, to

$$\mathbf{Y}^H(k)\mathbf{Y}(k)\hat{\mathbf{s}}_{\text{str,LS}}(k) = \mathbf{Y}^H(k)\mathbf{p}(k). \quad (6.16)$$

If the matrix  $\mathbf{Y}(k)$  is of full rank (i.e. if it has independent columns), the matrix  $\mathbf{Y}^H(k)\mathbf{Y}(k)$  is also of full rank so that its inverse exists. The least squares estimate is then

$$\begin{aligned} \hat{\mathbf{s}}_{\text{str,LS}}(k) &= (\mathbf{Y}^H(k)\mathbf{Y}(k))^{-1} \mathbf{Y}^H(k)\mathbf{p}(k) \\ &= \mathbf{Y}^+(k)\mathbf{p}(k) \end{aligned} \quad (6.17)$$

$\mathbf{Y}^+(k) = (\mathbf{Y}^H(k)\mathbf{Y}(k))^{-1} \mathbf{Y}^H(k)$  is the pseudoinverse of  $\mathbf{Y}(k)$  and is a  $M \times L$  matrix.

Using the least squares estimate, the source model is

$$\begin{aligned} \mathbf{s}_{\text{mo,LS}}(k) &= \mathbf{Y}(k)\hat{\mathbf{s}}_{\text{str,LS}}(k) \\ &= \mathbf{Y}(k)\mathbf{Y}^+(k)\mathbf{p}(k) \\ &= \mathbf{Y}(k)(\mathbf{Y}^H(k)\mathbf{Y}(k))^{-1} \mathbf{Y}^H(k)\mathbf{p}(k). \end{aligned} \quad (6.18)$$

This is a deterministic computation from the mode vector  $\mathbf{Y}(k)$  and the microphone output vector  $\mathbf{p}(k)$ . The computation does not involve the statistics of  $\mathbf{p}(k)$  or a stochastic instantaneous gradient computation like for the least-mean-square (LMS) algorithm.

### 6.3.2 Difference between the exact source signal and the source model

Because of the noise, of components from modes with orders above  $N_{\text{order}}$  not being included in the model, and of errors caused by the difference between the exact mode vectors and the mode vectors used for beamforming, there is a difference between the exact source signal and the source model, which is

$$\begin{aligned} \mathbf{s}(k) - \mathbf{s}_{\text{mo, LS}}(k) &= \mathbf{Y}'(k) \mathbf{s}'_{\text{str}}(k) + \sum_{n'=N_{\text{order}}+1}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_n^{m'}(k) \\ &\quad - \mathbf{Y}(k) (\mathbf{Y}^H(k) \mathbf{Y}(k))^{-1} \mathbf{Y}^H(k) \mathbf{p}(k) \end{aligned} \quad (6.19)$$

Substituting for  $\mathbf{p}(k)$ , the difference becomes

$$\begin{aligned} \mathbf{s}(k) - \mathbf{s}_{\text{mo, LS}}(k) &= \mathbf{Y}'(k) \mathbf{s}'_{\text{str}}(k) + \sum_{n'=N_{\text{order}}+1}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_n^{m'}(k) \\ &\quad - \mathbf{Y}(k) (\mathbf{Y}^H(k) \mathbf{Y}(k))^{-1} \mathbf{Y}^H(k) \\ &\quad \times \left[ \mathbf{Y}'(k) \mathbf{s}'_{\text{str}}(k) + \sum_{n'=N_{\text{order}}+1}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_n^{m'}(k) + \mathbf{e}(k) \right], \end{aligned} \quad (6.20)$$

which, regrouping similar terms, leads to

$$\begin{aligned} \mathbf{s}(k) - \mathbf{s}_{\text{mo, LS}}(k) &= \mathbf{Y}'(k) \mathbf{s}'_{\text{str}}(k) - \mathbf{Y}(k) (\mathbf{Y}^H(k) \mathbf{Y}(k))^{-1} \mathbf{Y}^H(k) \mathbf{Y}'(k) \mathbf{s}'_{\text{str}}(k) \\ &\quad + \sum_{n'=N_{\text{order}}+1}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \left[ \mathbf{Y}_n^{m'}(k) - \mathbf{Y}(k) (\mathbf{Y}^H(k) \mathbf{Y}(k))^{-1} \mathbf{Y}^H(k) \mathbf{Y}_n^{m'}(k) \right] \\ &\quad - \mathbf{Y}(k) (\mathbf{Y}^H(k) \mathbf{Y}(k))^{-1} \mathbf{Y}^H(k) \mathbf{e}(k) \end{aligned} \quad (6.21)$$

If  $N_{\text{order}}$  is sufficiently large so that all the spherical harmonic modes present in the source are included in  $\mathbf{Y}(k)$ ,  $s_{n'm'}(k)$  is zero for all values of  $n'$  and  $m'$  in the double summations so that the difference is

$$\begin{aligned} \mathbf{s}(k) - \mathbf{s}_{\text{mo, LS}}(k) &= \mathbf{Y}'(k) \mathbf{s}'_{\text{str}}(k) - \mathbf{Y}(k) (\mathbf{Y}^H(k) \mathbf{Y}(k))^{-1} \mathbf{Y}^H(k) \mathbf{Y}'(k) \mathbf{s}'_{\text{str}}(k) \\ &\quad - \mathbf{Y}(k) (\mathbf{Y}^H(k) \mathbf{Y}(k))^{-1} \mathbf{Y}^H(k) \mathbf{e}(k) \end{aligned} \quad (6.22)$$

Furthermore, if the source position is known exactly,  $\mathbf{Y}(k) = \mathbf{Y}'(k)$  and the difference becomes

$$\begin{aligned}
\mathbf{s}(k) - \mathbf{s}_{\text{mo,LS}}(k) &= \mathbf{Y}(k)\mathbf{s}'_{\text{str}}(k) - \mathbf{Y}(k)(\mathbf{Y}^H(k)\mathbf{Y}(k))^{-1}\mathbf{Y}^H(k)\mathbf{Y}(k)\mathbf{s}'_{\text{str}}(k) \\
&\quad - \mathbf{Y}(k)(\mathbf{Y}^H(k)\mathbf{Y}(k))^{-1}\mathbf{Y}^H(k)\mathbf{e}(k) \\
&= \mathbf{Y}(k)\mathbf{s}'_{\text{str}}(k) - \mathbf{Y}(k)\mathbf{s}'_{\text{str}}(k) - \mathbf{Y}(k)(\mathbf{Y}^H(k)\mathbf{Y}(k))^{-1}\mathbf{Y}^H(k)\mathbf{e}(k) \\
&= -\mathbf{Y}(k)(\mathbf{Y}^H(k)\mathbf{Y}(k))^{-1}\mathbf{Y}^H(k)\mathbf{e}(k)
\end{aligned} \tag{6.23}$$

The noise vector causes the only remaining term.

In the case where there would be reverberation or interfering sources, those would also lead to terms in the difference between the source signal and the source model.

### 6.3.3 Overall beamforming weight vector

Substituting for the source model Eq. (6.18) in Eq. (2.17), the overall beamforming weight vector is

$$\begin{aligned}
\mathbf{w}_{\text{LS}}(k) &= \mathbf{R}_{\text{ee}}^{-1}(k)\mathbf{s}_{\text{mo,LS}}(k) \\
&= \mathbf{R}_{\text{ee}}^{-1}(k)\mathbf{Y}(k)(\mathbf{Y}^H(k)\mathbf{Y}(k))^{-1}\mathbf{Y}^H(k)\mathbf{p}(k) \\
&= \mathbf{R}_{\text{ee}}^{-1}(k)\mathbf{Y}(k)\mathbf{Y}^+(k)\mathbf{p}(k)
\end{aligned} \tag{6.24}$$

$\mathbf{p}(k)$  is the particular realization of the microphone output vector that was used in the estimation.

### 6.3.4 Computation of the pseudoinverse and the source model

The  $L \times M$  matrix  $\mathbf{Y}(k)$  has for singular value decomposition (SVD) [Therrien 1992]

$$\mathbf{Y}(k) = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \tag{6.25}$$

where  $\mathbf{U}$  is the  $L \times L$  unitary matrix of left singular vectors (column vectors),  $\mathbf{V}$  is the  $M \times M$  unitary matrix of right singular vectors (column vectors), and  $\mathbf{\Sigma}$  is the  $L \times M$  matrix of nonnegative real singular values. It has the form

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{S}_{r \times r} & \mathbf{0}_{r \times (M-r)} \\ \mathbf{0}_{(L-r) \times r} & \mathbf{0}_{(L-r) \times (M-r)} \end{bmatrix}, \quad (6.26)$$

where  $\mathbf{S}_{r \times r}$  is a diagonal matrix of the nonzero singular values

$$\mathbf{S}_{r \times r} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \quad (6.27)$$

and  $r$  is the rank of  $\mathbf{Y}(k)$  (the number of independent columns of  $\mathbf{Y}(k)$ ).

The pseudoinverse of  $\mathbf{Y}(k)$  is defined as

$$\mathbf{Y}^+(k) = \mathbf{V} \mathbf{\Sigma}^+ \mathbf{U}^H. \quad (6.28)$$

$\mathbf{\Sigma}^+$  is the pseudoinverse of  $\mathbf{\Sigma}$  and is the  $M \times L$  matrix

$$\mathbf{\Sigma}^+ = \begin{bmatrix} \mathbf{S}_{r \times r}^{-1} & \mathbf{0}_{r \times (L-r)} \\ \mathbf{0}_{(M-r) \times r} & \mathbf{0}_{(M-r) \times (L-r)} \end{bmatrix}, \quad (6.29)$$

where

$$\mathbf{S}_{r \times r}^{-1} = \begin{bmatrix} 1/\sigma_1 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_r \end{bmatrix}. \quad (6.30)$$

In the case where  $L \geq M$  and the matrix  $\mathbf{Y}(k)$  has full rank  $M$ , this definition of the pseudoinverse is equivalent to  $\mathbf{Y}^+(k) = (\mathbf{Y}^H(k)\mathbf{Y}(k))^{-1}\mathbf{Y}^H(k)$ . Furthermore, if  $\mathbf{Y}(k)$  is square and of full rank,  $\mathbf{Y}^+(k) = \mathbf{Y}^{-1}(k)$ .

The rank of the matrix  $\mathbf{Y}(k)$  depends on the array geometry and the position of the source. In a beamforming situation where the array has a limited view of the source and spherical harmonic modes look similar on the array (e.g. as in the measurement setup discussed in Chapter 5), we would expect  $\mathbf{Y}(k)$  not to be of full rank  $M$  if higher order

modes are included. In which case, computing the pseudoinverse with  $(\mathbf{Y}^H(k)\mathbf{Y}(k))^{-1}\mathbf{Y}^H(k)$  and with singular value decomposition do not give the same results, leading to different beamforming weight vectors. Even if a result is found for  $(\mathbf{Y}^H(k)\mathbf{Y}(k))^{-1}\mathbf{Y}^H(k)$ , it may not be accurate because the  $M \times M$  matrix  $\mathbf{Y}^H(k)\mathbf{Y}(k)$  may be ill conditioned. A necessary but not sufficient condition for the matrix  $\mathbf{Y}(k)$  to be of full rank  $M$  is that the number of modes included be smaller or equal to the number of microphones  $L$ .

The product of  $\mathbf{Y}(k)$  and  $\mathbf{Y}^+(k)$  appears in the development above for the source model. Properties of singular value decomposition can be used to reduce the computational complexity by avoiding multiplying the matrix  $\mathbf{Y}(k)$  with its pseudoinverse. Alternative forms to write the two matrices using the singular vectors are:

$$\mathbf{Y}(k) = \sum_{m=1}^r \sigma_m \mathbf{u}_m \mathbf{v}_m^H \quad (6.31)$$

and

$$\mathbf{Y}^+(k) = \sum_{m=1}^r \frac{1}{\sigma_m} \mathbf{v}_m \mathbf{u}_m^H. \quad (6.32)$$

The product is

$$\begin{aligned} \mathbf{Y}(k)\mathbf{Y}^+(k) &= (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H)(\mathbf{V}\mathbf{\Sigma}^+ \mathbf{U}^H) \\ &= \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^+ \mathbf{U}^H \\ &= \mathbf{U} \begin{bmatrix} \mathbf{S}_{r \times r} & \mathbf{0}_{r \times (M-r)} \\ \mathbf{0}_{(L-r) \times r} & \mathbf{0}_{(L-r) \times (M-r)} \end{bmatrix} \begin{bmatrix} \mathbf{S}_{r \times r}^{-1} & \mathbf{0}_{r \times (L-r)} \\ \mathbf{0}_{(M-r) \times r} & \mathbf{0}_{(M-r) \times (L-r)} \end{bmatrix} \mathbf{U}^H \\ &= \mathbf{U} \begin{bmatrix} \mathbf{I}_{r \times r} & \mathbf{0}_{r \times (L-r)} \\ \mathbf{0}_{(L-r) \times r} & \mathbf{0}_{(L-r) \times (M-r)} \end{bmatrix} \mathbf{U}^H \\ &= \sum_{m=1}^r \mathbf{u}_m \mathbf{u}_m^H \end{aligned} \quad (6.33)$$

$\mathbf{Y}(k)\mathbf{Y}^+(k)$  is an  $L \times L$  matrix. If  $L > M$  (more microphones than modes), the resulting  $\mathbf{Y}(k)\mathbf{Y}^+(k)$  matrix is rank deficient, that is  $r < L$ .

The source model can be found as

$$\begin{aligned} \mathbf{s}_{\text{mo, LS}}(k) &= \mathbf{Y}(k)\mathbf{Y}^+(k)\mathbf{p}(k) = \left( \sum_{m=1}^r \mathbf{u}_m \mathbf{u}_m^H \right) \mathbf{p}(k) \\ &= \mathbf{u}_1 (\mathbf{u}_1^H \mathbf{p}(k)) + \mathbf{u}_2 (\mathbf{u}_2^H \mathbf{p}(k)) + \dots + \mathbf{u}_r (\mathbf{u}_r^H \mathbf{p}(k)). \end{aligned} \quad (6.34)$$

Since the terms within parenthesis are complex scalars, the order of the multiplication can be changed to get

$$\begin{aligned} \mathbf{s}_{\text{mo, LS}}(k) &= (\mathbf{u}_1^H \mathbf{p}(k))\mathbf{u}_1 + (\mathbf{u}_2^H \mathbf{p}(k))\mathbf{u}_2 + \dots + (\mathbf{u}_r^H \mathbf{p}(k))\mathbf{u}_r \\ &= \sum_{m=1}^r (\mathbf{u}_m^H \mathbf{p}(k))\mathbf{u}_m \end{aligned} \quad (6.35)$$

The source model is a weighted sum of left singular vectors with each weight being the projection of the microphone output vector on the corresponding left singular vector.

### 6.3.5 Details on estimated mode strengths

It is also interesting to look at the product  $\mathbf{Y}^+(k)\mathbf{Y}(k)$  as this will inform us on what the estimated mode strengths are. With the help of the singular value decomposition, the product is

$$\begin{aligned} \mathbf{Y}^+(k)\mathbf{Y}(k) &= (\mathbf{V}\mathbf{\Sigma}^+\mathbf{U}^H)(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H) \\ &= \mathbf{V}\mathbf{\Sigma}^+\mathbf{\Sigma}\mathbf{V}^H \\ &= \mathbf{V} \begin{bmatrix} \mathbf{S}_{r \times r}^{-1} & \mathbf{0}_{r \times (L-r)} \\ \mathbf{0}_{(M-r) \times r} & \mathbf{0}_{(M-r) \times (L-r)} \end{bmatrix} \begin{bmatrix} \mathbf{S}_{r \times r} & \mathbf{0}_{r \times (M-r)} \\ \mathbf{0}_{(L-r) \times r} & \mathbf{0}_{(L-r) \times (M-r)} \end{bmatrix} \mathbf{V}^H \\ &= \mathbf{V} \begin{bmatrix} \mathbf{I}_{r \times r} & \mathbf{0}_{r \times (M-r)} \\ \mathbf{0}_{(M-r) \times r} & \mathbf{0}_{(M-r) \times (M-r)} \end{bmatrix} \mathbf{V}^H \\ &= \mathbf{I}_{r \times r} \end{aligned} \quad (6.36)$$

If  $\mathbf{Y}(k)$  has a full rank  $M$ ,  $\mathbf{Y}^+(k)\mathbf{Y}(k) = \mathbf{I}_{M \times M}$ . This means that the vector product of one line of  $\mathbf{Y}^+(k)$  by one column of  $\mathbf{Y}(k)$  is equal to zero unless the line number is the same as the column number.

The microphone output vector is written in Eq. (6.7) as

$$\mathbf{p}(k) = \mathbf{Y}'(k)\mathbf{s}'_{\text{str}}(k) + \sum_{n'=N_{\text{order}}+1}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k)\mathbf{Y}_n^{m'}(k) + \mathbf{e}(k). \quad (6.37)$$

Assuming, as was done in the middle of Sub-Section 6.3.2, that the source position is known exactly,  $\mathbf{Y}'(k) = \mathbf{Y}(k)$ , and that all the modes present in the source are included in the matrix  $\mathbf{Y}(k)$ , the microphone output vector becomes

$$\mathbf{p}(k) = \mathbf{Y}(k)\mathbf{s}'_{\text{str}}(k) + \mathbf{e}(k). \quad (6.38)$$

The modes strengths are estimated all at once with Eq. (6.17) which, after substituting for the microphone output vector, is

$$\begin{aligned} \hat{\mathbf{s}}_{\text{str,LS}}(k) &= \mathbf{Y}^+(k)\mathbf{Y}(k)\mathbf{s}'_{\text{str}}(k) + \mathbf{Y}^+(k)\mathbf{e}(k) \\ &= \mathbf{s}'_{\text{str}}(k) + \mathbf{Y}^+(k)\mathbf{e}(k) \end{aligned} \quad (6.39)$$

To go from the first to the second line, we have assumed that  $\mathbf{Y}(k)$  has a full rank  $M$  so that  $\mathbf{Y}^+(k)\mathbf{Y}(k) = \mathbf{I}_{M \times M}$ . We note that even if the spherical harmonic modes are not orthogonal, with the least-squares estimation they do not bias the strength estimation of other modes. However, there is contamination of the  $m^{\text{th}}$  estimated mode strength by the vector product of the  $m^{\text{th}}$  line of the matrix  $\mathbf{Y}^+(k)$  with the noise vector  $\mathbf{e}(k)$ .

### 6.3.6 Least squares with source and interferer models (LSSI)

Minimizing the sum of the square of the errors between the microphone output vector and the source model will cause additional errors in the estimated mode strengths when there is an interfering source.

If the location of the interferer is known, it is possible to use the least squares method to estimate both the source model and the interferer model, and then use the source model to realize the beamforming.

A first matrix is constructed for the mode vectors at the source location

$$\mathbf{Y}_s(k) = [\mathbf{Y}_0^0(k), \mathbf{Y}_1^{-1}(k), \mathbf{Y}_1^0(k), \mathbf{Y}_1^{-1}(k), \dots, \mathbf{Y}_{N_{\text{order}}}^{N_{\text{order}}}(k)] , \quad (6.40)$$



where the subscript “s” indicates that this is for the source location. A second matrix is constructed for the mode vectors at the interferer location

$$\mathbf{Y}_I(k) = [\mathbf{Y}_{0,I}^0(k), \mathbf{Y}_{1,I}^{-1}(k), \mathbf{Y}_{1,I}^0(k), \mathbf{Y}_{1,I}^{-1}(k), \dots, \mathbf{Y}_{N_{\text{order}},I}^{N_{\text{order}}}(k)] , \quad (6.41)$$

where the subscript “I” indicates that this is for the interferer location. Those two matrices are combined into a single matrix

$$\mathbf{Y}(k) = [\mathbf{Y}_s(k), \mathbf{Y}_I(k)] . \quad (6.42)$$

A column vector containing the mode strengths we wish to determine for the source is built

$$\hat{\mathbf{s}}_{\text{str},s}(k) = [s_{0,0}(k), s_{1,-1}(k), s_{1,0}(k), s_{1,1}(k), \dots, s_{N_{\text{order}},N_{\text{order}}}(k)]^T \quad (6.43)$$

as is a similar column vector for the interferer mode strengths

$$\hat{\mathbf{s}}_{\text{str},I}(k) = [s_{0,0,I}(k), s_{1,-1,I}(k), s_{1,0,I}(k), s_{1,1,I}(k), \dots, s_{N_{\text{order}},N_{\text{order}},I}(k)]^T \quad (6.44)$$

which are then combined into a single vector

$$\hat{\mathbf{s}}_{\text{str},\text{LSSI}}(k) = \begin{bmatrix} \hat{\mathbf{s}}_{\text{str},s}(k) \\ \hat{\mathbf{s}}_{\text{str},I}(k) \end{bmatrix} . \quad (6.45)$$

The computation of the least squares solution is then performed as

$$\begin{aligned} \hat{\mathbf{s}}_{\text{str},\text{LSSI}}(k) &= (\mathbf{Y}^H(k)\mathbf{Y}(k))^{-1} \mathbf{Y}^H(k)\mathbf{p}(k) \\ &= \mathbf{Y}^+(k)\mathbf{p}(k) . \end{aligned} \quad (6.46)$$

Finally, the source model is built using the mode vector for the source location and the estimated source mode strengths

$$\mathbf{s}_{\text{mo},s}(k) = \mathbf{Y}_s(k)\hat{\mathbf{s}}_{\text{str},s}(k) . \quad (6.47)$$

The principle is not limited to a single interferer. In the situation where there would be multiple interferers at known locations, additional mode vectors and estimated

mode strengths could be included in  $\mathbf{Y}(k)$  and  $\hat{\mathbf{s}}_{\text{str, LSSI}}(k)$  respectively, before computing the least squares solution.

We use the abbreviation LSSI to refer to this method.

## 6.4 Proposed beamforming method to estimate the source model for MVDR beamforming

Linearly Constrained Minimum Variance (LCMV) beamforming and Minimum Variance Distortionless Response (MVDR) beamforming are popular methods which were briefly discussed in Sub-Section 2.4.2. It was mentioned that, with those methods, a source model is required to compute the beamforming weight vector.

It could be possible to use the proposed beamforming method to estimate a source model, given by Eq. (4.51) if the GO estimator is used, or given by Eq. (6.18) if least squares is used instead. This estimated source model could then be used to compute the beamforming vector with Eq. (2.26) for the MVDR beamformer. The beamforming performance of MVDR will then depend on how good the estimation of the source model is, in the presence of noise, interferers, or reverberation. This will be further investigated in Chapter 7. The same concept could also be applied for the LCMV beamformer.

## 6.5 Estimation using many frames

Until now, the estimation of the mode strengths was realized after performing an FFT on a single frame of time data, for each microphone output. We will now examine the possibility of improving the estimation by using many consecutive frames.

When performing power spectral estimation with non parametric methods such as the Bartlett or the Welch methods [Proakis 1996], the time data is segmented into (possibly overlapping) frames and an FFT is computed on each frame. The power spectral estimate is obtained by averaging together the square of each normalized FFT. This is possible because only the magnitude (or the magnitude square) is of interest. In our case however, the phase information is also necessary to obtain the mode strengths.

As seen in Section 2.2, the time-domain output signal for microphone  $l$  is  $p_l(n)$ , where  $n$  is the sample index (not a spherical harmonic order). This time sequence is first segmented into frames of  $N$  time samples

$$p_l(n, fr) = p_l(n + [fr - 1]N), \quad 0 \leq n \leq N - 1, fr \geq 1, \quad (6.48)$$

where  $fr$  is the frame index. The normalized FFT, Eq. (2.2), is then applied on each frame

$$p_l(k, fr) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} p_l(n, fr) e^{-i2\pi kn/N}. \quad (6.49)$$

It is not possible to average the FFTs together (e.g.  $p_l(k) = \frac{1}{N_{\text{frames}}} \sum_{fr=1}^{N_{\text{frames}}} p_l(k, fr)$ )

since for a given frequency bin  $k$  the phase will be different among frames (unless the signal is composed of sinusoids whose frequencies are exactly at the center of frequency bins). As a side note, for the unrealistic case where the signal is composed of a single sinusoid whose frequency is known precisely, it would be possible to compensate for the phase and then average.

A possible alternative is to compute the mode strengths estimator  $d_{nm}(k, fr)$  for each frame. Then, the mode strength having the largest magnitude is found and used as a reference, calling it  $d_{n_{\text{ref}}m_{\text{ref}}}(k, fr)$ . (This assumes that the mode strength having the largest magnitude is the same in all frames, which may not be the case for low input SNR, or in the presence of significant reverberation or interferences.) All the estimated mode strengths of frame  $fr$  are divided by  $d_{n_{\text{ref}}m_{\text{ref}}}(k, fr)$  to realize the phase alignment before averaging across frames as given by

$$d_{nm}(k) = \frac{1}{N_{\text{frames}}} \sum_{fr=1}^{N_{\text{frames}}} \frac{d_{nm}(k, fr)}{d_{n_{\text{ref}}m_{\text{ref}}}(k, fr)}. \quad (6.50)$$

Because the phase alignment uses  $d_{n_{\text{ref}},m_{\text{ref}}}(k, fr)$ , the estimated strength for mode  $(n_{\text{ref}}, m_{\text{ref}})$  after averaging is  $d_{n_{\text{ref}},m_{\text{ref}}}(k) = 1$ . The same weight is given to all frames even if the signal power is lower in a particular frame.

This can be illustrated with an example (dipole source and rectangular array geometry of Sub-Section 7.3.2). The source is a dipole, with a source frequency of 1000 Hz (not at the center of a frequency bin so that the phase changes from frame to frame), and with an input SNR of 50 dB (so that very little additional phase changes are caused by the noise). In the tables, the first number is the magnitude (linear scale) while the number following “ $\angle$ ” is the phase in degrees.

Tables 6.1 and 6.2 show the estimated mode strengths for frame 1 and 2, respectively. The maximum of the estimated mode strengths is for  $n_{\text{ref}} = 4, m_{\text{ref}} = 0$  even though only modes  $(n, m) = (1, -1)$  and  $(n, m) = (1, 1)$  are present in the source. This happens because the array has a limited view of the dipole lobes. The source looks somewhat similar to other modes. (This would not be the case for an array more completely surrounding the source.)

For frame 1,  $d_{n_{\text{ref}},m_{\text{ref}}}(k,1)$  equals  $18.41 \angle -30.01^\circ$  while for frame 2,  $d_{n_{\text{ref}},m_{\text{ref}}}(k,2)$  equals  $18.43 \angle 89.98^\circ$ . Table 6.3 shows the mode strengths for frame 1 after dividing by  $d_{n_{\text{ref}},m_{\text{ref}}}(k,1)$  and Table 6.4 shows the mode strengths for frame 2 after dividing by  $d_{n_{\text{ref}},m_{\text{ref}}}(k,2)$ .

Table 6.1. Estimated mode strengths, frame 1.

n	0	1	2	3	4
m	$d_{nm}(k,1)$				
-4					2.76 $\angle$ -134.90°
-3				2.41 $\angle$ 109.16°	2.13 $\angle$ 23.00°
-2			2.12 $\angle$ 45.58°	4.54 $\angle$ -82.76°	17.24 $\angle$ 168.54°
-1		15.88 $\angle$ 41.13°	15.62 $\angle$ -45.17°	9.39 $\angle$ 22.93°	18.03 $\angle$ -40.11°
0	3.89 $\angle$ 156.41°	10.70 $\angle$ 55.74°	9.27 $\angle$ -58.72°	5.70 $\angle$ 70.17°	<b>18.41 <math>\angle</math> -30.01°</b>
1		15.43 $\angle$ -125.81°	15.48 $\angle$ 150.21°	7.58 $\angle$ -161.70°	17.54 $\angle$ 158.33°
2			1.30 $\angle$ -55.66°	4.20 $\angle$ -144.57°	15.85 $\angle$ 130.03°
3				2.58 $\angle$ -160.94°	2.26 $\angle$ 105.77°
4					2.34 $\angle$ 33.82°

Table 6.2. Estimated mode strengths, frame 2.

n	0	1	2	3	4
m	$d_{nm}(k,2)$				
-4					2.76 $\angle$ -15.04°
-3				2.41 $\angle$ -130.75°	2.13 $\angle$ 143.20°
-2			2.12 $\angle$ 165.09°	4.56 $\angle$ 37.09°	17.23 $\angle$ -71.52°
-1		15.87 $\angle$ 161.11°	15.62 $\angle$ 74.83°	9.37 $\angle$ 142.66°	18.02 $\angle$ 79.88°
0	3.92 $\angle$ -83.59°	10.74 $\angle$ 175.76°	9.23 $\angle$ 61.06°	5.73 $\angle$ -169.81°	<b>18.43 <math>\angle</math> 89.98°</b>
1		15.43 $\angle$ -5.88°	15.49 $\angle$ -89.87°	7.54 $\angle$ -41.91°	17.54 $\angle$ -81.74°
2			1.30 $\angle$ 64.25°	4.20 $\angle$ -24.67°	15.86 $\angle$ -110.07°
3				2.58 $\angle$ -40.75°	2.26 $\angle$ -134.17°
4					2.35 $\angle$ 153.94°

Table 6.3. Estimated mode strengths after phase alignment, frame 1.

n	0	1	2	3	4
m	$d_{nm}(k,1)/d_{n_{ref}m_{ref}}(k,1)$				
-4					0.15 $\angle$ -104.88°
-3				0.13 $\angle$ 139.18°	0.12 $\angle$ 53.02°
-2			0.11 $\angle$ 75.59°	0.25 $\angle$ -52.75°	0.94 $\angle$ -161.44°
-1		0.86 $\angle$ 71.15°	0.85 $\angle$ -15.15°	0.51 $\angle$ 52.95°	0.98 $\angle$ -10.1°
0	0.21 $\angle$ -173.58°	0.58 $\angle$ 85.75°	0.5 $\angle$ -28.71°	0.31 $\angle$ 100.19°	<b>1 <math>\angle</math> 0°</b>
1		0.84 $\angle$ -95.8°	0.84 $\angle$ -179.77°	0.41 $\angle$ -131.69°	0.95 $\angle$ -171.66°
2			0.07 $\angle$ -25.65°	0.23 $\angle$ -114.55°	0.86 $\angle$ 160.04°
3				0.14 $\angle$ -130.93°	0.12 $\angle$ 135.78°
4					0.13 $\angle$ 63.83°

Table 6.4. Estimated mode strengths after phase alignment, frame 2.

n	0	1	2	3	4
m	$d_{nm}(k,2)/d_{n_{ref}m_{ref}}(k,2)$				
-4					0.15 $\angle$ -105.02°
-3				0.13 $\angle$ 139.27°	0.12 $\angle$ 53.22°
-2			0.12 $\angle$ 75.11°	0.25 $\angle$ -52.89°	0.94 $\angle$ -161.5°
-1		0.86 $\angle$ 71.13°	0.85 $\angle$ -15.15°	0.51 $\angle$ 52.68°	0.98 $\angle$ -10.11°
0	0.21 $\angle$ -173.57°	0.58 $\angle$ 85.78°	0.5 $\angle$ -28.92°	0.31 $\angle$ 100.21°	<b>1 <math>\angle</math> 0°</b>
1		0.84 $\angle$ -95.87°	0.84 $\angle$ -179.85°	0.41 $\angle$ -131.89°	0.95 $\angle$ -171.72°
2			0.07 $\angle$ -25.73°	0.23 $\angle$ -114.65°	0.86 $\angle$ 159.95°
3				0.14 $\angle$ -130.73°	0.12 $\angle$ 135.84°
4					0.13 $\angle$ 63.96°

Finally, Table 6.5 presents the mode strengths after averaging across frames using Eq. (6.50).  $d_{n_{\text{ref}}, m_{\text{ref}}}(k)$  is indeed equal to 1.

Table 6.5. Estimated mode strengths after averaging across frames.

n	0	1	2	3	4
m	$d_{nm}(k)$				
-4					0.15 $\angle$ -104.95°
-3				0.13 $\angle$ 139.22°	0.12 $\angle$ 53.12°
-2			0.12 $\angle$ 75.35°	0.25 $\angle$ -52.82°	0.94 $\angle$ -161.47°
-1		0.86 $\angle$ 71.14°	0.85 $\angle$ -15.15°	0.51 $\angle$ 52.81°	0.98 $\angle$ -10.10°
0	0.21 $\angle$ -173.57°	0.58 $\angle$ 85.76°	0.5 $\angle$ -28.81°	0.31 $\angle$ 100.20°	<b>1 <math>\angle</math> 0°</b>
1		0.84 $\angle$ -95.83°	0.84 $\angle$ -179.81°	0.41 $\angle$ -131.79°	0.95 $\angle$ -171.69°
2			0.07 $\angle$ -25.69°	0.23 $\angle$ -114.60°	0.86 $\angle$ 160.00°
3				0.14 $\angle$ -130.83°	0.12 $\angle$ 135.81°
4					0.13 $\angle$ 63.90°

Another option to compute the mode strengths across frame would be

$$d_{nm}(k) = \frac{1}{N_{\text{frames}}} \sum_{fr=1}^{N_{\text{frames}}} \frac{d_{nm}(k, fr)}{e^{i\angle d_{n_{\text{ref}}, m_{\text{ref}}}(k, fr)}}, \quad (6.51)$$

where the phase is adjusted but not the magnitude. A frame with a lower signal power would be given a lower weight in the sum.

Inspired by the Multiple Scatterer Algorithm (MSA) [Steinberg 1991] and the Recursive Multiple Scatterer Algorithm (RMSA) [Wu 1995] [Bouchard 2000] used to focus an Inverse Synthetic Aperture Radar (ISAR) on a target for image processing, it would also be possible to use more than one mode to realize the phase alignment. A reference phase is first built from the phase of the modes  $(n_1, m_1), (n_2, m_2), \dots, (n_{N_{\text{ref}}}, m_{N_{\text{ref}}})$  having the largest strengths as

$$\angle d_{n_{\text{ref}}m_{\text{ref}}}(k, fr) = \frac{1}{N_{\text{ref}}} \sum_{r=1}^{N_{\text{ref}}} \angle d_{n_{\text{ref}}m_{\text{ref}}}(k, fr). \quad (6.52)$$

The average across frames is then given by

$$d_{nm}(k) = \frac{1}{N_{\text{frames}}} \sum_{fr=1}^{N_{\text{frames}}} \frac{d_{nm}(k, fr)}{e^{i\angle d_{n_{\text{ref}}m_{\text{ref}}}(k, fr)}}. \quad (6.53)$$

Comparison of beamforming performance when a single frame is used for the estimation and when several frames are used with Eq. (6.50), will be presented in the next chapter.

## 6.6 Chapter conclusion

Chapter 6 has presented a number of estimators for the source mode strengths, required in the proposed beamforming method for directional sources. Because of the difficulty to directly evaluate analytically the beamforming performance from the equations, investigation of the performance of the proposed beamforming method with the different estimators is realized using simulations in Chapter 7.



---

**CHAPTER**  
**7**

---

**Performance of the Proposed Beamforming Method  
in the Presence of Diffuse Noise, Reverberation, or  
Interfering Sources****7.1 Introduction**

The proposed beamforming method for directional sources and the gain-optimized (GO) estimator of the mode strengths were presented in Chapter 4. Chapter 5 proved that experimental and simulation results were comparable. Chapter 6 revisited the GO estimator and introduced a few additional estimators. Instead of going back to the experiments of Chapter 5 with all the estimators, we choose in Chapter 7 to continue with simulations because they provide more control over parameters like source directivity, signal-to-noise ratio, signal-to-interference ratio, etc., than do the experimental data.

**7.1.1 Organization**

The simulations are divided in three categories:

- A source in the presence of diffuse noise (Sections 7.2 and 7.3);
- A source in the presence of diffuse noise and reverberation (Sections 7.4 and 7.5);
- A source in the presence of diffuse noise and an interferer (Sections 7.6 and 7.7).

Each category is discussed in two sections, the first introduces the simulation method while the second presents the simulation results and analysis.

More details are provided in the respective sections but in all cases we assume that the diffuse noise is uncorrelated with the source. On the other hand, reverberation is correlated with the source, but uncorrelated with the noise. Source, interferer, and noise are mutually uncorrelated. An interferer is a source other than the source on which we are beamforming. It has its own location and its own directivity.

### **7.1.2 Array geometries**

Simulations are done with two different array geometries:

- (1) A linear array with 16 uniformly spaced microphones with 8 cm spacing. This is the same geometry as for the experiment in Chapter 5.
- (2) 32 microphones positioned on the perimeter of a rectangle parallel to the  $x$ - $y$  plane, 1 meter above the origin. Parallel to the  $x$ -axis, the microphones are located between -3.5 and 3.5 m, with a spacing of 0.778 m. Parallel to the  $y$ -axis, the microphones are located between -2.5 and 2.5 m, with a spacing of 0.714 m. This was chosen as an array that surrounds part of the source while still maintaining a low number of microphones.

The microphones are omnidirectional for both arrays. The geometries are shown in Fig. 7.1, with a dipole source.

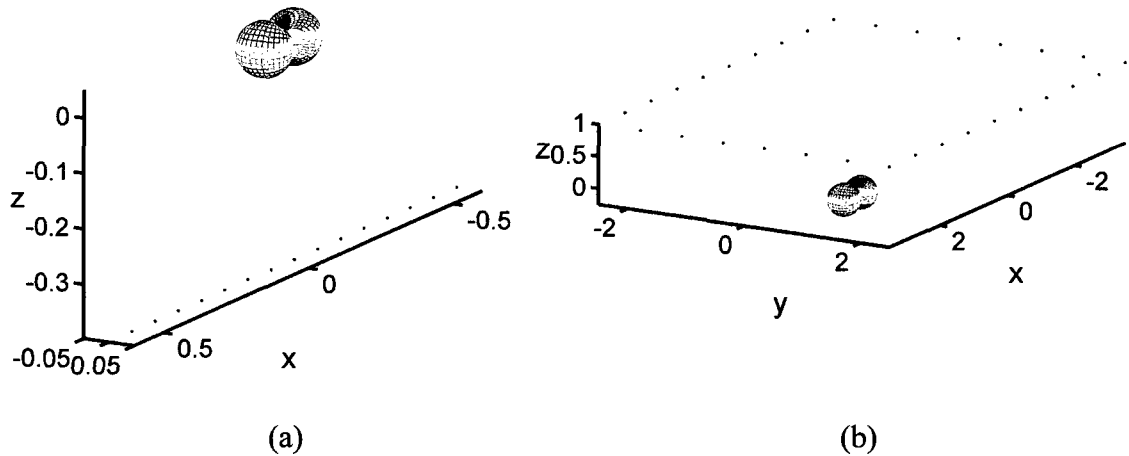


Figure 7.1. The two array geometries (with a dipole source) used for simulations: (a) Linear array; (b) Rectangular array.

### 7.1.3 Synopsis of the results

Chapter 7 is extensive and there are numerous figures and tables. It may seem repetitive but we wanted to document the three scenario categories we investigated: a source in the presence of diffuse noise, a source in the presence of diffuse noise and reverberation, and a source in the presence of diffuse noise and an interferer. The page header indicates the simulation scenario, including source and array, for that page.

The highlights of the results are:

- When beamforming on directional sources, the proposed beamforming method outperforms conventional beamforming when the input SNR is equal or larger than 0 dB.
- For lower input SNR (i.e. lower than 0 dB), there is a statistical distribution of the array gain for the proposed beamforming method but, depending on the value of the input SNR, a large portion of the realizations provides good array gains.
- If the source is a monopole but this is unknown when performing the mode strength estimation, the proposed beamforming method works and provides an array gain close to the optimum value for input SNR values of 0 dB or higher.

- It is possible to use the proposed beamforming method optimizing for the array gain subject to a constraint on the white-noise gain (end of Sub-Section 7.3.1), which adds robustness to the method.
- Estimating a source model with the proposed beamforming method and using this model for MVDR beamforming does not perform adequately (middle of Sub-Section 7.3.1, end of Sub-Section 7.5.1, and second half of Sub-Section 7.7.1).
- In the presence of reverberation (Section 7.5), of an interferer (Section 7.7), or beamforming position errors (Sub-Section 7.7.2), there is degradation of performance but the proposed beamforming method still provides a reasonable performance.

## 7.2 Simulations: Diffuse noise

A spherically isotropic noise field can be built with an infinite number of uncorrelated plane waves, arriving from all directions, and having all the same auto-correlation function [Rafaely 2000]. It is assumed that the acoustic field is spatially homogeneous so that the correlation function between two microphones does not depend on the absolute position of the two microphones [Elko 2001b] and that the (self) power-spectral density is the same for all microphones. In room acoustics, a spherically isotropic noise field is often referred to as a “diffuse” sound field [Elko 2001a].

The next two sub-sections describe the simulation of diffuse noise and give a description of the simulation method. Analysis of the simulation results to evaluate the performance of the proposed beamforming method in the presence of diffuse noise is presented in Section 7.3.

### 7.2.1 Simulation of diffuse noise

The cross-power spectral density (cross-PSD) function between two omnidirectional microphones for diffuse noise is [Cook 1955] [Elko 2001b]

$$S_{l_1, l_2}(f) = S_{xx}(f) \frac{\sin(2\pi f d_{l_1, l_2}/c)}{2\pi f d_{l_1, l_2}/c}, \quad (7.1)$$

where  $d_{i,j_2}$  is the distance separating the two microphones and  $S_{xx}(f)$  is the (self) power-spectral density of the diffuse noise at the output of the microphones.

To evaluate the proposed beamforming method in the presence of diffuse noise, it is necessary to generate noise vectors  $\mathbf{e}(k)$  having a cross-power spectral density function between microphone pairs according to Eq. (7.1).

A computationally intensive method to supply the noise vector  $\mathbf{e}(k)$  would be to first generate  $L$  time-domain noise sequences ( $e_1(n), e_2(n), \dots, e_L(n)$ ) which would then go through  $L$  normalized FFT. To have the proper cross-power spectral density function between each pair of time-domain noise sequences, the sequences would need to be created as the sum of a large number of uncorrelated random sequences, corresponding to plane waves arriving from different directions. Those random sequences would then have to be delayed at each microphone based on the distance they would have traveled to reach each microphone.

Since at this point we are evaluating the performance of the proposed beamforming method at a single frequency bin, a much more computationally efficient method is to generate the vector  $\mathbf{e}(k)$  directly using the cross-power spectral density for diffuse noise.

The normalized noise cross-power spectral density matrix  $\mathbf{R}_{ee}(f)$  for diffuse noise and the array geometry is built using Eq. (7.1) and is written as

$$\mathbf{R}_{ee}(k) = \begin{bmatrix} \frac{\sin(2\pi k F_s d_{1,1}/(Nc))}{2\pi k F_s d_{1,1}/(Nc)} & \frac{\sin(2\pi k F_s d_{1,2}/(Nc))}{2\pi k F_s d_{1,2}/(Nc)} & \dots & \frac{\sin(2\pi k F_s d_{1,L}/(Nc))}{2\pi k F_s d_{1,L}/(Nc)} \\ \frac{\sin(2\pi k F_s d_{2,1}/(Nc))}{2\pi k F_s d_{2,1}/(Nc)} & \frac{\sin(2\pi k F_s d_{2,2}/(Nc))}{2\pi k F_s d_{2,2}/(Nc)} & \dots & \frac{\sin(2\pi k F_s d_{2,L}/(Nc))}{2\pi k F_s d_{2,L}/(Nc)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sin(2\pi k F_s d_{L,1}/(Nc))}{2\pi k F_s d_{L,1}/(Nc)} & \frac{\sin(2\pi k F_s d_{L,2}/(Nc))}{2\pi k F_s d_{L,2}/(Nc)} & \dots & \frac{\sin(2\pi k F_s d_{L,L}/(Nc))}{2\pi k F_s d_{L,L}/(Nc)} \end{bmatrix}, \quad (7.2)$$

where  $F_s$  is the sampling frequency and  $N$  is the number of time samples used in the normalized FFT.

We also specify the required average noise power spectral density over the array,  $\sigma_e^2(k)$ , so that the noise vector  $\mathbf{e}(k)$ , has a correlation matrix  $E\{\mathbf{e}(k)\mathbf{e}^H(k)\} = \sigma_e^2(k)\mathbf{R}_{ee}(k)$ . The vector  $\mathbf{e}(k)$  having such a correlation matrix is obtained from a linear transformation applied on a noise vector  $\mathbf{e}_w$  whose elements are uncorrelated random complex numbers, with correlation matrix  $E\{\mathbf{e}_w\mathbf{e}_w^H\}$  equal to the identity matrix. To find the necessary linear transformation, eigenvalue decomposition is performed on  $\mathbf{R}_{ee}(k)$  such that it is written as

$$\mathbf{R}_{ee}(k) = \mathbf{E}(k)\mathbf{\Lambda}(k)\mathbf{E}^H(k). \quad (7.3)$$

$\mathbf{E}(k)$  is a matrix whose columns are the eigenvectors of  $\mathbf{R}_{ee}(k)$  and  $\mathbf{\Lambda}(k)$  is a diagonal matrix whose elements are the eigenvalues  $\lambda_1, \dots, \lambda_L$ . Because  $\mathbf{R}_{ee}(k)$  is Hermitian symmetric, there are  $L$  orthonormal eigenvectors and  $L$  real eigenvalues.

It is easy to demonstrate that the vector  $\mathbf{e}(k)$  for diffuse noise, having the desired correlation matrix  $E\{\mathbf{e}(k)\mathbf{e}^H(k)\}$ , is given by the linear transformation of the vector  $\mathbf{e}_w$

$$\mathbf{e}(k) = \sqrt{\sigma_e^2(k)}\mathbf{E}(k) \begin{bmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\lambda_L} \end{bmatrix} \mathbf{e}_w. \quad (7.4)$$

As will be seen, the simulations are done either with a source frequency of 1000 Hz (which, at a sampling rate of 48 kHz, falls in the frequency bin  $k = 85$ ) or at 996.1 Hz (which is the center of the frequency bin  $k = 85$  for a sampling rate of 48 kHz). Figure 7.2 shows the normalized cross-power spectral density as a function of continuous distance for this frequency bin, for the two array geometries. The dots show the normalized cross-power spectral density between actual pairs of microphones (values present in the matrix  $\mathbf{R}_{ee}(k)$ ).

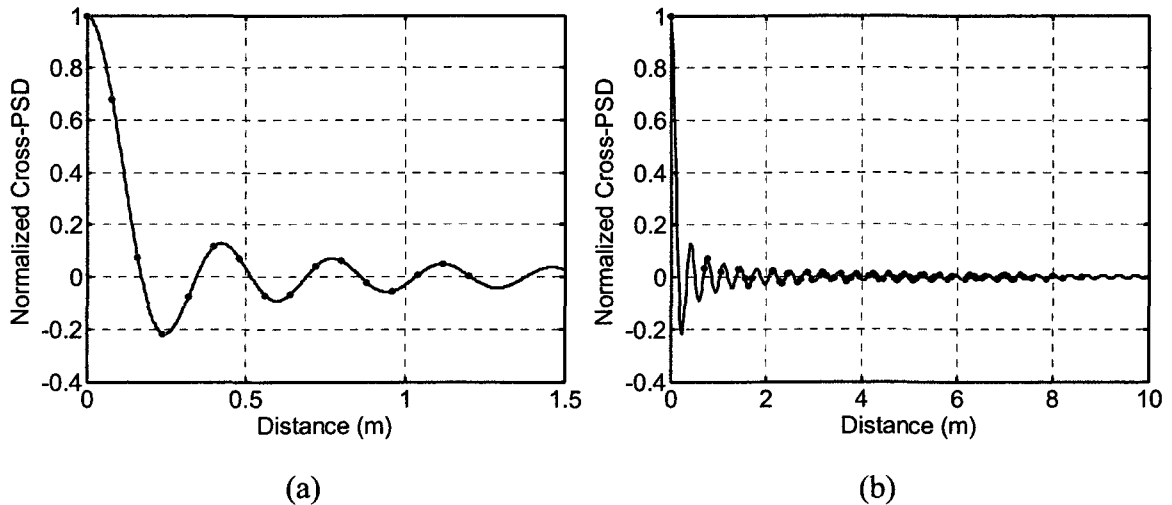


Figure 7.2. Normalized cross-power spectral density as a function of continuous distance for the frequency bin  $k = 85$ , at a sampling rate of 48 kHz. Dots show values present in  $\mathbf{R}_{ee}(k)$ . (a) Linear array; (b) Rectangular array.

We observe on Fig. 7.2(a), for the linear array, that there are microphone pairs where the cross-PSD is large (close to 0.7, for a distance of 8 cm). This is the cross-PSD between adjacent microphones of the linear array whose spacing between microphones is 8 cm. On Fig. 7.2(b), we observe that with the rectangular array, there are no microphone pairs where the cross-correlation is large. The maximum is 0.07. This is because the spacing between microphone pairs is much larger, the smallest distance being 71 cm. Values of  $\mathbf{R}_{ee}(k)$  outside of the main diagonal are closer to zero for the rectangular array than for the linear array. The matrix  $\mathbf{R}_{ee}(k)$  is closer to an identity matrix for the rectangular array. We will come back to this issue when looking at array gain versus white-noise gain in Sub-Section 7.3.1.

## 7.2.2 Simulation description

The microphone output vector is given by

$$\mathbf{p}(k) = \mathbf{s}(k) + \mathbf{e}(k). \quad (7.5)$$

The average signal power spectral density over the  $L$  microphones of the array is

$$\sigma_s^2(k) = \mathbf{s}^H(k)\mathbf{s}(k)/L \quad (7.6)$$

while the average noise power spectral density over the  $L$  microphones of the array is

$$\sigma_e^2(k) = E\{\mathbf{e}^H(k)\mathbf{e}(k)\}/L. \quad (7.7)$$

The normalized signal cross-power spectral density matrix is

$$\mathbf{R}_{ss}(k) = \mathbf{s}(k)\mathbf{s}^H(k)/\sigma_s^2(k) \quad (7.8)$$

and the normalized noise cross-power spectral density matrix is

$$\mathbf{R}_{ee}(k) = E\{\mathbf{e}(k)\mathbf{e}^H(k)\}/\sigma_e^2(k). \quad (7.9)$$

The optimization criterion for the simulation is the improvement in SNR with respect to diffuse noise, not white-noise. So the criterion is optimization of the array gain with respect to  $\mathbf{R}_{ee}(k)$ , not the optimization of the white-noise gain. We will compute the resulting white-noise gain because it is a measure of the sensitivity of the beamformer to random errors. The beamforming weight vector is based on Eq. (2.17) but with the source model replacing the unknown source signal.

For the gain-optimized (GO) estimator, the source mode strengths are estimated with Eq. (4.41)

$$d_{nm}(k) = \frac{1}{C_{nm}^{nm}(k)} \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{p}(k), \quad (7.10)$$

which is then used to get the source model with Eq. (4.51)

$$\mathbf{s}_{\text{mo, GO}}(k) = \sum_{n=0}^{N_{\text{order}}} \sum_{m=-n}^n d_{nm}(k) \mathbf{Y}_n^m(k), \quad (7.11)$$

leading to the overall beamforming weight vector

$$\mathbf{w}_{\text{GO}}(k) = \mathbf{R}_{ee}^{-1}(k) \sum_{n=0}^{N_{\text{order}}} \sum_{m=-n}^n d_{nm}(k) \mathbf{Y}_n^m(k). \quad (7.12)$$

Unless specified otherwise,  $N_{\text{order}}$  is set to 4.



For the least squares approach, the pseudoinverse of the matrix  $\mathbf{Y}(k)$  can be either computed directly or using singular value decomposition (SVD). If the matrix  $\mathbf{Y}(k)$  is of full rank, the pseudoinverse is theoretically the same for both methods, although singular value decomposition has better numerical properties [Therrien 1992]. We will compare results using both methods. The source model, as given by Eq. (6.18), is

$$\mathbf{s}_{\text{mo,LS}}(k) = \mathbf{Y}(k)\mathbf{Y}^+(k)\mathbf{p}(k) \quad (7.13)$$

and the beamforming weight vector is Eq. (6.24)

$$\mathbf{w}_{\text{LS}}(k) = \mathbf{R}_{\text{ee}}^{-1}(k)\mathbf{Y}(k)\mathbf{Y}^+(k)\mathbf{p}(k). \quad (7.14)$$

For an MVDR beamformer using the GO estimator for the source model (as mentioned in Section 6.4), the beamforming weight vector is

$$\mathbf{w}_{\text{MVDR,GO}}(k) = \frac{\mathbf{R}_{\text{pp}}^{-1}(k)\mathbf{s}_{\text{mo,GO}}(k)}{\mathbf{s}_{\text{mo,GO}}^H(k)\mathbf{R}_{\text{pp}}^{-1}(k)\mathbf{s}_{\text{mo,GO}}(k)}. \quad (7.15)$$

For an MVDR beamformer using the LS estimator for the source model, the beamforming weight vector is

$$\mathbf{w}_{\text{MVDR,LS}}(k) = \frac{\mathbf{R}_{\text{pp}}^{-1}(k)\mathbf{s}_{\text{mo,LS}}(k)}{\mathbf{s}_{\text{mo,LS}}^H(k)\mathbf{R}_{\text{pp}}^{-1}(k)\mathbf{s}_{\text{mo,LS}}(k)}. \quad (7.16)$$

The performance measures to evaluate the different approaches are the array gain and the white-noise gain. The array gain, Eq. (2.12), is

$$G(k) = \frac{\mathbf{w}^H(k)\mathbf{R}_{\text{ss}}(k)\mathbf{w}(k)}{\mathbf{w}^H(k)\mathbf{R}_{\text{ee}}(k)\mathbf{w}(k)} = \frac{|\mathbf{w}^H(k)\mathbf{s}(k)|^2 / \sigma_s^2(k)}{\|\mathbf{w}^H(k)[\mathbf{E}(k)\sqrt{\Lambda(k)}]\|^2}. \quad (7.17)$$

As in Eq. (7.4),  $\mathbf{E}(k)$  is a matrix whose columns are the eigenvectors of  $\mathbf{R}_{\text{ee}}(k)$ .  $\sqrt{\Lambda(k)}$  is a diagonal matrix whose elements are the square root of the eigenvalues of  $\mathbf{R}_{\text{ee}}(k)$ . The computation is done with the right-hand side of Eq. (7.17) to avoid the non-

insignificant imaginary part of  $\mathbf{w}^H(k)\mathbf{R}_{ee}(k)\mathbf{w}(k)$  when MVDR is used caused by the finite numerical precision. The white-noise gain is

$$G_w(k) = \frac{\mathbf{w}^H(k)\mathbf{R}_{ss}(k)\mathbf{w}(k)}{\mathbf{w}^H(k)\mathbf{w}(k)}. \quad (7.18)$$

Array gain and white-noise gain are evaluated for varying input SNR. For each input SNR, 10000 realizations are used. A random noise vector  $\mathbf{e}(k)$  is generated for each realization. Since the source signal vector  $\mathbf{s}(k)$  does not depend on the noise vector, it is only computed once, as is the average signal power spectral density over the array,  $\sigma_s^2(k)$ , and the normalized signal cross-power spectral density matrix,  $\mathbf{R}_{ss}(f)$ . The microphone output cross-power spectral density matrix is necessary to find the beamforming weight vector for MVDR. It is given by

$$\mathbf{R}_{pp}(k) = \sigma_s^2(k)\mathbf{R}_{ss}(k) + \sigma_e^2(k)\mathbf{R}_{ee}(k). \quad (7.19)$$

For each realization, a microphone output vector  $\mathbf{p}(k) = \mathbf{s}(k) + \mathbf{e}(k)$  is computed. For the different approaches, and for each realization, the beamforming weight vector is computed using  $\mathbf{p}(k)$ . The array gain and white-noise gain are then computed with Eqs. (7.17) and (7.18), respectively.

For comparison purposes, the array gain and white-noise gain for conventional beamforming (point source model) are also computed, using the beamforming weight vector

$$\mathbf{w}_{\text{conv}}(k) = \mathbf{R}_{ee}^{-1}(k)\mathbf{Y}_0^0(k). \quad (7.20)$$

The optimum array gain and resulting white-noise gain are also computed. In this case, the beamforming weight vector is

$$\mathbf{w}_{\text{opt}}(k) = \mathbf{R}_{ee}^{-1}(k)\mathbf{s}(k). \quad (7.21)$$

Of course, in practice, the exact source signal would not be available and it would not be possible to have this weight vector.

The beamforming weight vectors for conventional beamforming and for the optimum solution do not depend on the realization of the noise vector  $\mathbf{e}(k)$  and do not change with input SNR. They are computed only once as are their array gain and white-noise gain.

### 7.3 Simulation results: Diffuse noise

This section presents simulation results for a source in the presence of diffuse noise. The array and source scenarios are:

- Linear array and dipole source, as in the measurements of Chapter 5;
- Rectangular array and dipole source;
- Rectangular array and quadrupole source;
- Rectangular array and monopole source;
- Rectangular array and dipole source, estimation using a varying number of FFT frames.

#### 7.3.1 Linear array and dipole source as in the experimental measurements

In this section, the simulations use the same linear array of  $L = 16$  sensors as for the measurement setup of Chapter 5. The source position is also the same as for the measurements but instead of building a source resembling a dipole by putting two point sources close to each other and 180 degrees out of phase, the dipole is built by adding the spherical harmonic of order and degree  $(n, m) = (1, -1)$  to the spherical harmonic of order and degree  $(n, m) = (1, 1)$  multiplied by -1. This approach is used so that the exact source mode strengths  $s_{1,-1}(k)$  and  $s_{1,1}(k)$  are known and are respectively 1 and -1. The source frequency is 1000 Hz which, with a sampling rate of 48 kHz, falls in the frequency bin  $k = 85$ . The frequency at the center of this frequency bin is 996.1 Hz. The array as well as the source location and directivity are shown in Fig. 7.3. The source signal vector  $\mathbf{s}(k)$  is computed from the mode strengths, the array geometry, and array/source positions.

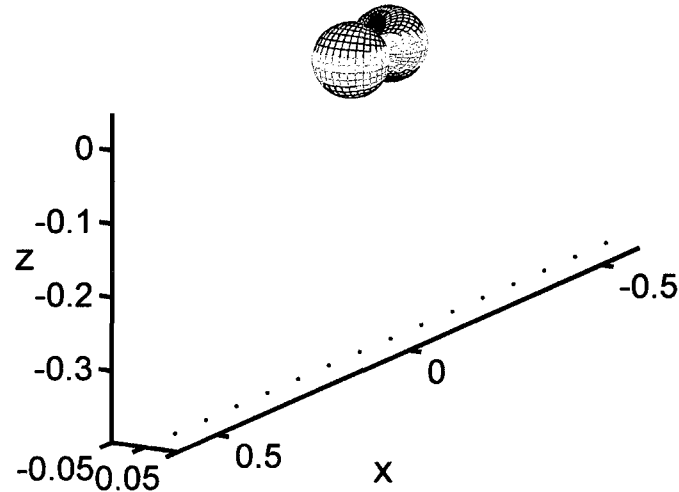


Figure 7.3. Linear array and dipole source.

To allow for comparison with the proposed beamforming method and the different mode strength estimators, the optimum array gain and the gain for conventional beamforming are given in Table 7.1. The point source is a poor model to beamform on a dipole source at this position and orientation with respect to the array. This explains why the array gain is so low for conventional beamforming. The optimum array gain is quite large but the white-noise gain is extremely low. This is an indication of the sensitivity of the array gain to random errors, as discussed in Sub-Section 2.4.2.

Table 7.1. Array gain and white-noise gain for the dipole source and linear array in the same configuration as in the measurements.

Source model	Array gain	White-noise gain
	(dB)	
Exact source signal (optimum)	45.77	-69.12
Point source (conventional beamforming)	-66.73	-158.10

The explanation for the low white-noise gain is simple. With noise uncorrelated from microphone to microphone (white-noise), the cross-power spectral density is zero for all microphone pairs. However, we saw on Fig. 7.2(a) that with diffuse noise and the linear array, there are microphone pairs where the cross-PSD is important. This means that optimizing for the array gain leads to a very different beamforming weight vector than optimizing for white-noise gain. In such a case, the beamforming weight vector optimized for the array gain provides a very low white-noise gain. We will come back to this issue at the end of the present sub-section, to verify if the array gain can be optimized subject to a constraint on the white-noise gain with the proposed beamforming method as suggested in Sub-Section 4.7.

We will now discuss the array gain for varying value of input SNR for the proposed beamforming method when the mode strengths are computed with the different estimators, or with exact mode strengths for comparison. Some of the figures show the mean, median, minimum, and maximum values of the array gain over 10000 realizations for varying values of input SNR while other figures show the histogram of the array gain for 10000 realizations for specific input SNR values.

### ***Gain-optimized estimator***

The spread of the array gain with the GO estimator and  $N_{\text{order}} = 4$  is shown in Fig. 7.4(a). For low input SNR, the spread is large and the resulting array gain for a realization depends on the noise vector  $\mathbf{e}(k)$  for that realization. For input SNR above -25 dB, the spread of the array gain is very small and the array gain is only approximately 0.01 dB below the optimum gain. The proposed beamforming method performs properly where conventional beamforming does not, with its array gain of -66.7 dB.

Figure 7.4(b) presents the case when the exact source mode strengths are used instead of an estimation (possible here because, in simulations, we know the source directivity). Since there is no estimation from the microphone outputs, the array gain does not depend on the input SNR or on the realization of the vector  $\mathbf{e}(k)$ . The optimum array gain is met. This demonstrates that the principle of using a weighted sum of sub-

beamformers can provide the optimum gain. However the issue is then how good is the estimation of the source mode strengths for an unknown source directivity.

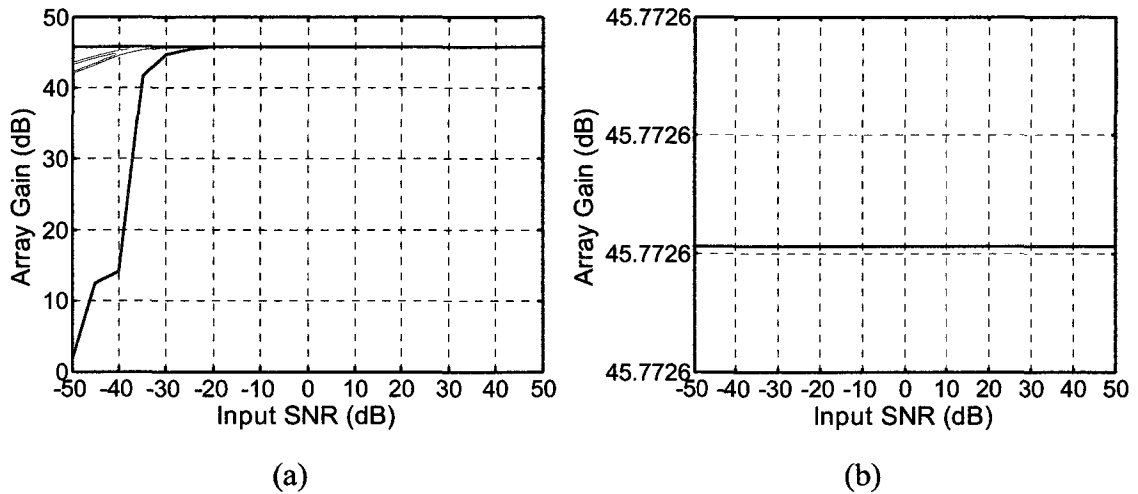


Figure 7.4. Linear array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations. (a) GO estimator; (b) Exact mode strengths.

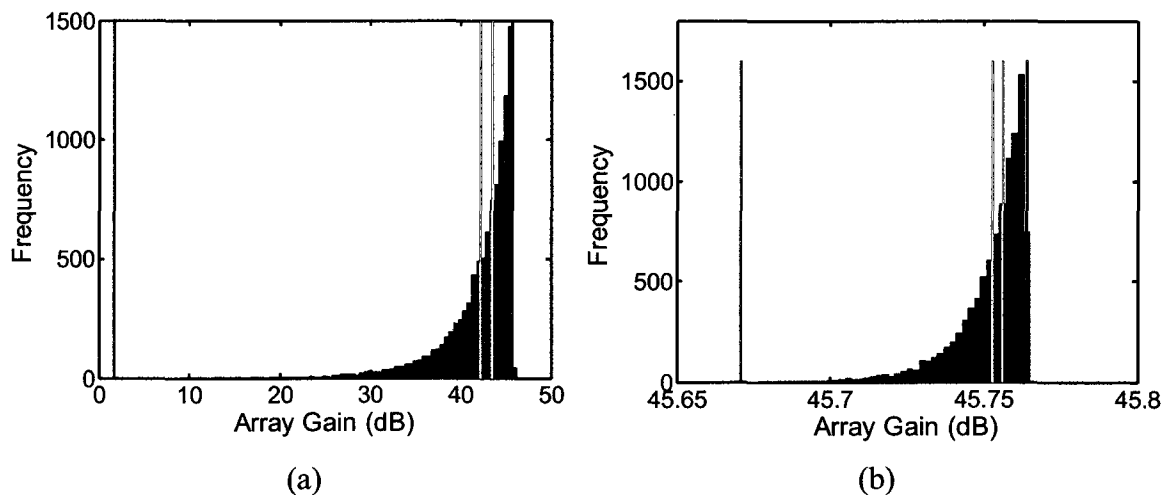


Figure 7.5. Linear array, dipole source, and diffuse noise. Histogram of array gain for 10000 realizations with the GO estimator for input SNR of (a) -50 dB, bar width: 0.5 dB; (b) -20 dB, bar width: 0.002 dB. Mean (green), median (red), minimum (blue), and maximum (blue) are also displayed.

Histograms of the array gain for 10000 realizations, with the GO estimator, for input SNR of -50 and -20 dB are shown in Fig. 7.5. We observe on Fig. 7.5(a) that, even with an input SNR of -50 dB, the array gain is above 30 dB for most of the realizations. However, there are cases where the array gain is almost as low as 0 dB. The spread of the array gain is small for an input SNR of -20 dB. Figure 7.5(b) shows the distribution of the array gain for the 10000 realizations.

The spread of the resulting white-noise gain for the GO estimator is displayed in Fig. 7.6. Even for low input SNR, the mean and median are very close to the maximum value. They can hardly be seen on the figure. For all input SNR, the white-noise gain is very low and comparable to the value of -69 dB for the optimum solution given in Table 7.1.

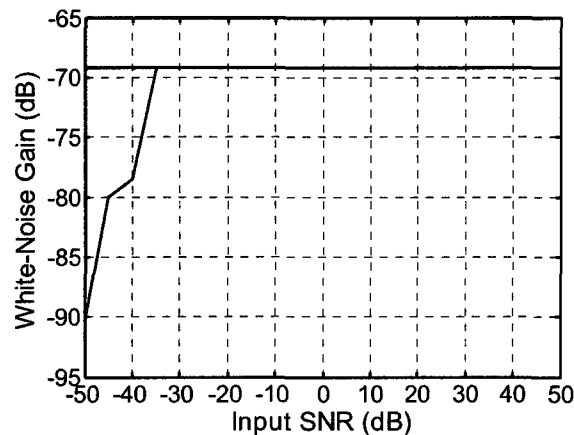


Figure 7.6. Linear array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the white-noise gain for 10000 realizations with the GO estimator.

### ***Least-squares estimators***

Array gain results with LS estimation when including spherical harmonic modes up to order  $n = 3$  (a total of 16 modes) into the matrix  $\mathbf{Y}(k)$  are shown in Fig. 7.7.  $\mathbf{Y}(k)$  is then a  $16 \times 16$  matrix. Results for when the pseudoinverse is computed with singular value decomposition and when it is computed directly are shown in Fig. 7.7(a) and Fig. 7.7(b), respectively.

We first note that the minimum array gain with varying input SNR is not the same for both methods of computing the pseudoinverse.  $\mathbf{Y}(k)$  is not full rank, its rank being 7. Matlab does provide the warning “Matrix is close to singular or badly scaled” when computing the inverse of  $\mathbf{Y}^H(k)\mathbf{Y}(k)$ , necessary to computing directly the pseudoinverse of  $\mathbf{Y}(k)$ . The issue of the rank of  $\mathbf{Y}(k)$  was discussed in Sub-Section 6.3.4. The rank of the matrix  $\mathbf{Y}(k)$  depends on the array geometry and the position of the source. In a beamforming situation as with the linear array where the array has a limited view of the source, spherical harmonic modes look similar on the array and  $\mathbf{Y}(k)$  is rank deficient.

In both cases, the spread of the array gain remains large below input SNR of approximately 35 dB. At 10 dB input SNR, the mean array gain is still 1.7 dB below the optimum array gain. This is much worse than with the GO estimator and we attribute this behavior with the LS estimator to the matrix  $\mathbf{Y}(k)$  being rank deficient.

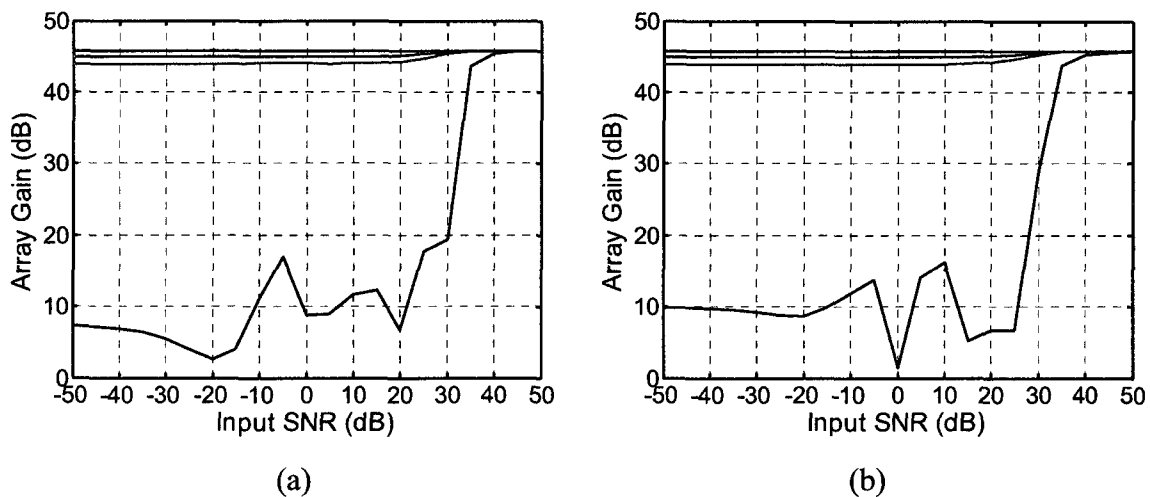


Figure 7.7. Linear array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations. (a) LS (SVD) estimator; (b) LS (Direct) estimator.



Histograms of the array gain for 10000 realizations, with the LS (SVD) estimator for input SNR of -50 and -20 dB are shown in Fig. 7.8. We observe on Fig. 7.8(a) that, even with an input SNR of -50 dB, the array gain is above 40 dB for most of the realizations. The main part of the distribution may be tighter than it was on Fig. 7.5(a) for the GO estimator. Once again however, there are cases where the array gain is almost as low as 0 dB. With an input SNR of -20 dB, we see on Fig. 7.8 (b) that the behavior is similar to what it is on Fig. 7.8(a), which is for an input SNR of -50 dB. With the GO estimator and an input SNR of -20 dB, the array gain stayed within 0.1 dB of the optimum gain for the 10000 realizations, as seen on Fig. 7.5(b).

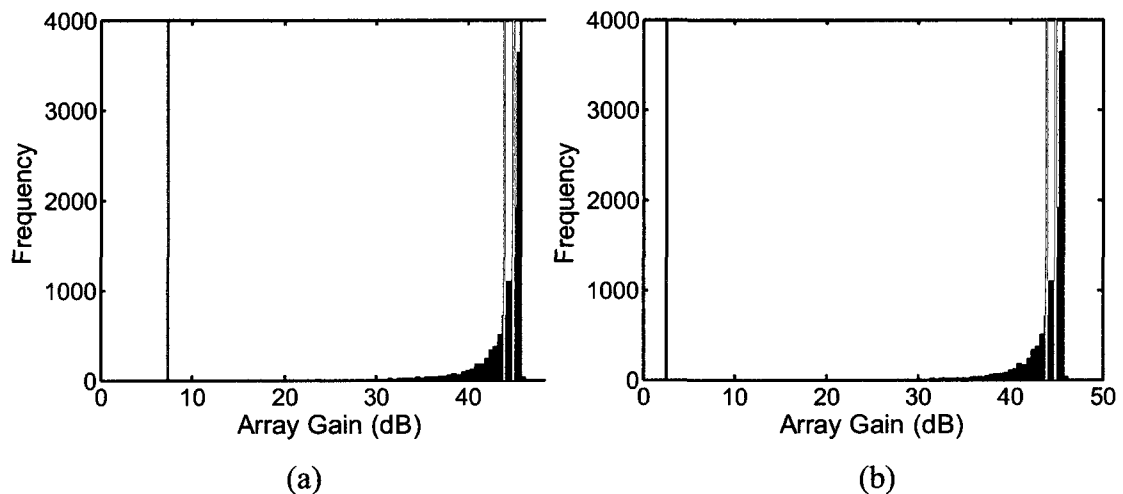


Figure 7.8. Linear array, dipole source, and diffuse noise. Histogram of array gain for 10000 realizations with the LS (SVD) estimator for input SNR of (a) -50 dB, bar width: 0.5 dB; (b) -20 dB, bar width: 0.5 dB.

The spread of the resulting white-noise gain with the LS (SVD) estimator is shown in Fig. 7.9. The spread remains large up to approximately 30 dB input SNR, as does the array gain in Fig. 7.7(a). Generally, the spread of the resulting white-noise gain resembles that of the array gain, albeit at a different decibel level.

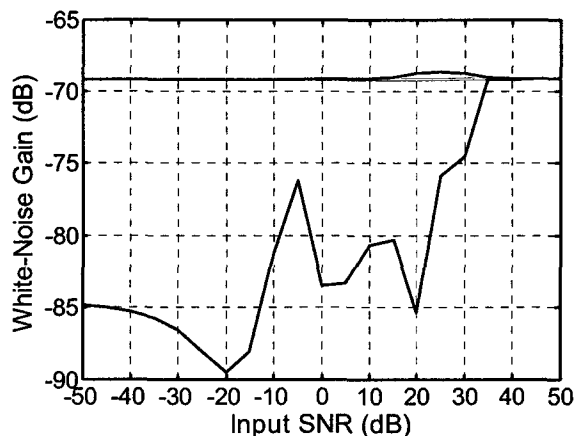


Figure 7.9. Linear array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the white-noise gain for 10000 realizations with the LS (SVD) estimator.

### ***Microphone output vector as the source model***

Two of the estimation approaches presented in Appendix C lead to the microphone output vector as the source model. The resulting spread of the array gain is shown in Fig. 7.10. This is somewhat similar to the spread when using the GO estimator. However the spread is wider for input SNR of -50 dB and the mean is lower. We will see later that in the presence of reverberation or of an interferer, the microphone output as the source model performs much worse than the GO estimator.

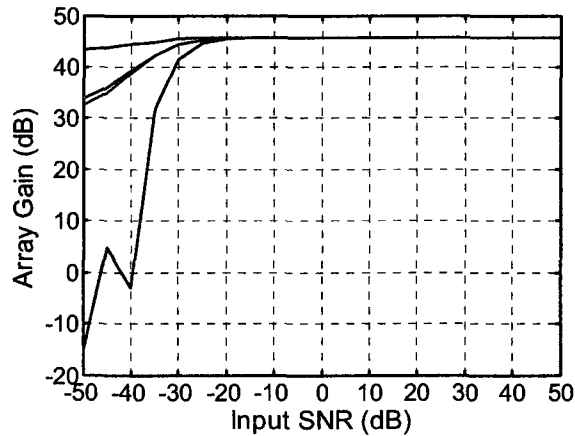


Figure 7.10. Linear array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations, microphone output vector as source model.

### ***Estimated source model and MVDR beamforming***

The array gain spread when the GO estimator is used to obtain a source model which is then used with an MVDR beamformer is shown in Fig. 7.11(a). For comparison, the result when the exact source model is used with MVDR is shown in Fig. 7.11(b). For input SNR above 35 dB, the rank of  $\mathbf{R}_{pp}(k)$  is 15 instead of 16. Since the inverse of  $\mathbf{R}_{pp}(k)$  is necessary to compute the beamforming weight vector, we have thus chosen an upper limit of input SNR of 35 dB for the figures.

On Fig. 7.11(b), we observe that MVDR with the source signal as the source model essentially provides the optimum array gain for all input SNR. On Fig. 7.11(a) for the GO estimator and MVDR beamforming, we note a spread of the array gain for low input SNR. However for input SNR above -20 dB, there is a linear decrease of the array gain with increasing input SNR. Our hypothesis is that if the estimated source model is even only slightly off from the exact source signal, MVDR sees the exact source signal as an interfering signal and tries to minimize it, along with the noise, thus leading to a poor array gain.

For low input SNR, the estimated source model changes with the noise realization (sometimes the estimation is better than others, this explained the spread of

the array gain in Fig. 7.4(a) when the GO estimator was used without MVDR beamforming). The noise is the dominant term in the microphone output vector. Using the estimated source model, MVDR beamforming decreases the noise at the output of the array but it also decreases the signal component to some extent, leading to a spread of the array gain similar to when the GO estimator is used without MVDR beamforming.

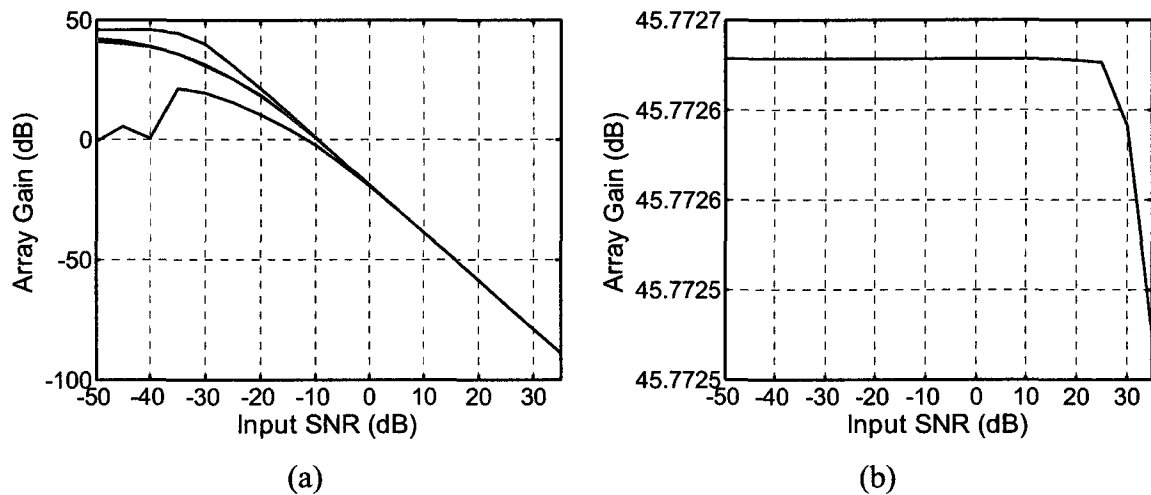


Figure 7.11. Linear array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations. (a) GO estimator with MVDR; (b) Exact source model and MVDR.

At higher input SNR, the estimation of the source model is better (the proof is that the GO estimator without MVDR provided an array gain very close to the optimum value, on Fig. 7.4(a)). However, since it is still different from the exact source signal, MVDR treats the exact source signal as an interfering signal and tries to decrease the response of the array to it. As the input SNR increases, this “interfering” signal becomes more and more dominant over the noise. So for MVDR to minimize the output power subject to the source model, it must decrease ever more the power at the output of the array for the source signal.

To verify this explanation, we ran a simulation where the source model is the sum of the source signal and a complex random vector. We compute the array gain as a function of the input SNR for different values of the magnitude of the random vector. The

result is displayed in Fig 7.12. The linear decrease of the array gain for an increase in input SNR behaves in a similar way as on Fig. 7.11(a).

This behavior is actually not limited to the use of the proposed beamforming method with estimation of the mode strengths. The topic of mismatched MVDR is discussed in Section 6.6 of [Van Trees 2002]. He discusses, among other things, the case of beamforming on a plane wave whose direction of arrival is different than the direction used for the source model. He also presents the case where there are errors on the position of the sensors. Figure 6.36 of [Van Trees 2002] displays the same behavior as our Fig 7.12.

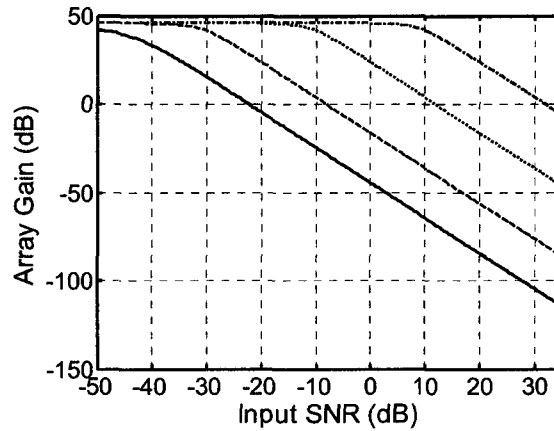


Figure 7.12. Linear array, dipole source, and diffuse noise. Array gain with MVDR and exact source signal as source model to which we add a random vector scaled by a factor (—)  $10^{-2}$ , (---)  $10^{-4}$ , (- · -)  $10^{-6}$ , (· · ·)  $10^{-8}$ .

Below is an explanation for the slope of -20 dB of array gain per 10 dB of input SNR seen on the two figures. The array gain is given by

$$G(k) = \frac{\mathbf{w}^H(k) \mathbf{R}_{ss}(k) \mathbf{w}(k)}{\mathbf{w}^H(k) \mathbf{R}_{ee}(k) \mathbf{w}(k)} = \frac{1}{\sigma_s^2(k)} \frac{|\mathbf{w}^H(k) \mathbf{s}(k)|^2}{\mathbf{w}^H(k) \mathbf{R}_{ee}(k) \mathbf{w}(k)}. \quad (7.22)$$

The MVDR beamforming weight is computed with

$$\mathbf{w}_{\text{MVDR}}(k) = \frac{\mathbf{R}_{pp}^{-1}(k) \mathbf{s}_{\text{mo}}(k)}{\mathbf{s}_{\text{mo}}^H(k) \mathbf{R}_{pp}^{-1}(k) \mathbf{s}_{\text{mo}}(k)} = \frac{\mathbf{R}_{pp}^{-1}(k) [\mathbf{s}(k) + \mathbf{s}_e(k)]}{[\mathbf{s}(k) + \mathbf{s}_e(k)]^H \mathbf{R}_{pp}^{-1}(k) [\mathbf{s}(k) + \mathbf{s}_e(k)]}, \quad (7.23)$$

where  $\mathbf{s}_e(k)$  is the difference between the source model and the source signal. The denominator of  $\mathbf{w}_{\text{MVDR}}(k)$  is a real scalar. We see that when substituting for  $\mathbf{w}_{\text{MVDR}}(k)$  in the array gain equation, the real scalar will cancel out from the numerator and denominator of the array gain equation, thus the important part of  $\mathbf{w}_{\text{MVDR}}(k)$  is its numerator, which we now examine.

The microphone output cross-power spectral density matrix  $\mathbf{R}_{\text{pp}}(k)$  is

$$\begin{aligned}\mathbf{R}_{\text{pp}}(k) &= \sigma_s^2(k)\mathbf{R}_{\text{ss}}(k) + \sigma_e^2(k)\mathbf{R}_{\text{ee}}(k) \\ &= \mathbf{s}(k)\mathbf{s}^H(k) + \sigma_e^2(k)\mathbf{R}_{\text{ee}}(k) \quad .\end{aligned}\tag{7.24}$$

To compute the inverse of  $\mathbf{R}_{\text{pp}}(k)$ , we use the matrix inversion lemma [Van Trees 2002] [Åström 1995]

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1},\tag{7.25}$$

where  $\mathbf{A}$ ,  $\mathbf{C}$ , and  $(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})$  are non singular square matrices. With  $\mathbf{A} = \sigma_e^2(k)\mathbf{R}_{\text{ee}}(k)$ ,  $\mathbf{B} = \mathbf{s}(k)$ ,  $\mathbf{D} = \mathbf{s}^H(k)$ , and  $\mathbf{C} = 1$ , we find the inverse of  $\mathbf{R}_{\text{pp}}(k)$  to be

$$\mathbf{R}_{\text{pp}}^{-1}(k) = \frac{1}{\sigma_e^2(k)}\mathbf{R}_{\text{ee}}^{-1}(k) - \frac{\frac{1}{(\sigma_e^2(k))^2}\mathbf{R}_{\text{ee}}^{-1}(k)\mathbf{s}(k)\mathbf{s}^H(k)\mathbf{R}_{\text{ee}}^{-1}(k)}{1 + \frac{1}{\sigma_e^2(k)}\mathbf{s}^H(k)\mathbf{R}_{\text{ee}}^{-1}(k)\mathbf{s}(k)}.\tag{7.26}$$

We now drop the frequency bin index to make the expressions shorter. It should be understood that the computation is done for each frequency bin. The numerator of the beamforming weight vector is

$$\begin{aligned}\text{NUM}(\mathbf{w}_{\text{MVDR}}) &= \mathbf{R}_{\text{pp}}^{-1}[\mathbf{s} + \mathbf{s}_e] \\ &= \frac{\frac{1}{\sigma_e^2} \left[ \left\{ 1 - \frac{\mathbf{s}^H \mathbf{R}_{\text{ee}}^{-1} \mathbf{s}_e}{\sigma_e^2} \right\} \mathbf{R}_{\text{ee}}^{-1} \mathbf{s} + \left\{ 1 + \frac{\mathbf{s}^H \mathbf{R}_{\text{ee}}^{-1} \mathbf{s}}{\sigma_e^2} \right\} \mathbf{R}_{\text{ee}}^{-1} \mathbf{s}_e \right]}{1 + \frac{1}{\sigma_e^2} \mathbf{s}^H \mathbf{R}_{\text{ee}}^{-1} \mathbf{s}}\end{aligned}\tag{7.27}$$

whose denominator is also a real scalar which will cancel out when substituted for in the array gain equation. We keep only the numerator

$$\text{NUM}\{\text{NUM}(\mathbf{w}_{\text{MVDR}})\} = \frac{1}{\sigma_e^2} \left[ \left\{ 1 - \frac{\mathbf{s}^H \mathbf{R}_{ee}^{-1} \mathbf{s}_e}{\sigma_e^2} \right\} \mathbf{R}_{ee}^{-1} \mathbf{s} + \left\{ 1 + \frac{\mathbf{s}^H \mathbf{R}_{ee}^{-1} \mathbf{s}}{\sigma_e^2} \right\} \mathbf{R}_{ee}^{-1} \mathbf{s}_e \right]. \quad (7.28)$$

The array gain is

$$G = \frac{\frac{1}{\sigma_s^2} \left| \text{NUM}\{\text{NUM}(\mathbf{w}_{\text{MVDR}})\}^H \mathbf{s} \right|^2}{\text{NUM}\{\text{NUM}(\mathbf{w}_{\text{MVDR}})\}^H \mathbf{R}_{ee} \text{NUM}\{\text{NUM}(\mathbf{w}_{\text{MVDR}})\}}. \quad (7.29)$$

The numerator of this expression is equal to

$$\text{NUM}(G) = \frac{1}{(\sigma_e^2)^2} \frac{1}{\sigma_s^2} \left| \mathbf{s}^H \mathbf{R}_{ee}^{-1} \mathbf{s} + \mathbf{s}_e^H \mathbf{R}_{ee}^{-1} \mathbf{s} \right|^2. \quad (7.30)$$

For a fixed source model error  $\mathbf{s}_e(k)$ , the magnitude-square term is fixed. We do not write the complete expression of the denominator because of its length, we only write its form, which is

$$\text{DENOM}(G) = \frac{1}{(\sigma_e^2)^2} \left[ \text{terms in } \frac{1}{(\sigma_e^2)^0} + \text{terms in } \frac{1}{\sigma_e^2} + \text{terms in } \frac{1}{(\sigma_e^2)^2} \right]. \quad (7.31)$$

Again, for a fixed source model error, the terms multiplying the different powers of  $1/\sigma_e^2$  in the square bracket are fixed so that the expression for the array gain has the form

$$G = \frac{\frac{1}{\sigma_s^2} \left| \mathbf{s}^H \mathbf{R}_{ee}^{-1} \mathbf{s} + \mathbf{s}_e^H \mathbf{R}_{ee}^{-1} \mathbf{s} \right|^2}{\text{terms in } \frac{1}{(\sigma_e^2)^0} + \text{terms in } \frac{1}{\sigma_e^2} + \text{terms in } \frac{1}{(\sigma_e^2)^2}}. \quad (7.32)$$

For small values of the noise power spectral density over the array,  $\sigma_e^2$ , the terms in  $1/(\sigma_e^2)^2$  are dominant so that a decrease in  $\sigma_e^2$  by a factor 10 leads to a decrease in the

array gain by a factor 100. This explains the linear decrease of 20 dB of array gain per 10 dB of increase in input SNR, at high input SNR.

Because MVDR with an estimated source model does not perform better than the GO or the LS estimators at low input SNR, and there is the linear decrease in the array gain with increasing input SNR, we will not include results for MVDR in the rest of this section. We will come back to MVDR in the following sections where reverberation or interferers are included in the simulations.

### ***Gain-optimized estimator and different values of $N_{\text{order}}$***

The spread of the array gain for varying input SNR with the GO estimator and  $N_{\text{order}}$  equal to 1, 2, 3, and 4 is shown in Fig. 7.13. It is not necessary to show the case with  $N_{\text{order}} = 0$  since it is conventional beamforming and the resulting array gain was given in Table 7.1 as -66.73 dB.

We note from the figure that the general behavior is the same for the 4 values of  $N_{\text{order}}$ . For an input SNR of 0 dB, the average array gain is approximately 45.772, 45.770, 45.767, and 45.763 dB respectively for  $N_{\text{order}}$  going from 1 to 4. The source being a dipole,  $N_{\text{order}} = 1$  would be sufficient to obtain the optimum array gain. The gain decreases very slightly as  $N_{\text{order}}$  increases, which we can attribute to the bias of the estimator. The remainder of the simulations will be performed with  $N_{\text{order}} = 4$ .



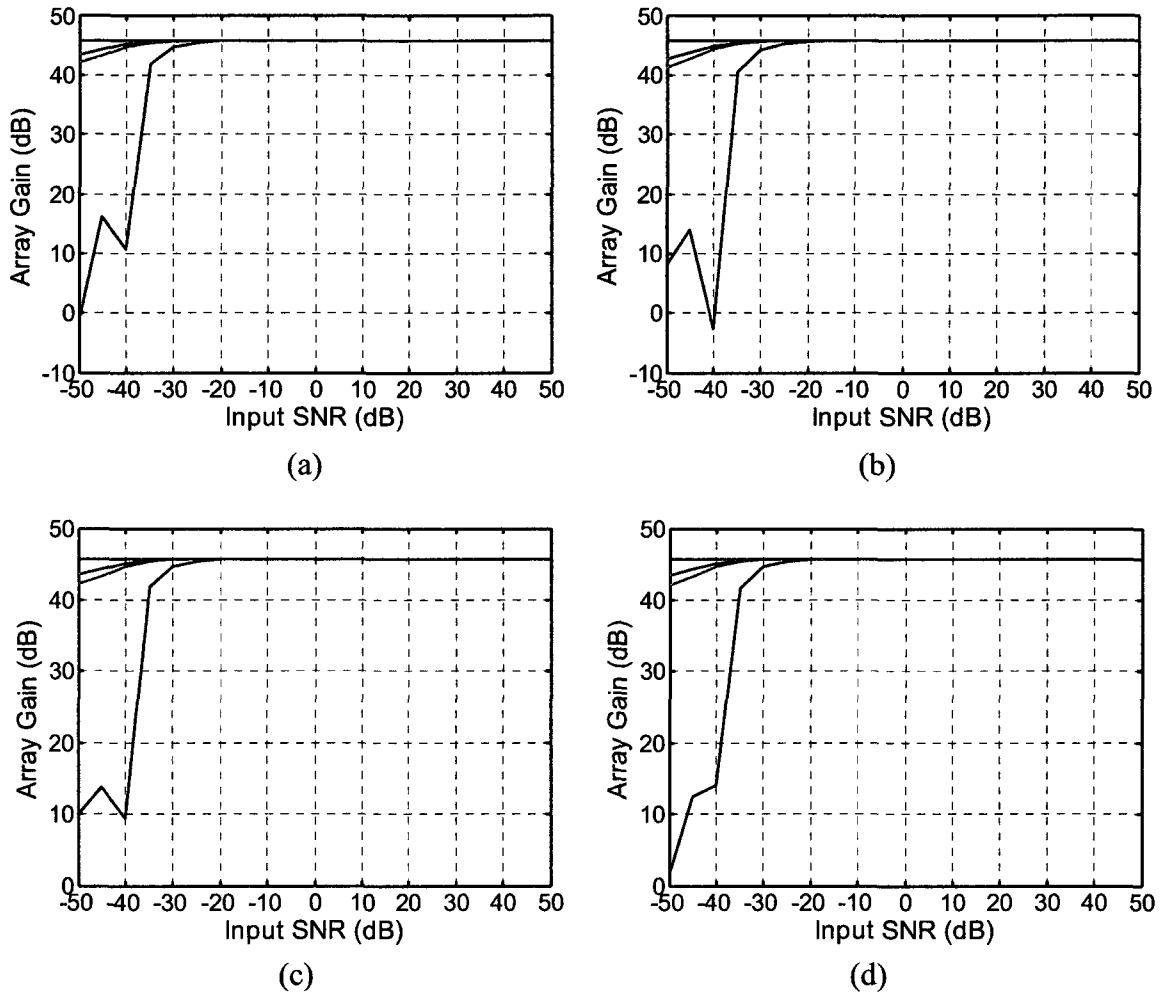


Figure 7.13. Linear array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations. GO estimator with (a)  $N_{\text{order}} = 1$ ; (b)  $N_{\text{order}} = 2$ ; (c)  $N_{\text{order}} = 3$ ; (d)  $N_{\text{order}} = 4$ .

### ***Gain optimization with constraint on white-noise gain, using the gain-optimized estimator***

As mentioned in Sub-Section 2.4.2, a low white-noise gain is an indication of the sensitivity of the array gain to random errors. The solution often used it to maximize the array subject to a constraint on the white-noise gain. The array gain will not be as high as it would be without constraint but it will be much less sensitive. Sub-Section 4.7 indicated that because the source model is unknown, it is not possible to compute the

exact white-noise gain while going through the iterative process until the constraint is met. This section investigates how accurately the white-noise gain constraint is met while using the estimated source model with the GO estimator to compute an approximate white-noise gain.

We set the constraint on the white-noise gain at 0 dB (as done on the fourth page of [Cox 1987]). We use the equations provided in Sub-Section 4.7 for the estimation and for the beamforming weight vector. To keep things simple, we begin with an  $\varepsilon$  value of 0 and increase  $\varepsilon$  by 0.001 at each iteration. We stop the iterative process when the white-noise gain is bigger than 0 dB. This way, even with the exact source model, the resulting white-noise gain will not be exactly equal to the constraint. Nonetheless, this is precise enough for our purpose of evaluating whether or not the proposed beamforming method could be used while maximizing the array gain subject to a constraint on the white-noise gain.

To allow for comparison with the proposed beamforming method, Table 7.2 contains the array gain and white-noise gain when the source signal is used as the source model, for two different optimization criteria: maximization of the array gain (as before) and maximization of the array gain subject to a constraint on the white-noise gain. With the constraint, the array gain is 12 dB instead of 45.77 dB without constraint. However, the resulting white-noise gain is much better than the -69 dB without constraint. We note that the resulting white-noise gain is slightly higher than the constraint of 0 dB because of our simple manner of finding  $\varepsilon$ .

Table 7.2 Array gain and white-noise gain for the dipole source and linear array in the same configuration as in the measurement, exact source model, without and with constraint of 0 dB on white-noise gain.

Optimization	Array gain	White-noise gain
	(dB)	
Maximize array gain	45.77	-69.12
Max. array gain, constraint 0 dB on W.-N.	12.03	0.19

Figure 7.14 shows the spread of the array gain with the proposed beamforming method and the GO estimator, without and with a constraint of 0 dB on the white-noise gain. Without constraint, Figure 7.14(a) is a copy of Fig. 7.4(a). With the constraint, we see on Fig. 7.14(b) that, again, there is a spread of the array gain for low input SNR. This time, the spread is for input SNR below 0 dB. At higher input SNR, the spread is small and the array gain is close to 12 dB, which is the value when the source signal is used as the source model, as shown in Table 7.2.

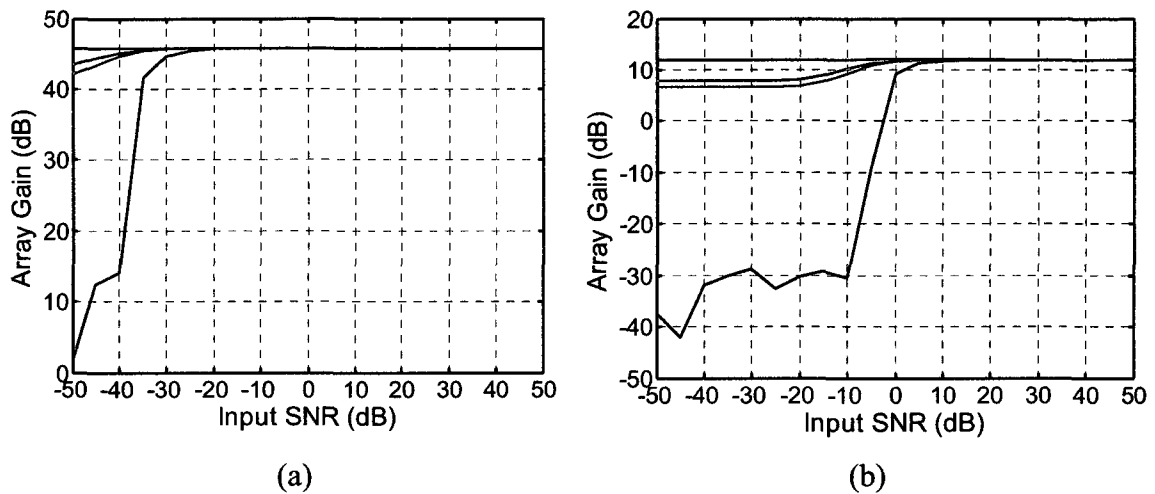


Figure 7.14. Linear array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations. (a) GO estimator (b) GO estimator and gain optimization with constraint on white-noise gain of 0 dB.

The spread of the white-noise gain with the proposed beamforming method and the GO estimator, without and with a constraint of 0 dB on the white-noise gain, is displayed in Fig. 7.15. Without constraint, Figure 7.15(a) is a copy of Fig. 7.6. With the constraint, we see on Fig. 7.15(b) that, again, there is a spread of the white-noise gain for input SNR below 0 dB. At higher input SNR, the white-noise gain is slightly above the constraint of 0 dB, still because of our simple way of finding  $\epsilon$ .

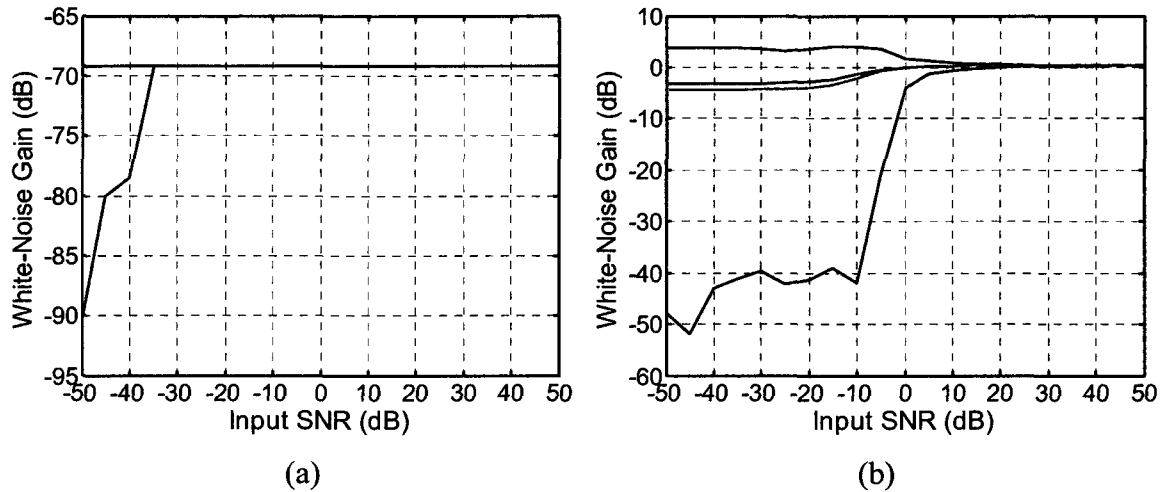


Figure 7.15. Linear array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the white-noise gain for 10000 realizations. (a) GO estimator (b) GO estimator and gain optimization with constraint on white-noise gain of 0 dB.

To go into more details than the figures showing the spread can provide, Figure 7.16 presents the distribution of the array gain for the 10000 realization with the GO estimator and gain optimization with a 0 dB constraint on the white-noise gain, for input SNR of -50, -20, 0, and 30 dB. Figure 7.17 shows the distribution for the white-noise gain for the same input SNR values.

We see on Fig. 7.16(a) and 7.16(b) that, even at low input SNR, a good portion of the realizations provides an array gain larger than 0 dB. However, the tails of the distributions do extend to very low array gains. At higher input SNR, the distribution becomes narrower (e.g. Fig. 7.16(c)). For 30 dB input SNR, on Fig. 7.16(d), the array gain values for the 10000 realizations are all within the same bar of width 0.02 dB.

The same general behavior with input SNR is seen for the white-noise gain on Fig. 7.17.

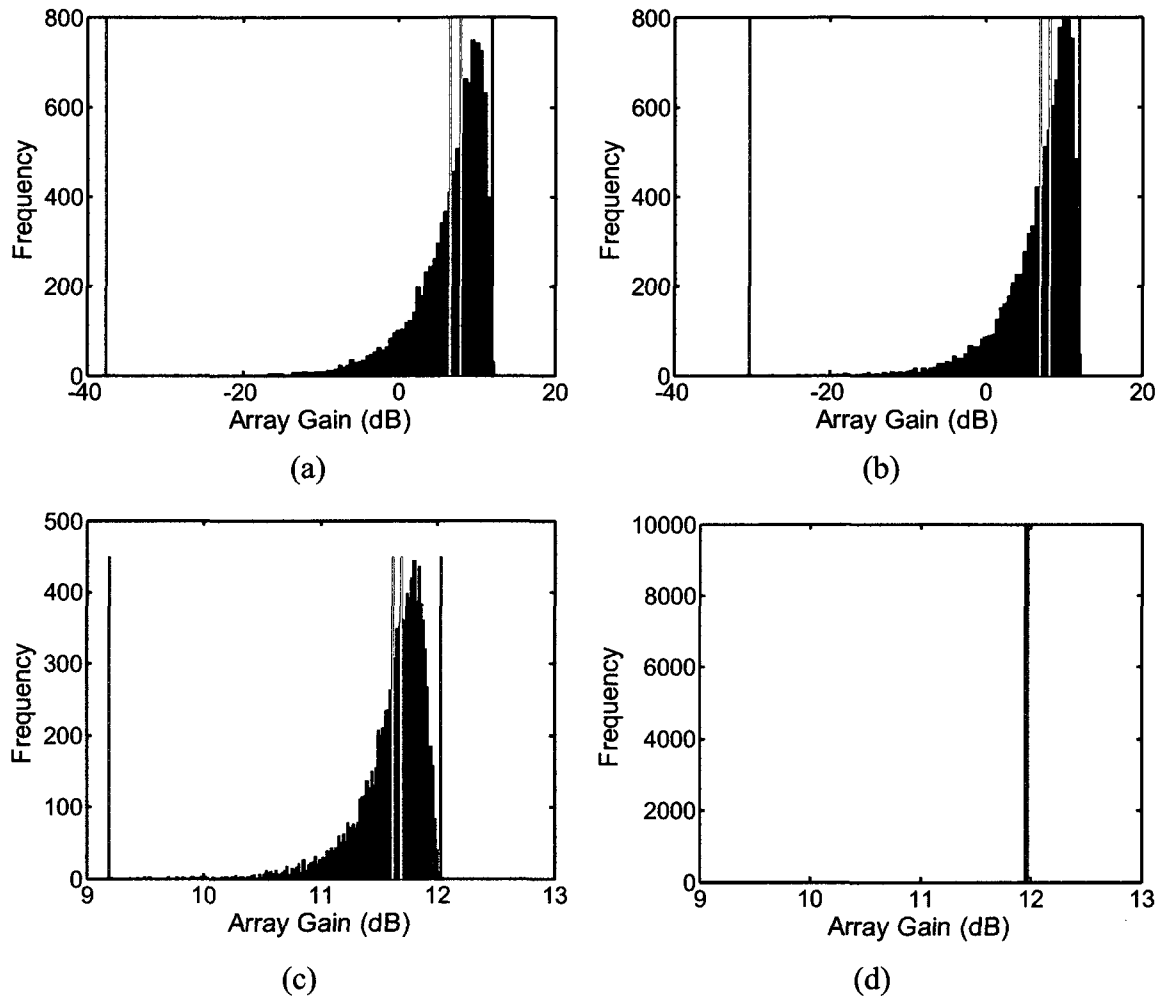


Figure 7.16. Linear array, dipole source, and diffuse noise. Histogram of array gain for the 10000 realizations with the GO estimator and gain optimization with a 0 dB constraint on the white-noise gain, for input SNR of (a) -50 dB, bar width: 0.5 dB; (b) -20 dB, bar width: 0.5 dB; (c) 0 dB, bar width: 0.02 dB; (d) 30 dB, bar width: 0.02 dB.

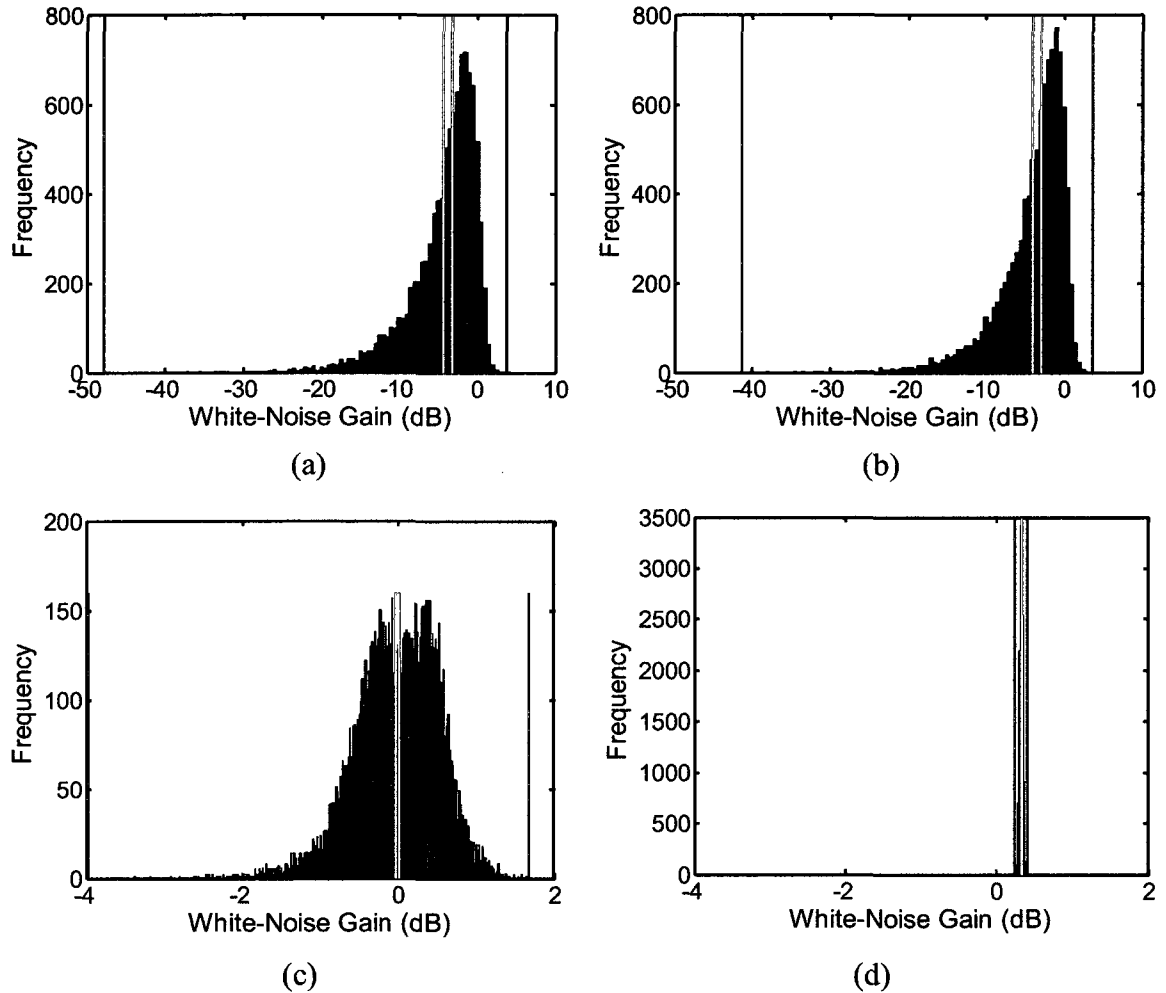


Figure 7.17. Linear array, dipole source, and diffuse noise. Histogram of white-noise gain for the 10000 realizations with the GO estimator and gain optimization with a 0 dB constraint on the white-noise gain, for input SNR of (a) -50 dB, bar width: 0.5 dB; (b) -20 dB, bar width: 0.5 dB; (c) 0 dB, bar width: 0.02 dB; (d) 30 dB, bar width: 0.02 dB.

Figure 7.18 shows the white-noise gain versus the array gain for the same four input SNR values. A dot marks the array gain and white-noise gain for each of the 10000 realizations. For input SNR of -50 and -20 dB, we see on Fig. 7.18(a) and Fig. 7.18(b) a large spread as well as the linear increase of white-noise gain with array gain. This suggests that when the source model is closer to the source signal, the iterative process is stopped to a white-noise gain value closer to the constraint and, because the source model is better, the array gain is also larger.

For an input SNR of 0 dB, Figure 7.18(c) demonstrates the decrease in spread of both the array gain and the white-noise gain. With an input SNR of 30 dB, we observe on Fig. 7.18(d) the much smaller spread in the array gain. The white-noise gain is not exactly at the constraint value of 0 dB and also presents a spread. We notice an increase in the array gain for a decrease in white-noise gain. This is caused by the trade off between the array gain and the white-noise gain in the optimization.

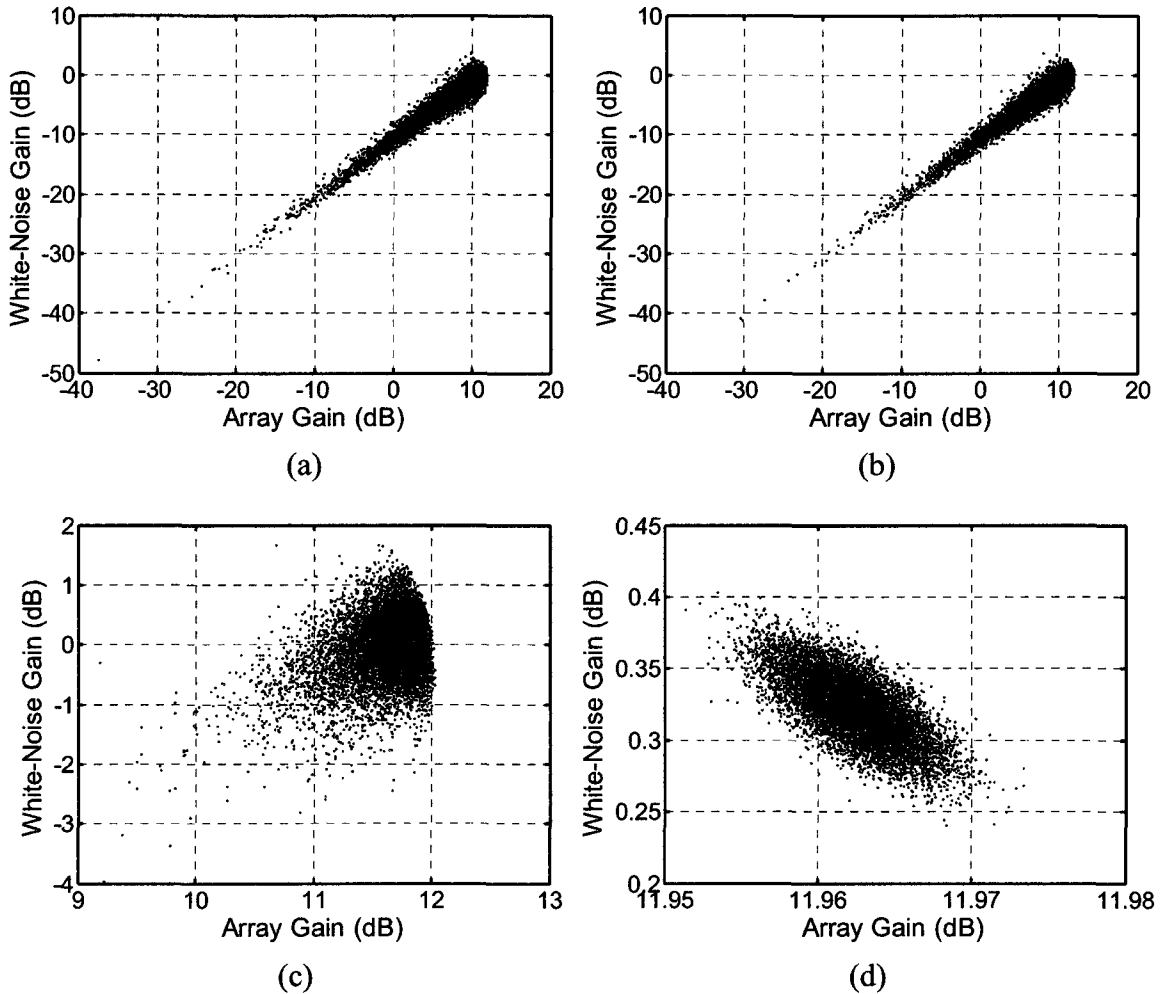


Figure 7.18. Linear array, dipole source, and diffuse noise. White-noise gain vs. array gain for the 10000 realizations with the GO estimator and gain optimization with a 0 dB constraint on white-noise gain, for input SNR of (a) -50 dB; (b) -20 dB; (c) 0 dB; (d) 30 dB.

We also performed simulations with a constraint on the white-noise gain of 5 dB and observed the same behavior as we have just discussed for a constraint of 0 dB.

For a source in the presence of diffuse noise, it thus appears that the proposed beamforming method could be used while optimizing the array gain subject to a constraint on the white-noise gain, in particular for higher input SNR.

Of the two array geometries considered in the simulations, the cross-power spectral density between microphone pairs is significant only for the linear array. Because of space consideration, we will not include the linear array in the simulations with reverberation or with an interferer. We will thus not investigate optimization of the array gain subject to a constraint on the white-noise gain for reverberation or interferers. Since the presence of reverberation or interferers has for effect to deteriorate the source model estimation, it would deteriorate the accuracy of the estimation of the white-noise gain necessary in the iterative process.

### 7.3.2 Rectangular array and dipole source

For this simulation, the array is composed of 32 microphones located on the perimeter of a rectangle. The source is a dipole located at  $(x, y, z) = (2, 1, 0)$  with its axis parallel to the  $x$ -axis. We purposely avoided  $(x, y)$  coordinates equal to  $(0, 0)$  so that the response of conventional beamforming would not be excessively small (theoretically zero). The exact source mode strengths are still  $s_{1,-1}(k) = 1$  and  $s_{1,1}(k) = -1$ . The array as well as the source location and directivity are displayed in Fig. 7.19.

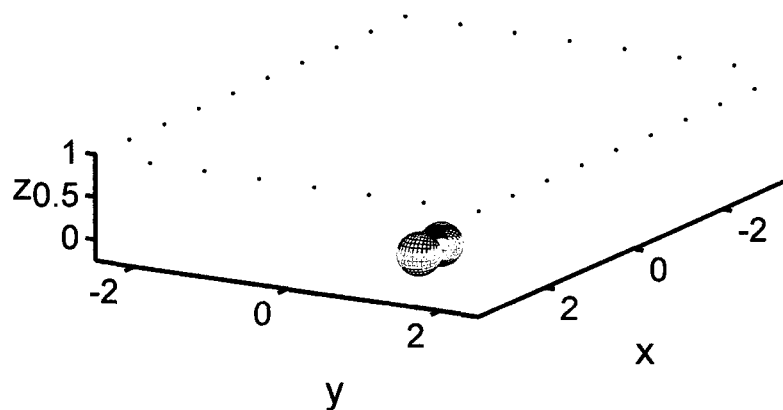


Figure 7.19. Rectangular array and dipole source.



The optimum array gain and the gain for conventional beamforming are given in Table 7.3. As we noted earlier concerning the rectangular array, the elements of the matrix  $\mathbf{R}_{ee}(k)$  outside of the main diagonal are closer to zero than for the linear array. For the rectangular array, the beamforming weight vector found when optimizing the array gain subject to diffuse noise does not lead to a poor white-noise gain as it did for the linear array. The resulting white-noise gain when the exact source signal is used as the source model is 15 dB.

Table 7.3. Array gain and white-noise gain for the dipole source and the rectangular array

Source model	Array gain	White-noise gain
	(dB)	
Exact source signal (optimum)	15.06	15.01
Point source (conventional beamforming)	-0.89	-1.00

Conventional beamforming still provides a poor array gain with the rectangular array and the dipole source at this location. The simulation results show that the proposed beamforming method with, for example, the GO estimator, provides an array gain close to the optimum array gain for an input SNR above 0 dB, and an array gain of 5 dB or more in 74% of the 10000 realizations, for an input SNR of -20 dB.

### Gain-optimized estimator

The case with the GO estimator with  $N_{\text{order}} = 4$  is shown in Fig. 7.20. The spread of the array gain is shown in Fig. 7.20(a). The array gain penalty with respect to the optimum gain is approximately 0.16 dB at high input SNR. The spread becomes small for input SNR above 0 dB. The spread of the white-noise gain is displayed in Fig. 7.20(b). We note how similar it is to the spread of the array gain, because of how close  $\mathbf{R}_{ee}(k)$  is to the identity matrix.

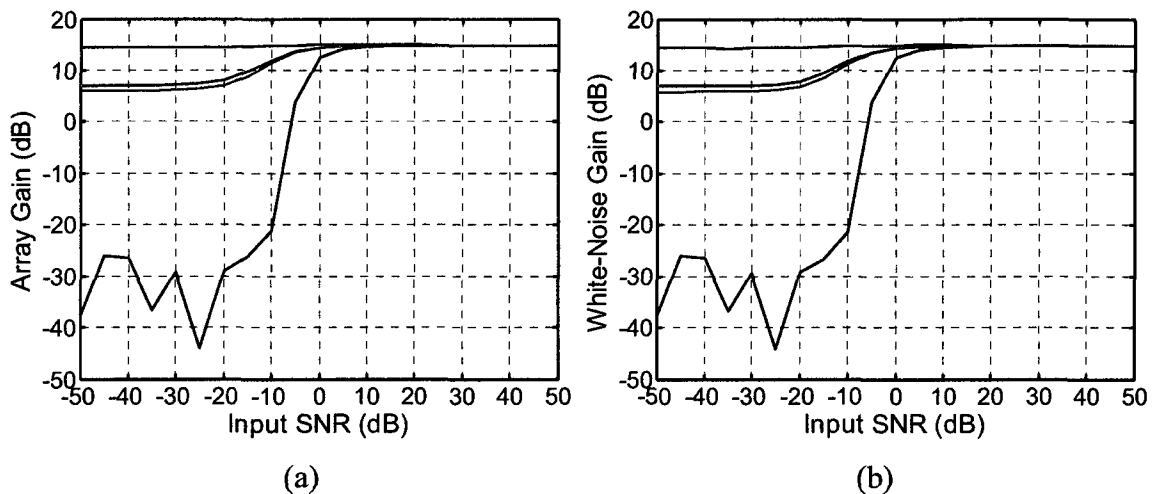


Figure 7.20. Rectangular array, dipole source, and diffuse noise. With the GO estimator, mean (green), median (red), minimum (blue), and maximum (blue) values for 10000 realizations of (a) Array gain; (b) White-noise gain.

By only observing the minimum and maximum array gain values obtained over the 10000 realizations, it may seem that the GO estimator gives proper gain only for input SNR above 0 dB. Histograms of the array gain for the 10000 realizations, with input SNR of -50, -20, 0, and 10 dB are shown in Fig. 7.21. Even for input SNR of -50 and -20 dB, a large number of realizations provide a good array gain, but the tail of the distribution does extend to very low values. At an input SNR of -50 dB, 66% of the realizations have an array gain of 5 dB or more. With an input SNR of -20 dB, the array gain is 5 dB or more in 74% of the realizations.

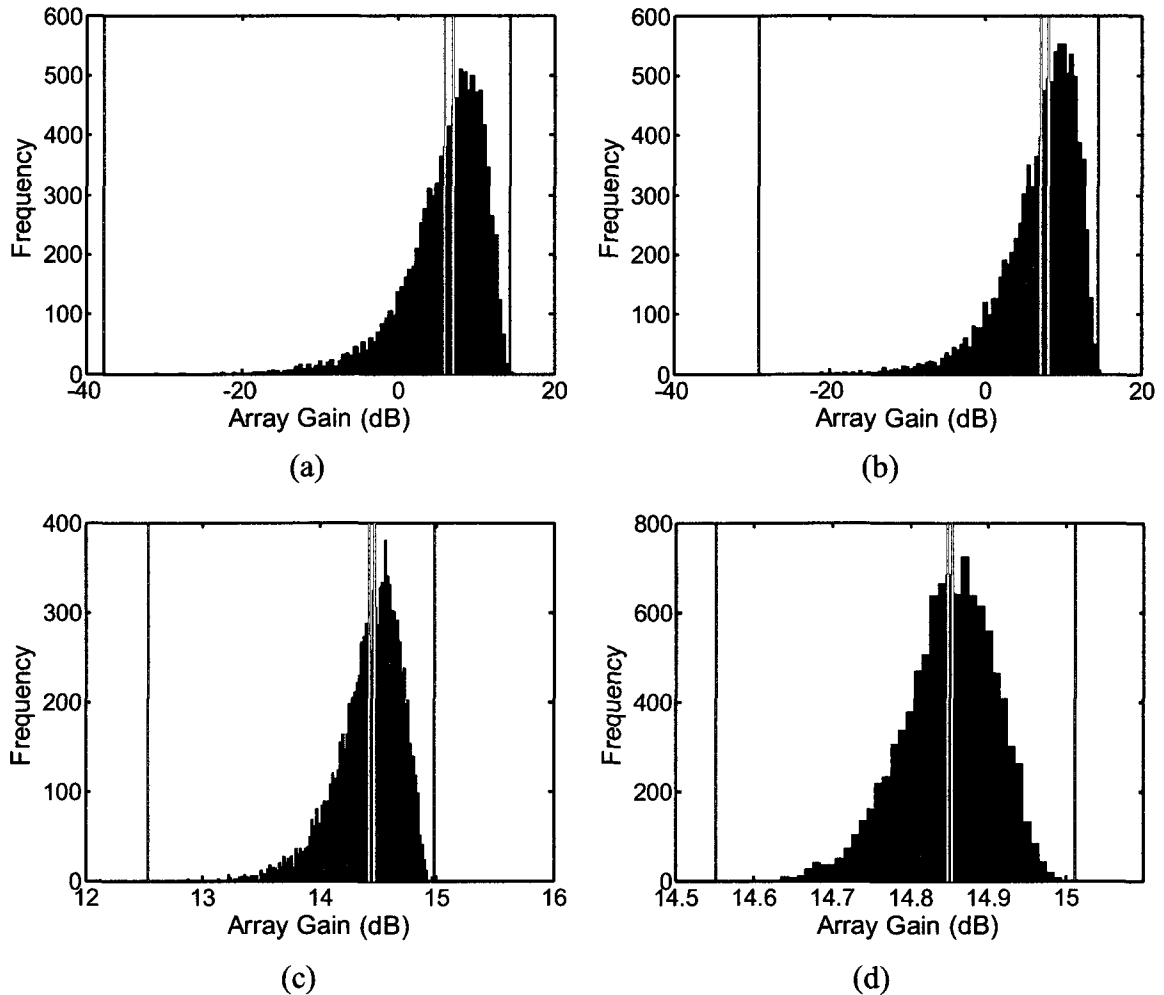


Figure 7.21. Rectangular array, dipole source, and diffuse noise. Histogram of array gain for the 10000 realization with the GO estimator for input SNR of (a) -50 dB, bar width: 0.5 dB; (b) -20 dB, bar width: 0.5 dB; (c) 0 dB, bar width: 0.02 dB; (d) 10 dB, bar width: 0.01 dB.

### ***Least squares estimators***

Array gain results with LS estimation when including spherical harmonic modes up to order  $n = 3$  (16 modes) into the matrix  $\mathbf{Y}(k)$  are shown in Fig. 7.22(a).  $\mathbf{Y}(k)$  is a  $16 \times 16$  matrix and its rank is 16. Matlab does not give a warning when computing the inverse of  $\mathbf{Y}^H(k)\mathbf{Y}(k)$ . Results when the pseudoinverse is computed with singular value decomposition are the same as when it is computed directly. Therefore, from this point forward, LS estimation is computed only with singular value decomposition.

At low input SNR, the mean of the array gain is only slightly above 0 dB while it was approximately 7 dB with the GO estimator. The spread of the array gain becomes small above 5 dB input SNR. At high input SNR, the array gain penalty with respect to the optimum gain is less than 0.01 dB.

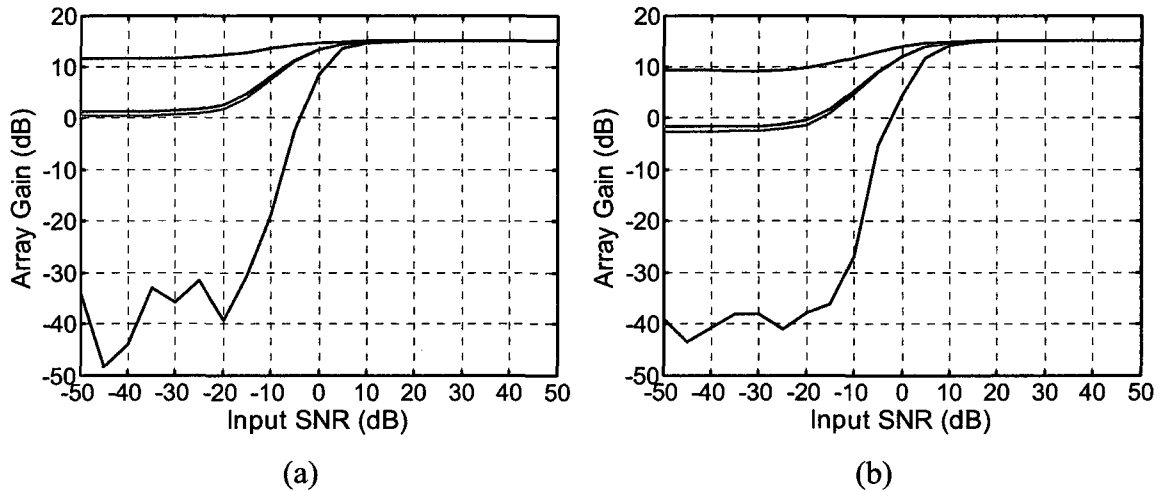


Figure 7.22. Rectangular array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations. (a) LS (SVD); (b) microphone output vector as the source model.

### ***Microphone output vector as the source model***

The spread of the array gain when the microphone output vector is used as the source model is presented in Fig. 7.22(b). The average array gain is below 0 dB for low input SNR. The spread of the array gain becomes small for input SNR above 5 dB. The optimum gain is almost met at high input SNR since the source model is then very close to the source signal.

### 7.3.3 Rectangular array and quadrupole source

Simulations were also done with the rectangular array and a quadrupole source located at  $(x, y, z) = (2, 1, 0)$ . We kept the same source location as for the dipole, which still allows to avoid an excessively small (theoretically zero) response for conventional beamforming. The exact source mode strengths are  $s_{2,-2}(k) = 1$  and  $s_{2,2}(k) = 1$ . The array as well as the source location and directivity are displayed in Fig. 7.23.

The optimum array gain and the gain for conventional beamforming are given in Table 7.4. This is another situation where conventional beamforming performs poorly.

We do not include figures for this case as the behavior of the different estimators is the same as they were in the previous sub-section: at low input SNR, a large number of realizations provide a good array gain but the tail of the distribution does extend to very low gain values; the spread of the array gains becomes small for input SNR above 0 dB. However, we do include the array gain and the resulting white-noise gain for the GO and LS estimators at high input SNR in Table 7.4. This is another case where the proposed beamforming method performs much better than conventional beamforming.

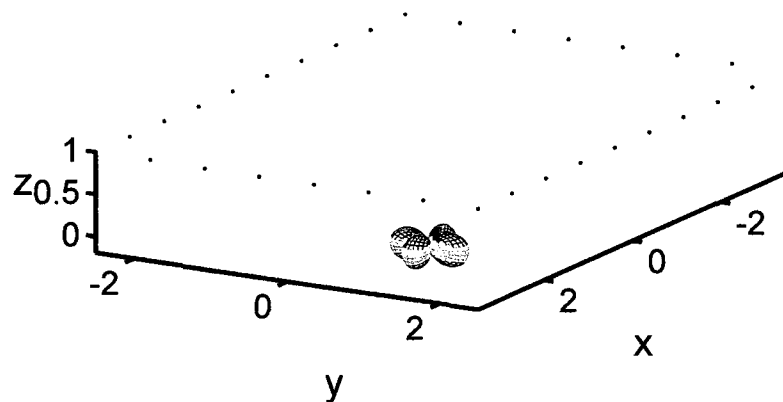


Figure 7.23. Rectangular array and quadrupole source.

Table 7.4. Array gain and white-noise gain for the quadrupole source and rectangular array.

Source model	Array gain	White-noise gain
	(dB)	
Exact source signal (optimum)	15.27	15.00
Point source (conventional beamforming)	-17.81	-17.92
GO estimator, high input SNR	14.77	14.39
LS (SVD) estimator, high input SNR	15.27	15.00

#### 7.3.4 Rectangular array and monopole source

The proposed beamforming method should still perform properly when the source is a monopole. For this case, the simulations are done with the rectangular array and a monopole source located at  $(x, y, z) = (2, 1, 0)$ . It was not necessary to avoid  $(x, y)$  coordinates equal to  $(0, 0)$  with the monopole source but we chose to keep the same source location as for the dipole and the quadrupole. There is only one mode and the exact source mode strength is  $s_{0,0}(k) = 1$ . The array as well as the source location and directivity are shown in Fig. 7.24.

The optimum array gain and the gain for conventional beamforming are given in Table 7.5. Of course for a monopole source, the point source is the exact source model so that conventional beamforming does provide the optimum array gain.

Table 7.5. Array gain and white-noise gain for monopole source and rectangular array.

Source model	Array gain	White-noise gain
	(dB)	
Exact source signal (optimum)	15.13	15.01
Point source (conventional beamforming)	15.13	15.01

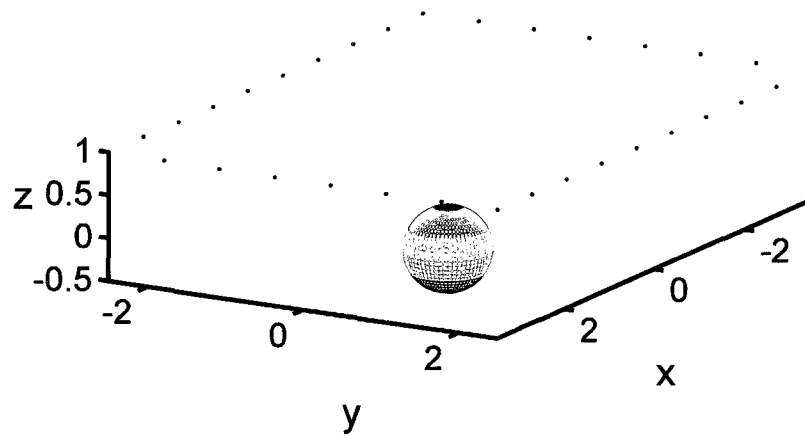


Figure 7.24. Rectangular array and monopole source.

The spread of the array gain for the GO estimator is shown in Fig. 7.25(a). The behavior is similar as to what it was for the dipole source. At high input SNR, the penalty with respect to the optimum gain is 0.1 dB. The spread of the array gain for the LS estimator is shown in Fig. 7.25(b). Again, the behavior is similar to what it was for the dipole source. At high input SNR, the penalty with respect to the optimum gain is less than 0.01 dB.

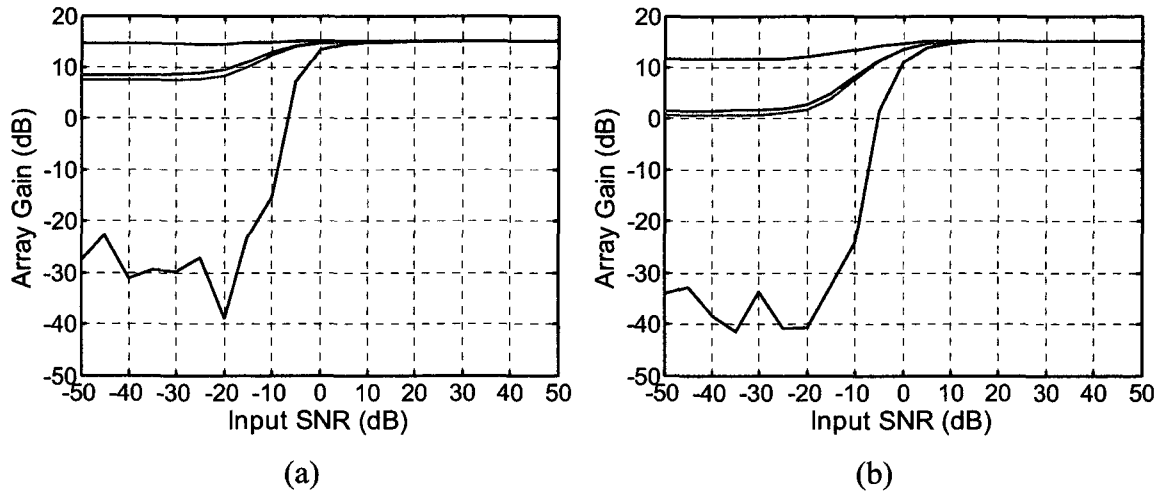


Figure 7.25. Rectangular array, monopole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with (a) GO estimator; (b) LS (SVD).

### 7.3.5 Gain-optimized estimator with varying number of frames

We will now revisit the simulation of Sub-Section 7.3.2, i.e. rectangular array and dipole source, with the GO estimator and for a varying number of frames used for the estimation. We will limit the simulation to input SNR of 0 dB and below, as this is where the spread in the array gain is larger and where it would be most useful if averaging could reduce the spread.

Equation (6.50) is used to compute the GO estimator with more than one frame. When the mode  $(n, m)$  having the largest strength is not the same for all frames, the mode having the largest strength in the first frame is chosen as  $(n_{\text{ref}}, m_{\text{ref}})$ .

The spread of the array gain for the GO estimator with 1, 5, and 15 frames used in the estimation are displayed in Fig. 7.26.



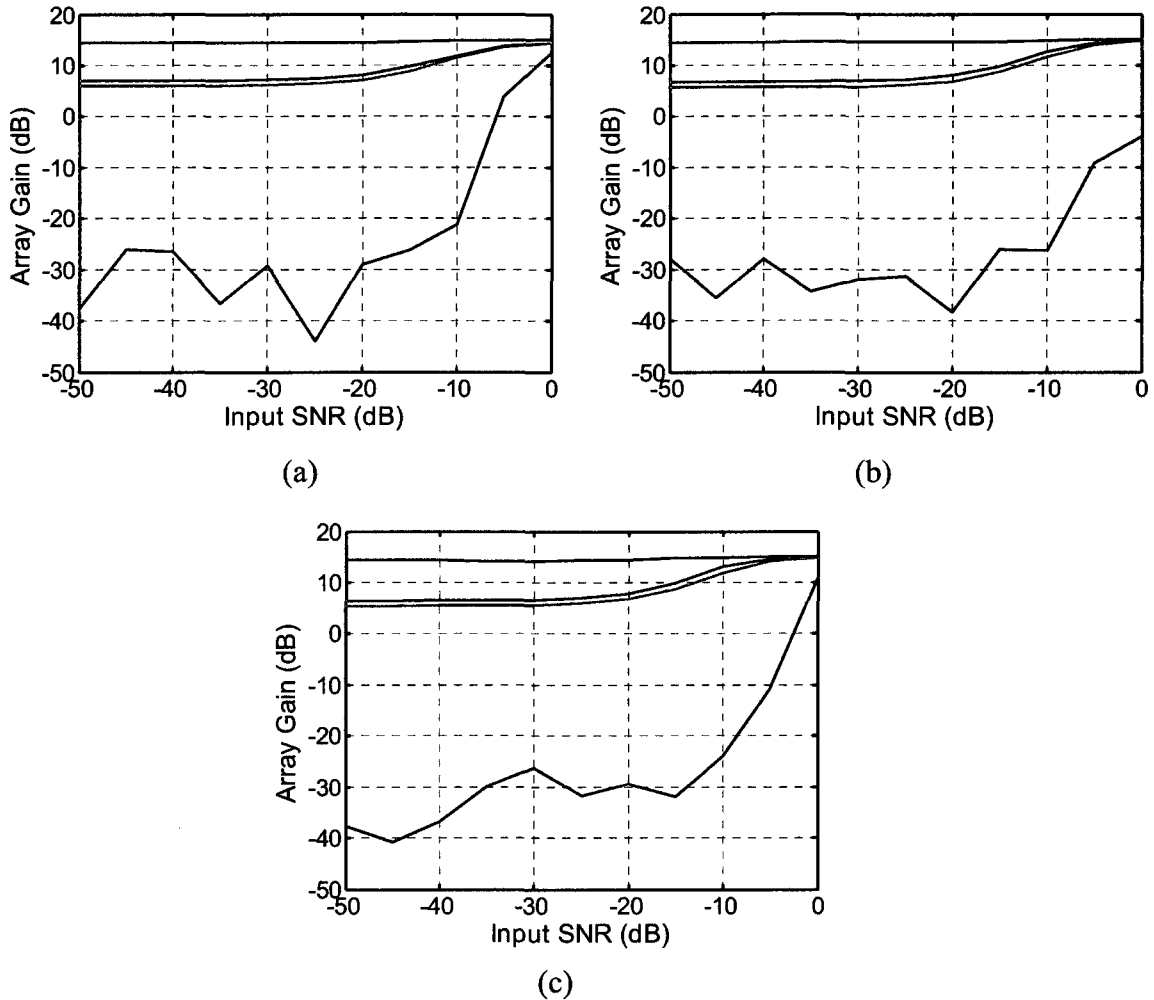


Figure 7.26. Rectangular array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the GO estimator and estimation done with (a) 1 frame; (b) 5 frames; (c) 15 frames.

Figure 7.27 shows the standard deviation of the array gain over 10000 realizations for varying input SNR while estimating mode strengths with 1, 5, or 15 frames. For input SNR of -50 to -20 dB, the standard deviation of the array gain is very close for 1, 5, and 15 frames. The case with 15 frames is a little worse than the other two. For input SNR of -20 to -5 dB, using a single frame provides a slightly smaller standard deviation, while at 0 dB input SNR the standard deviation is somewhat better with 15 frames.

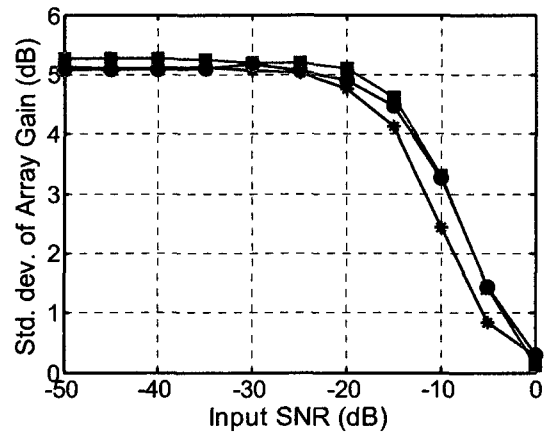


Figure 7.27. Rectangular array, dipole source, and diffuse noise. Standard deviation of the array gain for the 10000 realizations with the GO estimator and estimation with (\*) 1 frame, (●) 5 frames, (■) 15 frames.

Figure 7.28 shows that the distribution of the array gain for an input SNR of -20 dB is similar when using 1, 5, or 15 frames to estimate the mode strengths.

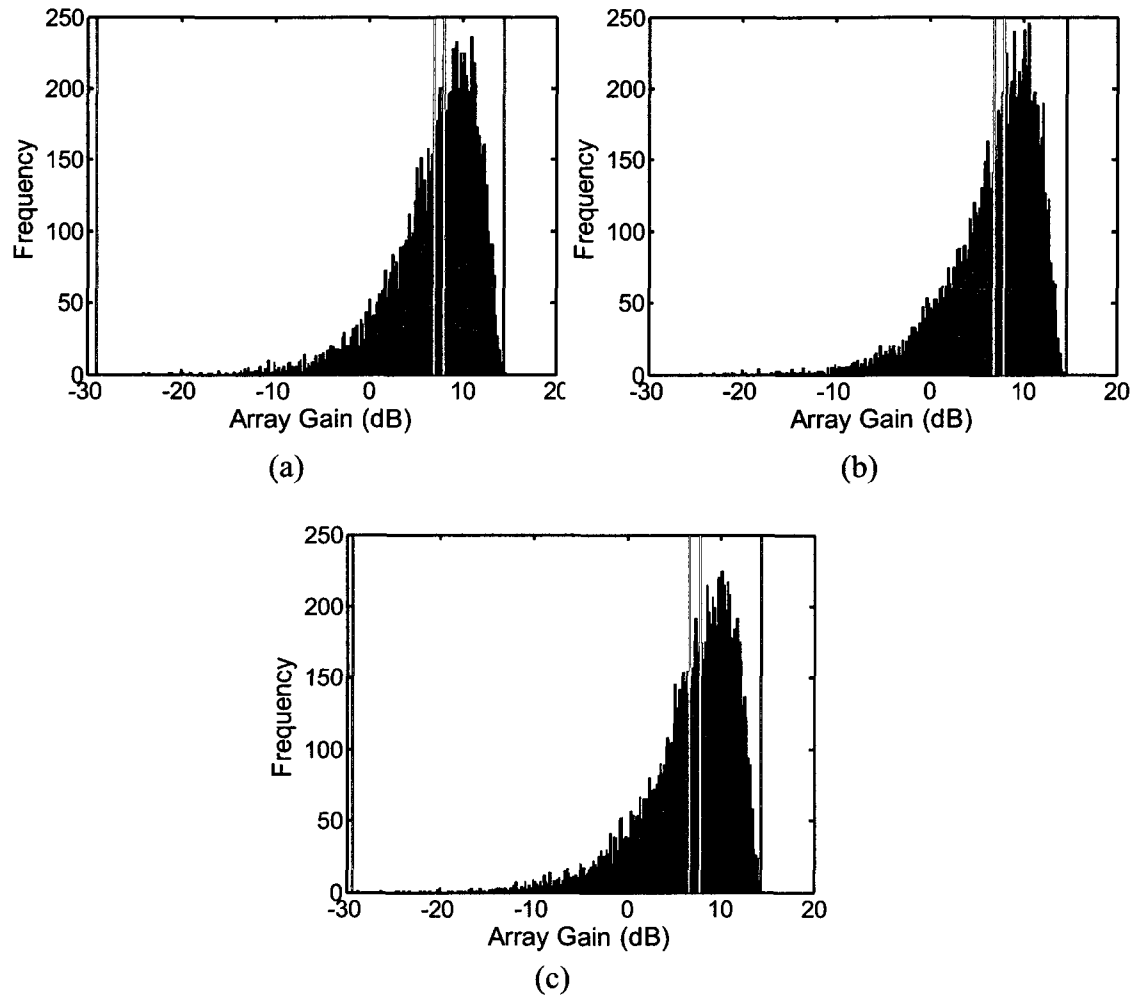


Figure 7.28. Rectangular array, dipole source, and diffuse noise. Histogram of array gain for the 10000 realizations with the GO estimator for input SNR of -20 dB and estimation with (a) 1 frame; (b) 5 frames; (c) 15 frames. Bar width: 0.2 dB. Mean (green), median (red), minimum (blue), and maximum (blue) are also displayed.

In Fig. 7.29, for an input SNR of -10 dB, the peak of the distribution of the array gain gets sharper with an increase in the number of frames. However on Fig. 7.27, the standard deviation was lower for a single frame than for 5 or 15 frames. This may be due to the minimum value of the array gain over the 10000 realizations being lower (worse) for 5 and 15 frames than for a single frame at input SNR of -10 dB, as seen in Fig. 7.26.

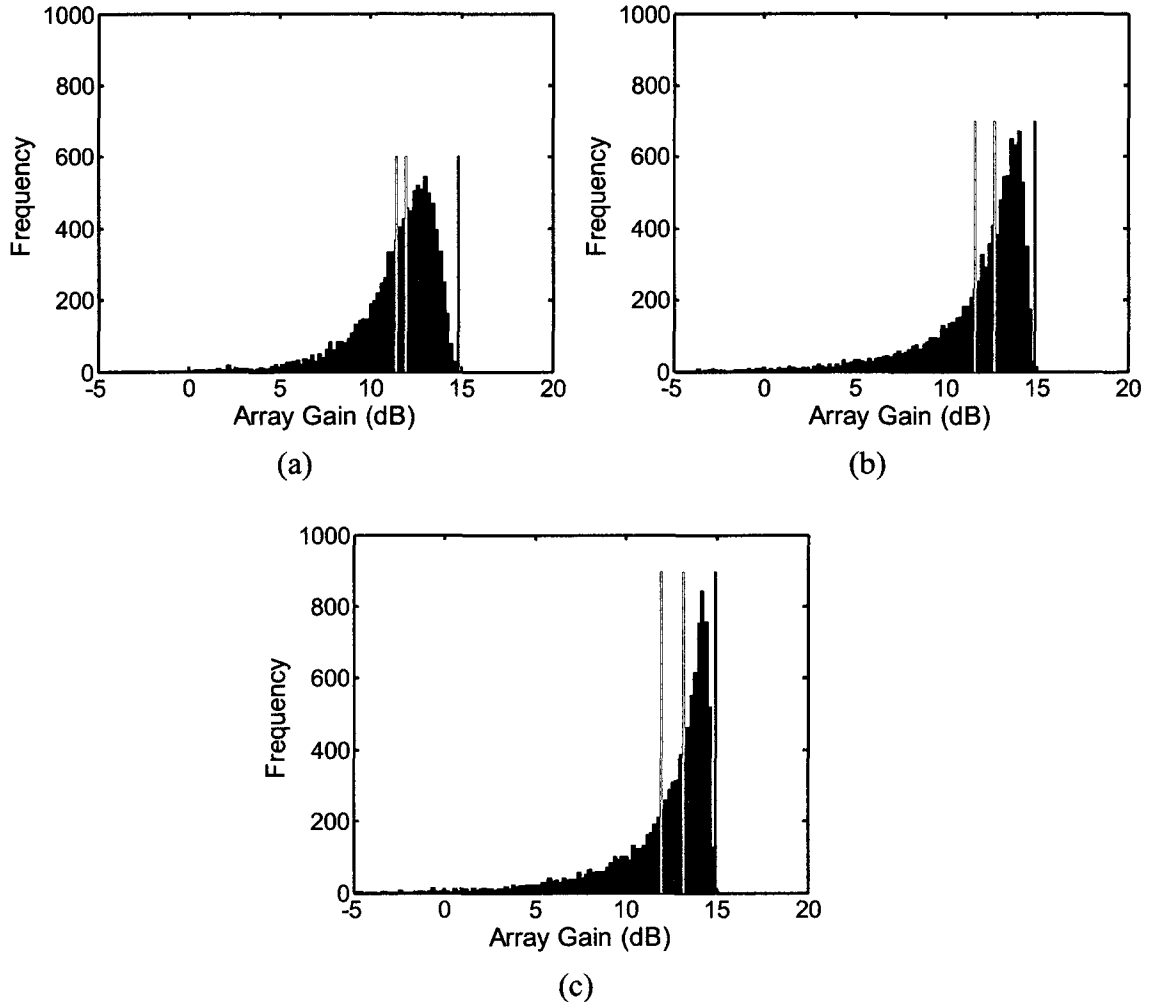


Figure 7.29. Rectangular array, dipole source, and diffuse noise. Histogram of array gain for the 10000 realizations with the GO estimator for input SNR of -10 dB and estimation with (a) 1 frame; (b) 5 frames; (c) 15 frames. Bar width: 0.2 dB. Mean (green), median (red), minimum (blue), and maximum (blue) are also displayed.

For 0 dB input SNR, Figure 7.30 shows the peak of the distribution becoming again sharper with the number of frames used in the estimation. In this case, the standard deviation was also lower with 15 frames. However, over the 10000 realizations, the minimum array gain was higher (better) when using a single frame than when using 5 or 15 frames, as shown in Fig. 7.26. This would indicate that even though the peak of the distribution is tighter when 15 frames are used instead of 1, the tail does extend further.

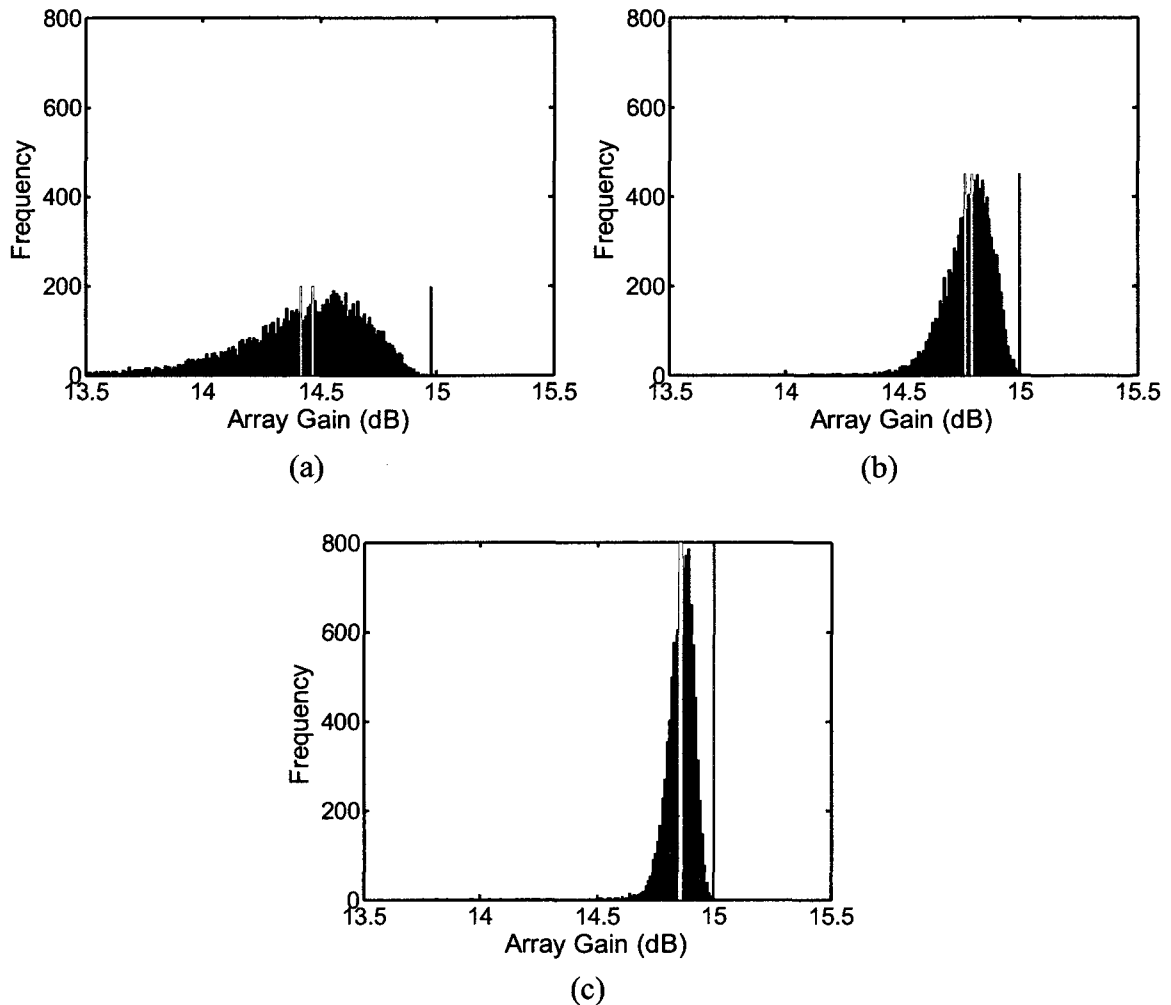


Figure 7.30. Rectangular array, dipole source, and diffuse noise. Histogram of array gain for the 10000 realizations with the GO estimator for input SNR of 0 dB and estimation with (a) 1 frame; (b) 5 frames; (c) 15 frames. Bar width: 0.01 dB. Mean (green), median (red), minimum (blue), and maximum (blue) are also displayed.

Though we will not include figures because of the space it would require, we have also investigated the possibility of using more than one mode to perform the phase alignment and compute the GO estimator with a number of frames, using Eqs (6.52) and (6.53). For a mode to be included in Eq. (6.52), the estimated strength of the mode, in frame 1, had to be 50% or more of the largest mode strength, also in frame 1. We have tried with 75% also. In the two cases (50% and 75%), the standard deviation of the array gain did not decrease compared to values in Fig 7.27. In fact, for input SNR values from -10 to 0 dB, the standard deviation increased. In addition, the distributions did not narrow

down as much with an increase in the number of frames as it did using Eq. (6.50), as was seen in Fig. 7.29 and Fig. 7.30.

A reduction in the array gain spread would have been the most welcome for low input SNR but using the GO estimator and several frames did not provide such an improvement. We attribute this to the phase alignment not performing adequately when the noise power at the output of the microphones is large. We will not investigate further in this thesis the use of many frames in the estimation.

### 7.3.6 Concluding remarks: Diffuse noise

In the simulation scenarios run with a directional source in the presence of diffuse noise, the proposed beamforming method performed much better than conventional beamforming. For example, with the GO estimator and an input SNR of 0 dB or higher, the array gain was close to the optimum value.

The proposed beamforming method also worked when the source was a monopole, providing again an array gain close to the optimum value for input SNR of 0 dB or higher.

We saw that for low input SNR, there is a spread of the array gain. A large portion of the realizations provides good array gain values but the tail of the distribution does extend to very low values.

At high input SNR, the proposed beamforming method with least-squares estimation provides an array gain extremely close to the optimum gain. This is because when the mode vectors for beamforming, contained in the matrix  $\mathbf{Y}(k)$ , are built for the exact source position, the source mode strength estimation of a mode is not biased by other modes present in the source. In the absence of reverberation or interfering sources, the error in the estimation then only comes from the noise, and at high input SNR, the effect of the noise is negligible.

The estimation of the strength of a mode with the GO estimator is biased by the other modes present in the source. However, the array gain penalty is not major. It was, for example, of 0.5 dB for the quadrupole source. In this scenario, using conventional beamforming causes a penalty of more than 30 dB, the array gain being -17.8 dB.

Using the proposed beamforming method to provide a source model for MVDR beamforming did not prove to be an interesting option. At low input SNR, the spread in the array gain is similar to the proposed beamforming method without MVDR. At higher input SNR, there is a linear decrease in the array gain with an increase in input SNR, caused by the estimated source model being different from the source signal. This phenomenon is also seen with conventional beamforming using MVDR, for example when the beamforming location is different from the source position or when there are errors on the sensor positions.

## **7.4 Simulations: Diffuse noise + reverberation**

The goal in considering reverberation is not to have an exact model of the reverberation in a specific room, but rather to investigate if the proposed beamforming method stops working when there are multiple reflections, since the mode strengths are estimated from the microphone output signals.

The first sub-section describes the simulation of reverberation. The second sub-section gives a description of the simulation method. It will be explained why the simulations are done with the source frequency at the center of a frequency bin instead of using a source frequency of 1000 Hz. The analysis of the simulation results to evaluate the performance of the proposed beamforming method in the presence of reverberation is presented in Section 7.5.

### **7.4.1 Simulation of reverberation**

The method of images is used [Allen 1979] [Kuttruff 2000]. The room is rectangular to make it easier to keep track of the image locations. The reflection coefficient does not change with angle of incidence and is equal to a real number. For simplicity, the same reflection coefficient is used for the entire surface of all 6 walls.

Since the source is directional, the images are also directional. For every wall reflection, the source must be “flipped” to obtain the resulting image. Because of the simplified reflection model, the only information we need is the distance from the image to the microphone and at which angle the microphone observes the image. It is possible to simplify the simulation by taking “images” of the microphone (which are omnidirectional) while keeping the source at its position (instead of taking images of the source and keeping the microphone fixed) thus avoiding the issue of taking into account “source flips”.

The order of an image is the number of walls encountered by the ray to go from the source to the microphone. For a rectangular room, the number of images of order  $i_0$  is [Kuttruff 2000, p.304]

$$N_r(i_0) = 4i_0^2 + 2. \quad (7.33)$$

The simulation of reverberation is done order by order (all images of 1<sup>st</sup> order, all images of 2<sup>nd</sup> order, and so on). We stop including higher image orders in the reverberation signal when the change in magnitude by adding all the images of the current order is less than -50 dB and the change in phase is less than 0.01 degree. For each microphone, a complex number is obtained whose magnitude is applied to the amplitude of a time sinusoid and whose phase becomes the phase of the sinusoid. The sinusoid is the time output of the microphone (we are back to the actual microphone, not images of the microphone) from the reverberation component. A normalized FFT is then performed to obtain the reverberation component,  $s_r(k)$ , for the frequency bin  $k$ . We generate a time sinusoid going through a normalized FFT for the simulation of the reverberation component since this is how the signal for the direct path is generated.

The locations of the walls of the rectangular room are given below. The walls which are parallel to the  $y$ - $z$  plane are located at  $x$  equal -5 and 5 meters, the walls which are parallel to the  $x$ - $z$  plane are located at  $y$  equal -4 and 4 meters, and finally, the walls which are parallel to the  $x$ - $y$  plane are located at  $z$  equal to -2 and 3 meters. The absorption coefficient  $\alpha$  of the wall surface is 0.25. With these room dimensions and absorption coefficient, the Sabine formula [Kuttruff 2000, p.119]



$$T = 0.163 \frac{V}{\alpha S}, \quad (7.34)$$

gives a reverberation time of 0.77 second. The reverberation time is the time interval necessary for the sound decay level to drop by 60 dB. In the formula,  $V$  is the volume of the room and  $S$  is the total surface of the walls. However, we can not read much from this reverberation time with respect to our simulations because the assumptions for the Sabine formula (direct sound and all reflections from the walls are mutually incoherent, diffuse reflections, sound field isotropy) are different from our assumptions for the image method (specular reflections, coherence between all reflections, and coherence between reflections and the direct sound).

#### 7.4.2 Simulation description

The simulations are again done with varying input SNR, and as before, for each input SNR, 10000 realizations of the noise vector  $\mathbf{e}(k)$  are used. For each realization, the received microphone output vector is given by

$$\mathbf{p}(k) = \mathbf{s}(k) + \mathbf{s}_r(k) + \mathbf{e}(k). \quad (7.35)$$

The average reverberation power spectral density over the  $L$  microphones of the array is

$$\sigma_{s_r}^2(k) = \mathbf{s}_r^H(k) \mathbf{s}_r(k) / L. \quad (7.36)$$

The input signal-to-reverberation ratio (averaged over the  $L$  microphones) is

$$\text{SRR}_{\text{in}}(k) = \frac{\sigma_s^2(k)}{\sigma_{s_r}^2(k)}. \quad (7.37)$$

We do not specify the input signal-to-reverberation ratio. It results from the geometry of the array, source, and walls, as well as on the wall surface absorption coefficient. The normalized reverberation cross-power spectral density matrix is

$$\mathbf{R}_{s_r, s_r}(k) = \mathbf{s}_r(k) \mathbf{s}_r^H(k) / \sigma_{s_r}^2(k). \quad (7.38)$$

The average signal plus reverberation power spectral density over the  $L$  microphones of the array is

$$\sigma_{s+s_r}^2(k) = (\mathbf{s}(k) + \mathbf{s}_r(k))^H (\mathbf{s}(k) + \mathbf{s}_r(k)) / L \quad (7.39)$$

while the normalized signal plus reverberation cross-power spectral density matrix is

$$\mathbf{R}_{(s+s_r)(s+s_r)}(k) = (\mathbf{s}(k) + \mathbf{s}_r(k))(\mathbf{s}(k) + \mathbf{s}_r(k))^H / \sigma_{s+s_r}^2(k). \quad (7.40)$$

The microphone output cross-power spectral density is

$$\mathbf{R}_{pp}(k) = \sigma_{s+s_r}^2(k) \mathbf{R}_{(s+s_r)(s+s_r)}(k) + \sigma_e^2(k) \mathbf{R}_{ee}(k), \quad (7.41)$$

With the MVDR beamformer,  $\mathbf{R}_{pp}(k)$  is part of the computation to find the beamforming weight vector, so reverberation is taken into account in the optimization. For all the other cases, the optimization criterion remains the improvement in signal-to-noise ratio in the presence of diffuse noise, and the optimization criterion is not the improvement in signal-to-reverberation ratio. The beamforming weight vectors for the different approaches are computed as they were in the section on diffuse noise.

In addition to observing the array gain and the white-noise gain, we also compute the gain in signal-to-reverberation ratio (SRR), defined as

$$G_{\text{SRR}}(k) = \frac{\mathbf{w}^H(k) \mathbf{R}_{ss}(k) \mathbf{w}(k)}{\mathbf{w}^H(k) \mathbf{R}_{s_r s_r}(k) \mathbf{w}(k)}, \quad (7.42)$$

which is equivalent to

$$G_{\text{SRR}}(k) = \frac{|\mathbf{w}^H(k) \mathbf{s}(k)|^2 / \sigma_s^2(k)}{|\mathbf{w}^H(k) \mathbf{s}_r(k)|^2 / \sigma_{s_r}^2(k)}. \quad (7.43)$$

### **Source frequency**

For the simulations with reverberation (or with an interferer, in Section 7.7), the signal frequency is at the center of the frequency bin  $k = 85$  (corresponding to 996.1 Hz) instead of being at 1000 Hz. This is to avoid a discrepancy in the gain in signal-to-reverberation ratio (SRR) (or in the gain in signal-to-interference ratio (SIR)) when the source is a monopole. For this case of a monopole source, the optimum source model is the mode vector  $(m, n) = (0, 0)$ . Using either the optimum source model or the mode

vector  $\mathbf{Y}_0^0(k)$  should provide the same gain in SRR (or the same gain in SIR). But the way that the simulations are built, the gain in SRR (or the gain in SIR) is not the same when the source frequency is not at the center of a frequency bin.

To understand this, we recall that the optimum source model is obtained from a sampled time signal generated with the specified source frequency, on which an FFT is performed. The magnitude and phase of the signal component in the frequency bin where the peak of the FFT is located (frequency bin  $k = 85$  for a source at 1000 Hz) is used as the optimum source model. The magnitude and phase for the  $L$  sensors are combined into a vector. This is the vector  $\mathbf{s}(k)$  used to compute the normalized signal cross-power spectral density matrix  $\mathbf{R}_{ss}(k)$  and to compute the different gains.

On the other hand, the mode vector  $\mathbf{Y}_0^0(k)$  is computed directly with the spherical harmonic of order 0 and the Hankel function of second kind of order 0, using the wavelength corresponding to the center of the frequency bin  $k = 85$ . A sampled time signal is not created or used to compute this.

To summarize,  $\mathbf{s}(k)$  is computed for a specified source frequency (not necessarily at the center of a frequency bin) while  $\mathbf{Y}_0^0(k)$  is computed for the center of the bin.

When the source has a frequency of 1000 Hz, using  $\mathbf{s}(k)$  or  $\mathbf{Y}_0^0(k)$  lead to a difference in the array gain (or the white-noise gain) of only a few hundredth of a decibel. For the gain in SRR (or the gain in SIR), the difference can be in the order of 1 dB. Therefore, to avoid this issue, we choose a source frequency at the center of the frequency bin  $k = 85$ .

## 7.5 Simulation results: Diffuse noise + reverberation

This section presents simulation results for a source in the presence of reverberation. The scenarios are:

- Rectangular array and dipole source;
- Rectangular array and monopole source.

### 7.5.1 Rectangular array and dipole source

The array is composed of 32 microphones located on the perimeter of a rectangle. The source is a dipole located at  $(x, y, z) = (2, 1, 0)$  with its axis parallel to the  $x$ -axis. The exact source mode strengths are  $s_{1,-1}(k) = 1$  and  $s_{1,1}(k) = -1$ . The array as well as the source location and directivity were displayed in Fig. 7.19.

For this particular room, array geometry, and source, reflections up to order 35 were included to obtain the reverberation component at the output of the microphones,  $s_r(k)$ . The resulting input signal-to-reverberation ratio,  $\text{SRR}_{\text{in}}(k)$ , is -3.46 dB.

The optimum array gain and the gain for conventional beamforming are given in Table 7.6 along with the resulting white-noise gain and gain in SRR. As it was the case in Section 7.3.2, conventional beamforming provides a low array gain for this array geometry and the dipole at its location. Now we also see that conventional beamforming provides a low gain in SRR.

Table 7.6. Array gain, white-noise gain, and gain in SRR for dipole source and rectangular array.

Source model	Array gain	W.-N. gain	Gain in SRR
		dB	
Exact source signal	15.06	15.01	32.41
Point source (conv. beamf.)	-1.55	-1.66	-4.27

#### ***Gain-optimized estimator***

The spread in the array gain with the GO estimator, without and with reverberation, is shown in Fig. 7.31(a) and Fig. 7.31(b), respectively. The two plots are very similar but the array gain at high input SNR is slightly lower when there is reverberation, at a value of 13.7 dB, compared to 14.9 dB when there is no reverberation.

Those values are in Table 7.7 along with the values for the other estimators. The optimum array gain is 15 dB.

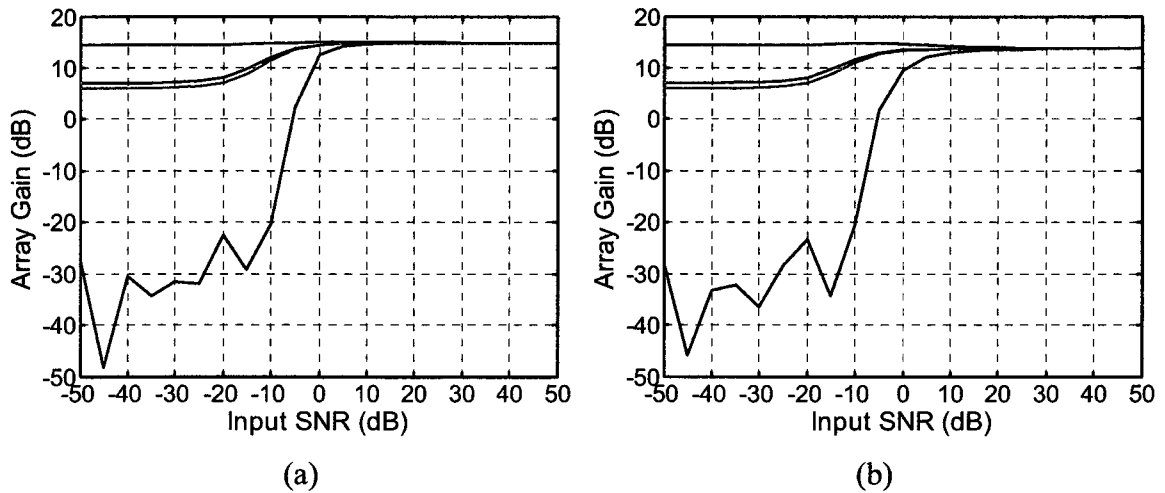


Figure 7.31. Rectangular array, dipole source, diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the GO estimator. (a) Without reverberation; (b) With reverberation.

Table 7.7. Array gain and gain in SRR at high input SNR for dipole source and rectangular array.

Approach	Array gain		Gain in SRR
	No reverb	Reverb	Reverb
	dB		
GO estimator	14.9	13.7	11.5
LS	15.0	12.2	3.9
Microphone output	~ 15	~10	-3
GO + MVDR	-	11.1	3.4

The spread of the resulting gain in SRR is displayed in Fig. 7.32. The average gain in SRR is above 0 dB for all input SNR. However the spread is large at low input SNR. For input SNR above approximately -4 dB, the gain in SRR remains above 0 dB for the 10000 realizations. The comparison for the four estimators is in Table 7.8. The spread

of the gain in SRR becomes small above 10 dB input SNR. At high input SNR, the resulting gain in SRR is 11.5 dB. This value is in Table 7.7, along with the values for the other 3 approaches. The gain in SRR with the optimum source model is 32.4 dB.

Histograms of the resulting gain in SRR for the 10000 realizations with input SNR of -50, -20, 0, and 10 dB are shown in Fig. 7.33.

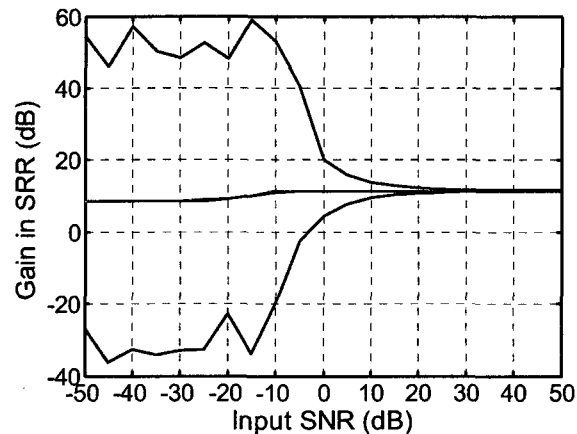


Figure 7.32. Rectangular array, dipole source, diffuse noise, and reverberation. Mean (green), median (red), minimum (blue), and maximum (blue) values of the resulting gain in SRR for the 10000 realizations with the GO estimator.

Table 7.8. Minimum input SNR for which the gain in SRR is 0 dB or above for all 10000 realizations, for dipole source and rectangular array.

Approach	Input SNR dB
GO estimator	-4
LS	5
Microphone output	-
GO + MVDR	-6

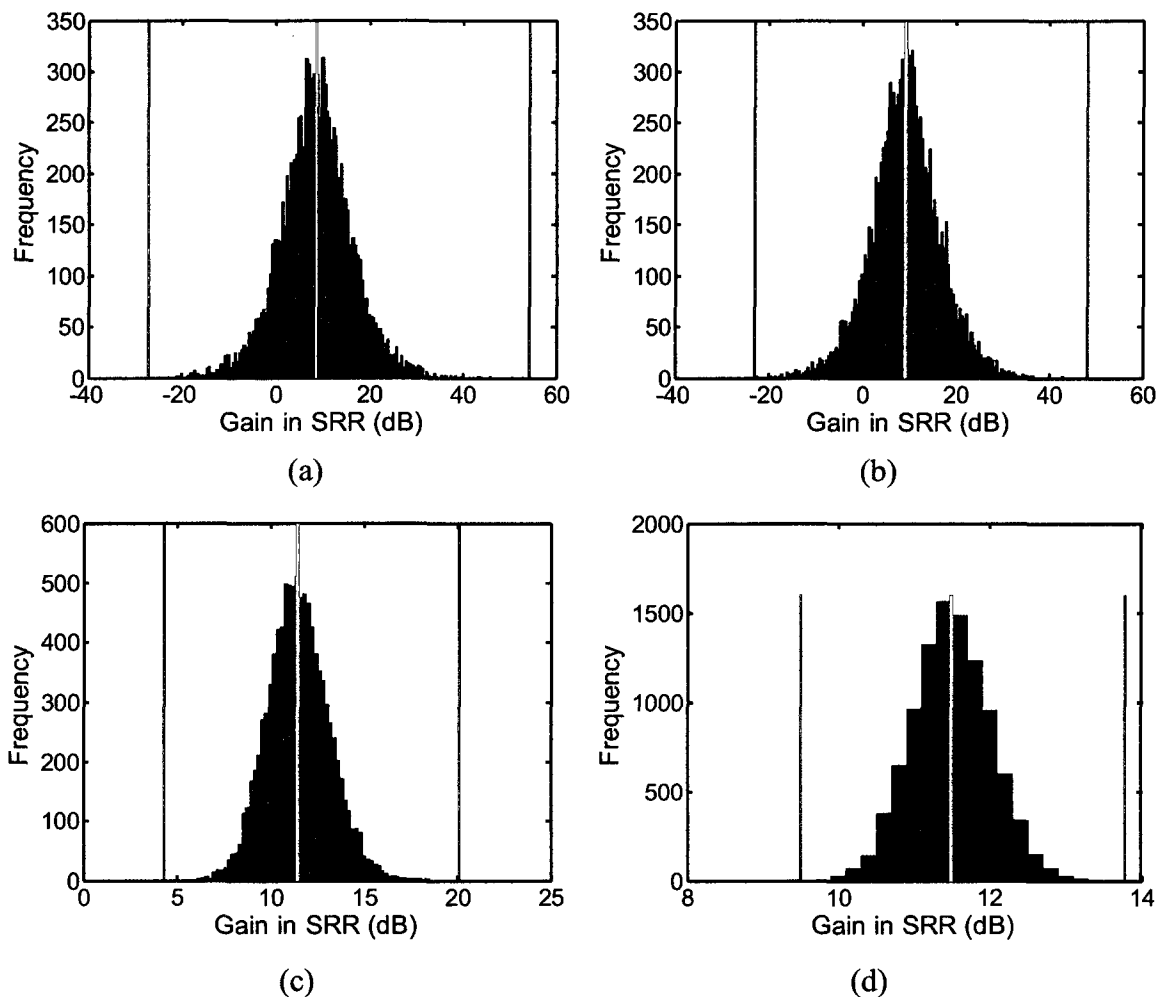


Figure 7.33. Rectangular array, dipole source, diffuse noise, and reverberation. Histograms of resulting gain in SRR for the 10000 realizations with the GO estimator for input SNR of (a) -50 dB, bar width: 0.5 dB; (b) -20 dB, bar width: 0.5 dB; (c) 0 dB, bar width: 0.02 dB; (d) 10 dB, bar width: 0.2 dB.

Figures of the spread of the array gain versus the input SNR, the spread of the resulting gain in SRR versus the input SNR, or even the histograms of the gain in SRR at specific input SNR do not indicate, for a given realization of the noise vector, whether the resulting gain in SRR is large if the array gain is large. For that, plots of the resulting gain in SRR versus array gain for input SNR of -50, -20, 0, and 10 dB are included in Fig. 7.34. A dot marks the array gain and the resulting gain in SRR for each of the 10000 realizations.

We observe, as in the previous figures, the larger spread for lower input SNR. We also note the general trend of higher resulting gain in SRR for higher array gain.

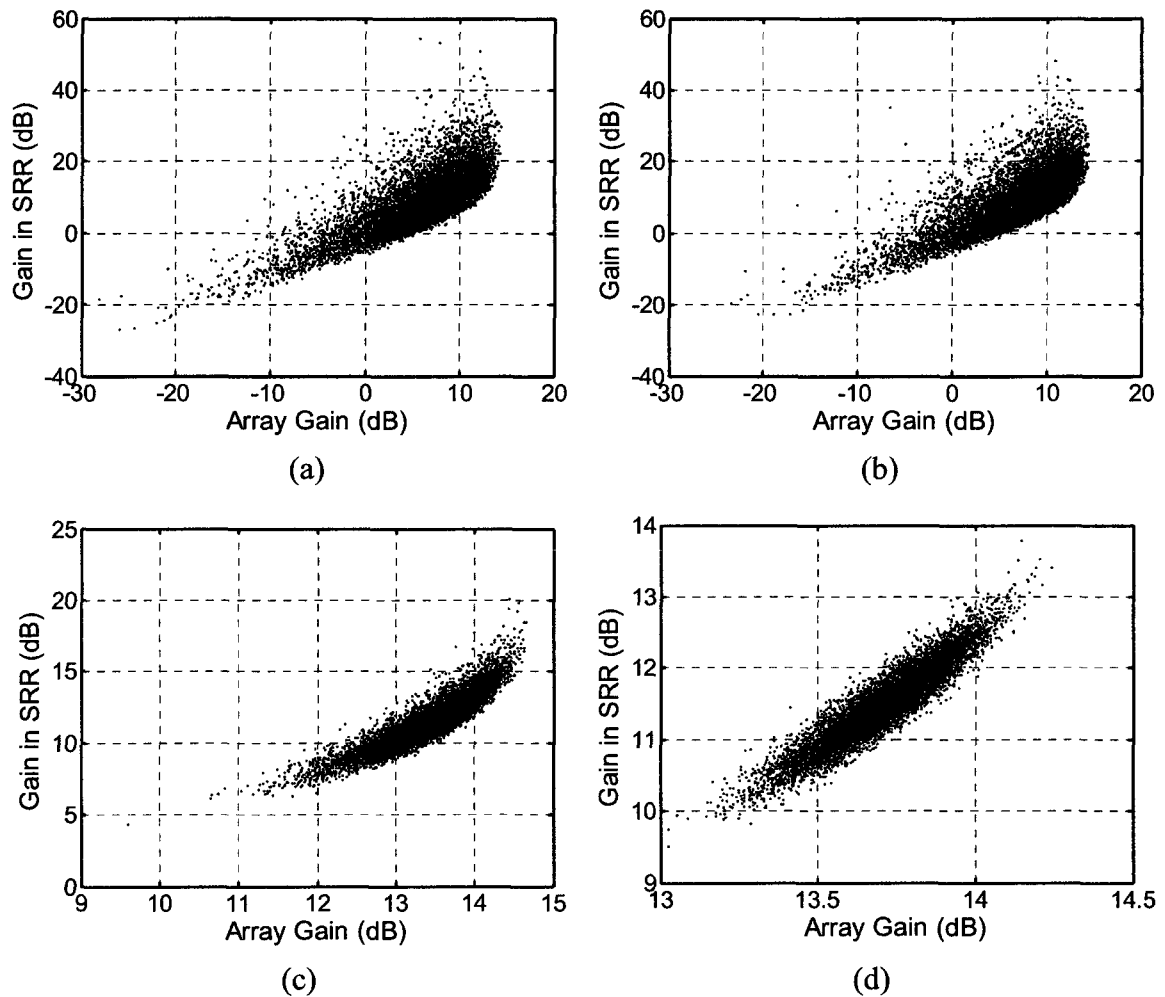


Figure 7.34. Rectangular array, dipole source, diffuse noise, and reverberation. Gain in SRR vs. array gain for the 10000 realizations with the GO estimator for input SNR of (a) -50 dB; (b) -20 dB; (c) 0 dB; (d) 10 dB.

### ***Least squares estimator***

The spread in the array gain with the LS estimator, without and with reverberation, is shown in Fig. 7.35(a) and Fig. 7.35(b), respectively. The two plots are very similar but the array gain at high input SNR is lower when there is reverberation, at a value of 12.2 dB, while it is 15 dB when there is no reverberation. Compared to



Fig. 7.31(a) and Fig. 7.31(b), the average array gain is lower with the LS estimator than for the GO estimator at lower input SNR. This is also the case for the maximum array gain value over the 10000 realizations, at low input SNR. With reverberation, at 12.2 dB the array gain at high input SNR with the LS estimator is lower than with the GO estimator, which is 13.7 dB.

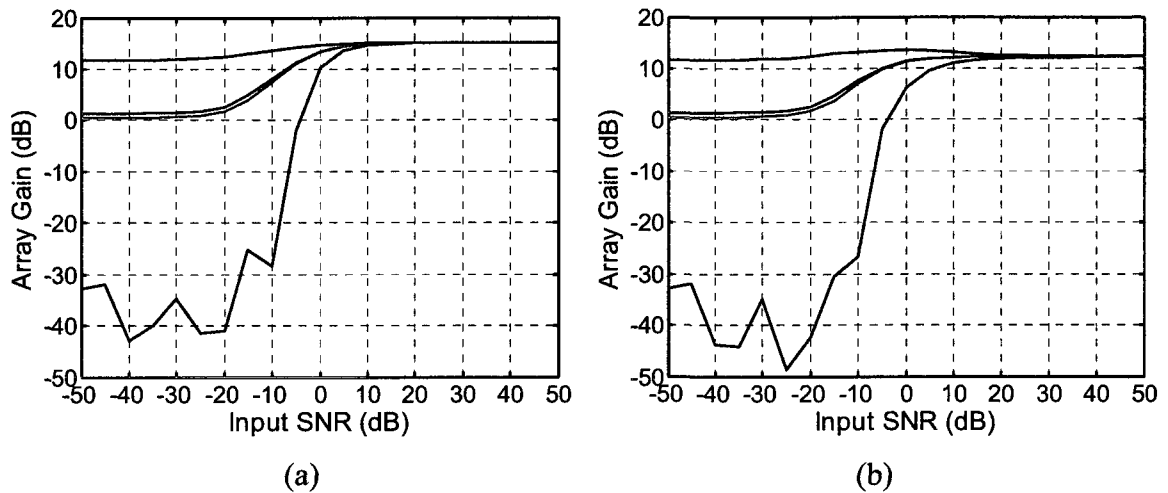


Figure 7.35. Rectangular array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the LS estimator. (a) Without reverberation; (b) With reverberation.

The spread of the resulting gain in SRR is displayed in Fig. 7.36. The average gain in SRR is above 0 dB for all input SNR. As with the GO estimator, the spread is large at low input SNR. The average gain in SRR is lower than with the GO estimator, for all input SNR. For input SNR above approximately 5 dB, the gain in SRR remains above 0 dB for the 10000 realizations, while this was observed for input SNR above -4 dB with the GO estimator. The spread becomes small above 10 dB input SNR, as it did with the GO estimator, but the resulting gain in SRR tends toward a value of 3.87 dB, compared to 11.5 dB for the GO estimator.

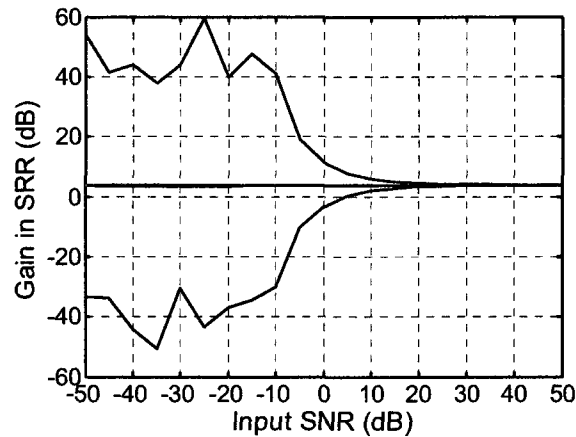


Figure 7.36. Rectangular array, dipole source, diffuse noise, and reverberation. Mean (green), median (red), minimum (blue), and maximum (blue) values of the resulting gain in SRR for 10000 realizations with the LS estimator.

### ***Microphone output vector as the source model***

The spread in the array gain with the microphone output vector used as the source model, without and with reverberation is shown in Fig. 7.37(a) and Fig. 7.37(b), respectively. The two plots are fairly similar but the array gain at high input SNR is about 5 dB lower when there is reverberation. In both cases, the average array gain is below 0 dB for input SNR below -17 dB.

The spread of the resulting gain in SRR is displayed in Fig. 7.38. The average gain in SRR is between 0 and 0.05 dB for input SNR between -50 and -35 dB, and it is below 0 dB for higher input SNR. As with the GO estimator, the spread of the gain in SRR is large at low input SNR. It becomes small above 10 dB input SNR, but around a gain in SRR value of -3 dB.

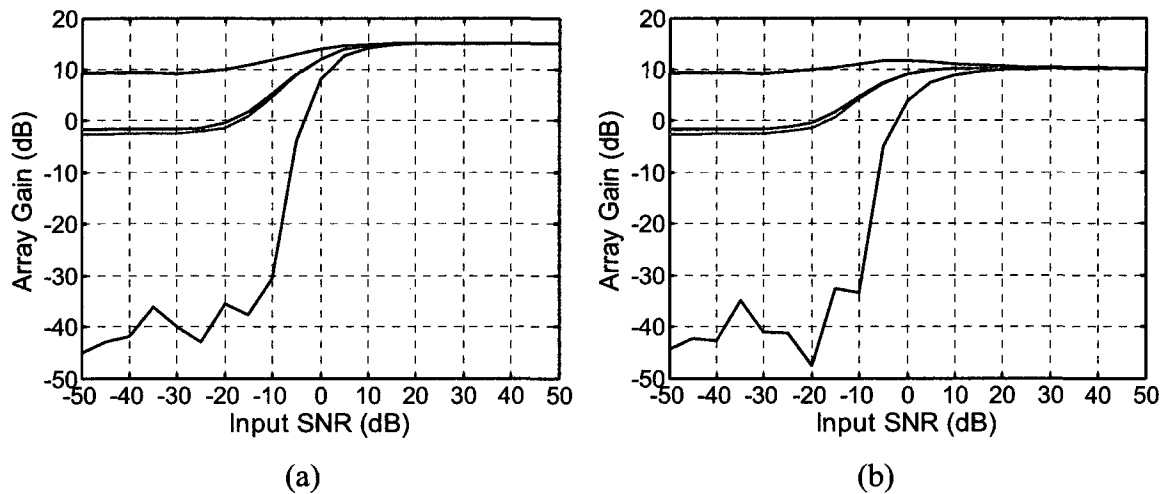


Figure 7.37. Rectangular array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the microphone output vector as the source model. (a) Without reverberation; (b) With reverberation.

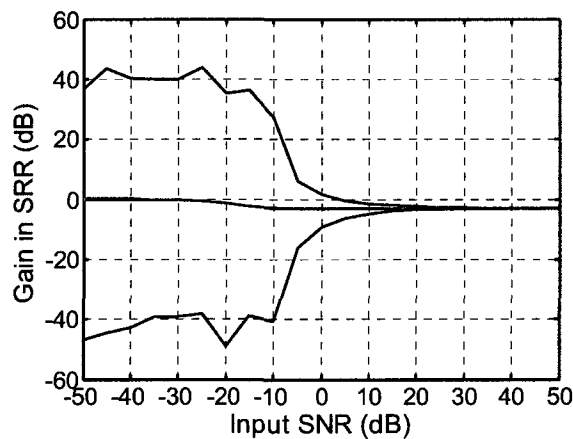


Figure 7.38. Rectangular array, dipole source, diffuse noise, and reverberation. Mean (green), median (red), minimum (blue), and maximum (blue) values of the resulting gain in SRR for 10000 realizations with the microphone output vector as the source model.

### ***Estimated source model and MVDR beamforming***

The spread in the array gain, when the GO estimator and our proposed beamforming method are used to obtain a source model which is then used with an MVDR beamformer, is shown for the scenarios without and with reverberation in Fig. 7.39(a) and Fig. 7.39(b), respectively. Without reverberation, we observe the same behavior as we did with the linear array, in Sub-Section 7.3.1: a large spread of the array gain at low input SNR and a linear decrease of the array gain with an increase in the input SNR at higher input SNR. With reverberation, there is a large spread at low input SNR, with an average similar as to when the GO estimator is used in our proposed beamforming method, without the MVDR beamformer. At high input SNR, the array gain is 11.1 dB, compared to 13.7 dB when the GO estimator is used without MVDR.

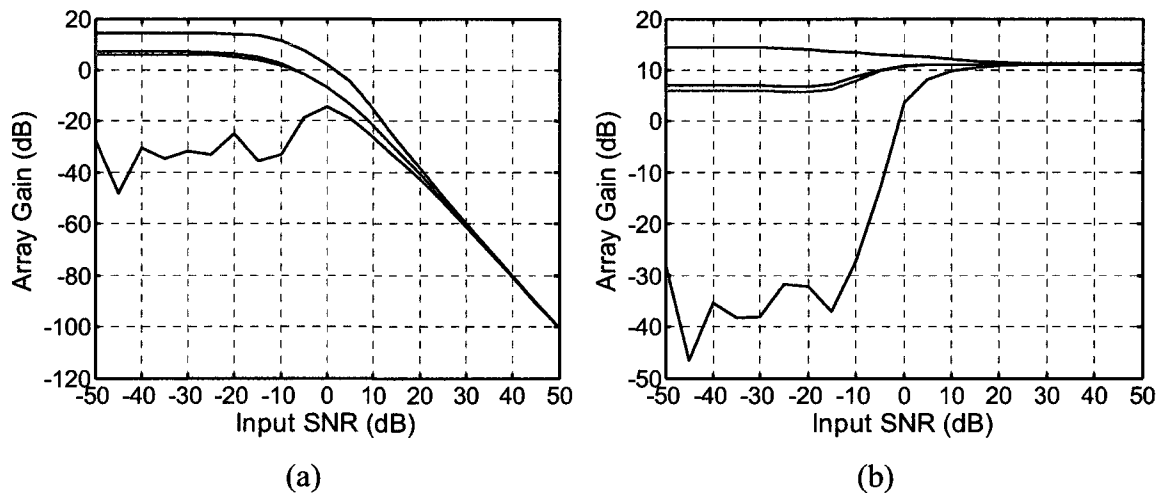


Figure 7.39. Rectangular array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the GO estimator and MVDR. (a) Without reverberation; (b) With reverberation.

The spread of the resulting gain in SRR is displayed in Fig. 7.40. The average gain in SRR is 8.7 dB at low input SNR, similar as to when the GO estimator is used without the MVDR beamformer. However, the gain in SRR decreases to a value of 3.4 dB at higher input SNR, while instead it increased to a value of 11.5 dB for the GO estimator used without MVDR.

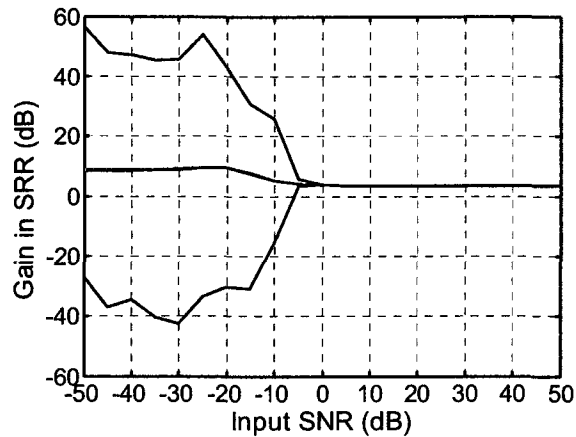


Figure 7.40. Rectangular array, dipole source, diffuse noise, and reverberation. Mean (green), median (red), minimum (blue), and maximum (blue) values of the resulting gain in SRR for 10000 realizations with the GO estimator and MVDR.

As comparison, the spread in the array gain with the exact source signal as the source model in the MVDR beamformer, for scenarios without and with reverberation is shown in Fig. 7.41(a) and Fig. 7.41(b), respectively. Without reverberation, the optimum array gain of 15.06 dB is obtained for all input SNR. With reverberation, the optimum array gain is obtained at low input SNR but not at higher input SNR, where the array gain is 13.3 dB. We attribute this behavior to the correlation between the direct signal and the reverberation component [Van Veen 1988; p.17].

The spread of the resulting gain in SRR is displayed in Fig. 7.42. For an input SNR of -50 dB, the resulting gain in SRR is similar to the one in Table 7.6 for the optimum source model. The resulting gain in SRR then increases with input SNR, before decreasing to a value of 3.4 dB for input SNR above 20 dB. This value of 3.4 dB is the same as when the GO estimator is used to obtain a source model then used with an MVDR beamformer. We attribute this behavior again to the correlation between the direct signal and the reverberation component.

As a side note, for the rectangular array geometry, the matrix  $\mathbf{R}_{pp}(k)$  had a full rank of 32, even for an input SNR of 50 dB.  $\mathbf{R}_{pp}(k)$  could then be inverted to find the MVDR beamforming weight vector so the plots could extend to input SNR of 50 dB, as opposed to the linear array case where the plots were limited to input SNR up to 35 dB.

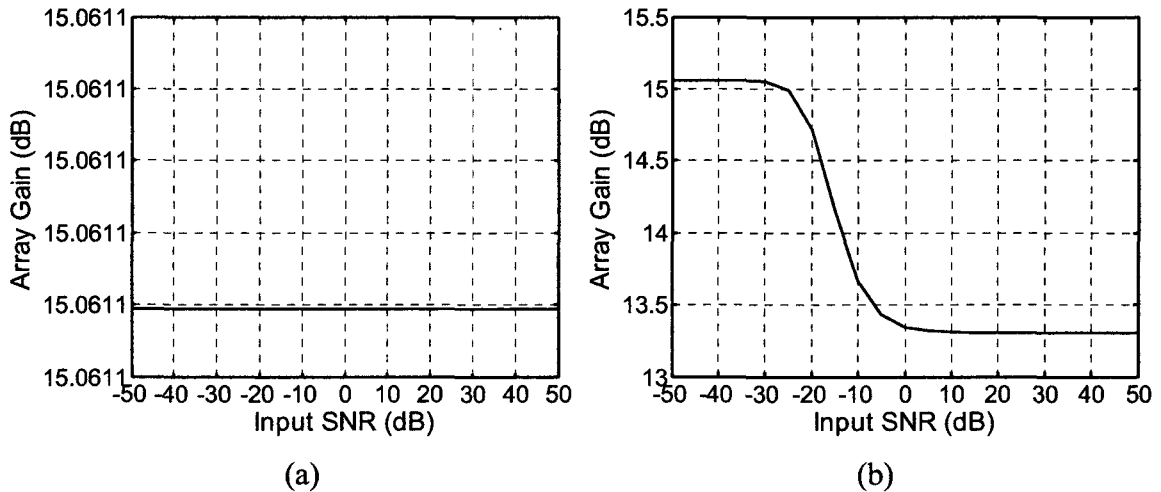


Figure 7.41. Rectangular array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the exact source signal and MVDR. (a) Without reverberation; (b) With reverberation.

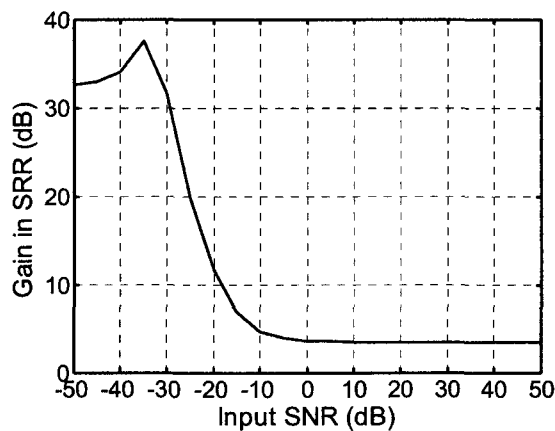


Figure 7.42. Rectangular array, dipole source, diffuse noise, and reverberation. Mean (green), median (red), minimum (blue), and maximum (blue) values of the resulting gain in SRR for 10000 realizations with the exact source signal and MVDR.

## 7.5.2 Rectangular array and monopole source

Again, the proposed beamforming method should still perform properly when the source is a monopole. The simulation is done with the rectangular array and a monopole source located at  $(x, y, z) = (2, 1, 0)$ . The exact source mode strength is  $s_{0,0}(k) = 1$ . The array as well as the source location and directivity were shown in Fig. 7.24.

The optimum array gain, the white-noise gain, and the resulting gain in SRR are given in Table 7.9. They are the same for conventional beamforming since the source is a monopole

Table 7.9. Array gain, white-noise gain, and gain in SRR for monopole source and rectangular array.

Source model	Array gain	W.-N. gain	Gain in SRR
	dB		
Exact source signal	15.12	15.01	10.12
Point source (conv. beamf.)	15.12	15.01	10.12

The general behavior of the different approaches is the same as in Sub-Section 7.5.1. Here are the details.

### ***Gain-optimized estimator***

The spread in the array gain with the GO estimator without and with reverberation is shown in Fig. 7.43(a) and Fig. 7.43(b), respectively. The two plots are very similar. When the GO estimator is used, the array gain at high input SNR, without reverberation, is 15 dB and, in the presence of reverberation, is 14.4 dB. As a reminder, the optimum array gain is 15.1 dB.

The spread of the resulting gain in SRR is displayed in Fig. 7.44. The average gain in SRR is above 0 dB for all input SNR. However the spread is large at low input

SNR. For input SNR above approximately -7 dB, the gain in SRR remains above 0 dB for the 10000 realizations. The spread becomes small above 0 dB input SNR. At high input SNR, the resulting gain in SRR is 5.8 dB compared to the 10.1 dB if the point source model (optimum) is used. Values with the GO estimator and other approaches are compiled in Tables 7.10 and 7.11.

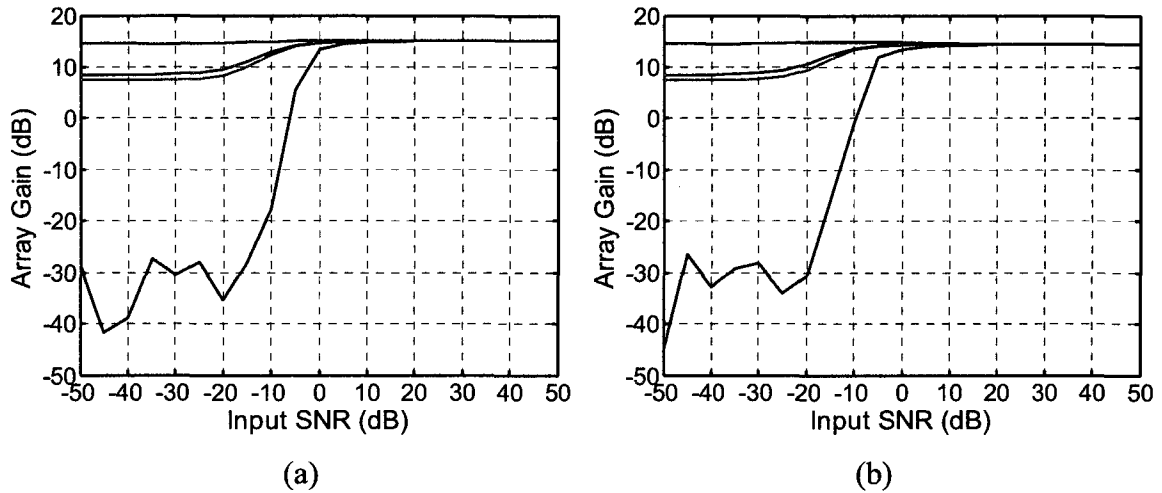


Figure 7.43. Rectangular array, monopole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the GO estimator. (a) Without reverberation; (b) With reverberation.



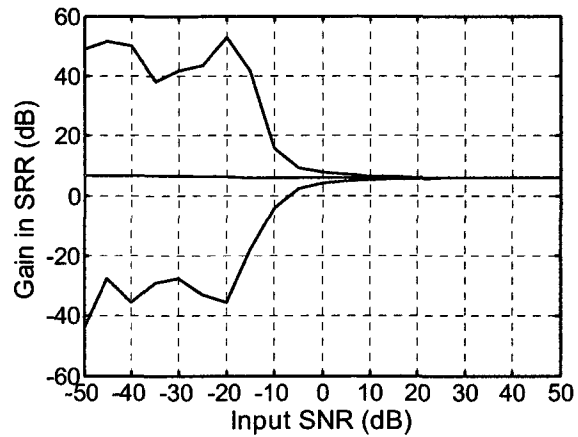


Figure 7.44. Rectangular array, monopole source, diffuse noise, and reverberation. Mean (green), median (red), minimum (blue), and maximum (blue) values of the resulting gain in SRR for 10000 realizations with the GO estimator.

Table 7.10. Array gain and gain in SRR at high input SNR for monopole source and rectangular array.

Approach	Array gain	Array gain	Gain in SRR
	No reverb	Reverb	Reverb
	dB		
GO estimator	15	14.4	5.8
LS	15.1	13.2	1.8
Microphone output	~ 15	~ 12	~ -2
GO + MVDR	-	10.7	3.5

Table 7.11. Minimum input SNR for which the gain in SRR is 0 dB or above for all 10000 realizations, for monopole source and rectangular array.

Approach	Input SNR
	dB
GO estimator	-7
LS	0
Microphone output	-
GO + MVDR	-10

### ***Least squares estimator***

The spread in the array gain with the LS estimator without and with reverberation is shown in Fig. 7.45(a) and Fig. 7.45(b), respectively. The two plots are very similar but the array gain at high input SNR is lower when there is reverberation, at a value of 13.2 dB while it is 15.1 dB without reverberation. Compared to Fig. 7.43(a) and Fig. 7.43(b), the average array gain is lower with the LS estimator than for the GO estimator at lower input SNR. This is also the case for the maximum array gain value over the 10000 realizations. With reverberation, at 13.2 dB the array gain at high input SNR with the LS estimator is lower than with the GO estimator, which is 14.4 dB.

The spread of the resulting gain in SRR is displayed in Fig. 7.46. The average gain in SRR is above 0 dB for all input SNR. As with the GO estimator, the spread is large at low input SNR. The average gain in SRR is lower than with the GO estimator, for all input SNR. For input SNR above approximately 0 dB, the gain in SRR remains above 0 dB for the 10000 realizations, while it was above -7 dB input SNR with the GO estimator. The spread becomes small above 10 dB input SNR, while it was above 0 dB with the GO estimator. At high input SNR, the resulting gain in SRR is 1.8 dB, compared to 5.8 dB with the GO estimator.

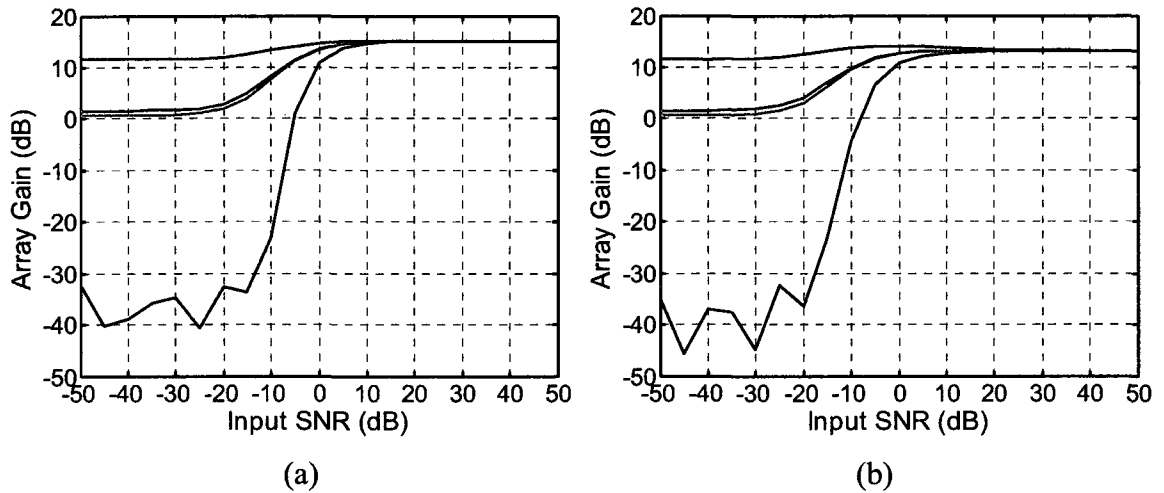


Figure 7.45. Rectangular array, monopole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the LS estimator. (a) Without reverberation; (b) With reverberation.

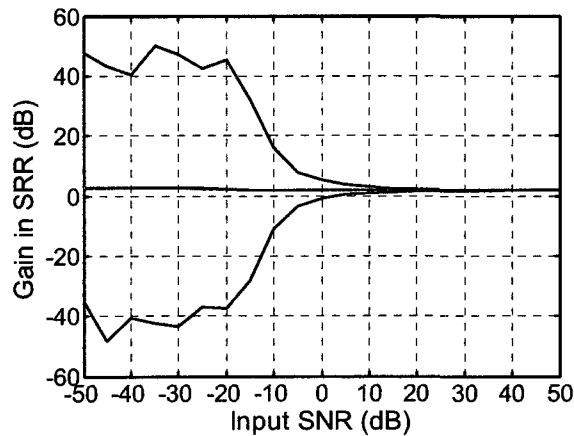


Figure 7.46. Rectangular array, monopole source, diffuse noise, and reverberation. Mean (green), median (red), minimum (blue), and maximum (blue) values of the resulting gain in SRR for 10000 realizations with the LS estimator.

### ***Microphone output vector as the source model***

The spread in the array gain with the microphone output vector used as the source model, for the scenarios without and with reverberation is shown in Fig. 7.47(a) and Fig. 7.47(b), respectively. The two plots are fairly similar but the array gain at high input SNR is lower when there is reverberation. In both cases, the average array gain is below 0 dB for input SNR below -20 dB.

The spread of the resulting gain in SRR is displayed in Fig. 7.48. The average gain in SRR is between 0 and 0.1 dB for input SNR between -50 and -40 dB and below 0 dB for higher input SNR. As with the GO estimator, the spread in gain in SRR is large at low input SNR. It becomes small above 10 dB input SNR, but around a gain in SRR value of -2 dB.

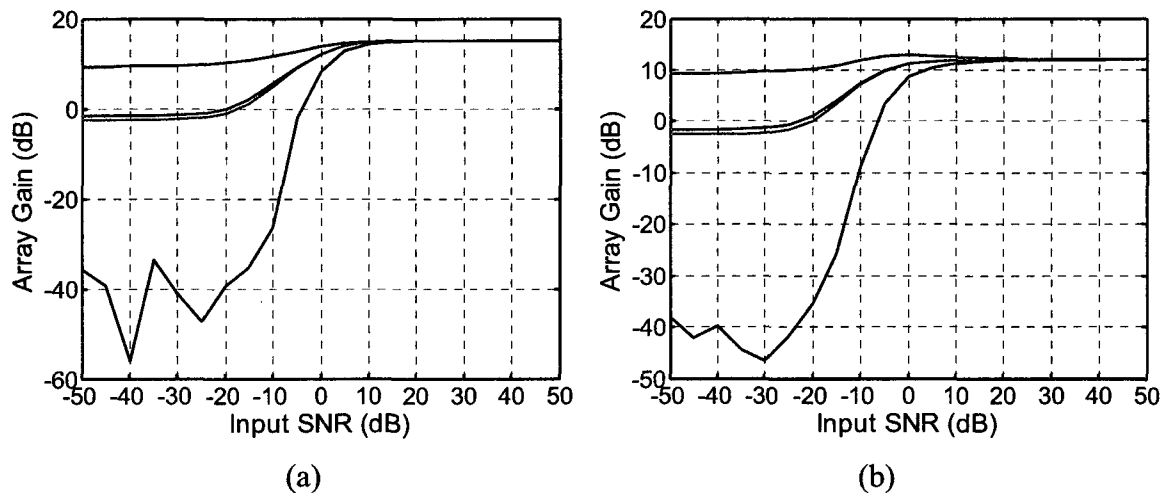


Figure 7.47. Rectangular array, monopole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the microphone output vector as the source model. (a) Without reverberation (b) With reverberation.

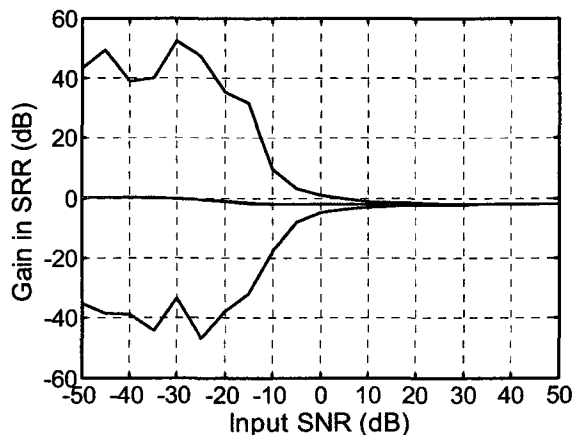


Figure 7.48. Rectangular array, monopole source, diffuse noise, and reverberation. Mean (green), median (red), minimum (blue), and maximum (blue) values of the resulting gain in SRR for 10000 realizations with the microphone output vector as the source model.

### ***Estimated source model and MVDR***

The spread in the array gain with the GO estimator used to obtain a source model which is then used with an MVDR beamformer, for the scenarios without and with reverberation is shown in Fig. 7.49(a) and Fig. 7.49(b), respectively. Without reverberation, we observe the same behavior as we did for the linear array, in Sub-Section 7.3.1, and for the rectangular array and the dipole source in Sub-Section 7.5.1: a large spread of the array gain at low input SNR and a linear decrease of the array gain with an increase in the input SNR at higher input SNR. With reverberation, there is a large spread at low input SNR, with an average similar as to when the GO estimator and our proposed beamforming method are used without MVDR. At high input SNR, the array gain is 10.7 dB, compared to 14.4 dB when the GO estimator is used without MVDR.

The spread of the resulting gain in SRR is displayed in Fig. 7.50. The average gain in SRR is 6.5 dB at -50 dB input SNR, similar to when the GO estimator is used without MVDR. While without MVDR the average of the gain in SRR decreases down to 5.9 dB between -30 to -10 dB input SNR, the average of the gain in SRR with MVDR

presents a bump up to 15.5 dB at -20 dB input SNR. The average of the gain in SRR then goes down to 3.5 dB above -10 dB input SNR. The resulting gain in SRR is 5.8 dB when the GO estimator is used without MVDR.

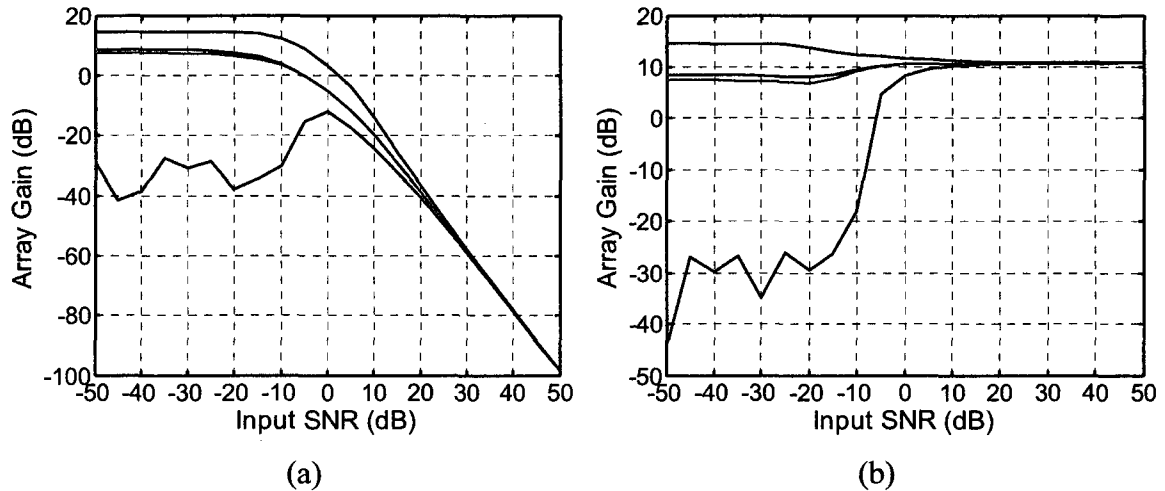


Figure 7.49. Rectangular array, monopole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the GO estimator and MVDR. (a) Without reverberation; (b) With reverberation.

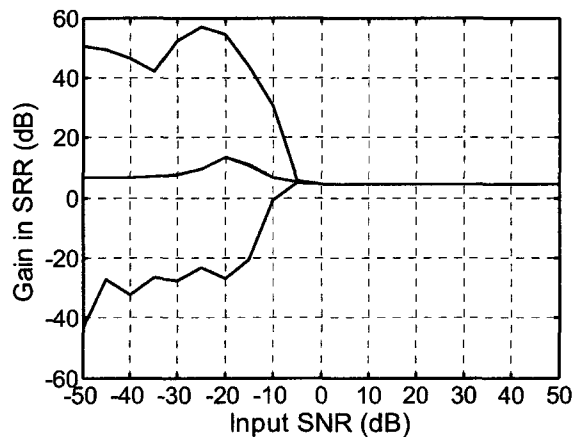


Figure 7.50. Rectangular array, monopole source, diffuse noise, and reverberation. Mean (green), median (red), minimum (blue), and maximum (blue) values of the resulting gain in SRR for 10000 realizations with the GO estimator with MVDR.

### 7.5.3 Concluding remarks: Diffuse noise + reverberation

Overall, the performance is decreased by reverberation but the proposed beamforming method still provides a reasonable performance, even though the mode strengths are estimated from the microphone output vector.

Of the four approaches (GO estimator, LS, microphone output as the source model, and GO with MVDR), reverberation causes less penalty with the GO estimator. For example, at high input SNR, the array gain and the resulting gain in SRR are higher with the GO estimator. Also, the resulting gain in SRR is above 0 dB for all 10000 realizations at a lower input SNR with the GO estimator than with the other approaches.

The GO estimator with MVDR beamforming did not show the linear decrease in the array gain with increasing input SNR in the simulations with reverberation, unlike the simulations performed without reverberation in a previous section. However, the performance of the GO estimator with MVDR beamforming is still worse than for the GO estimator without MVDR.

## 7.6 Simulations: Diffuse noise + an interferer

### 7.6.1 Simulation of an interferer

An interferer is treated as a source, with its own location and its own directivity. The interferer is assumed to be uncorrelated with the source on which we are beamforming (simply referred to as the “source”) and uncorrelated with the noise.

Similar as it is for the source, the component of the interferer at the output of the microphones is expressed in a vector  $\mathbf{s}_I(k)$  whose spherical harmonic decomposition is

$$\mathbf{s}_I(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm,I}(k) \mathbf{Y}_{n,I}^m(k). \quad (7.44)$$

The subscript “I” indicates that this is for the interferer.  $s_{nm,I}(k)$  is the mode strength of the spherical harmonic mode  $(n, m)$  for the interferer in the frequency bin  $k$ . The mode vector for the source was defined with Eq. (4.11), here  $\mathbf{Y}_{n,I}^m(k)$  is the mode vector for the interferer and is defined as

$$\mathbf{Y}_{n,I}^m(k) = [h_n^{(2)}(\kappa_k r_{1,I}) Y_n^m(\theta_{1,I}, \phi_{1,I}) \dots h_n^{(2)}(\kappa_k r_{L,I}) Y_n^m(\theta_{L,I}, \phi_{L,I})]^T. \quad (7.45)$$

To avoid making the notation of the interferer mode vector needlessly heavier, the  $l^{\text{th}}$  microphone is located at  $(r_{l,I}, \theta_{l,I}, \phi_{l,I})$  in a new spherical coordinate system with the interferer at its origin.

## 7.6.2 Simulation description

The source frequency is 996.1 Hz, the center of a frequency bin (for more details please refer to the end of Sub-Section 7.4.2).

The simulations are done with the rectangular array for a few different values of input signal-to-interference ratio (SIR). For each value of input SIR, we vary the input SNR, and for each input SNR, 10000 realizations of the noise vector  $\mathbf{e}(k)$  are used. Since  $\mathbf{s}_I(k)$  is deterministic, as is  $\mathbf{s}(k)$ , it only needs to be computed once.

For each realization, the received microphone output vector is given by

$$\mathbf{p}(k) = \mathbf{s}(k) + \mathbf{s}_I(k) + \mathbf{e}(k). \quad (7.46)$$

The average interferer power spectral density over the  $L$  microphones of the array is

$$\sigma_{s_I}^2(k) = \mathbf{s}_I^H(k) \mathbf{s}_I(k) / L. \quad (7.47)$$

The input signal-to-interference ratio (averaged over the  $L$  microphones) is

$$\text{SIR}_{\text{in}}(k) = \frac{\sigma_s^2(k)}{\sigma_{s_I}^2(k)}. \quad (7.48)$$

The normalized interferer cross-power spectral density matrix is



$$\mathbf{R}_{s_1 s_1}(k) = \mathbf{s}_1(k) \mathbf{s}_1^H(k) / \sigma_{s_1}^2(k). \quad (7.49)$$

The microphone output cross-power spectral density is

$$\mathbf{R}_{pp}(k) = \sigma_s^2(k) \mathbf{R}_{ss}(k) + \sigma_{s_1}^2(k) \mathbf{R}_{s_1 s_1}(k) + \sigma_e^2(k) \mathbf{R}_{ee}(k), \quad (7.50)$$

where we have used the assumption that the source, the interferer, and the noise are mutually uncorrelated.

With the MVDR beamformer,  $\mathbf{R}_{pp}(k)$  is part of the computation to find the beamforming weight vector so the interferer is taken into account in the optimization. For all the other approaches, the optimization criterion remains the improvement in signal-to-noise ratio in the presence of diffuse noise, and the optimization criterion is not to maximize the improvement in signal-to-interference ratio.

The beamforming weight vectors for the different approaches are computed as they were in the section on simulations with diffuse noise. Because we have not included an interferer until this point in the thesis, the estimation approach described in Sub-Section 6.3.6 has not been investigated yet: least-squares with mode vectors for both the source and the interferer positions, i.e. LSSI for short. For this approach, the matrix  $\mathbf{Y}(k)$  is built from Eq. 6.42 with 8 modes for the source location and 8 modes for the interferer location, that is

$$\mathbf{Y}_s(k) = [\mathbf{Y}_0^0(k), \mathbf{Y}_1^{-1}(k), \mathbf{Y}_1^0(k), \mathbf{Y}_1^{-1}(k), \mathbf{Y}_2^{-2}(k), \mathbf{Y}_2^{-1}(k), \mathbf{Y}_2^0(k), \mathbf{Y}_2^1(k)] \quad (7.51)$$

and

$$\mathbf{Y}_I(k) = [\mathbf{Y}_{0,I}^0(k), \mathbf{Y}_{1,I}^{-1}(k), \mathbf{Y}_{1,I}^0(k), \mathbf{Y}_{1,I}^{-1}(k), \mathbf{Y}_{2,I}^{-2}(k), \mathbf{Y}_{2,I}^{-1}(k), \mathbf{Y}_{2,I}^0(k), \mathbf{Y}_{2,I}^1(k)] . \quad (7.52)$$

The sixteen mode strengths are estimated with Eq. (6.46). The first eight are then used to build the source model with Eq. (6.47).

In addition to observing the array gain and white-noise gain, we also compute the gain in signal-to-interference ratio, defined as

$$G_{\text{SIR}}(k) = \frac{\mathbf{w}^H(k) \mathbf{R}_{\text{ss}}(k) \mathbf{w}(k)}{\mathbf{w}^H(k) \mathbf{R}_{\text{s}_1 \text{s}_1}(k) \mathbf{w}(k)}, \quad (7.53)$$

which is equivalent to

$$G_{\text{SRR}}(k) = \frac{|\mathbf{w}^H(k) \mathbf{s}(k)|^2 / \sigma_s^2(k)}{|\mathbf{w}^H(k) \mathbf{s}_1(k)|^2 / \sigma_{s_1}^2(k)}. \quad (7.54)$$

## 7.7 Simulations results: Diffuse noise + an interferer

This section presents simulation results for a source in the presence of an interferer. Because of the number of pages required to present the results, only two scenarios are included:

- Rectangular array, dipole source, and monopole interferer;
- Rectangular array, dipole source, and beamforming position errors, without and with an interferer.

### 7.7.1 Rectangular array, dipole source, and monopole interference

The array is still composed of 32 microphones located on the perimeter of a rectangle. The source is a dipole located at  $(x, y, z) = (2, 1, 0)$  with its axis parallel to the  $x$ -axis, as before. The exact source mode strengths are  $s_{1,-1}(k) = 1$  and  $s_{1,1}(k) = -1$ . The interferer is a monopole, located at  $(x, y, z) = (-1, 0.5, -0.2)$ . The exact mode strength for the interferer is  $s_{0,0,1}(k) = 1$ . The array, the source location and directivity, as well as the interferer location and directivity are displayed in Fig. 7.51.

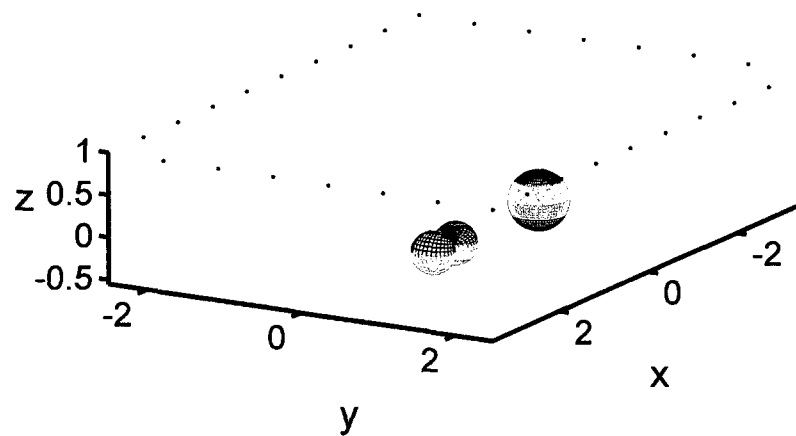


Figure 7.51. Rectangular array, dipole source, and monopole interferer.

The optimum array gain and the gain for conventional beamforming are given in Table 7.12 along with the resulting white-noise gain and gain in SIR. As it was the case in Section 7.3.2, conventional beamforming provides a low array gain for this array geometry and a dipole source at its location. Conventional beamforming also provides a fairly low gain in SIR.

Table 7.12. Array gain, white-noise gain, and gain in SIR for dipole source, monopole interferer, and rectangular array.

Source model	Array gain	W.-N. gain	Gain in SIR
	dB		
Exact source signal	15.06	15.01	17.14
Point source (conv. beamf.)	-1.55	-1.66	4.47

### ***Approaches without MVDR beamforming***

We will first analyze together the results for the approaches without MVDR: gain-optimized estimator (GO), least-squares estimator with modes for the source location only (LS), least-squares estimator with modes for both the source and the interferer locations (LSSI), and the microphone output vector as the source model.

Figures 7.52 to 7.55 shows the spread in the array gain versus the input SNR for 4 different values of SIR: -20, -10, 0, and 10 dB, for the 4 approaches.

At high input SIR, the array gain results should resemble the previous results when there was only diffuse noise. With the GO estimator, we see that with an input SIR of 10 dB, the maximum and mean of Fig. 7.52(d) are very similar to what they were in Fig. 7.31(a), which was for a dipole source in the presence of diffuse noise only. The minimum array gain crosses 0 dB between -10 and 0 dB input SNR, as it does in Fig. 7.31(a). In the same way, for the LS estimator, Fig. 7.53(d) is comparable to Fig. 7.35(a) and for the microphone output vector as the source model, Fig. 7.55(d) is comparable to Fig. 7.37(a).

On the range of input SIR from -20 to 10 dB, for input SNR below approximately -25 dB, the noise is dominant over the interferer. Thus the mean array gain does not vary much with the SIR level. This is shown for the four approaches in Fig. 7.56(a) for an input SNR of 50 dB. Figure 7.56(b) shows the case for high input SNR. In this case, the interferer is dominant over the noise. As expected, the array gain increases with the input SIR. In other words, for a fixed source power, the array gain increases as the interferer power decreases. This is because a lower interferer power causes a smaller error in the mode strengths estimation.

To quantify the middle region of the plots, Tables 7.13 and 7.14 indicate the input SNR necessary for the array gain to be above or equal to 0 dB and above or equal to 10 dB, respectively, for all 10000 realizations. A lower input SNR is preferable. The approach with the microphone output as the source model performs the worst, followed by the LS estimator. At high input SIR, the GO estimator and the LSSI estimator are comparable. The advantage is for the LSSI estimator at low input SIR.

An additional explanation is necessary for the LSSI estimator results. Figure 7.56 shows that for this approach at a given input SNR (-50 dB on Fig. 7.56(a) and 50 dB on Fig. 7.56(b)), the mean array gain does not appear to change with the input SIR. Furthermore, the four plots of Fig. 7.54 are identical.

The first bit of information needed to understand this behavior comes from Sub-Section 6.3.5 where it was explained how, with least-squares estimation, the strength estimation of a given mode is not biased by other modes. This is the case here also because modes for both the source and the interferer are included. So even if the interference strength is changed, this does not change the estimation of the mode strengths for the source. The second bit of information is about the 10000 noise vectors  $\mathbf{e}(k)$ . For each input SIR and input SNR, the 10000 noise vectors  $\mathbf{e}(k)$  are created with the same 10000 random vectors  $\mathbf{e}_w$ . Only  $\sqrt{\sigma_e^2(k)}$  in Eq. (7.4) is changed with the input SNR.

We can now combine these two bits of information. For a given input SNR, the 10000 vectors  $\mathbf{e}(k)$  lead to 10000 array gain values. Because the strength of the interferer does not affect the estimation of the mode strengths for the source, the change in input SIR does not change the 10000 array gain values. This explains why the four plots of Fig. 7.54 are identical.

Table 7.13. Approximate input SNR for which all realizations provide an array gain above or equal to 0 dB. (Bold font indicates the best value for each input SIR)

Approach	Input SIR			
	-20 dB	-10 dB	0 dB	10 dB
Input SNR (dB)				
GO estimator	5	2	<b>-2</b>	<b>-5</b>
LS estimator	-	8	-2	-4
LSSI	<b>-2</b>	<b>-2</b>	-2	-2
Mic. output as source model	-	20	-2	-4

Table 7.14. Approximate input SNR for which all realizations provide an array gain above or equal to 10 dB. (Bold font indicates the best value for each input SIR)

Approach	Input SIR			
	-20 dB	-10 dB	0 dB	10 dB
Input SNR (dB)				
GO estimator	-	13	<b>0</b>	<b>-1</b>
LS estimator	-	-	3	1
LSSI	<b>3</b>	<b>3</b>	3	3
Mic. output as source model	-	-	8	3

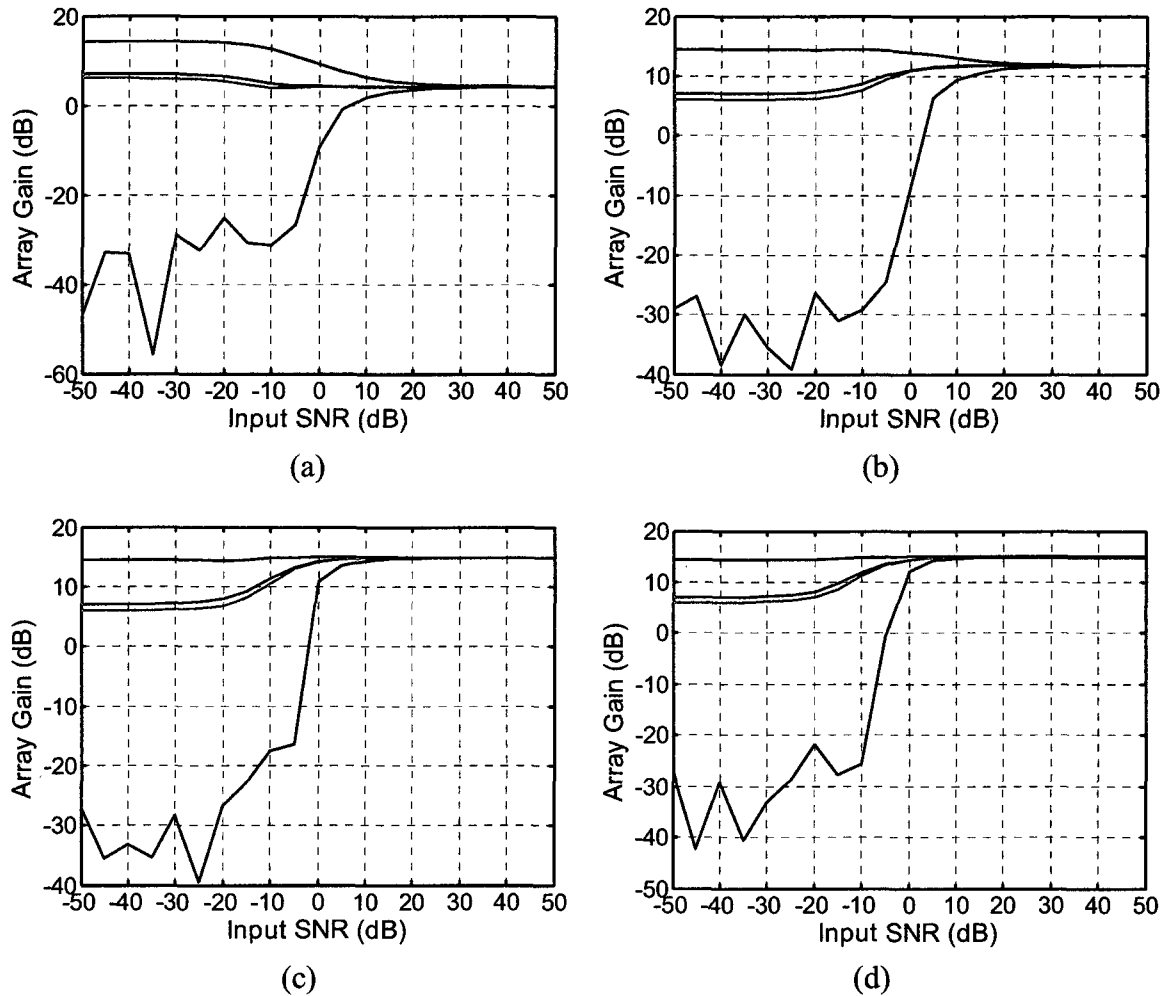


Figure 7.52. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the GO estimator for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.

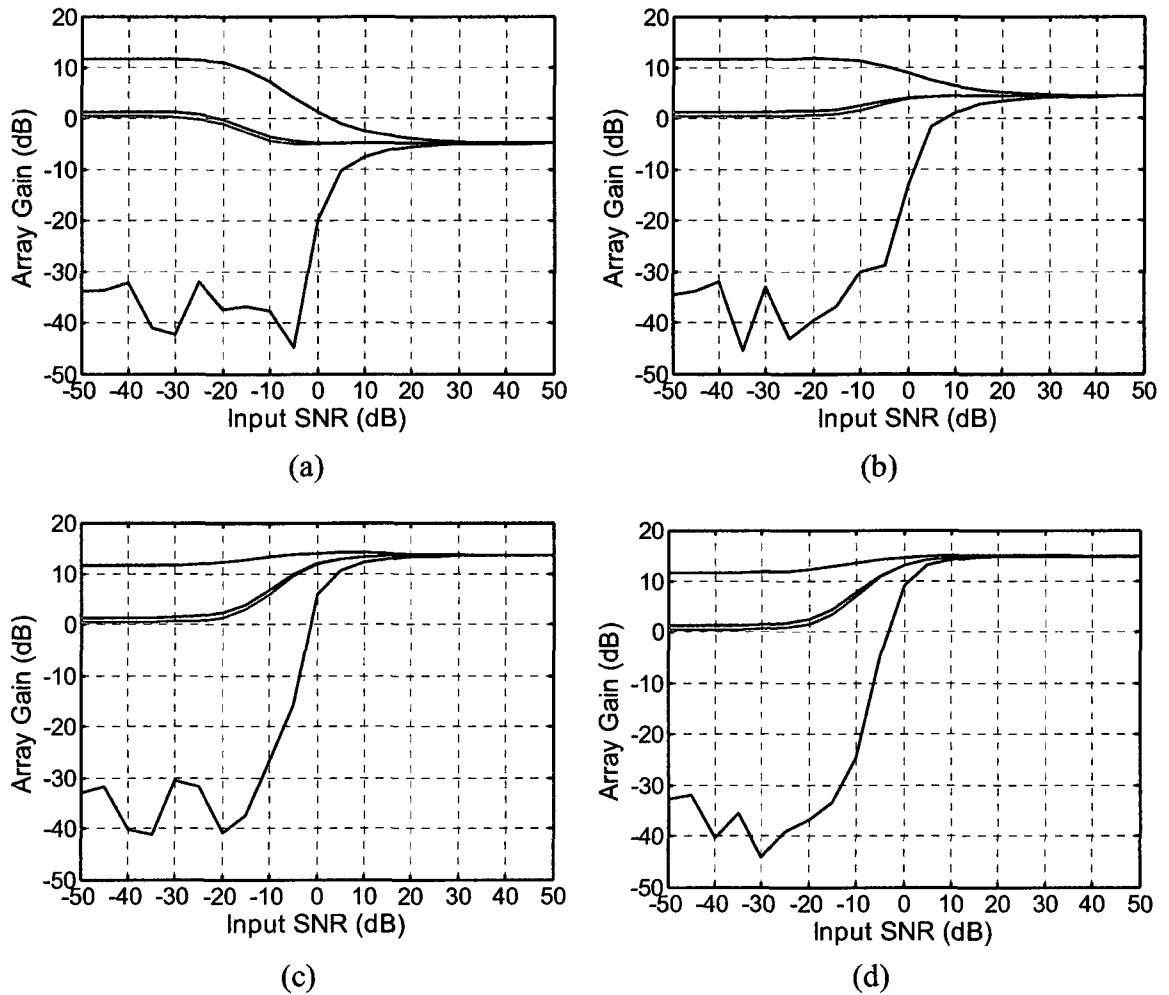


Figure 7.53. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the LS estimator for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.



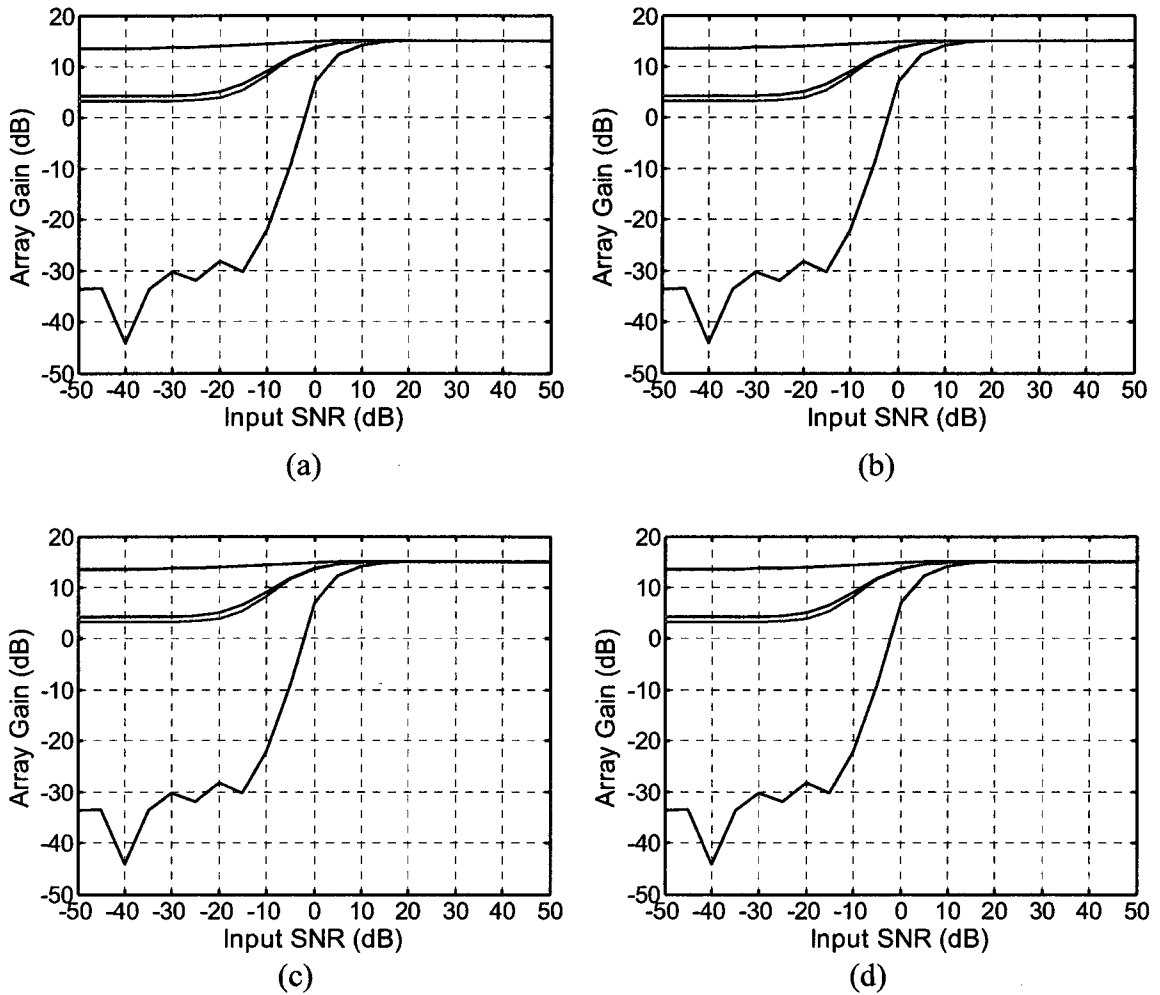


Figure 7.54. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with LSSI for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.

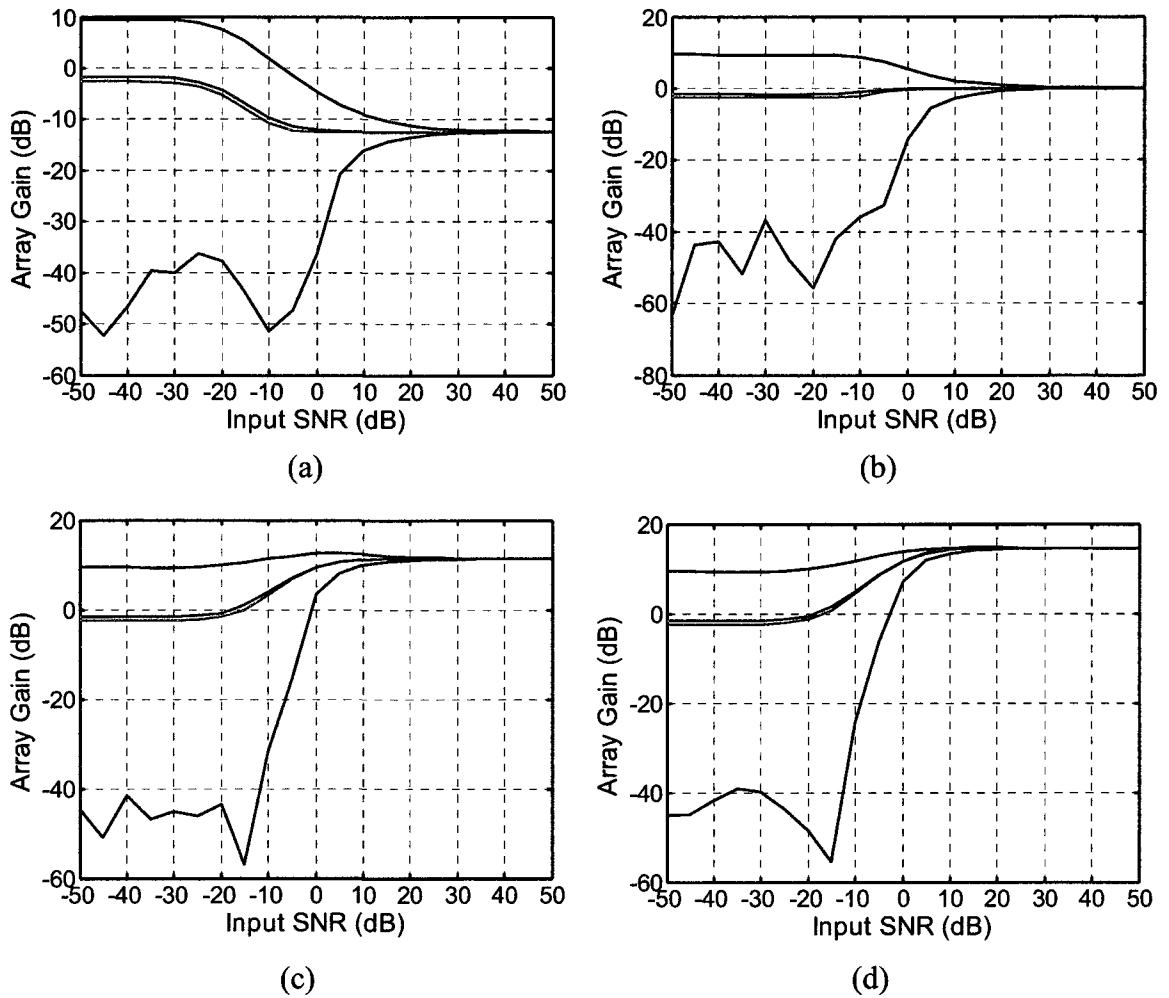


Figure 7.55. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the microphone output vector as the source model for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.

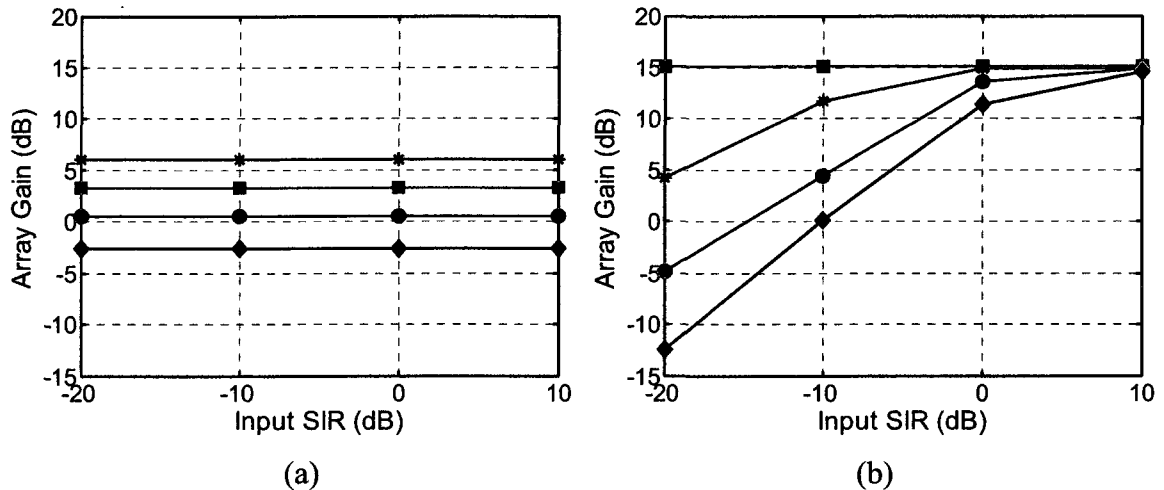


Figure 7.56. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean array gain over the 10000 realizations, for input SNR of (a) -50 dB; (b) 50 dB. (\*) GO estimator, (●) LS estimator, (■) LSSI, (◆) microphone output as the source model.

For the same four mode strength estimation approaches, the gain in SIR is shown in Fig. 7.57 to 7.60. Without MVDR beamforming, the optimization is with respect to the noise field, as characterized by  $\mathbf{R}_{ee}(k)$ , not to maximize the improvement in signal-to-interference ratio. In other words, the beamforming weight vector is computed to maximize the array gain with the estimated signal model, not to maximize the gain in SIR. The gain in SIR is a by-product.

The spread of the gain in SIR remains large for input SNR below 10 dB or more, depending on the approach. On the range of input SIR from -20 to 10 dB, for very low input SNR, the mean gain in SIR does not vary much with the SIR level. This is shown for the four approaches in Fig. 7.61(a) for an input SNR of -50 dB. Figure 7.61(b) shows the case for high input SNR where the gain in SIR increases with the input SIR, except for the LSSI estimator, for the same reason as explained above.

Tables 7.15 and 7.16 indicate the input SNR necessary for the gain in SIR to be above or equal to 0 and above or equal to 10 dB, respectively, for the 10000 realizations. A lower input SNR is preferable here as well. The approach with the microphone output as the source model still performs the worst, followed, once again, by the LS estimator. While at high input SIR, the GO estimator and the LSSI estimator were comparable for

the array gain, the GO estimator has a better performance than the LSSI estimator for the gain in SIR. At low input SIR, the LSSI estimator has the advantage over the GO estimator for the array gain and it also has the advantage for the gain in SIR.

Table 7.15. Approximate input SNR for which all realizations provide a gain in SIR above or equal to 0 dB. (Bold font indicates the best value for each input SIR)

Approach	Input SIR			
	-20 dB	-10 dB	0 dB	10 dB
	Input SNR (dB)			
GO estimator	-	3	<b>-3</b>	<b>-6</b>
LS estimator	-	-	0	-3
LSSI	<b>-2</b>	<b>-2</b>	-2	-2
Mic. output as source model	-	-	20	0

Table 7.16. Approximate input SNR for which all realizations provide a gain in SIR above or equal to 10 dB. (Bold font indicates the best value for each input SIR)

Approach	Input SIR			
	-20 dB	-10 dB	0 dB	10 dB
	Input SNR (dB)			
GO estimator	-	13	<b>-1</b>	<b>-3</b>
LS estimator	-	-	10	3
LSSI	<b>8</b>	<b>8</b>	8	8
Mic. output as source model	-	-	-	15

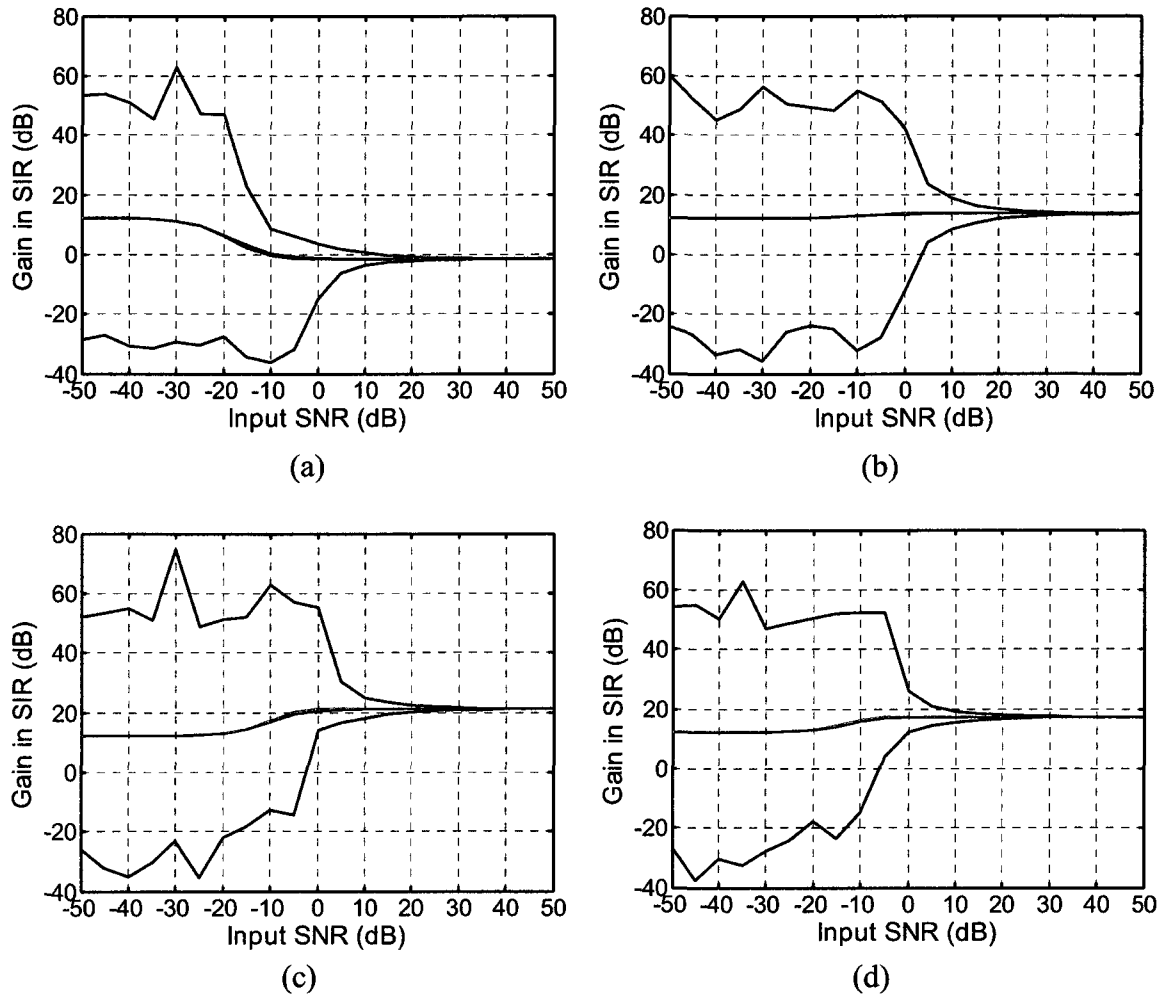


Figure 7.57. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the gain in SIR for 10000 realizations with the GO estimator for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.

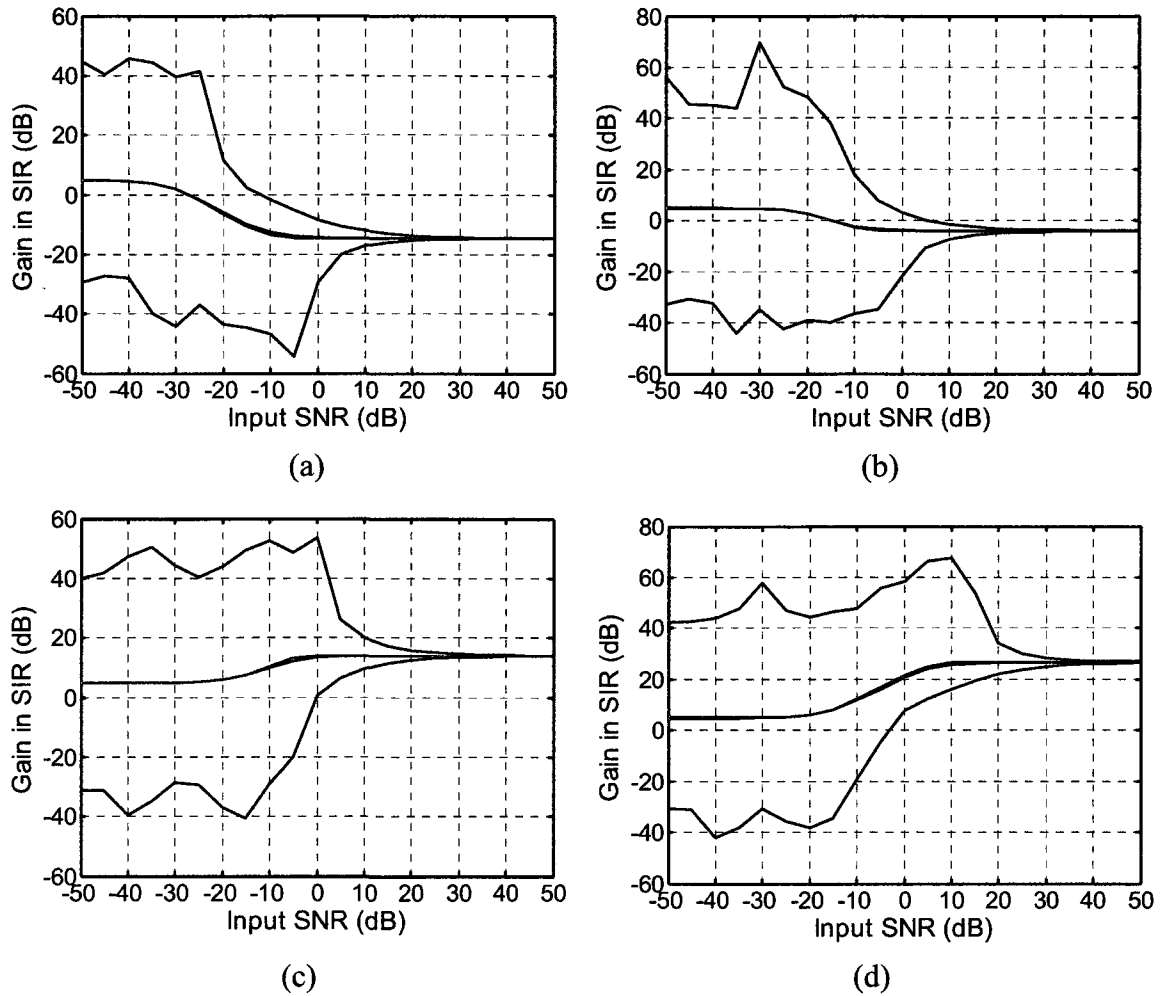


Figure 7.58. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the gain in SIR for 10000 realizations with the LS estimator for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.

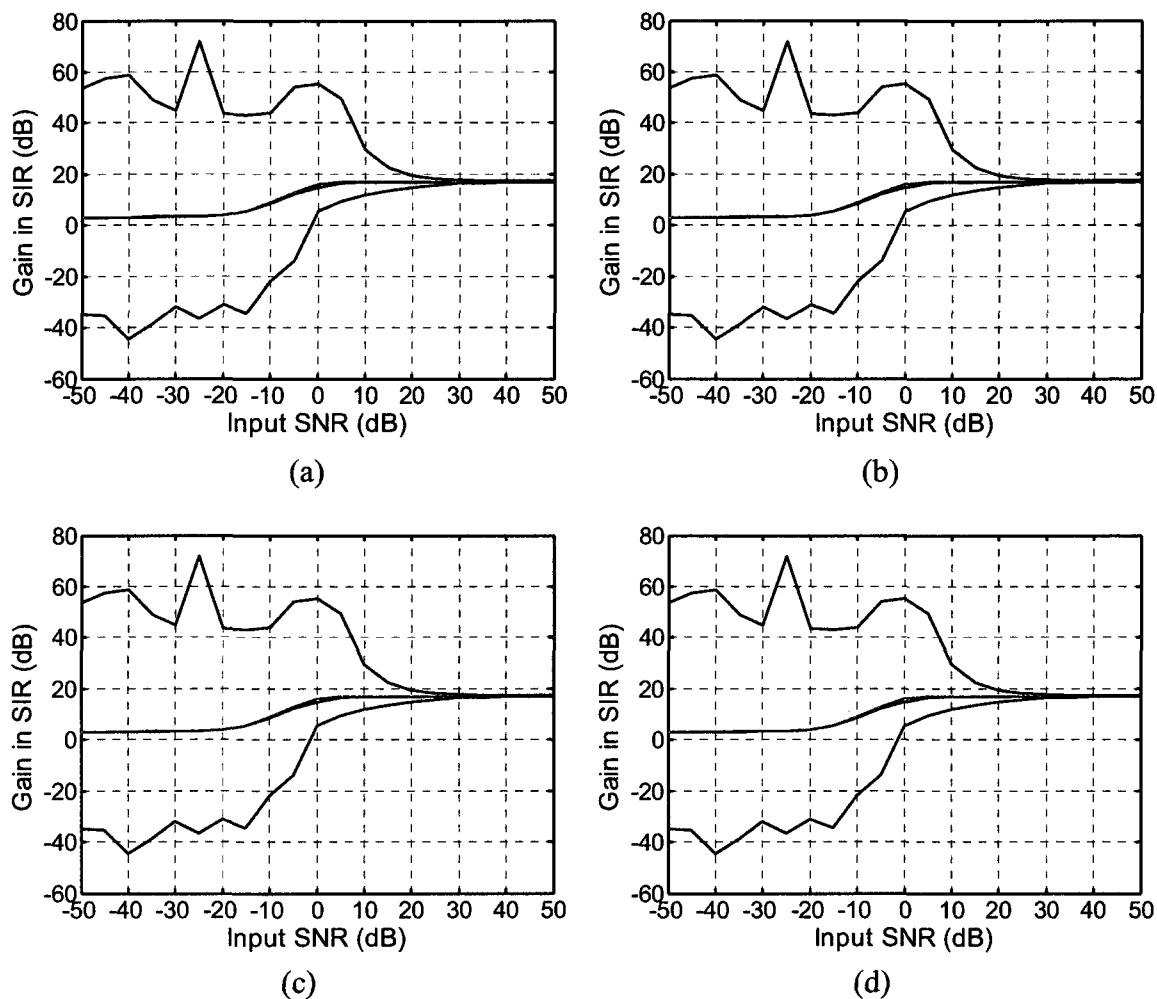


Figure 7.59. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the gain in SIR for 10000 realizations with LSSI for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.

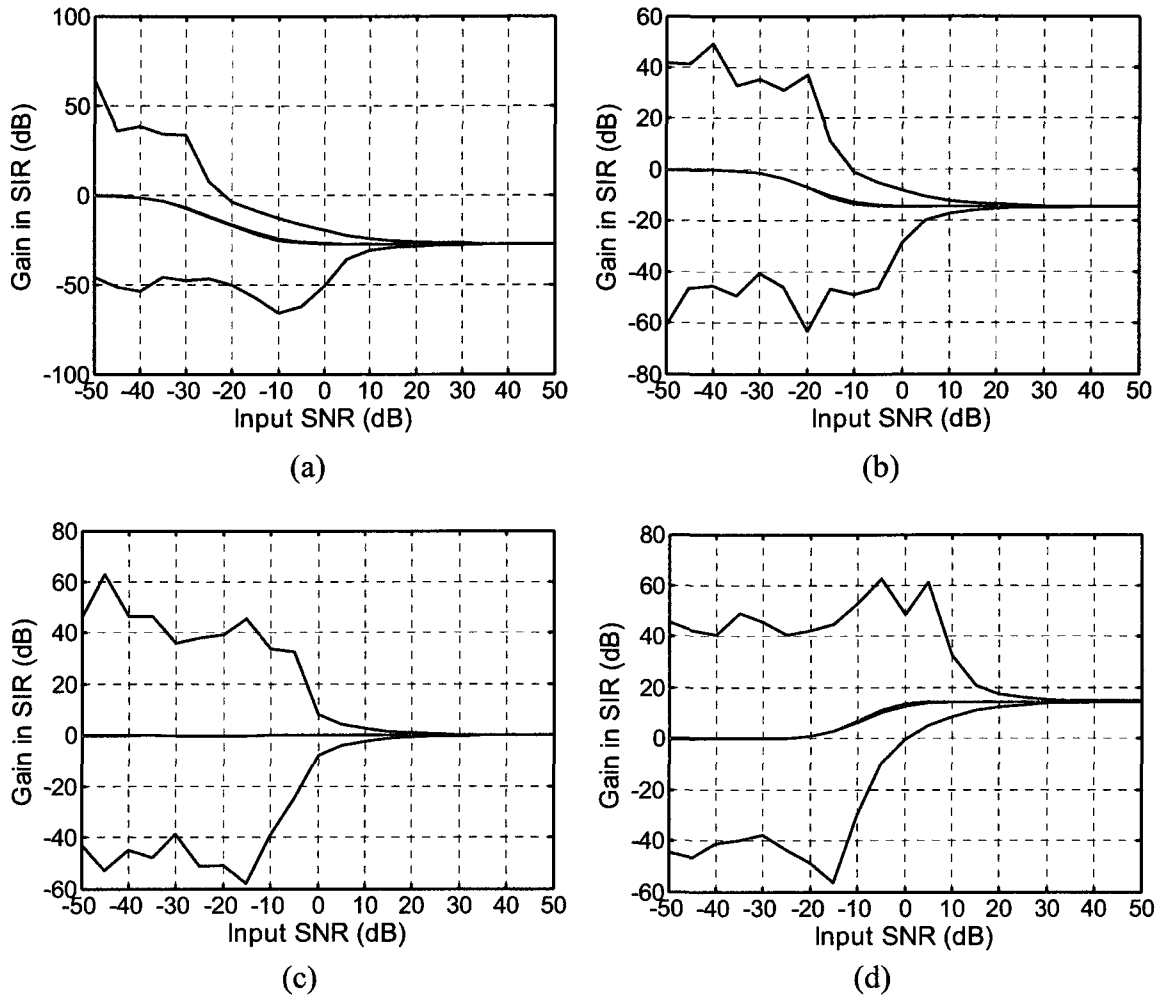


Figure 7.60. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the gain in SIR for 10000 realizations with the microphone output vector as source model for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.



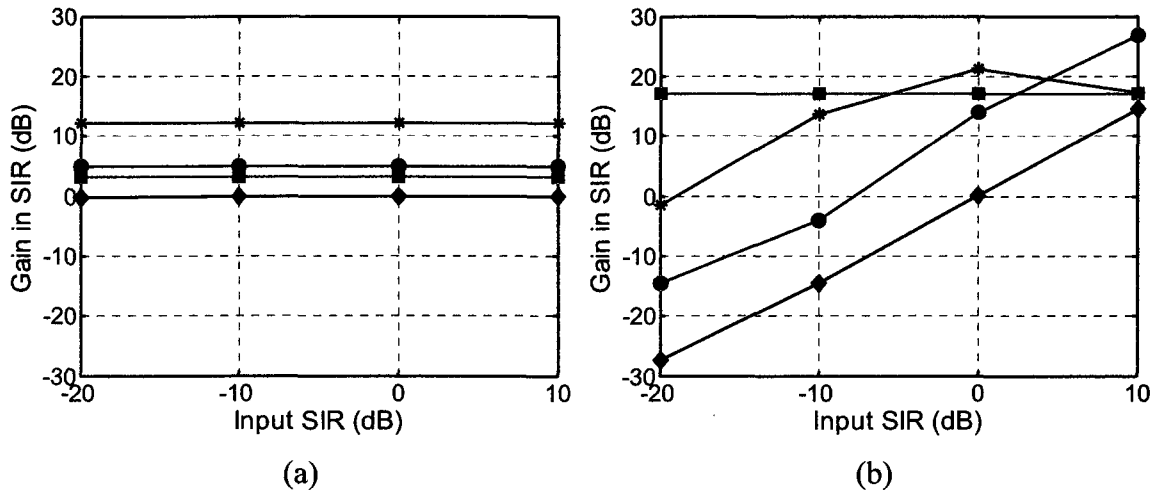


Figure 7.61. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean gain in SIR over the 10000 realizations, for input SNR of (a) -50 dB; (b) 50 dB. (\*) GO estimator, (●) LS estimator, (■) LSSI, (◆) microphone output as the source model.

Figure 7.62 shows the resulting gain in SIR versus the array gain with the GO estimator and an input SIR of 0 dB, for input SNR of -50, -10, 0, and 10 dB. A dot marks the array gain and the resulting gain in SIR for each of the 10000 realizations.

We observe, as in the figures above, the larger spread for lower input SNR. With input SNR of -50 and -10 dB, we note the general trend of higher gain in SIR for higher array gains. For an input SNR of 10 dB, the trend may be opposite: a beamforming weight vector providing an array gain close to the optimum value result in a slightly lower gain in SIR.

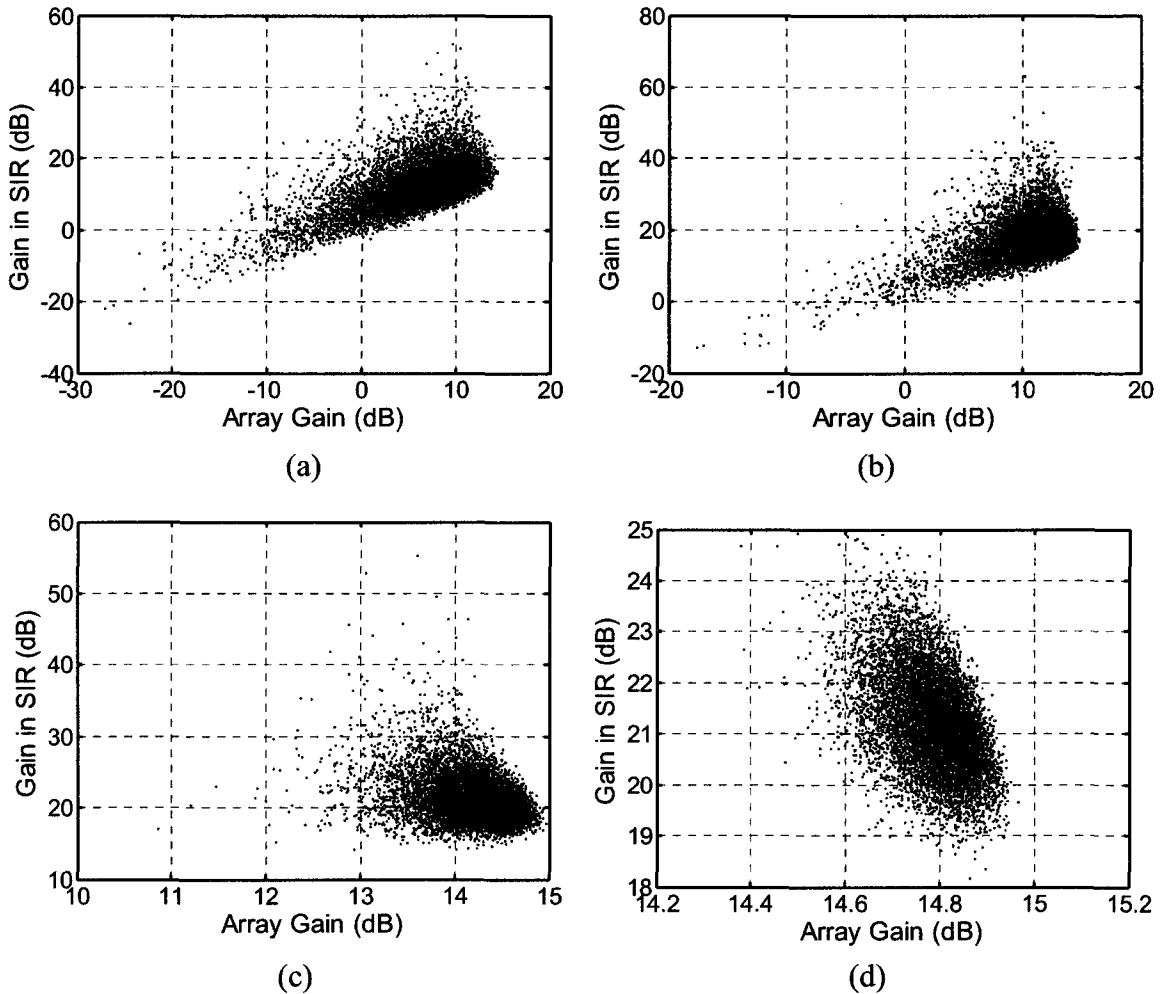


Figure 7.62. Rectangular array, dipole source, diffuse noise, and monopole interferer. Gain in SIR vs. array gain for the 10000 realizations with the GO estimator at an input SIR of 0 dB and for input SNR of (a) -50 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.

### ***Approaches with MVDR beamforming***

Figures 7.63 and 7.64 present the array gain when the GO estimator is used to provide a source model for MVDR beamforming and when the LSSI estimator provides a source model for MVDR beamforming, respectively.

Both figures show the same behavior as seen in Sub-Section 7.3.1: a large spread of the array gain at low input SNR and a linear decrease of the array gain with an

increase in the input SNR, at higher input SNR. With the GO estimator and an MVDR beamformer, the input SNR separating the two regions increases with an increase in input SIR. While the four plots of Fig. 7.63 show differences, the four plots of Fig. 7.64 for the LSSI estimator with an MVDR beamformer are more similar. However, they are not identical as when the LSSI estimator is used without MVDR (see Fig. 7.54). This is because the inverse of the microphone output vector cross-correlation matrix,  $\mathbf{R}_{pp}(k)$ , is part of the computation of the beamforming weight vector with MVDR, and  $\mathbf{R}_{pp}(k)$  changes with the interferer power.

For comparison, Fig. 7.65 shows the array gain when the exact source model is used with an MVDR beamformer. The array gain remains within 0.1 dB of the optimum gain. The shape of the plots can be explained by the respective powers of the interferer and the noise. When the power of the noise is much higher than the power of the interferer (input SNR much lower than input SIR), to minimize the array output power, the MVDR beamformer minimizes the contribution of the noise as much as possible. This is similar to maximizing the array gain. This is the part of the plot slightly above 15.06 dB that is at -50 dB input SNR for an input SIR of -20 dB in Fig. 7.65(a), that extends up to an input SNR of -40 dB for input SIR of -10 dB in Fig. 7.65(b), that extends up to an input SNR of -30 dB for an input SIR of 0 dB in Fig. 7.65(c), and finally that extends up to an input SNR of -20 dB for an input SIR of 10 dB in Fig. 7.65(d).

For input SNR much higher than the input SIR, the interferer power is dominant. To minimize the array output power, the contribution of the interferer must be minimized. The optimization is then more with respect to the interferer than to the noise, resulting in a lower array gain. This is the part of the plots that extends above input SNR of -10, 0, 10, and 20 dB for input SIR of -20 dB (Fig. 7.65(a)), -10 dB (Fig. 7.65(b)), 0 dB (Fig. 7.65(c)), and 10 dB (Fig. 7.65(d)), respectively.

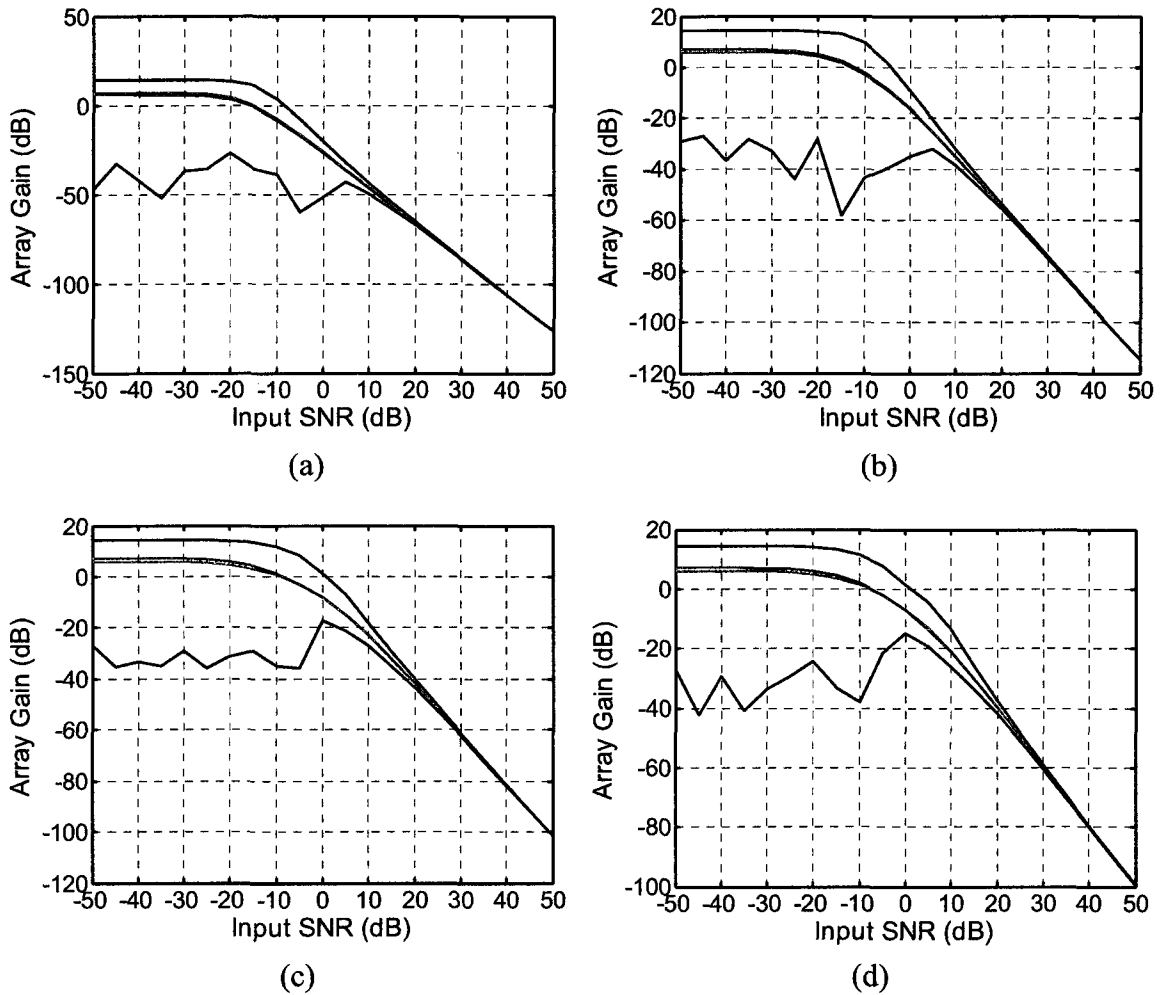


Figure 7.63. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the GO estimator and MVDR for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.

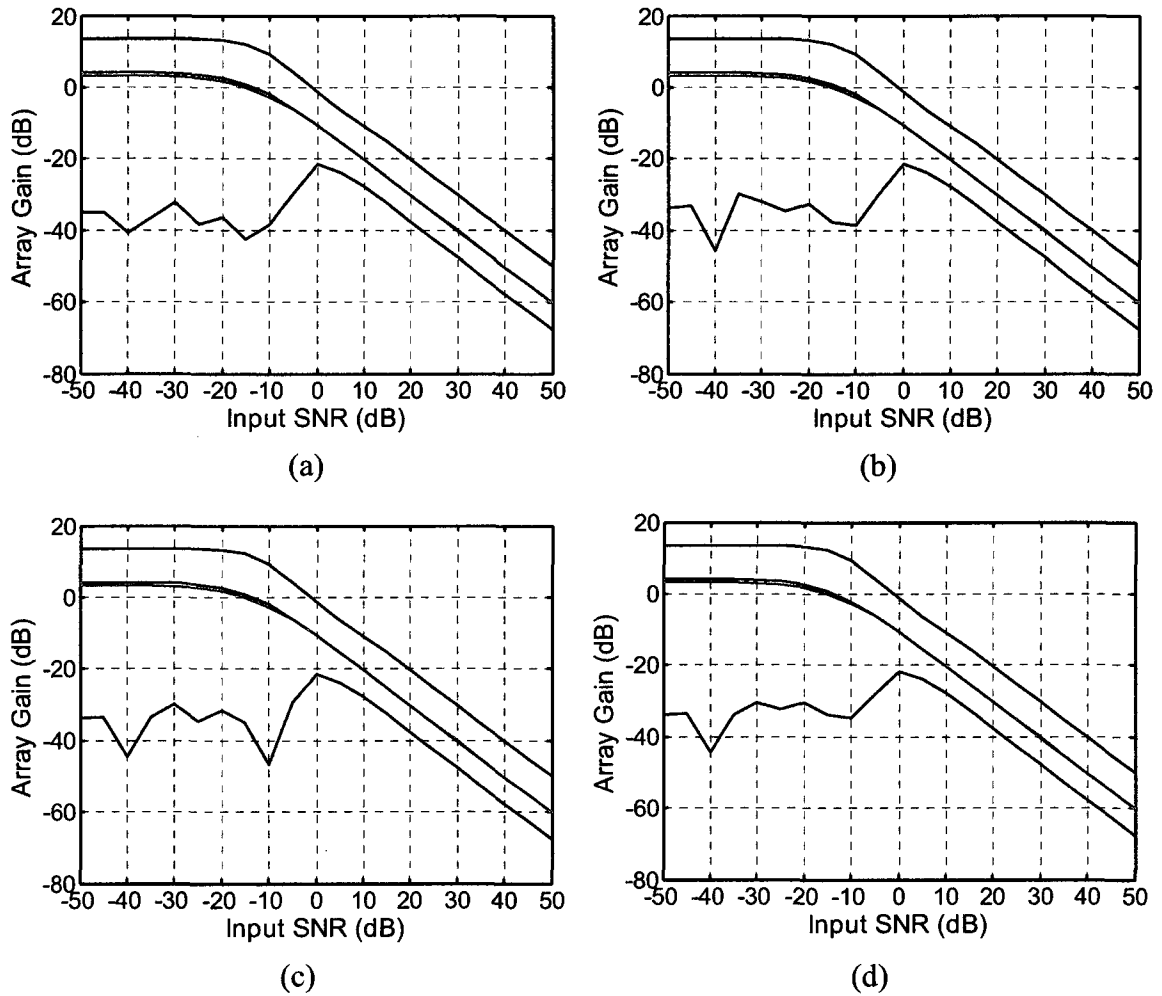


Figure 7.64. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (black) values of the array gain for 10000 realizations with LSSI and MVDR for input SNR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.

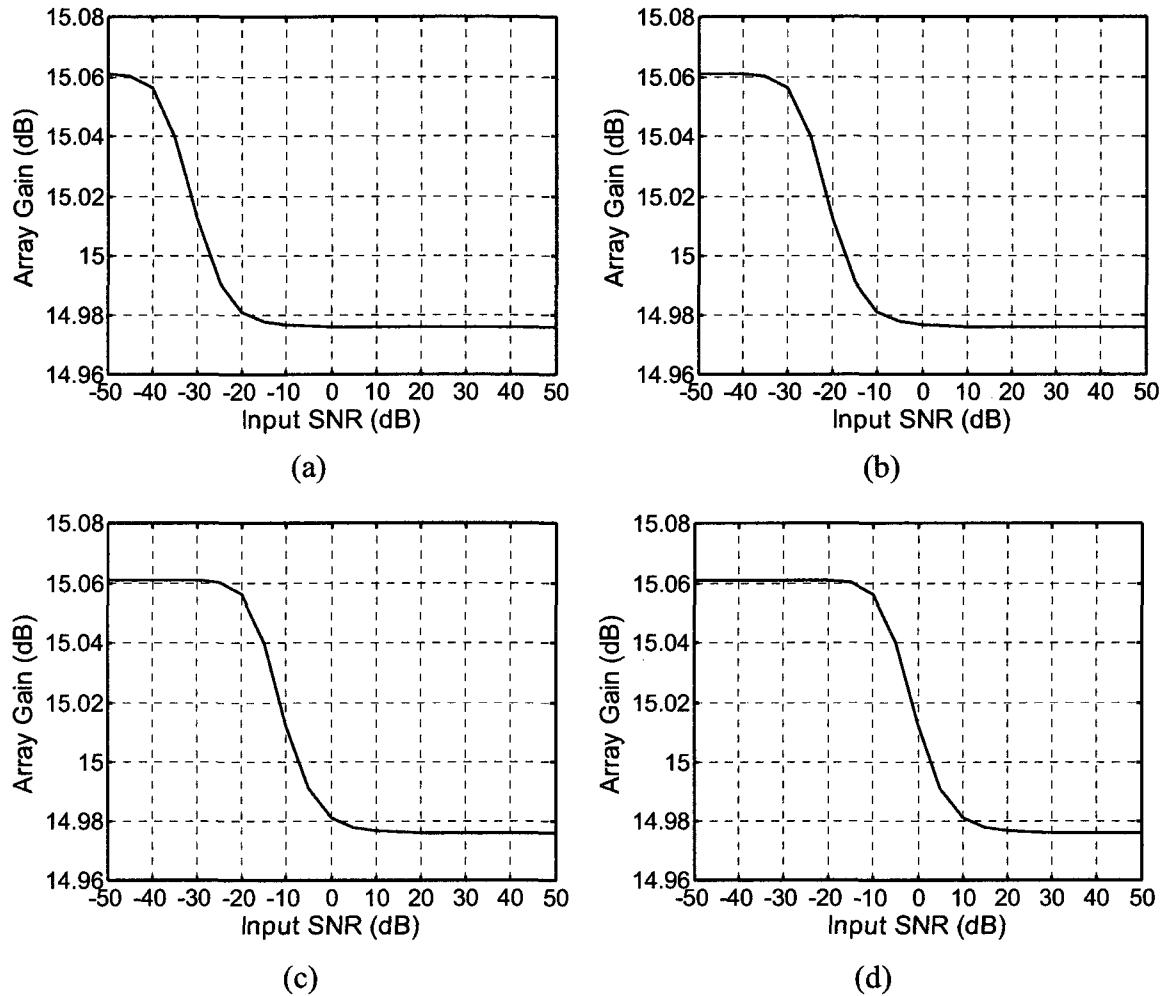


Figure 7.65. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations with the exact source model and MVDR for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.

Figures 7.66 presents the gain in SIR when the GO estimator is used to provide a source model to an MVDR beamformer. We observe two regions again, depending on the input SIR versus input SNR. The input SNR transition between the regions increases again with an increase in input SIR. The spread in the gain in SIR is large, as it was the case earlier. The input SNR at which the spread becomes small increases with input SIR also. Even for an input SNR sufficiently high to get a good gain in SIR, Fig. 7.63 showed a negative array gain. The good gain in SIR allows the signal to be dominant over the

interferer but the negative array gain indicates that the noise power increases, even to a point where the output SNR becomes negative.

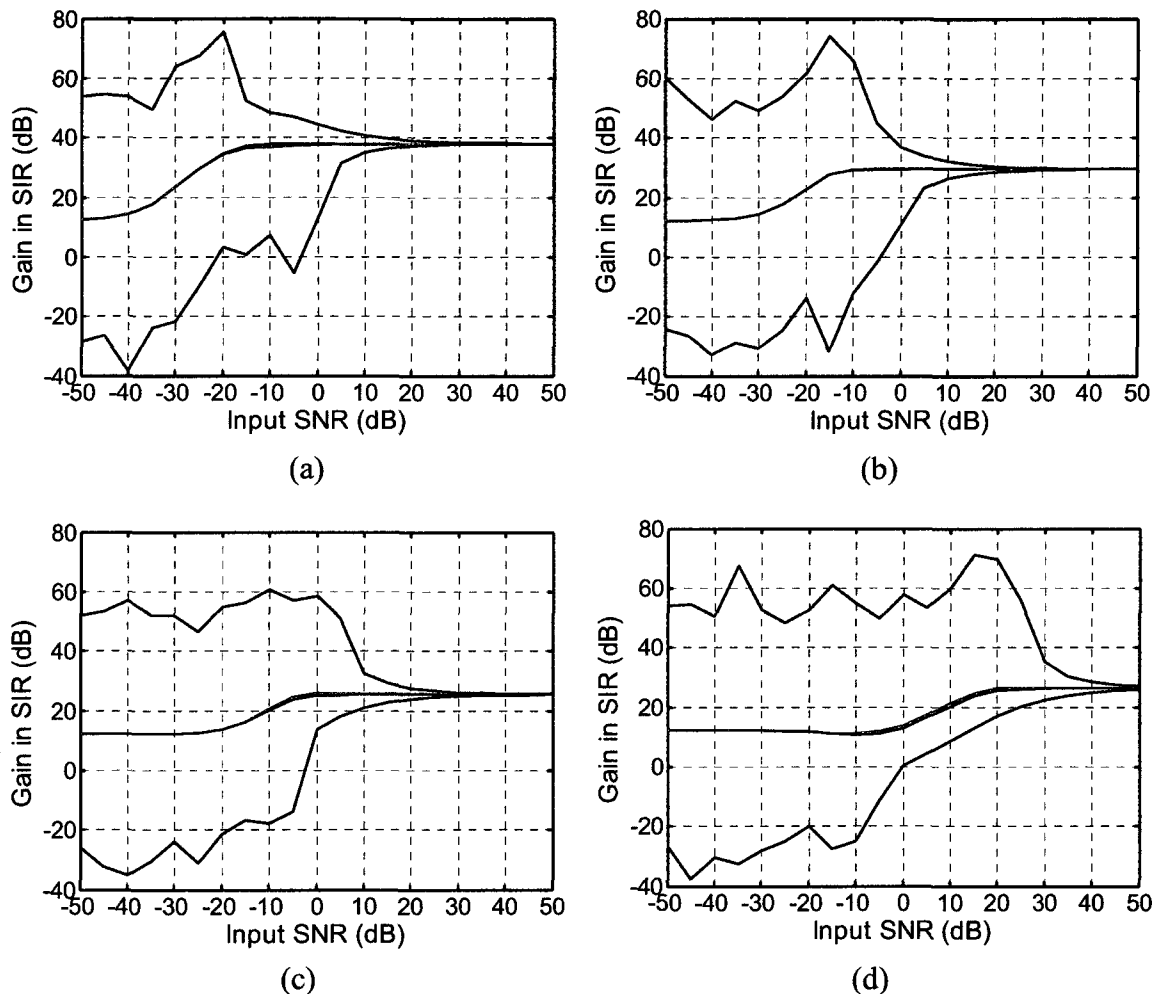


Figure 7.66. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the gain in SIR for 10000 realizations with the GO estimator and MVDR for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.

Figure 7.67 shows the gain in SIR when the LSSI estimator is used to provide a source model to an MVDR beamformer. We observe two regions again: a fairly constant gain in SIR when the input SNR is much lower than the input SIR, and an increase in gain in SIR with an increase in the input SNR when the input SNR is larger than the input SIR. As noted before, in that second region the interferer power is dominant over the noise and the MVDR beamformer has to decrease the interferer power to minimize the

array output power. However, this happens simultaneously with a decrease in the array gain as shown in Fig. 7.64. As it was the cases earlier with the GO estimator with MVDR beamforming, the output noise power can get stronger than the signal power.

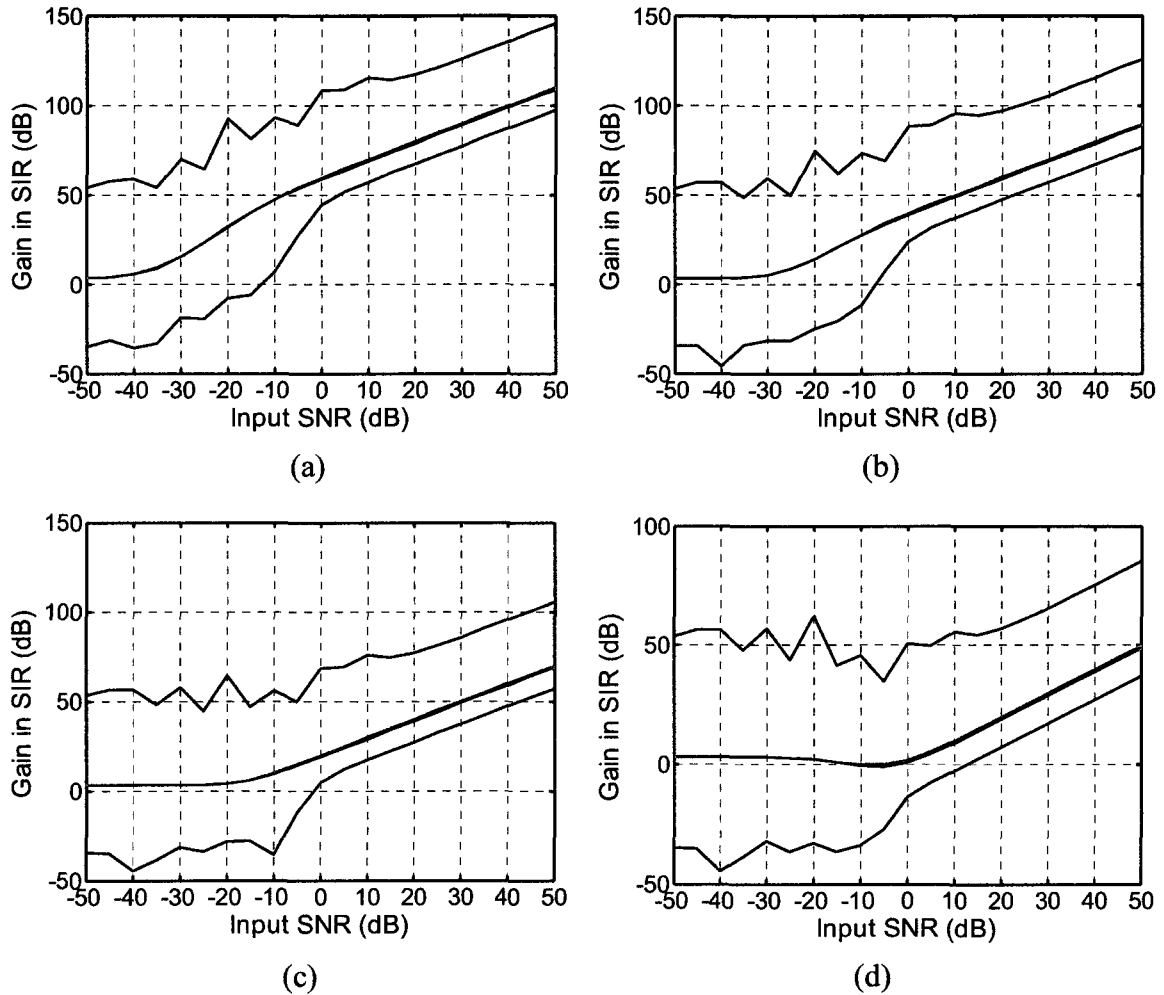


Figure 7.67. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the gain in SIR for 10000 realizations with LSSI and MVDR for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.

Figure 7.68 shows the gain in SIR when the exact source model is used with MVDR beamforming. The general behavior is as on Fig. 7.67 but of course without the spread.



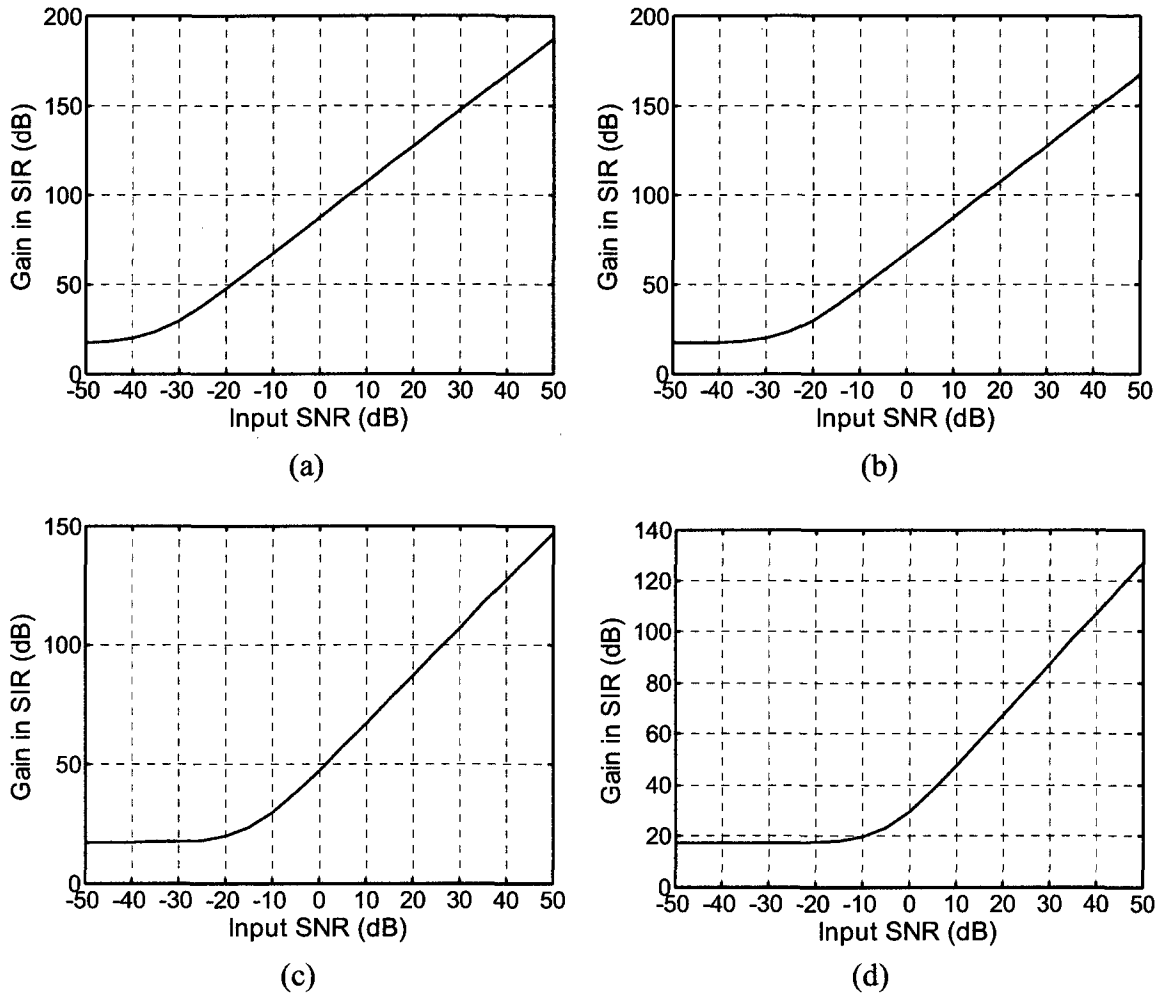


Figure 7.68. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the gain in SIR for 10000 realizations with the exact source model and MVDR for input SIR of (a) -20 dB; (b) -10 dB; (c) 0 dB; (d) 10 dB.

Overall, using MVDR beamforming with either the GO estimator or the LSSI estimator did not prove to be a good approach. MVDR beamforming requires a source model that is very close to the source signal. Otherwise, the MVDR beamformer reduces the power of the signal at the output of the array and this gets worse as the input SNR improves. As was noted in Sub-Section 7.3.1, this requirement on the source model and the resulting behavior of the MVDR beamformer is not limited to the use of our proposed beamforming method and estimators introduced in this thesis.

### 7.7.2 Beamforming position error, rectangular array, and dipole source

We define as *beamforming position error*, the error between the exact source position and the position used to compute the mode vectors necessary to beamform with the proposed beamforming method. This sub-section investigates through simulations the effect of a beamforming position error on the beamforming performance, for the case of a rectangular array and a dipole source, without and with a monopole interferer.

The dipole source is in the same position as for the previous simulations, at  $(x, y, z) = (2, 1, 0)$ . When the interfering monopole is present, it is still located at  $(x, y, z) = (-1, 0.5, -0.2)$ . The noise remains spherically isotropic (diffuse).

#### ***Source without interferer, beamforming position error***

100 beamforming position errors were generated. The error, along each of the 3 Cartesian coordinate axes, follows a normal distribution with a mean of zero and a standard deviation of 0.03 meter. The projection of the position errors in the  $x$ - $z$  and  $x$ - $y$  planes is shown in Fig. 7.69 while the distribution of the magnitude of the errors is displayed in Fig. 7.70. With a source frequency of 996.1 Hz, the wavelength is 0.34 meter. The maximum error magnitude is almost one third of the wavelength. This is thus a significant error.

The optimization criterion for the simulations remains the array gain. To allow comparison with the proposed beamforming method, we examine two beamforming methods. The first method is comparable to the optimum solution but beamforming at the wrong location. In this case, the beamforming weight vector is computed using the exact mode strengths, but beamforming is performed at the location of the source plus each of the 100 position errors. The second method is conventional beamforming but, again, beamforming at the wrong location because of the beamforming position errors. Tables 7.17 and 7.18 contain the statistics of the array gain and the white-noise gain for the 100 realizations of the position errors, for each method.

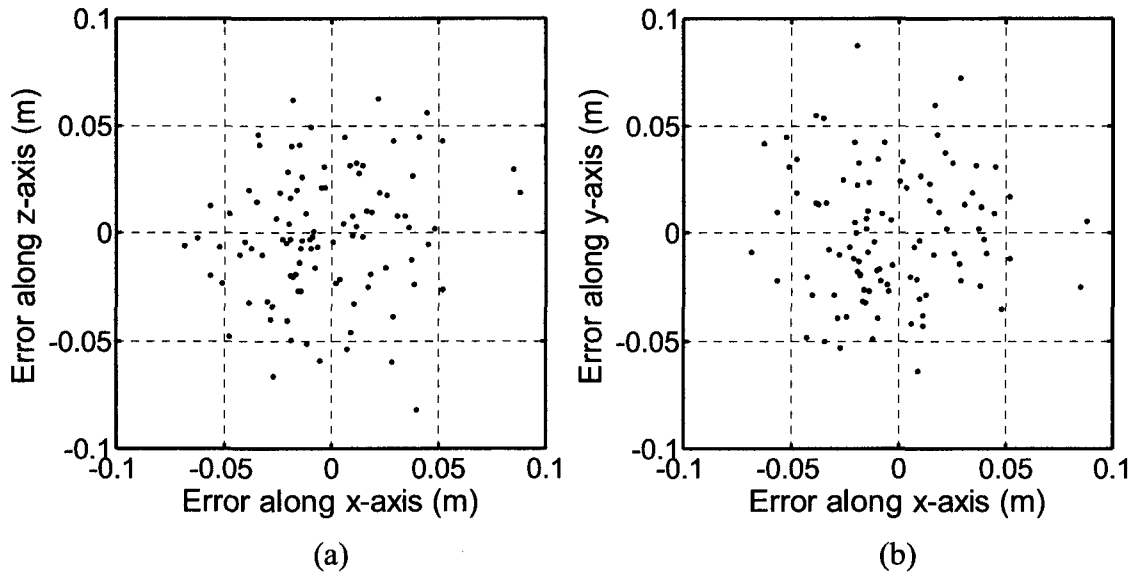


Figure 7.69. 100 Beamforming position errors (a)  $x$ - $z$  plane; (b)  $x$ - $y$  plane.

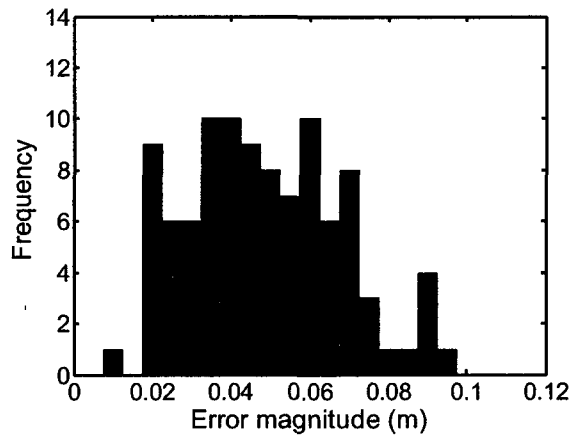


Figure 7.70. Histogram of the magnitude of the 100 beamforming position errors.

Table 7.17. Array gain and white-noise gain for dipole source and rectangular array, beamforming with exact mode strengths but with beamforming position error.

Statistic	Array gain	W.-N. gain
	dB	
Max	15.03	14.97
Median	14.32	14.28
Mean	13.76	13.71
Min	4.55	4.51

Table 7.18. Array gain and white-noise gain for dipole source and rectangular array, conventional beamforming but with beamforming position error.

Statistic	Array gain	W.-N. gain
	dB	
Max	13.00	12.90
Median	3.62	3.51
Mean	4.16	4.04
Min	-1.94	-2.07

For the dipole source and array geometry, the optimum array gain is 15.06 dB, as indicated in Table 7.12. We observe from Table 7.17, that the array gain can only be degraded because of the beamforming position error. When there are no beamforming position errors, conventional beamforming provides an array gain of -1.55 dB, again as presented in Table 7.12. With beamforming position errors, there is now a spread of array gain, which for the 100 beamforming position errors we have simulated, extends from -1.94 dB to 13 dB.

With the proposed beamforming method, for each input SNR, we evaluate the array gain while beamforming at the location of the source plus each of the position errors. For each position error, we go through 10000 realizations of the noise vector. This

means that for each input SNR, there are one million realizations when we use beamforming position errors.

Figure 7.71 and 7.72 show the spread of the array gain, without and with beamforming position errors, when using the GO and the LS estimators, respectively. The main difference is the lower minimum array gain for low input SNR when there are beamforming position errors compared to when beamforming at the exact source position.

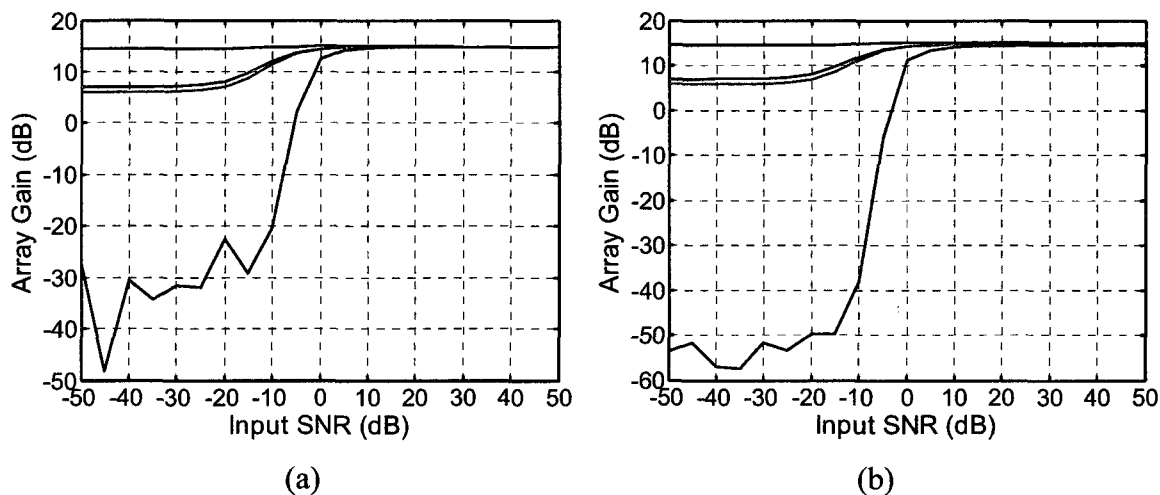


Figure 7.71. Rectangular array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations, the GO estimator (a) beamforming at source position; (b) with beamforming position errors.

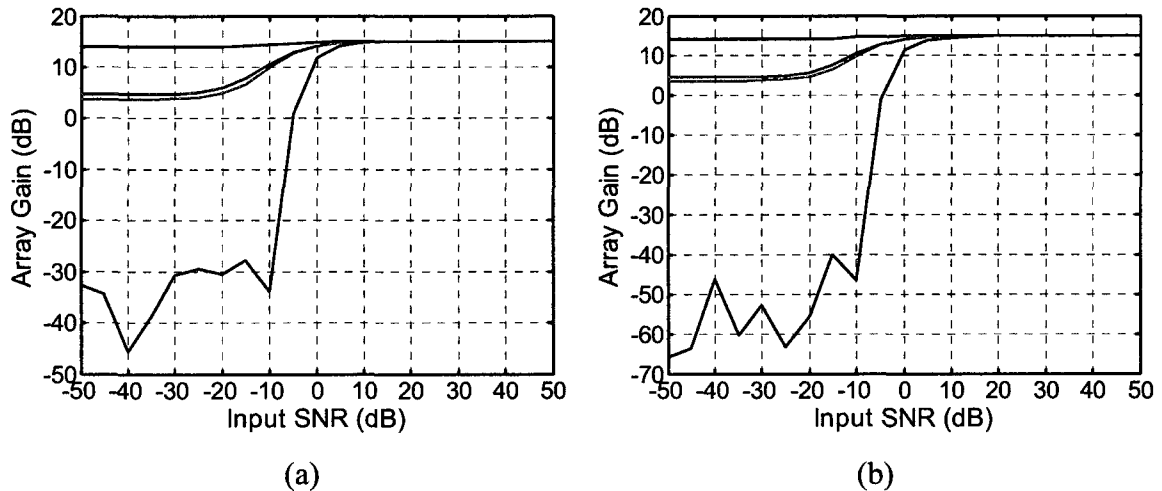


Figure 7.72. Rectangular array, dipole source, and diffuse noise. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations, LS estimation (a) beamforming at source position; (b) with beamforming position errors.

Without and with beamforming position errors, the spread becomes narrow at high input SNR. However, this happens at higher input SNR when there are beamforming position errors. For input SNR values 0 dB or above, there is a small decrease in the mean array gain when there are beamforming position errors. The spread does not get as small as when there are no position errors but it is much smaller than for beamforming with the exact source mode strength or with conventional beamforming when there are beamforming position errors, as seen in Tables 7.17 and 7.18.

### ***Source with interferer, beamforming position error***

The simulations now include a monopole interferer. The 100 beamforming position errors that were generated previously for the source are used again. Since the LSSI estimator also requires an interferer position to compute the mode vectors corresponding to the interferer, 100 interferer beamforming position errors were generated. The projection of the interferer beamforming position errors in the  $x$ - $z$  and  $x$ - $y$  planes is shown in Fig. 7.73 while the distribution of the magnitude of the errors is displayed in Fig. 7.74.

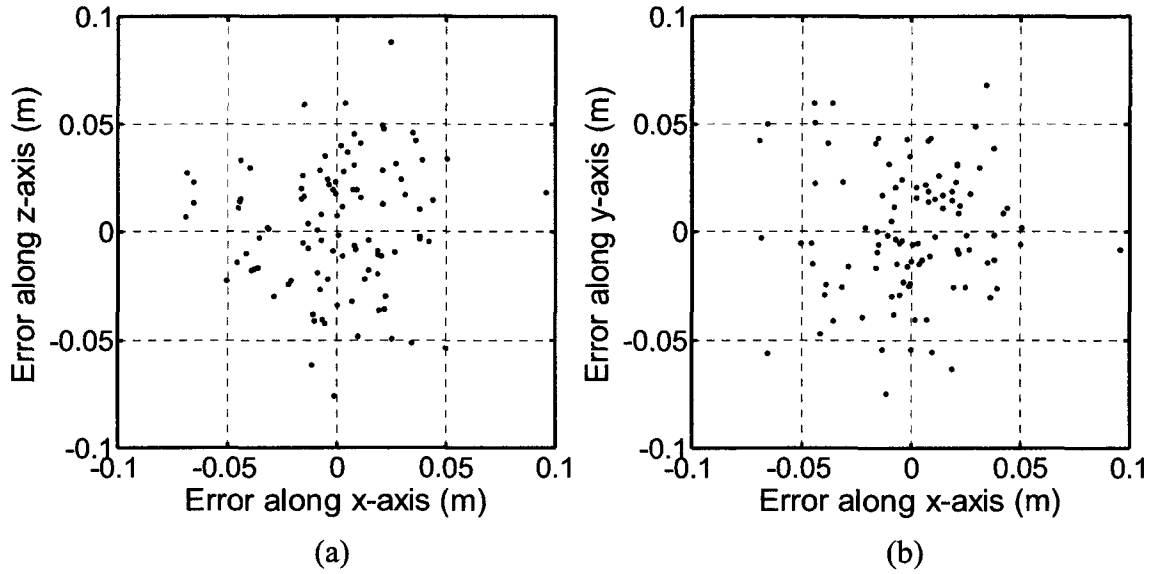


Figure 7.73. 100 interferer beamforming position errors for the LSSI estimator (a)  $x$ - $z$  plane; (b)  $x$ - $y$  plane.

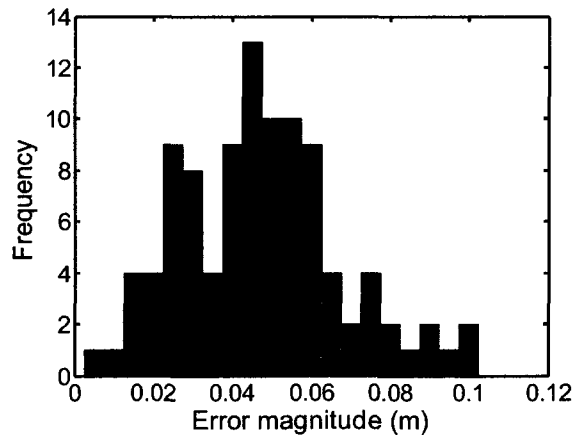


Figure 7.74. Histogram of the magnitude of the 100 interferer beamforming position errors for the LSSI estimator.

Tables 7.19 and 7.20 contain the statistics of the array gain, the white-noise gain, and the gain in SIR for the 100 realizations of the position errors, when beamforming with the exact mode strengths and with conventional beamforming, respectively. Since the optimization is with respect to diffuse noise and not the interferer, the statistics of the

array gain and the white-noise gain in Table 7.19 are the same as in Table 7.17 while the statistics of the array gain and the white-noise gain in Table 7.20 are the same as in Table 7.18. Beamforming at the exact source position with the exact source signal as the source model provided a gain in SIR of 17.14 dB (see Table 7.12). With the 100 beamforming position errors, beamforming with the exact source mode strengths provides a range of gain in SIR from 13.36 to 20.52 dB. Conventional beamforming without beamforming position errors provided a gain in SIR of 4.47 dB (again, see Table 7.12). With beamforming position errors, conventional beamforming provides a gain in SIR ranging from 3.35 to 30.40 dB.

Table 7.19. Array gain, white-noise gain, and gain in SIR for dipole source and rectangular array, beamforming with exact mode strengths but with beamforming position error.

Statistic	Array gain	W.-N. gain	Gain in SIR
	dB		
Max	15.03	14.97	20.52
Median	14.32	14.28	17.01
Mean	13.76	13.71	16.90
Min	4.55	4.51	13.36

Table 7.20. Array gain, white-noise gain, and gain in SIR for dipole source and rectangular array, conventional beamforming but with beamforming position error.

Statistic	Array gain	W.-N. gain	Gain in SIR
	dB		
Max	13.00	12.90	30.40
Median	3.62	3.51	10.04
Mean	4.16	4.04	10.71
Min	-1.94	-2.07	3.35



Figure 7.75 and 7.76 show the spread of the array gain, without and with beamforming position errors, when using the GO and the LSSI estimators, respectively. As when there was no interferer, the main difference is the lower minimum array gain for low input SNR when there are beamforming position errors compared to when beamforming at the exact source position. Without and with beamforming position errors, the spread becomes narrow at high input SNR. However, this happens at higher input SNR when there are beamforming position errors and the spread is larger than for beamforming with no position errors. This was also the case when there was no interferer. Again, there is a small decrease of the mean array gain when there are beamforming position errors.

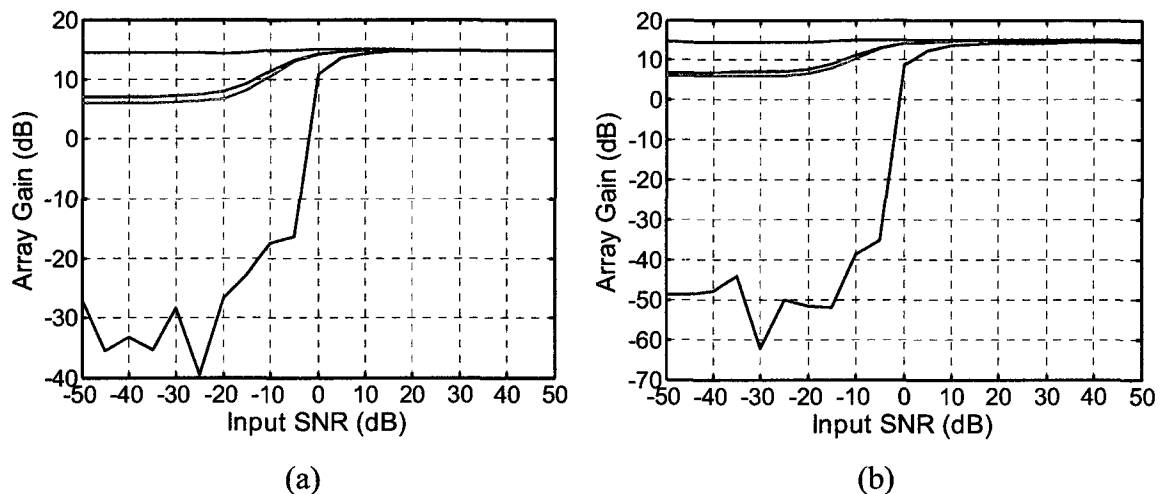


Figure 7.75. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations, the GO estimator, input SIR of 0 dB (a) beamforming at source position; (b) with beamforming position errors.

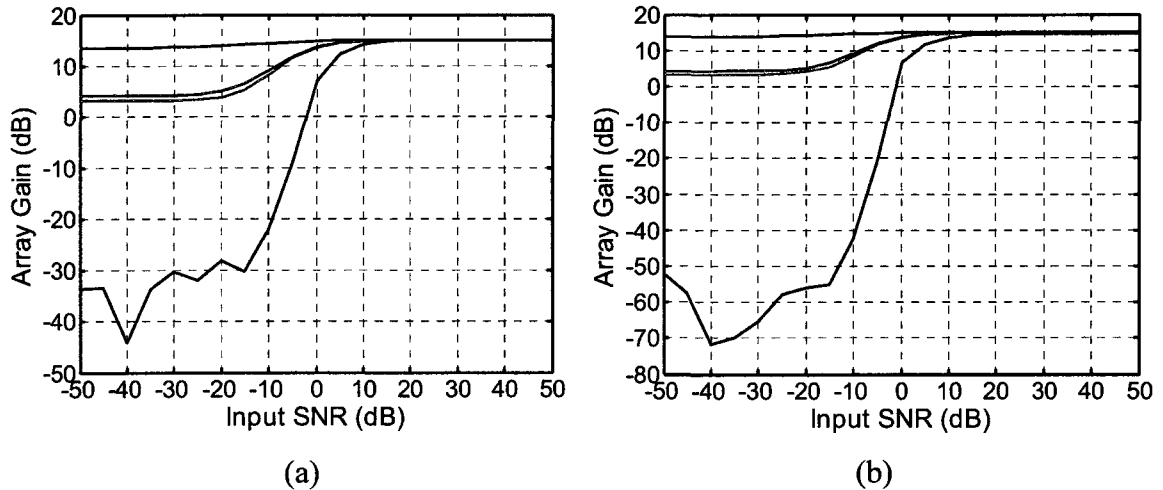


Figure 7.76. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the array gain for 10000 realizations, LSSI estimation, input SIR of 0 dB (a) beamforming at source position; (b) with beamforming position errors.

Figure 7.77 and 7.78 show the spread of the resulting gain in SIR, without and with beamforming position errors, when using the GO and the LSSI estimators, respectively. The spread of the gain in SIR is wider where there are beamforming position errors. Also, at high input SNR, the spread does not become as narrow when there are beamforming position errors as it does where beamforming at the exact source position. For these simulations, the proposed beamforming method provides a gain in SIR above 10 dB for the 1 millions realizations with input SNR values above approximately 1 and 7 dB, when the GO and the LSSI estimators are used, respectively.

We expect the degradation in performance with the proposed beamforming method to increase as the beamforming position error increases.

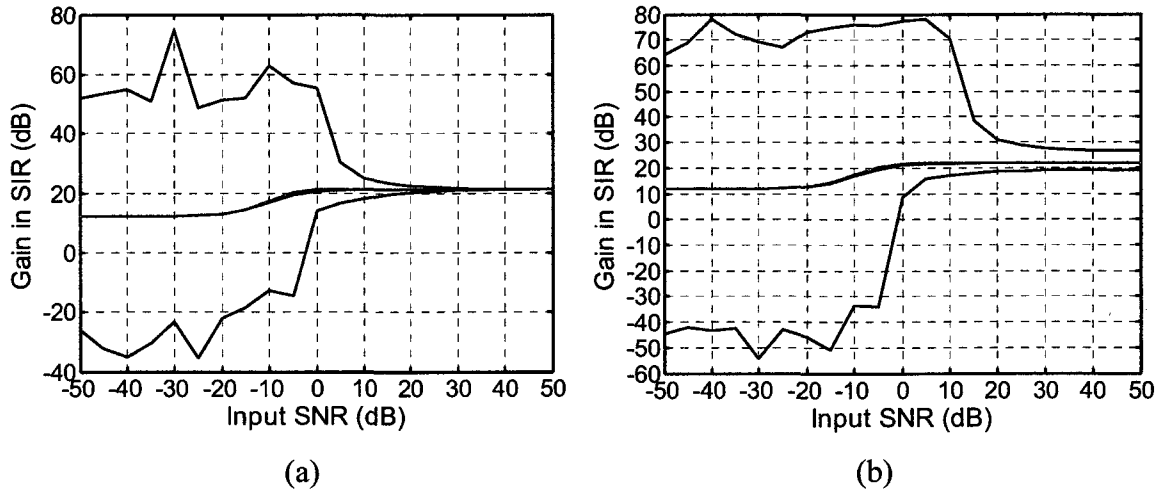


Figure 7.77. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the gain in SIR for 10000 realizations, the GO estimator, input SIR of 0 dB (a) beamforming at source position; (b) with beamforming position errors.

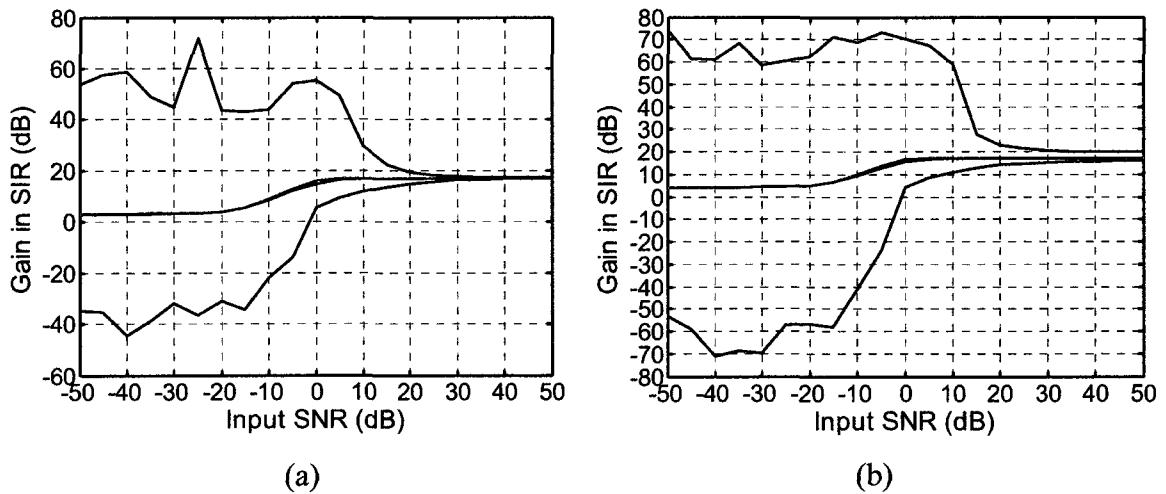


Figure 7.78. Rectangular array, dipole source, diffuse noise, and monopole interferer. Mean (green), median (red), minimum (blue), and maximum (blue) values of the gain in SIR for 10000 realizations, LSSI estimation, input SIR of 0 dB (a) beamforming at source position; (b) with beamforming position errors.

### 7.7.3 Concluding remarks: Diffuse noise + an interferer

Of the approaches that do not use MVDR beamforming (i.e. the approaches using our proposed beamforming method), the GO estimator and the LSSI estimator performed the best. While it is necessary to know the source position to use the GO estimator, it is necessary to have both the source and the interferer locations for the LSSI estimator.

With the GO estimator, the strength estimation of a mode is biased by other modes present in the source and by the interferer. For the LSSI estimator, the strength estimation of a mode of the source or the interferer is not biased by other modes of the source or the interferer, but this is only true if the locations are known precisely. Most likely in practice, the LSSI estimate would not perform as well as it did here, because there would be some resulting bias on the estimate. We saw a degradation in performance with the LSSI estimator when beamforming position errors were present, in Sub-Section 7.7.2.

In the previous sections, approaches using MVDR beamforming did not perform adequately when there was only diffuse noise or for diffuse noise and reverberation. In the present section, we saw that they do not perform adequately either in the presence of diffuse noise and an interferer. The reason is again that the estimated source model is different from the source signal.

The optimization was done with respect to the diffuse noise as characterized by the matrix  $\mathbf{R}_{ee}(k)$ . If the interferer-plus-noise cross-power spectral density matrix  $\sigma_{s_1}^2(k)\mathbf{R}_{s_1s_1}(k) + \sigma_e^2(k)\mathbf{R}_{ee}(k)$  was known somehow (for example, estimated while the source of interest is inactive), it could be used to find the beamforming weight vector, in order to obtain a better gain in SIR than optimizing with respect to  $\mathbf{R}_{ee}(k)$  can provide.

The proposed beamforming method still worked when beamforming position errors were introduced. We do expect the resulting performance penalty to increase with an increase in the error magnitude.

## 7.8 Chapter conclusion

This chapter has presented simulation methods and results for beamforming on directional sources (and monopoles) in the presence of diffuse noise, diffuse noise and reverberation, and diffuse noise and an interferer, with the proposed beamforming method, different source model estimation approaches, combinations of the source model estimates with MVDR beamforming, as well as conventional beamforming.

Section 7.3 showed that the proposed beamforming method could perform much better than conventional beamforming (point source model) when beamforming on directional sources, in particular for input SNR of 0 dB and above. It was also confirmed in that section that the proposed beamforming method can beamform on monopoles with a relatively small penalty with respect to conventional beamforming, again for input SNR of 0 dB and above. For lower input SNR, there is a statistical distribution of the array gain but, depending on the value of the input SNR, a large portion of the realizations provided good array gains.

Section 7.5 demonstrated that the proposed beamforming method still works when there is reverberation. However, there is some degradation of performance, as can be expected.

Finally, Section 7.7 presented results for beamforming on a dipole source in the presence of diffuse noise and an interfering monopole. The proposed beamforming method provided varying levels of array gain and gain in SIR, depending on the input SNR and the input SIR.

Generally when beamforming on directional sources, the proposed beamforming method outperformed conventional beamforming when the input SNR was 0 dB or higher.

---

# CHAPTER

# 8

---

## Conclusion

### 8.1 Important points

We covered many topics in this thesis: beamforming, example of directional sources, description of the proposed beamforming method including a number of mode strength estimators, and investigation of performance with simulations and measurements. We will now revisit some important points.

Chapter 2 presented the concepts of beamforming such that the beamforming weights are computed using the source model (Sub-Section 2.2.1). Performance measures were discussed in Section 2.3. The array gain (Sub-Section 2.3.1) is the performance measure we used the most in the thesis. It quantifies the improvement in signal-to-noise ratio between the input and the output of the array. Design methods to find the beamforming weights are mentioned in Section 2.4. In all cases, it is necessary to have a source model. The proposed beamforming method was developed while maximizing the array gain (Sub-Section 2.4.2), which is a statistically optimum beamforming method.

In the entire thesis, we have used the expression conventional beamforming (Sub-Section 2.2.2) to mean beamforming using the point source as the source model to compute the beamforming weights. Chapter 2 included simulation results (Section 2.5) to

demonstrate the effect on array gain of using conventional beamforming when the actual source is directional. If the array subtends a small angle about the source, the point source model may be acceptable, in particular if the array observes a small portion of a single lobe of the directivity pattern. If this is not the case, the optimum array gain is not obtained. Even more, if the array gain goes below 0 dB, the array decreases the signal-to-noise ratio between its input and output (the output SNR becomes lower than the input SNR).

To explain why the point source model has been used successfully to beamform on the human talker, we discussed its directivity in Section 2.6. The human talker directivity pattern magnitude is fairly constant in the frontal hemisphere of the talker. Furthermore, with arrays of small dimensions located in front of the talker, only a portion of the hemisphere is observed so that the point source is an acceptable model and does not lead to a much noticeable penalty.

Few phase measurements have been published on the human talker. In Sub-Section 2.6.2, using measurements for a life-size mannequin that includes some phase information [Flanagan 1960], we have tried and evaluated the effect of using a point source on the array gain. For an array covering the entire frontal hemisphere of a talker, the penalty in the array gain is 2.2 and 6.8 dB at 1 and 4 kHz, respectively. When the array only covers 20 degrees in front of the talker, the array gain penalty is much smaller, with a penalty of 0.3 and 1.3 dB at 1 and 4 kHz, respectively. As far as we can tell, phase information has never been included to evaluate the performance penalty of conventional beamforming on the human talker. This is a contribution, though minor, of this thesis.

Chapter 3 presented sources that are more directional than the human talker from two different fields: musical instruments and aeroacoustics. Musical instruments are of special interest because the thesis is mainly concerned with sound pickup. Aeroacoustics examples are more applicable to source localization applications than to sound pickup.

The main contributions of the thesis began with Chapter 4 which presented the development of the proposed beamforming method. It is aimed at sound pickup of directional sources. No assumption is made on the source directivity. It is treated as an unknown. The principal assumptions we made are that the source location is known as is

---

the array geometry. The array geometry does not change with time. Section 4.1 mentioned all the assumptions in detail.

We thought of the proposed beamforming method by observing the form of the spherical Fourier transform while reading on the phase-mode spherical array (“eigenbeamformer”). This is described in Section 4.2 for the abstract case of a continuous spherical array with the source located at its center. The development for the more practical case of a discrete array of arbitrary geometry is in Section 4.3. The design criterion is the maximization of the array gain. The source is decomposed into spherical harmonics modes and each mode is propagated independently up to the array microphones. It is shown that the proposed beamforming method provides the optimum array gain if the weight of each sub-beamformer in the linear sum is the strength of the corresponding mode composing the source. Though the full development is demonstrated with spherical harmonics, Sub-Section 4.3.1 showed that the method is also applicable if the source signal can be decomposed into some other set of modes.

The estimation method we called “gain-optimized (GO) estimator” was presented in Section 4.4. We showed how to derive the array gain for the proposed beamforming method with the GO estimator but we did not write the mathematical expression because of its complexity. The performance of the method was instead investigated with the processing of measurements and through simulations (Chapter 5 and 7).

Because the source mode strengths must be estimated, and because sub-beamformers operate on each mode, the proposed beamforming method is more computationally complex than beamforming directly with a known source model. However, if the source directivity, orientation, and position do not change with time, it is not necessary to keep estimating the source mode strengths. In this case, the sub-beamformers can be combined into a single beamformer, and the complexity of the proposed beamforming method becomes the same as for conventional beamforming.

As indicated in Sub-Section 2.4.2, obtaining the optimum array gain requires a highly tuned system and can be very sensitive to random errors such as amplitude and phase errors in the microphone channels, or to slightly inexact microphone positions. The



---

white-noise gain is a measure of robustness. So instead of maximizing the array gain, it may be preferable to maximize the array gain subject to a constraint on the white-noise gain. To apply this method however, the source directivity must be known so that the white-noise gain can be evaluated to see if the criterion is met. If it is not met, it is necessary to compute a new beamforming weight vector while giving more importance to the white-noise in the optimization. Section 4.7 showed how the proposed beamforming method and the GO estimator could be used to estimate the source model and the white-noise gain while going through an iterative process (simulation results for this are presented at the end of Sub-Section 7.3.1).

Chapter 4 ended with a comparison of the proposed beamforming method to other methods. In addition to the source model being known for the phase-mode spherical array, another big difference is that, with the proposed beamforming method, the source is decomposed into spherical harmonics modes and each mode is propagated independently up to the array microphones. For the phase-mode spherical array, the received signal on the spherical array, after propagation, is decomposed into spherical harmonics. For this to be possible, the array must be spherical. No such constraint is set on the array with the proposed beamforming method.

A comparison between the proposed beamforming method with the GO estimator and conventional beamforming was realized with simulations and with the processing of measurements, in Chapter 5. The optimization was done with respect to the white-noise gain. The proposed beamforming method provided a white-noise gain within 1 dB of the optimum value for the linear array of 16 microphones and a dipole source. As expected with this array geometry and source dipole, conventional beamforming provided a white-noise gain below 0 dB. Also, the array response was simulated and measured along the  $x$ -axis. The proposed beamforming method overcame the failure to sense the dipole source at the proper position, which is experienced by conventional beamforming. Even in a non-anechoic environment, there was a good match between measurements and simulations.

Chapter 6 went deeper into the source mode strengths estimation and provided additional approaches. The bias and variance of the GO estimator were provided, though

the complete mathematical development is in Appendix B. The development of a least-squares (LS) estimator was presented in Section 6.3, which also included the resulting estimated source model and the computation method using singular value decomposition. With this estimator, if the source location is known precisely and all the modes present in the source are included in the estimation, the estimation of the strength of one spherical harmonic mode will not be biased by other modes present in the source. This is not the case with the GO estimator. Sub-Section 6.3.6 showed that if there is an interferer whose location is known along with the source, it is possible to estimate the strengths of the modes composing the source and the strengths of the modes composing the interferer. We use the abbreviation LSSI for this estimator. Again, if the source and the interferer locations are known precisely and all the modes present in the source and the interferer are included in the estimation, the mode estimation of either the source or the interferer will not be biased by other modes.

The description of using the proposed beamforming method to provide a source model estimate for minimum variance distortionless response (MVDR) beamforming was mentioned in Section 6.4. The MVDR beamformer by itself was discussed in Sub-Section 2.4.2, as a special case of the linearly constrained minimum variance (LCMV) beamforming.

Up to that point, the estimation was done from a single FFT frame. Section 6.5 presented a method to perform the estimation using a number of FFT frames. Unfortunately, unlike in classical power-spectrum estimation, it is not possible to simply average FFT frames for this estimation, because the phase is required along with the magnitude.

It is difficult to analytically evaluate the performance of the proposed beamforming method because of its mathematical complexity. Instead, a number of simulations were realized with a linear array of 16 microphones and an array of 32 microphones located on the perimeter of a rectangle. The results are presented in tables and figures in Chapter 7, along with the corresponding analysis.

The simulation method for diffuse noise was described in Section 7.2, while the results for beamforming on a source in the presence of diffuse noise were presented in

Section 7.3. In the simulation scenarios run with a directional source in the presence of diffuse noise, the proposed beamforming method performed much better than conventional beamforming. For example, with the GO estimator and an input SNR of 0 dB or higher, the array gain was close to the optimum value. The proposed beamforming method also worked when the source was a monopole, providing again an array gain close to the optimum value for input SNR of 0 dB or higher. We saw that for low input SNR, there is a spread of the array gain. A large portion of the realizations provides good array gain values, but the tail of the distribution does extend to very low values.

The simulation of reverberation was detailed in Section 7.4. The results for beamforming on a source in the presence of diffuse noise and reverberation were presented and analyzed in Section 7.5. Overall, the performance is decreased by the reverberation, as expected since the mode strengths are estimated from the microphone output vector. However, the proposed beamforming method still provides a reasonable performance. Of the four approaches for the estimation, reverberation causes less penalty with the GO estimator.

Finally, simulation with an interferer was explained in Section 7.6. Results for beamforming on a dipole source in the presence of diffuse noise and a monopole interferer were presented and analyzed in Section 7.7. Of the approaches without MVDR beamforming, the GO and the LSSI estimators performed the best. While it is necessary to know the source position to use the GO estimator, it is necessary to have both the source and the interferer locations for the LSSI estimator. However, with the LSSI estimator, the strength estimation of a mode of the source or the interferer is not biased by other modes of the source or the interferer, with the condition that the source and interferer locations are known precisely. Most likely in practice, LSSI estimation would not be performed as well because there would be some bias. This decrease in performance was observed in Sub-Section 7.7.2 when beamforming position errors were included.

In all the simulations that were performed, using the proposed beamforming method to estimate a source model for MVDR beamforming did not perform adequately. At low input SNR, the spread in the array gain is as large as with the proposed

beamforming method used alone. At higher input SNR, in all the simulations except for the case with reverberation, there was a linear decrease in the array gain with an increase in input SNR caused by the estimated source model being different from the source signal. This last phenomenon is also seen for beamforming on a monopole with conventional point source MVDR beamforming, for example when the beamforming assumed source location is different from the actual source position or when there are errors on the sensor positions.

## **8.2 Future work**

The new beamforming method proposed in this thesis is a step toward beamforming on directional sources. However, there are still a number of issues and questions that need to be addressed by further work beyond this thesis. This section discusses some of those difficulties and research avenues.

### **8.2.1 Source directivity changes with frequency and time**

The source directivity may change with frequency and time. Section 3.2 presented the case for musical instruments where the directivity changes with frequency for a given note played, and it also changes with each note played. For such a case, in practice, the beamformer would need to use different source models for different frequencies, and to adapt all those models with time. The adaptation would have to be seamless so that a person listening to the audio output of the beamforming array would hear a natural output, without audible transitions or artifacts.

### **8.2.2 Which optimization criterion to use**

The development of the proposed beamforming method was presented for the case of optimizing the array gain for a known noise field. As was seen in Sub-Section 2.4.2, unconstrained array gain optimization may lead to a beamforming array which is sensitive to errors because of a resulting low white-noise gain. The solution mentioned in that section is to optimize the array gain subject to a constraint on the white-noise gain. The equations to use with the proposed beamforming method for such

an optimization were provided in Section 4.7 and an investigation of the performance with a simulation was presented at the end of Sub-Section 7.3.1.

However, the open question that remains is what would be the better optimization criterion depending on the application, on the source, etc?

### 8.2.3 Estimators of the source mode strengths

The first contribution of this thesis is the development of an extension to beamforming methods, applicable to directional sources whose directivity and orientation are unknown. The array output is a weighted sum of “sub-beamformers”, each beamforming on a different spatial mode of the source. With the appropriate weighting (source mode strengths), the sum of the sub-beamformers gives the same optimum array gain as beamforming directly with the exact source model.

Since the strength of the modes composing the source is unknown, it is necessary to have an estimator of the mode strengths. We have introduced a number of estimators in this thesis. However, the proposed beamforming method is not limited to these estimators. Other estimators should exist, whose performance may or may not be better.

### 8.2.4 Beamforming on real life directional sources

The performance of the proposed method should be investigated with real life directional sources, in scenarios more realistic than our experiment, described in Chapter 5. It is difficult to evaluate the beamforming performance without doing actual measurements. The source directivity information available in the literature very rarely includes information on the phase, required for beamforming.

At this point some questions still remain unanswered, such as:

- How many modes should be included with the proposed beamforming method (how high should  $N_{\text{order}}$  be)?
- For a source made up of a distribution of elementary sources (monopoles, dipoles, quadrupoles, etc.) not exactly at the same position, would it be preferable to apply

the proposed beamforming method on a single beamforming location or instead, apply the proposed beamforming method on each of the elementary sources (to obtain a beamforming output for each of the elementary sources) and then combine the beamformer outputs?

Processing of real life directional source measurements appears necessary to answer such questions.

---

## Appendix

# A

---

### Solution to the Wave Equation in the Spherical Coordinate System, Spherical Hankel Function, and Spherical Harmonics

This appendix describes the solution to the linear lossless wave equation for an outgoing wave, in the spherical coordinate system. Since they appear in the solution, the spherical Hankel function of the second kind and spherical harmonics will be presented. More details are available in Chapter 6 of [Williams 1999] and Chapter 10 of [Blackstock 2000].

#### A.1 Solution to the wave equation in the spherical coordinate system for an outgoing wave

The linear lossless wave equation is

$$\bar{\nabla}^2 s - \frac{1}{c^2} \frac{\partial^2 s}{\partial t^2} = 0, \quad (\text{A.1})$$

where  $s$  is the sound pressure,  $c$  is the speed of sound,  $t$  is the time, and  $\bar{\nabla}^2$  is the Laplacian. Explicitly writing the Laplacian in the spherical coordinate system, the previous equation becomes

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial s}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial s}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 s}{\partial \phi^2} - \frac{1}{c^2} \frac{\partial^2 s}{\partial t^2} = 0, \quad (\text{A.2})$$

whose solution  $s$  depends on the three spherical coordinates as well as time. Using separation of variables

$$s = R(r)\Theta(\theta)\Phi(\phi)\dots T(t), \quad (\text{A.3})$$

Eq. (A.2) can be separated into four ordinary differential equations, each depending on a single variable. Those equations are

$$\frac{1}{c^2} \frac{d^2 T}{dt^2} + \kappa^2 T = 0, \quad (\text{A.4})$$

$$\frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) + \kappa^2 R - \frac{n(n+1)}{r^2} R = 0, \quad (\text{A.5})$$

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) + \left[ n(n+1) - \frac{m^2}{\sin^2 \theta} \right] \Theta = 0, \quad (\text{A.6})$$

$$\frac{d^2 \Phi}{d\phi^2} + m^2 \Phi = 0. \quad (\text{A.7})$$

For an outgoing wave, the solution to the differential equation depending on time is

$$T(t) \propto e^{i2\pi f t}. \quad (\text{A.8})$$

The solution for the differential equation depending on distance  $r$  is the spherical Hankel functions of second kind

$$R(r) \propto h_n^{(2)}(\kappa r) = j_n(\kappa r) - iy_n(\kappa r), \quad (\text{A.9})$$

where  $\kappa = 2\pi f/c$  is the wave number.  $j_n(\kappa r)$  and  $y_n(\kappa r)$  are the spherical Bessel functions of the first and second kind respectively. The solution for the differential equation depending on angle  $\theta$  is the associated Legendre functions

$$\Theta(\theta) \propto P_n^m(\cos \theta). \quad (\text{A.10})$$

The solution for the differential equation depending on angle  $\phi$  is



$$\Phi(\phi) \propto e^{im\phi}. \quad (\text{A.11})$$

The general solution for an outgoing wave is the sum over all possible values of  $n$  and  $m$ . In the frequency-domain, this solution is

$$s(f, r, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n s_{nm}(f) h_n^{(2)}(kr) Y_n^m(\theta, \phi). \quad (\text{A.12})$$

$Y_n^m(\theta, \phi)$  is the spherical harmonic function of order  $n$  and degree  $m$ , obtained by combining the solutions to dependence on angles  $\theta$  and  $\phi$  as will be seen below. The factor  $s_{nm}(f)$  is a complex number equal to the contribution of the spherical harmonic mode  $(n, m)$  from the source at frequency  $f$ ; it is referred in the thesis as the mode strength.

## A.2 Spherical Hankel function of the second kind

The spherical Hankel functions of the second kind of order  $n$  can be expressed in terms of the spherical Bessel functions of the first and second kind of order  $n$ , as

$$h_n^{(2)}(x) = j_n(x) - iy_n(x). \quad (\text{A.13})$$

The spherical Bessel functions can themselves be expressed in terms of the ordinary Bessel functions as

$$\begin{aligned} j_n(x) &= \left(\frac{\pi}{2x}\right)^{1/2} J_{n+1/2}(x) \\ y_n(x) &= \left(\frac{\pi}{2x}\right)^{1/2} Y_{n+1/2}(x). \end{aligned} \quad (\text{A.14})$$

Numerical values for the ordinary Bessel functions of the first kind  $J$  and of the second kind  $Y$  are readily available from mathematical softwares such as Matlab.

It is also possible to compute the spherical Hankel functions of the second kind of order  $n$  by taking derivatives of the spherical Hankel functions of the second kind of order zero

$$h_n^{(2)}(x) = (-x)^n \left( \frac{1}{x} \frac{d}{dx} \right)^n h_0^{(2)}(x), \quad (\text{A.15})$$

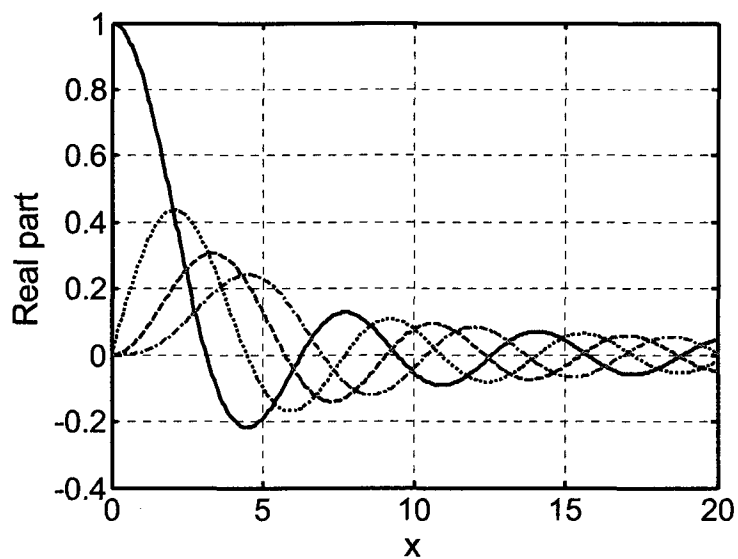
knowing that

$$h_0^{(2)}(x) = -\frac{e^{-ix}}{ix} = i \frac{e^{-ix}}{x}. \quad (\text{A.16})$$

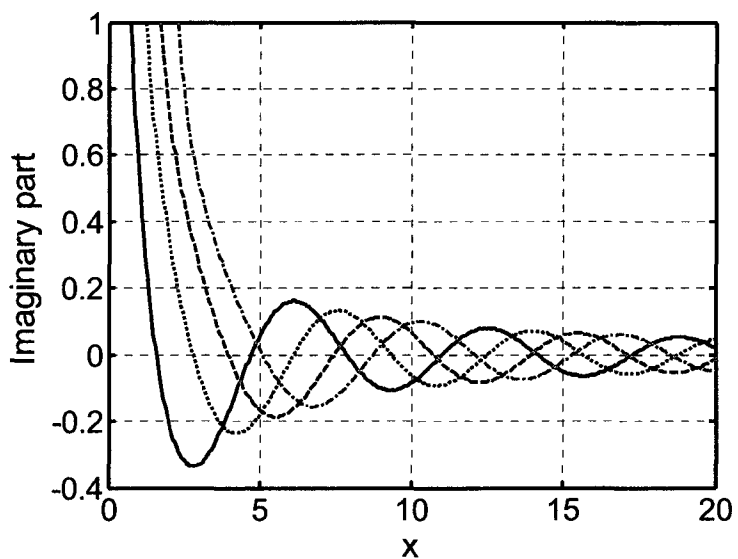
Expressions for the spherical Hankel functions of the second kind up to order three are given in Table A.1 and are plotted in Fig. A.1.

Table A.1. Spherical Hankel functions of the second kind of order zero to three.

$$\begin{aligned} h_0^{(2)}(x) &= i \frac{e^{-ix}}{x} \\ h_1^{(2)}(x) &= -\frac{e^{-ix}}{x^2} (-i + x) = e^{-ix} \left( -\frac{1}{x} + \frac{i}{x^2} \right) \\ h_2^{(2)}(x) &= -i \frac{e^{-ix}}{x^3} (-3 - 3ix + x^2) = e^{-ix} \left( -\frac{i}{x} - \frac{3}{x^2} + \frac{3i}{x^3} \right) \\ h_3^{(2)}(x) &= \frac{e^{-ix}}{x^4} (15i - 15x - 6ix^2 + x^3) = e^{-ix} \left( \frac{1}{x} - \frac{6i}{x^2} - \frac{15}{x^3} + \frac{15i}{x^4} \right) \end{aligned}$$



(a)



(b)

Figure A.1. Spherical Hankel function of the second kind  $h_n^{(2)}(x)$  (a) real part; (b) imag part; (—)  $n=0$ , ( $\cdots$ )  $n=1$ , (---)  $n=2$ , ( $\cdot-$ )  $n=3$ .

### A.3 Spherical harmonics

The solutions to dependence on angles  $\theta$  and  $\phi$  are combined into a single function called *spherical harmonic*

$$Y_n^m(\theta, \phi) \equiv \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos\theta) e^{im\phi}, \quad (\text{A.17})$$

where  $-n \leq m \leq n$ . For positive  $m$ , the associated Legendre function is

$$P_n^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_n(x), \quad (\text{A.18})$$

which is expressed as a function of the Legendre polynomials

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n. \quad (\text{A.19})$$

(This concise form of the Legendre polynomials is called Rodrigues' Formula.)

For negative  $m$ , the associated Legendre function is obtained from Eq. (A.18) with

$$P_n^{-m}(x) = (-1)^m \frac{(n-m)!}{(n+m)!} P_n^m(x), \quad (\text{A.20})$$

where  $m$  is positive in the equation. It is seen from this equation that  $P_n^{-m}(x)$  and  $P_n^m(x)$  differ only by a constant. Selected values of the associated Legendre functions are given in Table A.2.

Table A.2. Selected values of the associated Legendre functions  $P_n^m(\cos \theta)$ .

$m = 0$	$m = 1$
$P_0^0 = 1$	$P_1^1 = -\sin \theta$
$P_1^0 = \cos \theta$	$P_2^1 = -3 \cos \theta \sin \theta$
$P_2^0 = \frac{1 + 3 \cos 2\theta}{4}$	$P_3^1 = \frac{3(1 - 5 \cos^2 \theta) \sin \theta}{2}$
$P_3^0 = \frac{-3 \cos \theta + 5 \cos^3 \theta}{2}$	$P_3^1 = \frac{5(3 \cos \theta - 7 \cos^3 \theta) \sin \theta}{2}$
$m = 2$	$m = 3 \text{ and } 4$
$P_2^2 = 3 \sin^2 \theta$	$P_3^3 = -15 \sin^3 \theta$
$P_3^2 = 15 \cos \theta \sin^2 \theta$	$P_4^3 = -105 \cos \theta \sin^3 \theta$
$P_4^2 = \frac{15(5 + 7 \cos 2\theta) \sin^2 \theta}{4}$	$P_4^4 = 105 \sin^4 \theta$

The spherical harmonics have the property of being orthonormal to each others:

$$\int_0^{2\pi} \int_0^\pi Y_n^{m'}(\theta, \phi) Y_n^{m*}(\theta, \phi) \sin \theta d\theta d\phi = \begin{cases} 1 & n = n', m = m' \\ 0 & \text{otherwise} \end{cases}. \quad (\text{A.21})$$

The square root term in the spherical harmonic definition, Eq. (A.17), is there so the integral above equals one when  $n = n'$  and  $m = m'$ . As for the associated Legendre functions, there is a relation between spherical harmonics of positive and negative values of  $m$  and same order  $n$ ,

$$Y_n^{-m}(\theta, \phi) = (-1)^m Y_n^{m*}(\theta, \phi). \quad (\text{A.22})$$

It can be found that the real part of  $Y_n^{-m}(\theta, \phi)$  is the real part of  $Y_n^m(\theta, \phi)$  multiplied by  $(-1)^m$  while the imaginary part of  $Y_n^{-m}(\theta, \phi)$  is the imaginary part of  $Y_n^m(\theta, \phi)$  multiplied by  $-1(-1)^m = (-1)^{m+1}$ .

For  $m = 0$ , the spherical harmonic has no imaginary part and it is symmetrical about the  $z$ -axis. The magnitude of the spherical harmonics is always symmetrical about the  $z$ -axis.

Expressions for spherical harmonics of orders up to 3 are given in Table A.3. Examples are plotted in Fig. A.2 to A.4. The positive and negative signs indicate the lobe polarities.

Table A.3. Spherical harmonics of orders up to 3.

$n = 0 \text{ and } 1$	$n = 2$
$Y_0^0(\theta, \phi) = \frac{1}{\sqrt{4\pi}}$	$Y_2^{-2}(\theta, \phi) = 3\sqrt{\frac{5}{96\pi}} \sin^2 \theta e^{-2i\phi}$
$Y_1^{-1}(\theta, \phi) = \sqrt{\frac{3}{8\pi}} \sin \theta e^{-i\phi}$	$Y_2^{-1}(\theta, \phi) = \frac{3}{2}\sqrt{\frac{5}{24\pi}} \sin 2\theta e^{-i\phi}$
$Y_1^0(\theta, \phi) = \sqrt{\frac{3}{4\pi}} \cos \theta$	$Y_2^0(\theta, \phi) = \sqrt{\frac{5}{16\pi}} (-1 + 3 \cos^2 \theta)$
$Y_1^1(\theta, \phi) = -\sqrt{\frac{3}{8\pi}} \sin \theta e^{i\phi}$	$Y_2^1(\theta, \phi) = -\frac{3}{2}\sqrt{\frac{5}{24\pi}} \sin 2\theta e^{i\phi}$
	$Y_2^2(\theta, \phi) = 3\sqrt{\frac{5}{96\pi}} \sin^2 \theta e^{2i\phi}$
$n = 3$	
$Y_3^{-3}(\theta, \phi) = \sqrt{\frac{35}{64\pi}} \sin^3 \theta e^{-3i\phi}$	$Y_3^1(\theta, \phi) = \sqrt{\frac{21}{64\pi}} (1 - 5 \cos^2 \theta) \sin \theta e^{i\phi}$
$Y_3^{-2}(\theta, \phi) = 15\sqrt{\frac{7}{480\pi}} \cos \theta \sin^2 \theta e^{-2i\phi}$	$Y_3^2(\theta, \phi) = 15\sqrt{\frac{7}{480\pi}} \cos \theta \sin^2 \theta e^{2i\phi}$
$Y_3^{-1}(\theta, \phi) = \sqrt{\frac{21}{256\pi}} (3 + 5 \cos 2\theta) \sin \theta e^{-i\phi}$	$Y_3^3(\theta, \phi) = -\frac{5}{8}\sqrt{\frac{7}{5\pi}} \sin^3 \theta e^{3i\phi}$
$Y_3^0(\theta, \phi) = \sqrt{\frac{7}{16\pi}} (-3 \cos \theta + 5 \cos^3 \theta)$	

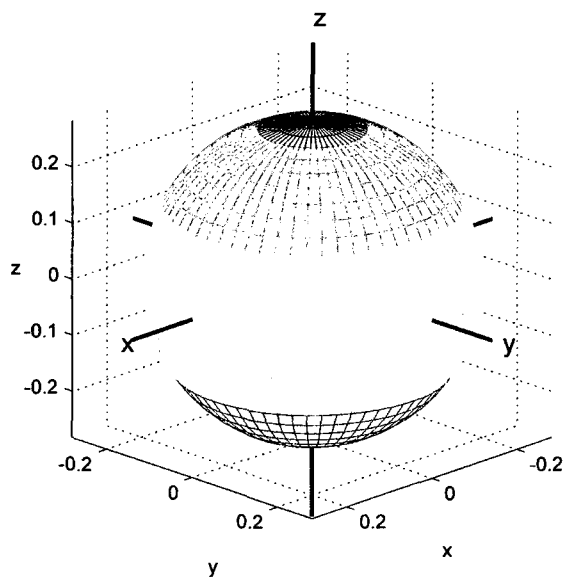


Figure A.2. Spherical harmonic  $Y_0^0(\theta, \phi) = \frac{1}{\sqrt{4\pi}}$ .

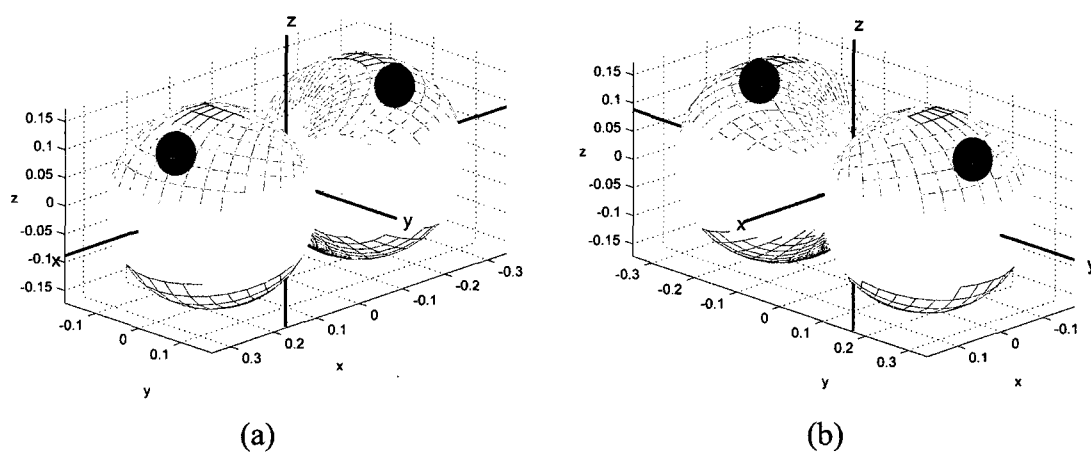


Figure A. 3. Spherical harmonic  $Y_1^{-1}(\theta, \phi) = \sqrt{\frac{3}{8\pi}} \sin \theta e^{-i\phi}$ . (a) real part; (b) imaginary part.

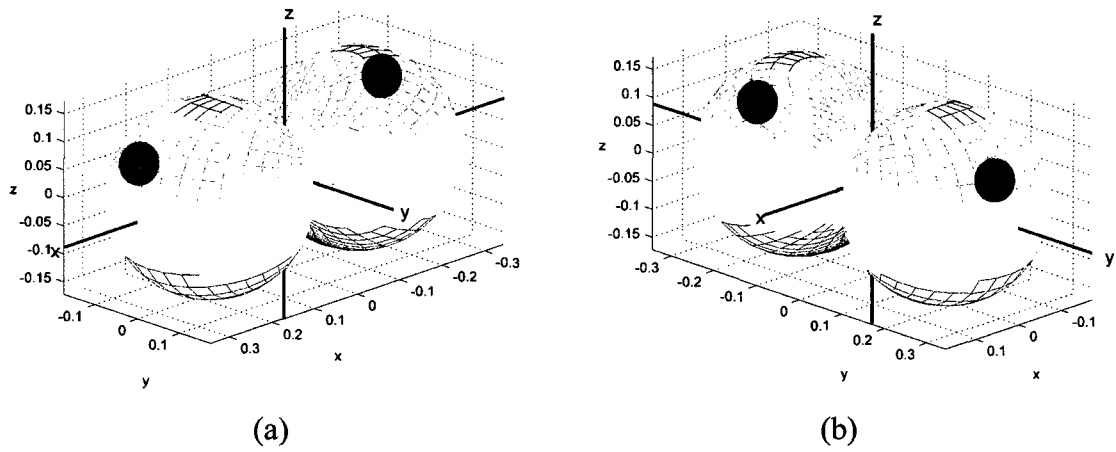


Figure A.4. Spherical harmonic  $Y_1^1(\theta, \phi) = -\sqrt{\frac{3}{8\pi}} \sin \theta e^{i\phi}$ . (a) real part; (b) imaginary part.



---

## Appendix *B*

---

### Theoretical Bias and Variance of the Gain-Optimized Estimator

To compute the bias and variance of the gain-optimized estimator, we assume that the noise field, and hence  $\mathbf{R}_{ee}(k)$ , is known. In practice however,  $\mathbf{R}_{ee}(k)$  would have to be estimated. In this case, the bias and variance would need to take into account the characteristics of the estimator of  $\mathbf{R}_{ee}(k)$ . We also assume that  $\mathbf{R}_{ee}(k)$  is constant.

#### B.1 Bias of the estimator

The expected value of the estimator must be calculated to then obtain its bias. The expected value is

$$E\{d_{nm}(k)\} = E\left\{\frac{1}{C_{nm}^{nm}(k)} y_{nm}(k)\right\}. \quad (\text{B.1})$$

Assuming that  $\mathbf{R}_{ee}^{-1}(k)$  is constant so that  $E\{C_{nm}^{nm}(k)\} = C_{nm}^{nm}(k)$ , we have

$$E\{d_{nm}(k)\} = \frac{1}{C_{nm}^{nm}(k)} E\{y_{nm}(k)\}. \quad (\text{B.2})$$

Replacing the sub-beamformer output by Eq. (4.40) leads to

$$E\left\{d_{nm}(k)\right\} = \frac{1}{C_{nm}^{nm}(k)} E\left\{ \begin{aligned} &C_{nm}^{nm}(k)s_{nm}(k) + \sum_{n'=0}^{\infty} \sum_{\substack{m'=-n' \\ (n',m') \neq (n,m)}}^{n'} C_{n'm'}^{nm}(k)s_{n'm'}(k) \\ &+ \mathbf{Y}_n^{mH}(k)\mathbf{R}_{ee}^{-1}(k)\mathbf{e}(k) \end{aligned} \right\}. \quad (\text{B.3})$$

Taking the expectation, we obtain

$$\begin{aligned} E\left\{d_{nm}(k)\right\} &= s_{nm}(k) + \frac{1}{C_{nm}^{nm}(k)} \sum_{n'=0}^{\infty} \sum_{\substack{m'=-n' \\ (n',m') \neq (n,m)}}^{n'} C_{n'm'}^{nm}(k)s_{n'm'}(k) \\ &+ \frac{1}{C_{nm}^{nm}(k)} \mathbf{Y}_n^{mH}(k)\mathbf{R}_{ee}^{-1}(k)E\{\mathbf{e}(k)\} \end{aligned} \quad (\text{B.4})$$

The element  $l$  of the noise vector  $\mathbf{e}(k)$  is

$$e_l(k) = \frac{1}{\sqrt{N}} \sum_{q=0}^{N-1} e_l(q)e^{-i2\pi kq/N}, \quad (\text{B.5})$$

where  $e_l(q)$  is the  $q^{\text{th}}$  time sample of the noise component at the output of the  $l^{\text{th}}$  microphone, and  $e$  is the exponential function. The expectation of element  $l$  in the vector  $\mathbf{e}(k)$  is

$$\begin{aligned} E\{e_l(k)\} &= E\left\{ \frac{1}{\sqrt{N}} \sum_{q=0}^{N-1} e_l(q)e^{-i2\pi kq/N} \right\} \\ &= \frac{1}{\sqrt{N}} \sum_{q=0}^{N-1} E\{e_l(q)\}e^{-i2\pi kq/N}. \end{aligned} \quad (\text{B.6})$$

The expected value of the noise at each microphone in the time-domain,  $E\{e_l(q)\}$ , is zero, so that  $E\{e_l(k)\} = 0$ . This leads to an expected value of the estimator

$$E\left\{d_{nm}(k)\right\} = s_{nm}(k) + \frac{1}{C_{nm}^{nm}(k)} \sum_{n'=0}^{\infty} \sum_{\substack{m'=-n' \\ (n',m') \neq (n,m)}}^{n'} C_{n'm'}^{nm}(k)s_{n'm'}(k). \quad (\text{B.7})$$

Using this result, the bias of the estimator is

$$E\left\{d_{nm}(k)\right\} - s_{nm}(k) = \frac{1}{C_{nm}^{nm}(k)} \sum_{n'=0}^{\infty} \sum_{\substack{m'=-n' \\ (n',m') \neq (n,m)}}^{n'} C_{n'm'}^{nm}(k) s_{n'm'}(k). \quad (\text{B.8})$$

This shows that the bias of the estimator for mode  $(n, m)$  is a weighted sum of the strengths  $s_{n'm'}(k)$  of the modes other than  $(n, m)$  present in the source. It also depends on the coefficients  $C_{n'm'}^{nm}(k)$  given by Eq. (4.39). Notice that the bias is independent of the average noise power over the array. In the specific case of the continuous spherical array with the source at its center discussed at the beginning of Chapter 4, and in the presence of white-noise so that  $\mathbf{R}_{ee}^{-1}(k)$  is the identity matrix, the coefficients  $C_{n'm'}^{nm}(k)$  (actually their equivalent for a continuous array which is a surface integral) would all be zero so that the estimator bias would also be zero.

## B.2 Variance of the estimator

The variance of the estimator is given by

$$\text{Var}\left\{d_{nm}(k)\right\} = E\left\{\left[d_{nm}(k) - E\left\{d_{nm}(k)\right\}\right]^2\right\}. \quad (\text{B.9})$$

Using Eqs. (4.40), (4.41), and (B.7) to replace the estimator and its mean, leads to

$$\text{Var}\left\{d_{nm}(k)\right\} = E\left\{\left[\begin{array}{l} \frac{1}{C_{nm}^{nm}(k)} \left( C_{nm}^{nm}(k) s_{nm}(k) + \sum_{\substack{n'=0 \\ (n',m') \neq (n,m)}}^{\infty} \sum_{m'=-n'}^{n'} C_{n'm'}^{nm}(k) s_{n'm'}(k) \right) \\ + \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k) \\ - \left( s_{nm}(k) + \frac{1}{C_{nm}^{nm}(k)} \sum_{\substack{n'=0 \\ (n',m') \neq (n,m)}}^{\infty} \sum_{m'=-n'}^{n'} C_{n'm'}^{nm}(k) s_{n'm'}(k) \right) \end{array} \right]^2\right\}. \quad (\text{B.10})$$

which simplifies, after collecting terms, to

$$\text{Var}\left\{d_{nm}(k)\right\} = E\left\{\left[\frac{1}{C_{nm}^{nm}(k)} \mathbf{Y}_n^{mH}(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k)\right]^2\right\}. \quad (\text{B.11})$$

Applying the square, the equation becomes

$$\text{Var}\left\{d_{nm}(k)\right\} = \frac{1}{C_{nm}^{nm}(k)C_{nm}^{nm*}(k)} E\left\{\left[\mathbf{Y}_n^{mH}(k)\mathbf{R}_{ee}^{-1}(k)\mathbf{e}(k)\right]\left[\mathbf{Y}_n^{mH}(k)\mathbf{R}_{ee}^{-1}(k)\mathbf{e}(k)\right]^*\right\}, \quad (\text{B.12})$$

which expands to

$$\text{Var}\left\{d_{nm}(k)\right\} = \frac{1}{C_{nm}^{nm}(k)C_{nm}^{nm*}(k)} \mathbf{Y}_n^{mH}(k)\mathbf{R}_{ee}^{-1}(k)E\left\{\mathbf{e}(k)\mathbf{e}^H(k)\right\}\mathbf{R}_{ee}^{-1}(k)\mathbf{Y}_n^m(k). \quad (\text{B.13})$$

Using Eq. (2.11),  $\mathbf{R}_{ee}(k) = E\left\{\mathbf{e}(k)\mathbf{e}^H(k)\right\}/\sigma_e^2(k)$ , and  $\mathbf{R}_{ee}(k)\mathbf{R}_{ee}^{-1}(k) = \mathbf{I}$  leads to

$$\text{Var}\left\{d_{nm}(k)\right\} = \frac{\sigma_e^2(k)}{C_{nm}^{nm}(k)C_{nm}^{nm*}(k)} \mathbf{Y}_n^{mH}(k)\mathbf{R}_{ee}^{-1}(k)\mathbf{Y}_n^m(k). \quad (\text{B.14})$$

By definition  $C_{nm}^{nm}(k) = \mathbf{Y}_n^{mH}(k)\mathbf{R}_{ee}^{-1}(k)\mathbf{Y}_n^m(k)$  and  $C_{nm}^{nm*}(k) = C_{nm}^{nm}(k)$  so that the variance finally is

$$\begin{aligned} \text{Var}\left\{d_{nm}(k)\right\} &= \frac{\sigma_e^2(k)}{C_{nm}^{nm}(k)} \\ &= \frac{\sigma_e^2(k)}{\mathbf{Y}_n^{mH}(k)\mathbf{R}_{ee}^{-1}(k)\mathbf{Y}_n^m(k)}. \end{aligned} \quad (\text{B.15})$$

This shows that the variance of the estimator is proportional to the average noise power over the array. To calculate the bias and variance, we have assumed that the noise field, and hence  $\mathbf{R}_{ee}(k)$ , is known; in practice however,  $\mathbf{R}_{ee}(k)$  would have to be estimated. In which case, the bias and variance would need to take into account the characteristics of the estimator of  $\mathbf{R}_{ee}(k)$ .

Note that, as shown in Eq. (2.2), a normalized FFT is used to obtain the frequency-domain microphone outputs so that the variance of the estimator is independent of the FFT length. If the regular FFT had been used instead,  $\sigma_e^2(k)$  would be proportional to the FFT length  $N$  while  $\mathbf{R}_{ee}(k)$  would be independent of  $N$ , leading to an estimator variance depending on the FFT length.

---

## Appendix C

---

### Less Promising Approaches to Mode Strength Estimation

This appendix presents three approaches that lead to estimators of the mode strengths that are deemed relatively unsatisfactory. The first two approaches lead to the microphone output vector as the source model and the third approach simply does not work because the modes composing the source are correlated.

#### C.1 Orthogonalization of spherical harmonic modes

With an arbitrary geometry array, the spherical harmonic modes  $\mathbf{Y}_n^m(k)$  are not orthogonal to each other. In this section, we will apply an orthonormalization to the matrix  $\mathbf{Y}(k)$  containing the spherical harmonic modes to obtain new modes. Those new modes will be used to perform the beamforming.

We will begin again with Eq. (6.7),

$$\mathbf{p}(k) = \mathbf{Y}'(k) \mathbf{s}'_{\text{str}}(k) + \sum_{n'=N_{\text{order}}+1}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_{n'}^{m'}(k) + \mathbf{e}(k). \quad (\text{C.1})$$

Assuming that the number of rows ( $L$ , the number of microphones) of the matrix  $\mathbf{Y}'(k)$  is at least equal to its number of column  $M$ , and that the columns are independent, it is

possible to derive a set of  $M$  orthonormal vectors using the QR factorization method (i.e. Gram-Schmidt) so that  $\mathbf{Y}'(k) = \mathbf{Q}'(k)\mathbf{R}'(k)$ . Those  $M$  orthonormal vectors are the columns of an  $L \times M$  matrix  $\mathbf{Q}'$ . The prime indicates that the orthonormalization is performed on  $\mathbf{Y}'(k)$ .  $\mathbf{R}'(k)$  is a  $M \times M$  upper triangular matrix (and it is not a correlation matrix even though the letter 'R' is used).

If  $\mathbf{Y}'(k)$  has a rank less than  $M$ , it spans a space with less than  $M$  dimensions. However, it is still possible to find  $M$  orthonormal vectors (the columns of  $\mathbf{Q}'$ ) that do span a space with  $M$  dimensions.  $\mathbf{Y}'(k)$  can still be expressed as  $\mathbf{Y}'(k) = \mathbf{Q}'(k)\mathbf{R}'(k)$ .

The matrix containing the  $M$  spherical harmonic mode vectors is written as

$$\mathbf{Y}'(k) = \mathbf{Q}'(k)\mathbf{R}'(k), \quad (\text{C.2})$$

which is then used in the microphone output vector whose expression becomes

$$\begin{aligned} \mathbf{p}(k) &= \{\mathbf{Q}'(k)\mathbf{R}'(k)\}\mathbf{s}'_{\text{str}}(k) + \sum_{n'=N_{\text{order}}+1}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_{n'}^{m'}(k) + \mathbf{e}(k) \\ &= \mathbf{Q}'(k)\{\mathbf{R}'(k)\mathbf{s}'_{\text{str}}(k)\} + \sum_{n'=N_{\text{order}}+1}^{\infty} \sum_{m'=-n'}^{n'} s_{n'm'}(k) \mathbf{Y}_{n'}^{m'}(k) + \mathbf{e}(k). \end{aligned} \quad (\text{C.3})$$

The columns of  $\mathbf{Q}'(k)$  are new modes (instead of spherical harmonics) and the vector  $\mathbf{R}'(k)\mathbf{s}'_{\text{str}}(k)$  contains the strength of those modes.

As in Section 6.3, the source model is defined as

$$\mathbf{s}_{\text{mo}}(k) = \mathbf{Y}(k)\hat{\mathbf{s}}_{\text{str}}(k). \quad (\text{C.4})$$

$\mathbf{Y}(k)$  is computed from the microphone positions and the prior knowledge (or estimation) of the source position. Using the QR factorization on  $\mathbf{Y}(k)$ , the source model becomes

$$\mathbf{s}_{\text{mo}}(k) = \mathbf{Q}(k)\mathbf{R}(k)\hat{\mathbf{s}}_{\text{str}}(k). \quad (\text{C.5})$$

Instead of estimating  $\hat{\mathbf{s}}_{\text{str}}(k)$  and then coming back to  $\mathbf{Y}(k)$  to perform the beamforming, a new estimate is defined as

$$\hat{\mathbf{s}}_{\text{str,QR}}(k) = \mathbf{R}(k)\hat{\mathbf{s}}_{\text{str}}(k) \quad (\text{C.6})$$

and the beamforming is realized with  $\mathbf{Q}(k)$ . The source model is then

$$\mathbf{s}_{\text{mo, QR}}(k) = \mathbf{Q}(k)\hat{\mathbf{s}}_{\text{str, QR}}(k). \quad (\text{C.7})$$

The estimation is performed from the received microphone output vector, which is now written as

$$\begin{aligned} \mathbf{p}(k) &= \mathbf{s}_{\text{mo, QR}}(k) + \boldsymbol{\varepsilon}(k) \\ &= \mathbf{Q}(k)\hat{\mathbf{s}}_{\text{str, QR}}(k) + \boldsymbol{\varepsilon}(k), \end{aligned} \quad (\text{C.8})$$

where  $\boldsymbol{\varepsilon}(k)$  is an error vector which regroups the noise vector, the components from modes of orders above  $N_{\text{order}}$  not included in the model, as well as errors caused by the difference between the exact mode vectors in  $\mathbf{Q}'(k)$  and the mode vectors in  $\mathbf{Q}(k)$  used for beamforming.

With the least-squares method as developed in Section 6.3 and using the fact that the columns of  $\mathbf{Q}(k)$  are orthonormal so that  $\mathbf{Q}^H(k)\mathbf{Q}(k)$  is the identity matrix, the solution for the estimate is

$$\hat{\mathbf{s}}_{\text{str, QR}}(k) = \mathbf{Q}^H(k)\mathbf{p}(k). \quad (\text{C.9})$$

Substituting for the estimate, the source model finally is

$$\begin{aligned} \mathbf{s}_{\text{mo, QR}}(k) &= \mathbf{Q}(k)\mathbf{Q}^H(k)\mathbf{p}(k) \\ &= \mathbf{p}(k) \end{aligned} \quad (\text{C.10})$$

Unfortunately, this is not an interesting result, and we would expect this source model to provide very poor performance when there are interferers or reflections, because they would be directly part of the source model. The noise is also directly a part of the source model with this approach.

## C.2 Modes set with zero-bias estimator

Section 4.2 showed how a continuous spherical array using the spherical harmonic function of order  $n$  and degree  $m$  to beamform on a source located at its center has for output only the component of the source with the same order and degree. If

the source does not have a component of this order and degree, the corresponding output would be zero (except for noise, interferers, and reverberation).

This result comes because of the orthogonality property of the spherical harmonics. It can not be used if the source is not located at the center of a continuous spherical array or if the gain optimization is with respect to a cross-power spectral density matrix other than the identity matrix.

As seen in Eq. (4.41), the estimator of the source mode strength for mode  $(n, m)$  uses the sub-beamformer output of the same mode. The components of this sub-beamformer are given in Eq. (4.40), which is repeated here for convenience,

$$y_{nm}(k) = C_{nm}^{nm}(k)s_{nm}(k) + \sum_{\substack{n'=0 \\ (n',m') \neq (n,m)}}^{\infty} \sum_{m'=-n'}^{n'} C_{n'm'}^{nm}(k)s_{n'm'}(k) + \mathbf{Y}_n^{mH}(k)\mathbf{R}_{ee}^{-1}(k)\mathbf{e}(k). \quad (\text{C.11})$$

As can be seen in the development of Appendix B, the complex coefficients  $C_{n'm'}^{nm}(k)$  in the second term causes the bias of the estimator. Those coefficients are defined in Eq. (4.39) as

$$C_{n'm'}^{nm}(k) = \mathbf{Y}_n^{mH}(k)\mathbf{R}_{ee}^{-1}(k)\mathbf{Y}_{n'}^{m'}(k). \quad (\text{C.12})$$

It was shown in Sub-Section 4.3.1 that the method is not restricted to using spherical harmonics. If the source signal can be decomposed into a set of modes  $\mathbf{M}_m(k)$  as

$$\mathbf{s}(k) = \sum_{m=0}^M s_m(k)\mathbf{M}_m(k), \quad (\text{C.13})$$

the  $m^{\text{th}}$  sub-beamformer is defined as

$$y_m(k) = \mathbf{M}_m^H(k)\mathbf{R}_{ee}^{-1}(k)\mathbf{p}(k) \quad (\text{C.14})$$

and the optimum array beamformer output is

$$y(k) = \sum_{m=0}^M s_m^*(k)y_m(k). \quad (\text{C.15})$$



In a development similar to what was done for spherical harmonics, the components of the  $m^{\text{th}}$  sub-beamformer output are

$$y_m(k) = C_m^m(k)s_m(k) + \sum_{\substack{m'=0 \\ m' \neq m}}^M C_{m'}^m(k)s_{m'}(k) + \mathbf{M}_m^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{e}(k), \quad (\text{C.16})$$

where a new complex coefficient is defined

$$C_{m'}^m(k) = \mathbf{M}_m^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{M}_{m'}(k). \quad (\text{C.17})$$

If a set of modes  $\mathbf{M}_{\text{ZB},m}(k)$  could be found such that

$$C_{m'}^m(k) = \begin{cases} 1 & m = m' \\ 0 & \text{otherwise} \end{cases}, \quad (\text{C.18})$$

the  $m^{\text{th}}$  sub-beamformer output would only have a component of the source of the same order (except for noise, interferers, or reverberation). The estimator would not be biased by modes  $m' \neq m$  present in the source. This is why we use the name “zero-bias estimator”. The estimator would still be subject to interferers and reverberation.

### **Computing the modes**

The task is then to find the modes  $\mathbf{M}_{\text{ZB},m}(k)$  such that the condition stated in Eq. (C.18) is met. The still-unknown modes are first combined in a  $L \times M$  matrix

$$\mathbf{M}_{\text{ZB}}(k) = [\mathbf{M}_{\text{ZB},1}(k), \mathbf{M}_{\text{ZB},2}(k), \dots, \mathbf{M}_{\text{ZB},M}(k)] \quad (\text{C.19})$$

and the condition of orthonormality is written in a matrix form

$$\mathbf{M}_{\text{ZB}}^H(k) \mathbf{R}_{ee}^{-1}(k) \mathbf{M}_{\text{ZB}}(k) = \mathbf{I}_{M \times M}. \quad (\text{C.20})$$

To solve for the matrix  $\mathbf{M}_{\text{ZB}}(k)$ , we write the inverse of the noise cross-power spectral density matrix as  $\mathbf{R}_{ee}^{-1}(k) = \mathbf{R}_{ee}^{-1/2}(k) \mathbf{R}_{ee}^{-1/2}(k)$ , the product of two identical matrices so that the condition becomes

$$\{\mathbf{M}_{\text{ZB}}^H(k) \mathbf{R}_{ee}^{-1/2}(k)\} \{\mathbf{R}_{ee}^{-1/2}(k) \mathbf{M}_{\text{ZB}}(k)\} = \mathbf{I}_{M \times M}. \quad (\text{C.21})$$

Using the fact that  $\mathbf{R}_{ee}^{-1/2H}(k) = \mathbf{R}_{ee}^{-1/2}(k)$ , we rewrite the condition as

$$\{\mathbf{R}_{ee}^{-1/2}(k)\mathbf{M}_{ZB}(k)\}^H \{\mathbf{R}_{ee}^{-1/2}(k)\mathbf{M}_{ZB}(k)\} = \mathbf{I}_{M \times M}. \quad (\text{C.22})$$

After defining a  $L \times M$  matrix  $\mathbf{A} = \mathbf{R}_{ee}^{-1/2}(k)\mathbf{M}_{ZB}(k)$ , we observe that the condition is

$$\mathbf{A}^H \mathbf{A} = \mathbf{I}_{M \times M}. \quad (\text{C.23})$$

Thus  $\mathbf{A}$  is a unitary matrix and its columns are orthonormal.

We know that using the QR factorization method (i.e. Gram-Schmidt), a matrix  $\mathbf{B}$  can be written as

$$\mathbf{B} = \mathbf{Q}\mathbf{R}, \quad (\text{C.24})$$

where  $\mathbf{Q}$  is a matrix whose columns are orthonormal and of the same size as  $\mathbf{B}$ .  $\mathbf{R}$  is a square, upper triangular matrix, with the same number of columns as  $\mathbf{B}$ . Note that  $\mathbf{R}$  is not a correlation matrix even though the letter ‘R’ is used. Post-multiplying both sides by  $\mathbf{R}^{-1}$ , we obtain

$$\mathbf{B}\mathbf{R}^{-1} = \mathbf{Q}. \quad (\text{C.25})$$

The matrix  $\mathbf{Q}$  is a unitary matrix as is the matrix  $\mathbf{A}$  above. So comparing  $\mathbf{R}_{ee}^{-1/2}(k)\mathbf{M}_{ZB}(k) = \mathbf{A}$  to  $\mathbf{B}\mathbf{R}^{-1} = \mathbf{Q}$ , we deduce that to find  $\mathbf{M}_{ZB}(k)$ , a QR factorization is performed on  $\mathbf{R}_{ee}^{-1/2}(k)$  from which  $\mathbf{M}_{ZB}(k)$  is obtained

$$\begin{aligned} \mathbf{R}_{ee}^{-1/2}(k) &= \mathbf{Q}\mathbf{R} \\ \mathbf{M}_{ZB}(k) &= \mathbf{R}^{-1}. \end{aligned} \quad (\text{C.26})$$

$\mathbf{M}_{ZB}(k)$  is a  $L \times L$  matrix so that the number of modes  $M$  is equal to the number of microphones  $L$ .

The modes  $\mathbf{M}_{ZB,m}(k)$  depend on the noise cross-power spectral density matrix and on the microphone positions through this same matrix. The modes do not depend on the source position. We could thus expect poor performance in the presence of interferers or reverberation with this approach.

### **Bias-free estimator**

To verify that those modes lead to a bias-free estimator, the  $M$  sub-beamformer outputs are combined into a column vector

$$\mathbf{y}_{\text{ZB}}(k) = [y_1(k), y_2(k), \dots, y_M(k)]^T \quad (\text{C.27})$$

so that the  $M$  sub-beamformer outputs are computed at once from the microphone output vector as

$$\mathbf{y}_{\text{ZB}}(k) = \mathbf{M}_{\text{ZB}}^H(k) \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{p}(k). \quad (\text{C.28})$$

The source signal is written using the matrix notation

$$\mathbf{s}(k) = \mathbf{M}'_{\text{ZB}}(k) \mathbf{s}'_{\text{str,ZB}}(k), \quad (\text{C.29})$$

where  $\mathbf{s}'_{\text{str,ZB}}(k)$  contains the  $M = L$  source mode strengths and the primes indicate that this is a decomposition for the source signal. The sub-beamformer outputs are

$$\begin{aligned} \mathbf{y}_{\text{ZB}}(k) &= \mathbf{M}_{\text{ZB}}^H(k) \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{p}(k) \\ &= \mathbf{M}_{\text{ZB}}^H(k) \mathbf{R}_{\text{ee}}^{-1}(k) \{ \mathbf{s}(k) + \mathbf{e}(k) \} \\ &= \mathbf{M}_{\text{ZB}}^H(k) \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{M}'_{\text{ZB}}(k) \mathbf{s}'_{\text{str,ZB}}(k) + \mathbf{M}_{\text{ZB}}^H(k) \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{e}(k). \end{aligned} \quad (\text{C.30})$$

Since the modes  $\mathbf{M}_{\text{ZB}}(k)$  do not depend on the source location and are computed with the same  $\mathbf{R}_{\text{ee}}^{-1}(k)$  matrix used in the beamforming,  $\mathbf{M}_{\text{ZB}}(k)$  equals  $\mathbf{M}'_{\text{ZB}}(k)$ . We can also drop the prime for the source mode strengths. The sub-beamformer outputs are

$$\begin{aligned} \mathbf{y}_{\text{ZB}}(k) &= \mathbf{M}_{\text{ZB}}^H(k) \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{M}(k) \mathbf{s}_{\text{str,ZB}}(k) + \mathbf{M}_{\text{ZB}}^H(k) \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{e}(k) \\ &= \mathbf{s}_{\text{str,ZB}}(k) + \mathbf{M}_{\text{ZB}}^H(k) \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{e}(k), \end{aligned} \quad (\text{C.31})$$

where we have used  $\mathbf{M}_{\text{ZB}}^H(k) \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{M}_{\text{ZB}}(k) = \mathbf{I}$  to go from the first to the second line. As was desired, the  $m^{\text{th}}$  sub-beamformer output does not depend on modes other than the  $m^{\text{th}}$  mode (the  $m^{\text{th}}$  element of  $\mathbf{y}_{\text{ZB}}(k)$  does not depend on any elements of  $\mathbf{s}_{\text{str,ZB}}(k)$  other than the  $m^{\text{th}}$  element).

The bias-free estimator of the source mode strength for mode  $\mathbf{M}_{\text{ZB},m}(k)$  is

$$d_{\text{ZB},m}(k) = \mathbf{M}_{\text{ZB},m}^H(k) \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{p}(k). \quad (\text{C.32})$$

The source model resulting from  $\mathbf{M}_{\text{ZB}}(k)$  and the estimation is

$$\begin{aligned} \mathbf{s}_{\text{mo,ZB}}(k) &= \mathbf{M}_{\text{ZB}}(k) \mathbf{y}_{\text{ZB}}(k) \\ &= \mathbf{M}_{\text{ZB}}(k) \mathbf{M}_{\text{ZB}}^H(k) \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{p}(k). \end{aligned} \quad (\text{C.33})$$

After substituting for  $\mathbf{M}_{\text{ZB}}(k) = \mathbf{R}^{-1}$  and using Eq. (C.26), the source model is

$$\mathbf{s}_{\text{mo,ZB}}(k) = \mathbf{p}(k). \quad (\text{C.34})$$

Unfortunately, this leads again to the microphone output vector as the source model. It is not an interesting result, and we would expect this source model to provide very poor performance when there are interferers or reflections, because they would be directly part of the source model. The noise would also directly be a part of the source model.

### C.3 MVDR to estimate the mode strengths

The use of Minimum Variance Distortionless Response (MVDR) beamformer was also attempted to estimate the mode strengths. As previously seen, MVDR is a special case of Linearly Constrained Minimum Variance beamforming (LCMV, see Section 2.4.2) when there is a single constraint with a desired response of one. The equations are the same as for the Maximum-Likelihood Method (MLM) of Capon [Capon 1969].

The idea to use the MVDR for weight estimation in the proposed beamforming method was based on the thought that while the strength of spatial mode  $(n, m)$  would be estimated, the power from all the other modes would be minimized. To estimate the strength of spatial mode  $(n, m)$ , the constraint would be

$$\mathbf{Y}_n^{mH}(k) \mathbf{w}_{\text{MVDR},mm}(k) = 1. \quad (\text{C.35})$$

The beamforming weight vector  $\mathbf{w}_{\text{MVDR},nm}(k)$  would then be found so as to minimize the array output power under the constraint

$$\min_{\mathbf{w}_{\text{MVDR},nm}(k)} \mathbf{w}_{\text{MVDR},nm}^H(k) \mathbf{R}_{\text{pp}}(k) \mathbf{w}_{\text{MVDR},nm}(k) \quad \text{subject to} \quad \mathbf{Y}_n^{mH}(k) \mathbf{w}_{\text{MVDR},nm}(k) = 1, \quad (\text{C.36})$$

whose solution is

$$\mathbf{w}_{\text{MVDR},nm}^H(k) = \frac{\mathbf{R}_{\text{pp}}^{-1}(k) \mathbf{Y}_n^m(k)}{\mathbf{Y}_n^{mH}(k) \mathbf{R}_{\text{pp}}^{-1}(k) \mathbf{Y}_n^m(k)}. \quad (\text{C.37})$$

The power in the mode would then be estimated by substituting Eq. (C.37) in

$$P_{\text{MVDR},nm}(k) = \mathbf{w}_{\text{MVDR},nm}^H(k) \mathbf{R}_{\text{pp}}(k) \mathbf{w}_{\text{MVDR},nm}(k), \quad (\text{C.38})$$

which gives

$$P_{\text{MVDR},nm}(k) = \frac{1}{\mathbf{Y}_n^{mH}(k) \mathbf{R}_{\text{pp}}^{-1}(k) \mathbf{Y}_n^m(k)}. \quad (\text{C.39})$$

As can be seen from this equation, the power could actually be estimated directly, without having to first compute  $\mathbf{w}_{\text{MVDR},nm}(k)$  with Eq. (C.37).

The first problem with this MVDR approach is that the mode strengths,  $s_{nm}(k)$ , are actually complex numbers while  $P_{\text{MVDR},nm}(k)$  is an estimated power. The phase of the complex number is lost.

However, even the power in each mode is not estimated properly. For example, at high input signal-to-noise ratio and when the source is composed of more than a single spatial mode, the estimated power of the modes actually present in the source are all very low, and similar to the estimated power of modes not composing the source. The cross-power spectral density matrix  $\mathbf{R}_{\text{pp}}(k)$  only has one eigenvalue much larger than the other eigenvalues associated with the noise. The modes present in the source are not independent ‘‘sources’’ which would lead to distinct eigenvectors in  $\mathbf{R}_{\text{pp}}(k)$ .

---

We have tried estimating  $\mathbf{R}_{pp}(k)$  with many FFT frames in the fictitious case where the magnitude of the mode strengths stays the same but the phase of each mode strength changes randomly from frame to frame (having for effect to decorrelate the modes). For example, with a source composed of two modes,  $\mathbf{R}_{pp}(k)$  had two eigenvalues much larger than the others. The estimation of the modes power with MVDR as described in this sub-section worked and gave an estimated power of the modes present in the source much larger than the estimated power of the modes not present. The issue of bias and modes looking similar on the array is not an issue here because MVDR causes a decrease of the power at the array output for a mode that is even only slightly different from the model. This is a fictitious case because if the phase of the spatial modes with respect to each other changed from frame to frame, then the directivity of the source would also change from frame to frame. This would indicate that the mode strengths need to be estimated independently for each frame before performing the beamforming.

---

## Appendix *D*

---

### **Comparison of Computational Complexity Between the Proposed Beamforming Method with the GO Estimator and Conventional Beamforming**

This appendix compares the computational complexity between the proposed beamforming method with the GO estimator and conventional beamforming. The complexity of computing the beamformer output with the mode vectors is compared first. This is followed by a comparison of the complexity of computing the mode vectors, which include the Hankel function of the second kind and the spherical harmonic. We also compare the run-time for our Matlab code for conventional beamforming method and the proposed beamforming method for varying values of  $N_{\text{order}}$ .

#### **D.1 Comparison of computational complexity to obtain the beamformer output**

This section examines the computational complexity to perform the beamforming part of the processing, assuming that the mode vectors are already available.

##### ***Conventional beamforming***

The beamforming weight vector for conventional beamforming is given by Eq. (7.20), repeated here for convenience

$$\mathbf{w}_{\text{conv}}(k) = \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{Y}_0^0(k). \quad (\text{D.1})$$

As a reminder, the array has  $L$  microphones,  $\mathbf{R}_{\text{ee}}^{-1}(k)$  is an  $L \times L$  matrix, and  $\mathbf{Y}_0^0(k)$  has  $L$  elements. To compute one element of  $\mathbf{w}_{\text{conv}}(k)$ ,  $L$  complex multiplications and  $L-1$  complex additions are required. To compute the entire beamforming weight vector  $\mathbf{w}_{\text{conv}}(k)$ ,  $L^2$  complex multiplications and  $L(L-1)$  complex additions are necessary.

The output for conventional beamforming is

$$y_{\text{conv}}(k) = \mathbf{w}_{\text{conv}}^H(k) \mathbf{p}(k), \quad (\text{D.2})$$

which requires  $L$  complex multiplications and  $L-1$  complex additions. The total number of operations to obtain both  $y_{\text{conv}}(k)$  and  $\mathbf{w}_{\text{conv}}(k)$  is  $L(L+1)$  complex multiplications and  $(L+1)(L-1)$  complex additions.

### ***Proposed beamforming method***

The output of the sub-beamformer of order  $n$  and degree  $m$  is defined by Eq. (4.16), repeated here

$$y_{nm}(k) = \mathbf{Y}_n^{mH}(k) \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{p}(k). \quad (\text{D.3})$$

The strength of this mode can be estimated with the GO estimator, provided by the first line of Eq. (4.41)

$$d_{nm}(k) = \frac{1}{\mathbf{Y}_n^{mH}(k) \mathbf{R}_{\text{ee}}^{-1}(k) \mathbf{Y}_n^m(k)} y_{nm}(k), \quad (\text{D.4})$$

where we have substituted for  $C_{nm}^{nm}(k)$  using Eq. (4.39).

The sub-beamformer output  $y_{nm}(k)$  appears in both equations and can be computed only once.  $\mathbf{Y}_n^{mH}(k) \mathbf{R}_{\text{ee}}^{-1}(k)$  appears in both equations and has to be calculated only once as well. It has the same computational complexity as computing  $\mathbf{w}_{\text{conv}}(k)$  in Eq. (D.1),  $L^2$  complex multiplications and  $L(L-1)$  complex additions.



Once  $\mathbf{Y}_n^{mH}(k)\mathbf{R}_{ee}^{-1}(k)$  is computed,  $L$  complex multiplications and  $L-1$  complex additions are required to obtain  $y_{nm}(k)$ . To compute the denominator of  $d_{nm}(k)$ ,  $L$  complex multiplications and  $L-1$  complex additions are also necessary. The division of  $y_{nm}(k)$  by the denominator, a real scalar, is then required to obtain  $d_{nm}(k)$ .

To summarize, for a given order  $n$  and degree  $m$ ,  $L(L+2)$  complex multiplications,  $(L+2)(L-1)$  complex additions, and 1 division by a scalar are required to obtain the sub-beamformer output and the GO estimation of the mode strength.

The overall output for the proposed beamforming method is given by Eq (4.50)

$$y(k) = \sum_{n=0}^{N_{\text{order}}} \sum_{m=-n}^n d_{nm}^*(k) y_{nm}(k). \quad (\text{D.5})$$

Including modes of order up to  $N_{\text{order}}$ , there are a total of  $(N_{\text{order}} + 1)^2$  modes. To obtain  $y(k)$ ,  $(N_{\text{order}} + 1)^2$  complex multiplications are required for the product of the estimated mode strengths and the sub-beamformer outputs.  $(N_{\text{order}} + 1)^2 - 1$  complex additions follow to realize the double summation. This leads to a total of  $[L(L+2)+1](N_{\text{order}} + 1)^2$  complex multiplications,  $[(L+2)(L-1)+1](N_{\text{order}} + 1)^2 - 1$  complex additions, and  $(N_{\text{order}} + 1)^2$  divisions by a scalar. Comparison of the number of operations for conventional beamforming and the proposed beamforming method with the GO estimator are shown in Table D.1. The approximate factor of increase of complexity is also included.

Table D.1. Comparison of computational complexity to obtain the beamformer output.

Operations	Conv. beamf.	Prop. method + GO	Factor
Complex mult.	$L(L+1)$	$[L(L+2)+1](N_{\text{order}}+1)^2$	$\approx (N_{\text{order}}+1)^2$
Complex add.	$(L+1)(L-1)$	$[(L+2)(L-1)+1](N_{\text{order}}+1)^2 - 1$	$\approx (N_{\text{order}}+1)^2$
Divisions		$(N_{\text{order}}+1)^2$	

### ***Run-time comparison***

Matlab is used here as for all simulations in the thesis. The scenario is for a dipole source and the rectangular array of 32 microphones. The geometry was shown in Fig. 7.19. Table D.2 shows the average time to compute 10000 beamformer outputs with conventional beamforming and the proposed method with the GO estimator for different values of  $N_{\text{order}}$ . In each case, five runs were performed to obtain the average. Also included is the relative run-time with respect to conventional beamforming as well as values for  $(N_{\text{order}}+1)^2$  (which, for  $L=32$  and  $N_{\text{order}}$  from 1 to 6, are within a value of 2 of the exact values given in Table D.1). We attribute the difference between the two factors to the way Matlab performs the operations.

Table D.2. Run-time comparison for the computation of 10000 beamformer outputs. The indicated run-time is the average of 5 runs.

Approach	Run-time	Factor	$(N_{\text{order}} + 1)^2$
	s		
Conventional beamforming	0.062	1	1
Prop. method, $N_{\text{order}} = 1$	0.49	7.9	4
Prop. method, $N_{\text{order}} = 2$	1.01	16.3	9
Prop. method, $N_{\text{order}} = 3$	1.77	28.5	16
Prop. method, $N_{\text{order}} = 4$	2.73	44.0	25
Prop. method, $N_{\text{order}} = 5$	3.86	62.3	36
Prop. method, $N_{\text{order}} = 6$	5.23	84.4	49

## D.2 Comparison of computational complexity to obtain the mode vectors

The mode vector was defined in Eq. (4.11), which is

$$\mathbf{Y}_n^m(k) = [h_n^{(2)}(\kappa_k r_1) Y_n^m(\theta_1, \phi_1) \quad \dots \quad h_n^{(2)}(\kappa_k r_L) Y_n^m(\theta_L, \phi_L)]^T. \quad (\text{D.6})$$

Each element of the vector is the product of the spherical Hankel function of the second kind (depending on the distance from the corresponding microphone to the source where the origin of the spherical coordinate system is located) and the spherical harmonic (depending on the coordinates  $(\theta, \phi)$  for the corresponding microphone). All degrees of mode vectors of a same order use the same Hankel function. For example, there are five mode vectors of order 2 but the Hankel function of second kind of order 2 only needs to be computed once.

### Spherical Hankel functions

The expressions for the Spherical Hankel functions of the second kind of orders 0 to 3 are provided in Table A.1. They are function of the parameter  $x$ , which in our case is equal to  $\kappa_k r_l$ .  $\kappa_k = 2\pi / \lambda_k$  is the wave number at the center of frequency bin  $k$ .

The complex exponential  $e^{-ix}$  is part of the spherical Hankel functions of all orders, including order 0. It would have been computed for conventional beamforming so there is no increase in computational complexity associated to it for the proposed beamforming method. The added computational complexity comes from the multiplicative factor. Furthermore, powers of  $1/x$  computed for a given order of an Hankel function would be available to Hankel functions of higher orders.

Once  $e^{-ix}$  is available, one complex multiplication and one division are required to compute  $h_0^{(2)}(x)$ . We will use this as the reference when comparing the proposed beamforming method. For  $h_1^{(2)}(x)$ , four complex multiplications, one division, and one addition are necessary. For  $h_2^{(2)}(x)$ , five complex multiplications, one division, and two addition are needed. For  $h_3^{(2)}(x)$ , six complex multiplications, one division, and three addition are required, and so on for higher orders. We need to multiply these values by  $L$ , the number of microphones. Table D.3 summarizes the number of operations.

Table D.3. Comparison of computational complexity to obtain the spherical Hankel functions of second kind.

Operations	Conv. beamf.	Prop. method + GO
Complex mult.	$L$	$L\{(N_{\text{order}} + 1)(N_{\text{order}} + 6)/2 - 2\}$
Complex add.		$LN_{\text{order}}(N_{\text{order}} + 1)/2$
Divisions	$L$	$L(N_{\text{order}} + 1)$

### **Spherical harmonics**

The definition of the spherical harmonic function is given in Eq. (A.17) and expressions for spherical harmonics of orders up to 3 are provided in Table A.3.

Conventional beamforming uses the spherical harmonic of order 0 which is independent of angle and does not have to be computed. The mode vector is formed directly with the  $L$  spherical Hankel function of the second kind of order 0.

To estimate the computational complexity for higher order spherical harmonics, we will make a number of assumptions. Constant factors are pre-computed. All operations, including the computation of trigonometric functions and the complex exponential, have the same computational complexity. Functions that appear a number of times with the same argument only have to be computed once. By observing the expressions in Table A.3, we will assume that 10 operations are required to obtain the spherical harmonics of order 1 for one microphone, that 20 are necessary for order 2, 30 for order 3, and so on. The number of operations needed to obtain the spherical harmonics for the proposed beamforming method would then be  $10LN_{\text{order}}(N_{\text{order}} + 1)/2$ , as indicated in Table D.4.

Table D.4. Comparison of computational complexity to obtain the spherical harmonic functions.

	Conv. beamf.	Prop. method + GO
Operations.		$10LN_{\text{order}}(N_{\text{order}} + 1)/2$

Finally, to build the mode vectors, multiplication of the Hankel functions and spherical harmonics for all orders are required, that is  $L(N_{\text{order}} + 1)^2$  multiplications.

With the assumptions we made, the number of operations to compute the mode vectors is proportional to  $L$  and  $N_{\text{order}}^2$ . From Table D.1, the number of operations to perform the beamforming is proportional to  $L^2$  and  $N_{\text{order}}^2$  so one would expect for the beamforming part of the proposed method to be the dominant component of the run-time.

### **Run-time comparison**

Using the same array and source geometry as in the previous section, Table D.5 shows the average time to compute the mode vectors required to beamform 1000 times. The beamforming part is not performed here. We note that it takes much more time to compute the mode vectors than it does to perform the beamforming. For example, it takes 0.38 s to compute the mode vector 1000 times for conventional beamforming while the beamforming part took 0.062 s for 10000 times (see Table D.2), a ratio of 61. This can be explained as follows. To compute the mode vectors, we wrote functions which must be called and these functions use Matlab functions such as “besselj”, “bessely”, “legendre”, and “repmat” in addition to one “if-end” condition. Also, each order is computed independently without using previous computations. Obviously, the assumptions we made in the theoretical development of the computational complexity for the spherical harmonics are not respected and computation of the mode vectors is the dominant factor.

Table D.5. Run-time comparison for the computation of the mode vectors necessary to beamforming 1000 times. The indicated run-time is the average of 5 runs.

Approach	Run-time	Factor
	s	
Conventional beamforming	0.38	1
Prop. method $N_{\text{order}} = 1$	1.19	3.1
Prop. method $N_{\text{order}} = 2$	2.14	5.6
Prop. method $N_{\text{order}} = 3$	3.15	8.3
Prop. method $N_{\text{order}} = 4$	4.20	11.1
Prop. method $N_{\text{order}} = 5$	5.27	13.9
Prop. method $N_{\text{order}} = 6$	6.38	16.8

The average run-time for the computation of the mode vectors and beamforming 1000 times are shown in Table D6. Comparing to Table D.5, we do see that the

computation of the mode vectors is the main component of the run-time in our Matlab implementation.

Table D6. Run-time comparison for the computation of the mode vectors and beamforming 1000 times. The indicated run-time is the average of 5 runs.

Approach	Run-time	Factor
	s	
Conventional beamforming	0.38	1
Prop. method $N_{\text{order}} = 1$	1.23	3.2
Prop. method $N_{\text{order}} = 2$	2.23	5.9
Prop. method $N_{\text{order}} = 3$	3.30	8.7
Prop. method $N_{\text{order}} = 4$	4.42	11.6
Prop. method $N_{\text{order}} = 5$	5.60	14.7
Prop. method $N_{\text{order}} = 6$	6.85	18.0

---

## Appendix

# *E*

---

### Email Exchange for Authorization to Reproduce Figures from IRCAM Technical Report [Rousseau 1996]

From: Bouchard, Christian  
Sent: Tue 10/14/2008 12:08 PM  
To: 'Claire.Marquet@ircam.fr'  
Cc:  
Subject: Demande autorisation pour reproduire figures d'un rapport de l'IRCAM  
Attachments:

Bonjour,

Mon nom est Christian Bouchard. Je suis étudiant au doctorat à l'Université d'Ottawa au Canada et j'effectue ma recherche au Conseil national de recherches du Canada.

Ma recherche porte sur les réseaux de microphones pour capter le son.

J'aimerais reproduire certaines figures du rapport

C. Lheureux, Simulation et mesure du rayonnement de différents instruments à vent à trous latéraux, Technical report, Institut de Recherche et de Coordination Acoustique/Musique (IRCAM), Paris (1997)

pour donner un exemple de patron de rayonnement pour une source directionnelle. Je reproduirais les figures dans une présentation à une conférence de l'Acoustical Society of America dont les "slides" seraient ensuite mises sur le site Internet de l'ASA, ainsi que dans ma thèse de doctorat. La source du matériel serait clairement indiquée dans tous les cas. En plus, dans le cas de la thèse, l'autorisation elle-même que vous me fourniriez serait aussi incluse.

Est-ce que l'IRCAM m'autorise à reproduire les figures?

Si vous n'êtes pas la personne à contacter, pourriez-vous m'indiquer ou m'adresser.

Merci,



---

Christian Bouchard  
University of Ottawa / National research council of Canada  
Ottawa, Ontario, K1A 0R6, Canada

---

From: Rene Causse  
Sent: Wed 10/15/2008 11:02 AM  
To: Bouchard, Christian  
Cc:  
Subject: Figures de rayonnement  
Attachments:

Bonjour,

Il n'y a pas de problèmes pour utiliser les figures du travail de Carole Lheureux pour une présentation à l'ASA et pour votre thèse à partir du moment où vous mentionnez le nom de Carole Lheureux et de son rattachement à l'époque : Equipe Acoustique instrumentale de l'Ircam.

N'hésitez pas à me contacter si vous souhaitez d'autres informations sur ce travail.

Bien cordialement,

René Caussé

---

From: Bouchard, Christian  
Sent: Wednesday, October 15, 2008 12:15 PM  
To: Rene Causse  
Subject: RE: Figures de rayonnement

Merci beaucoup Monsieur Caussé.

Christian Bouchard

---

From: Bouchard, Christian  
Sent: Fri 10/17/2008 10:38 AM  
To:  
Cc:  
Subject: RE: Figures de rayonnement  
Attachments:

Bonjour Mr. Caussé,

J'aimerais aussi reproduire des figures du rapport de stage de fin d'études de Antoine Rousseau intitulé "Modélisation du rayonnement des instruments à vent à trous latéraux", 1996, pour la présentation à l'Acoustical society of America et dans ma thèse.

Est-ce que vous m'y autorisez?

---

Est-ce que la modélisation et les mesures du rayonnement d'instruments se continuent encore à l'IRCAM?

Comme j'avais indiqué dans mon premier courrier électronique, je m'intéresse à la captation du son à l'aide de réseaux de microphones, en particulier de sources directionnelles. J'essaie donc d'inclure des exemples de patrons de rayonnement d'instruments. La très grande majorité de la littérature que j'ai vue décrit uniquement de façon qualitative le rayonnement des instruments. Les deux rapports de l'IRCAM auxquels j'ai accès (Mme Lheureux et Mr. Rousseau) font exception.

Encore merci,

Christian

Christian Bouchard  
University of Ottawa / National research council of Canada Ottawa, Ontario, K1A 0R6, Canada

---

From: Rene Causse  
Sent: Mon 10/20/2008 5:54 AM  
To: Bouchard, Christian  
Cc:  
Subject: Re: Figures de rayonnement  
Attachments: Image 1.png(32KB) Image 2.png(181KB) ATT1945786.txt(2KB)

Bonjour,

Il n'y a pas de problème pour reproduire les figures de Antoine Rousseau. C'est vraiment dommage que peu de personnes travaillent sur cet aspect là des instruments de musique.

Nous avons arrêté de travailler sur la modélisation et la mesure du rayonnement des instruments de musique pour passer à d'autres aspects. Cependant devant un besoin des compositeurs de l'Ircam, nous allons reprendre en 2009 dans le cadre d'un stage pour commencer. Nous allons bénéficier d'un dispositif de mesure du champ acoustique (antenne microphonique montée sur un portique mobile) mis au point par nos collègues qui travaillent en acoustique des salles. Il est prévu de travailler sur le saxophone, le cor de basset et la clarinette basse.

Je joins un croquis décrivant le nouveau système de mesure et une photo de l'ancien.

Bine cordialement,

René Caussé

---

## References

- Allen 1979 J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.* **65** (4), 943-950 (1979).
- Ando 1969/1970 Y. Ando, "On the sound radiation from semi-infinite circular pipe of certain wall thickness," *Acustica*, Vol. 22, 219-225 (1969/1970).
- Applebaum 1976 S. P. Applebaum and D. J. Chapman, "Adaptive arrays with main beam constraints," *IEEE Trans. on AP*, Vol. AP-24, pp. 650-662 (1976)
- Åström 1995 K. J. Åström and B. Wittenmark, *Adaptive control*, second edition (Addison Wesley, Reading Massachusetts, 1995)
- Balanis 1997 C. A. Balanis, *Antenna theory, analysis and design*, second edition (Wiley, New York, 1997).
- Betlehem 2003 T. Bethlehem and R. C. Williamson, "Acoustic beamforming exploiting directionality of human speech sources," *IEEE ICASSP* pp. V-365 - V-368 (2003)
- Bissinger 2008a G. Bissinger, "Structural acoustics of good and bad violins 2008," *J. Acoust. Soc. Am.* **124**, 1764-1773 (2008).
- Bissinger 2008b G. Bissinger, "Structural acoustics model of the violin radiativity profile," *J. Acoust. Soc. Am.* **124**, 4013-4023 (2008).
- Blackstock 2000 D. T. Blackstock, *Fundamentals of physical acoustics* (Wiley, New York, 2000).
- Blake 2002 W. K. Blake and D. A. Lynch III, "Source characterization by correlation techniques", in *Aeroacoustics measurements* (Springer, Berlin, 2002).
- Bouchard 2000 C. Bouchard, D. Grenier, and J.-R. Larocque, "ISAR-RMSA imaging radar with high-range-resolution algorithm and application to a linear phase antenna array," *Can. J. Elect. & Com. Eng.*, Vol. 25, No. 4 (2000).
- Bouchard 2006 C. Bouchard and D. I. Havelock, "Sound field phase deviations from a point source model for real acoustic sources," *Proceedings of Inter-Noise* (2006).
- Bouchard 2008 C. Bouchard, D. I. Havelock, and M. Bouchard, "Directional sources and beamforming," *Proceedings of Meetings on Acoustics* **5**, 055001 (2008).
- Bouchard 2009 C. Bouchard, D. I. Havelock, and M. Bouchard, "Beamforming with microphone arrays for directional sources," *J. Acoust. Soc. Am.* **125**, 2098-2104 (2009).

- Brooks 2005 T. F. Brooks and W. M. Humphreys Jr., "Three-dimensional application of DAMAS methodology for aeroacoustics noise source definition", 11th AIAA/CEAS Aeroacoustics Conf. (2005).
- Capon 1969 J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, 57, 119-129 (1969).
- Cho 2005 Y. T. Cho, J. S. Bolton, and J. Hald, "Source visualization by using statistically optimized near-field acoustical holography in cylindrical coordinates," *J. Acoust. Soc. Am.* **118**, 2355-2364 (2005).
- Chu 2002 W. T. Chu and A. C. C. Warnock, "Detailed directivity of sound fields around human talkers," Technical report IRC-RR-104, Institute for research in construction, National research council Canada (2002.)
- Cook 1955 R. K. Cook et al., "Measurement of correlation coefficients in reverberant sound fields," *J. Acoust. Soc. Am.*, vol. **27**, (6) pp. 1072-1077 (1955)
- Cox 1986 H. Cox, R. M. Zeskind, and T. Kooij, "Practical supergain," *IEEE Trans. on Acoust., Speech, and Signal Process.*, vol. ASSP-34, No. 3, 393-398 (1986).
- Cox 1987 H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. on Acoust., Speech, and Signal Process.*, Vol. ASSP-35, No. 10, 1365-1376 (1987)
- DiBiase 2001 J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone arrays* (Springer, Berlin, 2001).
- Dunn 1939 H. K. Dunn and D.W. Farnsworth, "Exploration of pressure field around the human head during speech", *J. Acoust. Soc. Am.*, vol. **10**, pp. 184-199 (1939)
- Elko 2001a G. W. Elko, "Superdirectional microphone arrays," in *Acoustic signal processing for telecommunication* (KAP, 2000, second printing 2001)
- Elko 2001b G. W. Elko, "Spatial coherence functions for differential microphones in isotropic noise fields," in *Microphone arrays* (Springer, Berlin, 2001).
- Flanagan 1960 J. L. Flanagan, "Analog measurements of sound radiation from the mouth," *J. Acoust. Soc. Am.*, vol. **32**, pp. 1613-1620 (1960)
- Frost 1972 O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, Vol. 60, pp. 926-935 (1972)
- Goldstein 1976 M. E. Goldstein, *Aeroacoustics* (McGraw-Hill, New York, 1976).
- Gover 2004 B. N. Gover, J. G. Ryan, and M. R. Stinson, "Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone array," *J. Acoust. Soc. Am.* **116**, 2138-2148 (2004).

- 
- Griffiths 1982 L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. on AP*, Vol. AP-30, pp. 27-34 (1982)
- Huang 2001 Y. Huang, J. Benesty, and G. W. Elko, "Microphone arrays for video camera steering," in *Acoustic signal processing for telecommunication* (KAP, 2000, second printing 2001)
- Jordan 2002 P. Jordan, J. A. Fitzpatrick, and J.-C. Valière, "Measurement of an aeroacoustic dipole using a linear microphone array," *J. Acoust. Soc. Am.* **111**, 1267-1273 (2002).
- Kalman 1960 R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.* 82, 35-45 (1960).
- Kuttruff 2000 H. Kuttruff, *Room acoustics*, 4<sup>th</sup> edition (Spon Press, London, 2000)
- Liu 2008 Y. Liu, A. R. Quayle, and A. P. Dowling, "Beamforming correction for dipole measurement using two-dimensional microphone arrays," *J. Acoust. Soc. Am.* **124** (1), 182-191 (2008).
- Meyer 1972 J. Meyer, "Directivity of the bowed stringed instruments and its effect on orchestral sound in concert halls," *J. Acoust. Soc. Am.* **51**, 1994-2009 (1972).
- Meyer 2002 J. Meyer and G. W. Elko, "A highly scalable spherical microphone array based on an orthogonal decomposition of the soundfield," *Proceedings of ICASSP 2002*, Vol. II, 1781-1784 (2002).
- Monzingo 1980 R. Monzingo and T. Miller, *Introduction to adaptive arrays* (Wiley and sons, New York, 1980).
- Park 2005 M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *J. Acoust. Soc. Am.* **118**, 3094-3103 (2005).
- Piet 2002 J. F. Piet, U. Michel, and P. Böhning, "Localization of the acoustic sources of the A340 with a large phase microphone array during flight tests," in 8<sup>th</sup> AIAA/CEAS Aeroacoustics conference and exhibit (2002)
- Powell 1990 A. Powell, "Some Aspects of Aeroacoustics: From Rayleigh Until Today," *J. Vibr. Acoust.* **112**, 145-159 (1990).
- Proakis 1996 J. G. Proakis and D. G. Manolakis, *Digital signal processing Principle, principles, algorithms, and applications*, 3<sup>rd</sup> edition (Prentice Hall, New Jersey 1996)
- Rafaely 2000 B. Rafaely, "Spatial-temporal correlation of a diffuse sound field," *J. Acoust. Soc. Am.* **107** (6), 3254-3258 (2000).
- Rafaely 2004 B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *J. Acoust. Soc. Am.* **116** (4), 2149-2157 (2004).

- Rafaely 2005 B. Rafaely, "Phase-mode versus delay-and-sum spherical microphone array processing," *IEEE Signal Process. Letters*, Vol. 12, No. 10 (2005).
- Rousseau 1996 A. Rousseau, *Modélisation du rayonnement des instruments à vent à trous latéraux (Model of radiation of lateral toneholes wind instruments)*, Technical report, Institut de Recherche et de Coordination Acoustique/Musique (IRCAM), Département d'acoustique instrumentale (IRCAM), Paris, May 1996.
- Ryan 1998 J. G. Ryan, "Criterion for the minimum source distance at which plane-wave beamforming can be applied," *J. Acoust. Soc. Am.* **104** (1), 595-598 (1998).
- Ryan 2003 J. G. Ryan and R. A. Goubran, "Application of near-field optimum microphone arrays to hands-free mobile telephony," *IEEE Trans. on veh. tech.* Vol. 52, No.2 (2003)
- Scavone 1998 G. P. Scavone and P. R. Cook, "Real-time computer modeling of woodwind instruments," *International symposium on musical acoustics*, Leavenworth, WA, 1998.
- Smith 2008 J. O. Smith, *Physical Audio Signal Processing*, December 2008 Edition, <http://ccrma.stanford.edu/~jos/pasp/>, online book, accessed March 04, 2009
- Steinberg 1991 B. D. Steinberg and H. M. Subbaram, *Microwave imaging techniques* (Wiley, New York, 1991).
- Teutsch 2001 H. Teutsch and G.W. Elko, "An adaptive close-talking microphone array," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 2001 (WASPAA)*, pp.163-166, 2001.
- Teutsch 2007 H. Teutsch, *Modal array signal processing: Principles and Applications of acoustic wavefield decomposition* (Springer-Verlag, Berlin 2007).
- Therrien 1992 C. W. Therrien, *Discrete random signals and statistical signal processing* (Prentice Hall, New Jersey, 1992).
- Van Trees 2002 H. L. Van Trees, *Optimum array processing – Part IV of Detection, Estimation, and Modulation Theory* (Wiley-Interscience, 2002).
- Van Veen 1988 B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, 4-24, April 1988
- Wang 2001 L. M. Wang and C. B. Burroughs, "Acoustic radiation from bowed violins," *J. Acoust. Soc. Am.* **110**, 543-555 (2001).
- Ward 1995 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. Am.* **97**, 1023-1034 (1995).

- 
- Warnock 2002 A. C. C. Warnock, W. T. Chu, and J.-C. Guy, "Directivity of human talkers," *Canadian acoustics*, Vol. 30 No. 3 (2002)
- Widrow 1967 B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode, "Adaptive antenna systems," *Proc. IEEE*, Vol. 55 pp. 2143-2159 (1967)
- Williams 1999 E. G. Williams, *Fourier Acoustics: sound radiation and nearfield acoustical holography* (Academic, New York, 1999).
- Wu 1995 H. Q. Wu, D. Grenier, G. Y. Delisle, and D.-G. Fang, "Translational motion compensation in ISAR image processing," *IEEE Trans. Image Process*, Vol. 4, No. 11, pp. 1561-1571 (1995).
- Wu 2000 S. F. Wu, "On reconstruction of acoustic pressure fields using the Helmholtz equation least squares method," *J. Acoust. Soc. Am.* **107**, 2511-2522 (2000).