Near-field Beamforming using Microphone Arrays

by

James G. Ryan, M.Eng., B. Eng.

**A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of**

Doctor of Philosophy

**Ottawa-Carleton Institute for Electrical and Computer Engineering
Faculty of Engineering
Department of Systems and Computer Engineering
Carleton University
Ottawa, Ontario, Canada, K1S 5B6**

May 31, 1999

© Copyright

1999, James G. Ryan

**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

**395 Wellington Street
Ottawa ON K1A 0N4
Canada**

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

**395, rue Wellington
Ottawa ON K1A 0N4
Canada**

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

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

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

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

0-612-48335-5

Canada

ABSTRACT

This thesis explores array optimization applied to acoustic targets located in the near field of a microphone array. The results contained herein illustrate the performance improvements obtained when differences in curvature between the desired and interfering wavefronts are exploited. In order to accomplish this task, conventional array optimization methods are extended to the case of near-field sources. Specifically, unconstrained, quadratically constrained and linearly constrained array optimization are considered. Gain improvements due to optimization are shown to increase for focal points closer to the array. Array gain can also be optimized for a near-field target subject to a constraint on the far-field response in the same direction. Enforcing a null in the response is shown to reduce performance substantially; however, a milder response constraint can be enforced that does not greatly degrade performance. Two different methods are then described for broadband implementation. The first method uses optimization at discrete frequency points and FIR filter design based on complex approximation. The second method is based on a reformulation of the design equations in the time domain assuming a real-coefficient FIR beamformer. Both techniques are then used to design broadband optimum beamformers for uniformly spaced linear arrays of microphones. Application of these concepts to mobile telephony yields a noise reduction approximately 3dB greater than an existing near-field optimum solution. When applied to computer telephony, the addition of a linear constraint to the optimum beamformer is shown to provide enhanced rejection of reverberation from walls parallel to the array axis.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my thesis supervisor, Dr. Rafik Goubran, for his encouragement and advice throughout the course of my studies. His enthusiastic support for my work will be fondly remembered.

I would also like to express my gratitude to the National Research Council of Canada for the support I received while working on this thesis. I would like to thank the members of the Acoustics and Signal Processing group of the National Research Council for their encouragement and countless helpful discussions during the course of my research. In particular, I would like to thank Dr. Michael Stinson whose steered microphone-array demonstration was the inspiration for my work. Special thanks also go to Mr. Rene St-Denis for his assistance in developing the experimental microphone array and Mr. John Quaroni for preparing and operating the equipment for the mobile array experiment.

Finally, the staff in the Department of Systems and Computer Engineering have always been a tremendous help to me throughout my years at Carleton. In particular, I would like to thank Ms. Darlene Hebert for her personal encouragement and assistance in making my off-campus studies a success.

To my family.

CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGEMENTS	iv
CONTENTS	vi
LIST OF FIGURES	ix
LIST OF TABLES	xiii
LIST OF NOTATION	xiv

Chapter 1

INTRODUCTION	1
1.1 Importance of Near-field Array Processing for Microphone Arrays.....	1
1.2 Problem Statement.....	6
1.3 Thesis Outline	7
1.4 Summary of Contributions.....	8

Chapter 2

SPEECH PICKUP IN REVERBERANT ROOMS.....	10
2.1 Monophonic Speech Pickup in Rooms	10
2.2 Using Directional Microphone Techniques to Improve Audio Quality	11
2.3 The Acoustics of Speech Radiation from the Mouth.....	14
2.4 Room Reverberation and Noise	17
2.4.1 Basic Room Acoustics	17
2.4.2 Modeling Room Acoustics	21
2.4.3 Relative Levels of Room, Microphone and Preamplifier Noise.....	23
2.4.4 Spatial Correlation of Noise	23
2.5 Objective Measures of Sound Pickup Systems.....	27
2.5.1 Room Acoustical Measurements	27

Chapter 3

INTRODUCTION TO ARRAY PROCESSING	29
3.1 Basic Concepts and Mathematical Formulation	29
3.1.1 Array Gain	30
3.1.2 White Noise Gain.....	32
3.1.3 Gain Optimization.....	32
3.1.3.1 Quadratically Constrained Gain Optimization	33
3.1.3.2 Linearly Constrained Gain Optimization.....	35
3.2 Near Field/Far Field Distinction	37

Chapter 4

CURRENT APPROACHES TO NEAR-FIELD BEAMFORMING.....	41
4.1 Delay-and-sum Beamforming for Spherical Waves	41
4.2 Near-field Beampattern Design	48
4.2.1 Near-field Compensation.....	48
4.2.2 Radial Beampattern Transformation.....	49
4.2.3 Multidimensional Filter Design	51

4.2.4 Nonlinear Goal Programming.....	51
4.2.5 Near-field Null Placement	51
4.3 Gain Optimization.....	52
4.3.1 Optimum Array Gain for Spatially Incoherent Noise.....	52
4.3.2 Gain Optimization using Sensor Weight and Position Adjustment.....	53
 Chapter 5	
NEAR-FIELD GAIN OPTIMIZATION	54
5.1 Unconstrained Optimization	55
5.1.1 Spatially Uncorrelated Noise	56
5.1.2 Partially Correlated Noise.....	58
5.1.3 Numerical Example	58
5.2 Quadratically Constrained Optimization	66
5.2.1 Numerical Example	69
5.3 Linearly Constrained Optimization	72
5.3.1 Numerical Example	77
 Chapter 6	
BROADBAND BEAMFORMER DESIGN.....	82
6.1 Sensor FIR Filter Design using Complex Approximation.....	82
6.1.1 Frequency Sampling Design	83
6.1.2 Least squares complex approximation.....	85
6.1.3 Complex Chebyshev Error Approximation	86
6.2 Direct Synthesis using Time-domain Problem Formulation	86
6.3 Design Example	94
6.3.1 Complex Approximation	96
6.3.2 Time-domain Formulation	99
6.3.3 Sensor Filters	102
 Chapter 7	
APPLICATIONS OF NEAR-FIELD BEAMFORMING	104
7.1 Mobile Telephony	104
7.1.1 Broadband Beamformer Design	105
7.1.2 Test Conditions	107
7.1.3 Measured Gain	109
7.1.4 Effect of Beamformers on Speech Signal	115
7.1.5 Discussion	116
7.2 Computer Telephony Integration.....	116
7.2.1 Room Impulse Response Measurements	118
7.2.2 Beamformer Designs	121
7.2.3 Performance in Reverberant Environment	125
7.2.3.1 Response to Near-end Signal	125
7.2.3.2 Response to Far-end Signal	126
 Chapter 8	
CONCLUSIONS	131

BIBLIOGRAPHY	133
Appendix A	
BROADBAND CORRELATION FUNCTIONS	140
Appendix B	
OPTIMIZATION SUBJECT TO LINEAR CONSTRAINTS	150

LIST OF FIGURES

Figure 1.	For talkers located at large distances from an array, acoustic wavefronts appear less curved and can be approximated by plane waves. For talkers located close to an array, acoustic wavefronts appear more curved and the assumption of plane wave propagation is not valid.	2
Figure 2.	Interaction between an array's far-field beam pattern and the image sources representing reflections from room boundaries. The shaded region indicates the 3 dB limits of the array beam pattern when the array is steered for plane waves originating from the sound source direction.	4
Figure 3.	Interaction between an array's near-field beam pattern and the image sources representing reflections from room boundaries for the same configuration as in Figure 2.	5
Figure 4.	Illustration of the multiple sound paths between a talker and a microphone in a reverberant room. The solid line indicates the path of the direct wave from the source to the receiver and the dashed lines indicate the paths travelled by several room reflections.	12
Figure 5.	Illustration of the effect of a directional microphone on the pickup of sound in a reverberant room. The shaded area indicates the receiving pattern of the directional microphone. Due to the difference in direction of propagation, the reflected waves are attenuated with respect to the direct wave.	13
Figure 6.	Sound radiation pattern for speech energy from the mouth measured at a distance of 30 cm in three different frequency bands: solid line 250-500 Hz, dashed line 700-1000 Hz, dotted line 2800-4000 Hz.	16
Figure 7.	Impulse response of a simulated rectangular room with a reverberation time of approximately 200 ms. Spacing between source and receiver is 1 m.	19
Figure 8.	Frequency response of the room impulse response shown in Figure 7.	20
Figure 9.	Location of image sources representing several horizontal reflections from the walls in a rectangular room. The solid lines indicate the actual paths travelled by the reflected waves; dashed lines indicate the paths travelled by the waves generated using the image sources.	22
Figure 10.	Relative levels of room noise, electret microphone noise and preamplifier noise measured in a quiet room. Room noise dominates at frequencies below 1 kHz.	24
Figure 11.	Background noise spectra obtained in a moving vehicle under different driving conditions.	25
Figure 12.	Acoustic source located in the near field of an array.	38
Figure 13.	Acoustic source located in the far field of an array.	39
Figure 14.	Illustration of three circular paths within the near field of a uniformly spaced linear array. The array length is $D = (M - 1)d$ where M is the number of elements and d is the sensor spacing. The radii of the three circular paths are: $r_1 = (D/2) + d$, $r_2 = D$ and $r_3 = 10D$. To maintain small details, the drawing is not to scale.	44
Figure 15.	Delay-and-sum beampatterns for a 15 element linear array with sensor spacing $d = \lambda/2$ focused at $(r_f, \theta_f) = ((D/2) + d, \pi/2)$. The beampattern is evalu-	

	ated along three different paths as shown in Figure 14: path ABCDA (top), path EFGHE (middle) and path JKMOJ (bottom).....	46
Figure 16.	Delay-and-sum beampatterns for a 15 element linear array with sensor spacing $d = \lambda/2$ focused at $(r_f, \theta_f) = (10D, \pi/2)$. The beampattern is evaluated along three different paths as shown in Figure 14: path ABCDA (top), path EFGHE (middle) and path JKMOJ (bottom).	47
Figure 17.	Contour plot of the optimum near-field gain for a 16-element uniformly spaced linear array with a microphone spacing of $\lambda/2$. Contour lines are spaced by 1dB and the x and y coordinates are normalized to the total length of the array. The positions of the array microphones are illustrated by the solid circles along the x axis at $y = 0$	60
Figure 18.	Contour plot of the net effect of near-field gain optimization for the array shown in Figure 17. The gain of an equivalent delay-and-sum beamformer has been removed to highlight the effect of the optimization.....	61
Figure 19.	Delay-and-sum beampatterns for a 16-element uniformly spaced linear array with $\lambda/2$ sensor spacing. The array is focused at $(r_f, \theta_f) = (D, \pi/2)$ and the beampattern is evaluated at three different distances: $r = (D/2) + d$ (top), $r = D$ (middle) and $r = \infty$ (bottom).....	63
Figure 20.	Optimum beampatterns for the array in Figure 19.	64
Figure 21.	Contour plot of the optimum near-field gain for the same conditions as in Figure 17 but with a microphone spacing of $\lambda/4$	65
Figure 22.	Delay-and-sum beampatterns for a 16-element uniformly spaced linear array with $\lambda/4$ sensor spacing. The array is focused at $(r_f, \theta_f) = (D, \pi/2)$ and the beampattern is evaluated at three different distances: $r = (D/2) + d$ (top), $r = D$ (middle) and $r = \infty$ (bottom).....	67
Figure 23.	Optimum beampatterns for the array in Figure 22. Note the change in scale for the top graph.	68
Figure 24.	Contour plot of the net effect of constrained near-field gain optimization for the array shown in Figure 21. White noise gain is constrained to a minimum of 0 dB. The gain of an equivalent delay-and-sum beamformer has been removed to highlight the effect of the optimization.....	70
Figure 25.	Quadratically constrained optimum beampatterns for the array in Figure 22. Array gain optimization is constrained to limit the minimum white noise to 0 dB.	71
Figure 26.	Increase in array gain as a function of white noise gain for a 16-element uniformly spaced linear array with $\lambda/4$ sensor spacing. The trade-off of white noise gain for array gain varies with the distance to the array focal point. For a given white noise constraint, higher array gain is observed for focal points close to the array.	73
Figure 27.	Gain loss due to linear constraint imposition as a function of $\cos(\gamma)^2$ for several values of α	76
Figure 28.	Variation in $\cos(\gamma)^2$ as a function of angle of incidence with respect to array axis θ . Solid line assumes uncorrelated noise (i.e. $R_{nn} = I$) whereas dashed line assumes isotropic noise.....	78
Figure 29.	Null-constrained optimum beampatterns for the array in Figure 22. Note the change in scale for the top graph.	80

Figure 30. Soft-constrained optimum beampatterns for the array in Figure 22. Array gain is optimized while suppressing the response in the broadside direction by an additional 6 dB compared to that shown in Figure 25. Note the change in scale for the top graph.	81
Figure 31. Broadband FIR beamformer structure.	88
Figure 32. Gain error statistics as a function of the number of FIR taps for the least-squared error design method : standard deviation (solid), peak error (dash-dot) and mean (dotted).	97
Figure 33. Response error statistics as a function of the number of FIR taps for the least-squared error design method: standard deviation (solid), peak error (dash-dot) and mean (dotted).	98
Figure 34. Gain error statistics as a function of the number of FIR taps for the time-domain design method: standard deviation (solid), peak error (dash-dot) and mean (dotted).	100
Figure 35. Response error statistics as a function of the number of FIR taps for the time-domain design method: standard deviation (solid), peak error (dash-dot) and mean (dotted).	101
Figure 36. Response of 28-tap sensor filter 3 for near-field beamformer: ideal response (solid), least-squares design (dash-dot) and time-domain design (dashed). ..	103
Figure 37. Physical layout showing relative positions of the microphone array and mannequin. Microphone number 3 is located closest to the mannequin's mouth and is taken as the reference.	106
Figure 38. Target (circles) and actual (solid line) frequency response for microphone filter number 3 in the FS beamformer.	108
Figure 39. Simplified block diagram of the multichannel recording system.	110
Figure 40. Vehicle noise-power spectra obtained under various conditions.	111
Figure 41. Predicted (dashed line) and measured (solid line) array gain for both the WDS beamformer (top graph) and the FS beamformer (bottom graph) as a function of frequency. These data are the average gains obtained for all conditions tested.	112
Figure 42. White noise gain as a function of frequency for both broadband beamformers.	114
Figure 43. Anechoic frequency response deviations produced by the broadband beamformers. The curves show the difference between the response provided by the beamformers compared to that of the reference microphone for speech signals from the mannequin.	117
Figure 44. Microphone array intended for computer telephony application. Array consists of ten microphones equispaced at intervals of 7cm. The array focal point is at a distance of 50cm along the y axis.	119
Figure 45. Physical layout of computer monitor and mannequin within the room.	120
Figure 46. Array gain provided by broadband beamformers designed for CTI. The curve labelled $\alpha = 1$ is the maximum gain solution, whereas the curves labelled $\alpha = 0.5$ and $\alpha = 0.1$ were obtained for constrained designs.	122
Figure 46. Array gain provided by broadband beamformers designed for CTI. The curve labelled $\alpha = 1$ is the maximum gain solution, whereas the curves labelled $\alpha = 0.5$ and $\alpha = 0.1$ were obtained for constrained designs.	122

$\alpha = 0.707$ and $\alpha = 0$ were obtained for constrained designs.	122
Figure 47. White noise gain provided by CTI beamformers.....	123
Figure 48. Target frequency response provided by the CTI beamformer optimized with $\alpha = 0$	127
Figure 49. Target frequency response provided by the CTI beamformer optimized with $\alpha = 0.707$	128
Figure 50. Energy decay curves for talker position provided by CTI beamformers.....	129
Figure 51. Acoustic point source located in the vicinity of two ideal point receivers. ...	141
Figure 52. Crosscorrelation function according to (81) for the physical arrangement shown in Figure 51. Both τ and $\Delta r = r_1 - r_2$ are normalized to the sampling interval $1/F_s$. The signal frequency range is 0 to $F_s/2$	145
Figure 53. Acoustic planewave propagating across two ideal point receivers.....	146
Figure 54. Crosscorrelation function according to (92) for the physical arrangement shown in Figure 53. Both τ and d are normalized to the sampling interval $1/F_s$. The signal frequency range is 0 to $F_s/2$	149

LIST OF TABLES

TABLE 1.	Directional gains of several common microphone types (derived from the results in [28]).	11
TABLE 2.	Acoustic pressure reflection coefficients of several common building materials measured at different frequencies (obtained from [36]).....	17
TABLE 3.	Total speech-bandwidth noise power reduction in decibels provided by the two broadband beamformers.....	115

LIST OF NOTATION

$\mathbf{a}_f$	Position vector of array focal point
$\mathbf{a}_m$	Position vector for microphone m
$\mathbf{a}_s$	Position vector for signal source
$\mathbf{a}_r$	Position vector for reference location
$b(r, \theta, \phi, f)$	Normalized near-field beampattern function
$\tilde{b}(r, \theta, \phi, f)$	Unnormalized near-field beampattern function
$b_c(\theta, f)$	Compensated near-field beampattern function
$b_d(\theta, f)$	Desired beampattern function
c	Speed of propagation of free-field sound wave (340m/s)
$\mathbf{c}_l$	Stacked array constraint vector
$\mathbf{C}$	Constraint matrix
d	Sensor spacing for uniformly spaced linear array
$d_m(n)$	Impulse response from near-field source location to sensor m
D	Maximum dimension of array
$e^{j\phi}$	Exponential function of argument ϕ
$\mathbf{e}$	Filter-design error vector
E	Sum-squared filter approximation error
$E\{x\}$	Expected value of random variable x
$\mathbf{F}$	Fourier coefficient matrix
F_s	Discrete-time sampling frequency
$\mathbf{g}$	Array response vector
g_m	Near-field compensation factor for sensor m
$G(\omega)$	Array gain
$G(\omega)_{\text{act}}$	Actual array gain for a given beamformer implementation

$G(\omega)_{\text{opt}}$	Optimum (maximum) array gain
$G_w(\omega)$	White noise gain
$G_w \text{ max}$	Maximum white noise gain
H	Superscript H denotes Hermitian (complex conjugate) matrix transpose
$H_{n+1/2}^{(1)}(\cdot)$	Half-integer order spherical hankel functions of the first kind
I	Identity matrix
j	Complex operator ($\sqrt{-1}$)
J	Number of linear constraints
k	Wavenumber ($k = 2\pi/\lambda$)
m	Index for sensors in array
M	Number of sensors in array
n	Discrete-time index
$n_m(n)$	Noise signal received by sensor m (time domain)
$\mathbf{n}(n)$	Stacked noise snapshot vector
N	Number of FIR taps per channel in broadband array
$N_m(\omega)$	Noise component received by sensor m at frequency ω normalized by σ_n
$\tilde{N}_m(\omega)$	Unnormalized noise component received by sensor m at frequency ω
$\mathbf{N}$	Noise component vector normalized by σ_n (frequency domain)
$\tilde{\mathbf{N}}$	Unnormalized noise component vector (frequency domain)
$P_n(\cdot)$	Legendre function of order n
$P_n^m(\cdot)$	Associated Legendre functions
$\mathbf{Q}(\omega)$	Noise spatial correlation matrix corresponding to environmental noise
r_{fm}	Distance from array focal point to sensor m
r_{sm}	Distance from source position to sensor m
(r, θ, ϕ)	Spherical coordinates (distance, azimuth, elevation)

(r_m, θ_m, ϕ_m)	Spherical coordinates relative to array sensor m
$R_{ss}(\omega)$	Signal spatial correlation matrix (frequency domain)
R_{ss}	Signal spatiotemporal correlation matrix (time domain)
$R_{nn}(\omega)$	Noise spatial correlation matrix (frequency domain)
R_{nn}	Noise spatiotemporal correlation matrix (time domain)
$R_{uu}(\omega)$	Composite sensor signal spatial correlation matrix (frequency domain)
R_{uu}	Composite sensor signal spatiotemporal correlation matrix (time domain)
$S_m(\omega)$	Signal component received by sensor m at frequency ω normalized by σ_s
$\tilde{S}_m(\omega)$	Unnormalized signal component received by sensor m at frequency ω
S	Signal propagation vector normalized by σ_s (frequency domain)
$\tilde{S}$	Unnormalized signal propagation vector (frequency domain)
$s_m(n)$	Signal received by sensor m (time domain)
$s(n)$	Stacked signal vector (time domain)
t	Continuous time variable
T	Superscript T denotes matrix transpose
T_{60}	Reverberation time (time required for 60 dB decay of sound field)
$u_m(n)$	Composite signal received by sensor m (time domain)
$u(n)$	Array stacked snapshot vector
$U_m(\omega)$	DFT of composite signal received by sensor m (frequency domain)
U	Composite sensor signal vector (frequency domain)
$V(\omega)$	FIR-design error weighting function
V	FIR-design error-weight matrix (diagonal)
w	Stacked sensor-weight vector (time domain)
$w_A(n)$	Actual array impulse response (time domain)
$w_m(n)$	Impulse response of filter for sensor m (time domain)

$\hat{\mathbf{w}}$	Vector of unknown filter coefficients $\hat{w}(n)$ (time domain)
$\hat{w}(n)$	Unknown FIR filter coefficients (time domain)
$\mathbf{W}$	Complex sensor-weight vector (frequency domain)
$W_A(\omega)$	Actual array frequency response (frequency domain)
$\mathbf{W}_d$	Vector of samples of desired filter response $W_d(\omega)$ (frequency domain)
$W_d(\omega)$	Desired filter frequency response (frequency domain)
$\mathbf{W}_I$	Vector of samples of ideal filter response (frequency domain)
$W_m(\omega)$	Complex weight for sensor m at frequency ω (DFT of $w_m(n)$ evaluated at ω)
$\mathbf{W}_{\max}$	Weight vector that maximizes white noise gain
$\mathbf{W}_{\text{opt}}$	Weight vector that maximizes array gain
$\hat{\mathbf{W}}$	Vector of samples of $\hat{W}(\omega)$ (frequency domain)
$\hat{W}(\omega)$	DFT of unknown FIR filter coefficients $\hat{w}(n)$ evaluated at frequency ω
(x, y, z)	Rectangular coordinates
(x_m, y_m, z_m)	Rectangular coordinates of sensor m
$y(n)$	Array output sequence (time domain)
$Y(\omega)$	DFT of array output (frequency domain)
α	arbitrary complex constant
$\alpha_n, \beta_{nm}, \gamma_{nm}$	See equations (19) through (21)
γ	Generalized angle between vectors
δ^2	White noise gain constrained level
$\epsilon_G(\omega)$	Array gain error
$\epsilon_W(\omega)$	Beamformer response error
κ	Lagrange multiplier
λ	Wavelength

$\mathbf{v}(n)$	Array snapshot vector
$\rho(d)$	Spatial correlation between points separated by a distance d
$\rho_{mp}(\omega)$	Crosscorrelation at frequency ω
σ_s^2	Signal variance
σ_n^2	Noise variance
σ_w^2	White noise variance
$\tau_{m,\theta}$	propagation delay for planewave arrival from angle θ at sensor m
$\Psi(\omega)$	Array white noise sensitivity
ω	Radian frequency ($\omega = 2\pi f$)
ω_l	Frequency sample points

Chapter 1

INTRODUCTION

Since microphone arrays are wave devices (that is, their operation exploits wave behavior) analysis and design can be complicated. Many techniques have been developed, however, based on the simplifying assumption that all acoustic sources are located far from the array. In this case wavefront curvature is very small and all waves impinging upon the array are assumed to be planar as is illustrated by the lower array in Figure 1. This *far-field* assumption greatly facilitates the analysis and design of microphone arrays and provides a realistic framework for certain applications [1][2][3][4][5].

There are circumstances, however, where the acoustic source is located very close to the array and the assumption of plane wave propagation is not valid. This situation arises, for instance, in mobile telephony [6][7] and conference telephony in small rooms [8][9][10]. In such cases, wavefronts impinging upon the array appear curved as illustrated by the upper array in Figure 1. Array analysis and design for such *near-field* sources is much more complicated than for the far-field and it has received much less attention in the research literature.

1.1 Importance of Near-field Array Processing for Microphone Arrays

The application of microphone arrays to speech pickup in small rooms is a realistic application where near-field array processing is required. In addition to eliminating direct-wave distortion, spherical-wave beamforming has the added advantage of providing a modest degree of range discrimination. This means that spherical-wave beamforming can separate two sound sources that lie along the same angle of incidence with respect to the array axis, as long as the distances are not identical. This is important in reducing the effects of room reverberation, for instance, since a spherical-wave focused array can eliminate reflections originating from parallel surfaces behind the talker. Consider, for example, a talker located in the near field of an array in a reverberant room.

Figure 1. For talkers located at large distances from an array, acoustic wavefronts appear less curved and can be approximated by plane waves. For talkers located close to an array, acoustic wavefronts appear more curved and the assumption of plane wave propagation is not valid.

Figure 2 and Figure 3 depict a single talker located in the near field of an array in a reverberant room. As discussed in Section 2.4.2 on page 21, the effects of the wall reflections can be modeled as image sources located behind the reflecting surfaces. These image sources are indicated by the shaded talkers.

Using plane-wave beamforming for near-field talkers corrupts the direct-wave response due to steering-delay errors. The peak array response is obtained only for sources located at a very large distance from the array. Consequently, the array response to reflected wavefronts from parallel walls is larger than that due to the desired source as depicted in Figure 2.

Using spherical-wave beamforming, however, the array can be focused in both angle and distance. The position of peak array response coincides with the location at which the array is focused and the array response is an oval-shaped region around the focal point as depicted in Figure 3.

Unfortunately, the range discrimination provided by conventional delay-and-sum beamforming is limited. Furthermore, current near-field design methods have not attempted to exploit range discrimination to improve audio performance. A review of existing near-field design techniques (Chapter 4) illustrates an almost exclusive focus on implementing predetermined beampatterns. In many applications, however, it is difficult to know what the optimum beampattern should be.

Instead, the approach in this thesis focuses on maximizing the signal-to-interference ratio at the output of the array. Additional gain improvements are obtained by exploiting the difference in curvature between the desired and interfering wavefronts. In order to accomplish this task, conventional array optimization methods are extended to the case of near-field sources. Furthermore, since speech signals are broadband, methods are examined for implementing the near-field optimum array processors over a frequency range suitable for telephony.

Figure 2. Interaction between an array's far-field beam pattern and the image sources representing reflections from room boundaries. The shaded region indicates the 3 dB limits of the array beam pattern when the array is steered for plane waves originating from the sound source direction.

Figure 3. Interaction between an array's near-field beam pattern and the image sources representing reflections from room boundaries for the same configuration as in Figure 2.

1.2 Problem Statement

This thesis addresses the near-field performance optimization of microphone arrays for application to speech pickup in reverberant and noisy environments. In this work, it is assumed that the microphones are located in fixed, predetermined positions and that the talker location is known. Consequently, the optimization procedure focuses solely on adjustment of the frequency-dependent microphone weights for performance enhancement. Furthermore, only fixed beamformers are considered. The reason for this is that the main applications envisioned by this work are for talkers in reverberant enclosures. Under such conditions, the application of adaptive beamforming techniques fail due to the problem of desired signal cancellation [11][12][14][13][15]. Such problems will be avoided by using a fixed but steerable beam which has been previously optimized for a given look direction.

A key element of the optimization is in choosing suitable signal and noise models to use in the performance criterion. By considering the acoustical properties of speech radiation from the mouth, reverberant energy distribution in rooms and also the relative levels of room, microphone and preamplifier noise, suitable models are chosen. It is demonstrated that optimizing array gain using the chosen models results in an improvement in audio performance of the array.

Broadband beamformers are designed using two different methods. The first method uses optimization at discrete frequency points and FIR filter design based on complex approximation. The second method is based on a reformulation of the design equations in the time domain assuming a real-coefficient FIR beamformer. The relative merits of the two approaches are discussed.

Finally, two applications that require near-field beamformers are presented. Broadband optimum beamformers are designed for specific cases and experimental performance is evaluated.

1.3 Thesis Outline

Chapter 1 describes the problem that this thesis addresses. It outlines the need for near-field beamforming in the application of microphone arrays to the problem of speech pickup in rooms.

Chapter 2 provides background material on the physical considerations for speech pickup in rooms. The discussion presents current knowledge that is used to justify the signal and noise models used in beamforming. The chapter concludes with a brief discussion of relevant objective measures that can be used to gauge system performance for this application.

A review of basic array processing techniques is presented in Chapter 3. The purpose of this chapter is not to present a complete survey of this field but rather to introduce key concepts and notation required for the developments in this thesis.

Chapter 4 presents a review of current approaches to the problem of near-field array design. It is shown that many existing techniques attempt to implement a specific near-field response pattern. Of the techniques that optimize near-field array gain, only one explicitly optimizes the response to spherical waves in a realistic noise field. The solution, however, is only optimum at a single frequency.

The main concepts of near-field array optimization are presented in Chapter 5. First, unconstrained gain optimization in isotropic noise is studied. While near-field gain enhancements can be demonstrated, for array with small sensor spacing relative to a wavelength the optimized beamformer exhibits poor robustness to uncorrelated noise and tolerance errors. Second, the problems of robustness can be reduced by using a quadratic optimization constraint. It is shown that realistic near-field gain improvements are possible. Finally, for discrete interference, constraining the beamformer response for specific far-field directions can provide some benefits.

Chapter 6 describes the broadband implementation of optimum arrays. Two methods for implementing a fixed-coefficient beamformer are presented: sensor FIR filter design using complex approximation and a direct time-domain synthesis using the methods from adaptive array processing.

Two specific applications that require near-field array processing are presented in Chapter 7: mobile telephony and computer telephony. The performance of broadband optimum beamformers implemented using the approaches in Chapter 6 are experimentally verified under realistic conditions.

Conclusions are presented in Chapter 8.

1.4 Summary of Contributions

This thesis contributes to the body of knowledge on near-field beamforming. Although the specific application considered is speech pickup using microphone arrays, many of the results can be applied to a more general array processing scenario.

The fundamental contribution of this work is to show how to exploit the difference in curvature between the desired and interfering acoustic wavefronts to enhance the audio performance of a microphone array. Specific contributions within this broad area are identified below.

1. Incorporation of a spherical-wave target model into the array-gain optimization procedure and demonstration of performance benefits in isotropic noise.
 - Illustrate the variation in array gain as a function of near-field position [16]
 - Calculation of array gain as a function of distance to the focal point [17]
 - Analytic optimum solution for the specific case of incoherent noise [18]
 - Exploitation of increases in near-field gain to reduce gain variation with frequency for a broadband array [18]
2. Addition of beamformer response constraints to further enhance performance in non-isotropic noise fields.
 - Derivation of optimum weights and array gain for a linearly constrained beamformer
 - Demonstration of the performance degradation when forcing an exact null for a far-field interferer.
 - Derivation of a response constraint that reduces performance degradation.
3. Broadband beamformer design to achieve near-field optimum performance.
 - Application of complex approximation techniques to design FIR sensor filters

- Application of time-domain problem formulation to broadband beamformer implementation
 - Derivation of broadband spatial correlation functions for spherical-wave source and isotropic noise.
4. Experimental evaluation of broadband, near-field optimum beamformers for realistic applications.
- In the application to mobile telephony, fixed optimum array provides performance comparable to several adaptive beamformers [19]
 - In the application to computer telephony, a far-field response constraint is used to exploit non-isotropy in the reverberant sound field

Chapter 2

SPEECH PICKUP IN REVERBERANT ROOMS

Before attempting to optimize the response of a microphone array for speech pickup applications, it is important to have an understanding of the physical factors that affect overall system performance. This chapter discusses the acoustical factors affecting speech pickup in rooms from generation to reception.

2.1 Monophonic Speech Pickup in Rooms

In many situations it is necessary to use a remote microphone to pickup speech from a talker in a reverberant room. A remote microphone permits sound pickup from multiple parties in a room and eliminates the need for cumbersome wiring or other devices necessary to support personal microphones. Unfortunately, locating a microphone remotely from the talker causes the audio signal quality to be degraded by the effects of room reverberation and background noise. These audio problems are manifested for two reasons.

First, since the speech signals picked up by the microphone are usually presented monaurally, many advantages supplied by our binaural hearing system are lost. This phenomenon can be demonstrated qualitatively in a simple experiment [20]. More detailed experiments have also been performed in an attempt to quantify this effect [21][22][23][24][25][26].

Second, for reasons of convenience, the microphone is usually located at a significant distance from the talker. As a result, the direct sound amplitude is diminished due to spherical wave spreading and the relative levels of reverberation and noise increase. Together with the lack of binaural processing, this dramatically reduces the perceived audio quality.

If either the transmission channel is converted to stereo or the microphone is moved close to the talker, then the above problems with remote sound pickup will largely disappear. Since these solutions are not always feasible, other means must be sought for improving audio quality.

2.2 Using Directional Microphone Techniques to Improve Audio Quality

When a person speaks, sound waves are emitted in all directions [27]. If a microphone is used to pick up these sounds in a reverberant room, the recorded signal is influenced by a single *direct wave* and many *reflected waves*. As the name implies, a direct wave is that which is received via the direct acoustic path from the talker to the microphone. Reflected waves reach the microphone via indirect paths due to reflections from surfaces in the room. This is illustrated in Figure 4.

A directional microphone can be used to improve the quality of sound pickup by exploiting the difference in direction of propagation between the direct and reflected sound waves. The directional microphone is oriented so that the direction of propagation of the direct wave coincides with the peak response direction of the microphone. The reflected waves arrive from different directions, so they are attenuated by an amount equal to the receiver response at that angle. This is illustrated in Figure 5.

Many directional microphones achieve their performance by responding to sound pressure gradient (a vector quantity) instead of sound pressure (a scalar quantity). Through appropriate design, a large variety of directional response patterns can be achieved [28][29][30][31]. The directional response characteristics of a receiver can be summarized using the directional gain. The directional gain is the ratio of main beam response amplitude to the average response amplitude over all directions. Table 1 summarizes the directional gains of several common microphone types.

TABLE 1. Directional gains of several common microphone types (derived from the results in [28]).

Microphone	Directional Gain (dB)
Second-order gradient	7.0
Hypercardioid	6.0
Cardioid	4.8
First-order gradient	4.8
Omnidirectional	0

Figure 4. Illustration of the multiple sound paths between a talker and a microphone in a reverberant room. The solid line indicates the path of the direct wave from the source to the receiver and the dashed lines indicate the paths travelled by several room reflections.

Figure 5. Illustration of the effect of a directional microphone on the pickup of sound in a reverberant room. The shaded area indicates the receiving pattern of the directional microphone. Due to the difference in direction of propagation, the reflected waves are attenuated with respect to the direct wave.

Directional microphones, alone, do not solve all the problems with speech pickup for remote talkers. First, the modest directivity provided by these devices is insufficient to yield large increases in audio quality. Second, a directional microphone has a single, fixed main-response direction. If the talker moves or if a new talker becomes active, the directional microphone must be physically rotated to a new position or another microphone with a different orientation must be switched on. The mechanical systems required to achieve physical rotation and the lack of prior knowledge of possible talker locations make these approaches unattractive. A more attractive method for achieving a spatially-selective response pattern is to use an array of (either directional or non-directional) microphones. The low cost and high performance of digital signal processors and electret microphones makes this approach increasingly attractive.

A microphone array consists of a number of individual microphones distributed in space in a known pattern. The microphone output signals are combined according to a beam-forming algorithm which is designed to create a directional response pattern. The basic delay-and-sum beamforming algorithm adds the microphone output signals after delaying each one by an amount that matches the propagation delay of a desired wave front. Wave fronts whose propagation delays match those of the beamformer add coherently and are reinforced. Wavefronts whose propagation delays do not match those of the beamformer add incoherently and are reduced relative to the desired wavefront.

Through appropriate signal processing, a microphone array can provide substantial increases in directive gain over a single, directional microphone. In addition, multiple peak response directions can be created and steered towards any location, electronically [32][33].

2.3 The Acoustics of Speech Radiation from the Mouth

For the purposes of telephony, the dominant energy in speech signals is generally assumed to lie within the 300Hz to 3000Hz frequency range. Assuming a sound speed in air of approximately 340ms^{-1} , the resulting acoustical wavelengths range from approximately 1 m to 0.1 m [34]. It is this decade change in wavelength and frequency that poses a chal-

lenge in the design of many speech-pickup systems.

Another important consideration is the physical mechanism of sound radiation from the mouth. To design an optimum sound pickup device, information is needed about the directional nature of sound energy radiation from the mouth.

Dunn and Farnsworth measured the directional properties of human talkers in an extensive set of measurements [27]. The average sound power level in a number of frequency bands from 62.5Hz to 12kHz was measured at various positions on a spherical surface centered on a talker. Measurements were performed for radial distances from 30cm to 100cm from the lips. As an example, Figure 6 shows the sound radiation pattern for speech energy from the mouth measured at a distance of 30cm in three different frequency bands.

As seen in the figure, the average sound power is fairly constant over the front hemisphere surrounding the talker. This suggests that a reasonable acoustical model for speech radiation is an acoustic point source. In fact, such a model has been tested against a more elaborate model and found to provide better agreement with measured data [35].

The acoustic field at a point (r, θ, ϕ) due to a unit-strength point source located at the coordinate origin is given by

$$\frac{e^{jkr}}{r}.$$

Note that the acoustic field due to a point source depends only on distance r between the source and observation point. This is indicative of the fact that a point source is omnidirectional, unlike true speech radiation which is only approximately omnidirectional over the front hemisphere surrounding the talker. For the applications considered in this thesis, use of the point source model for speech sources is restricted only to those applications covering the front hemisphere.

Figure 6. Sound radiation pattern for speech energy from the mouth measured at a distance of 30 cm in three different frequency bands: solid line 250-500 Hz, dashed line 700-1000 Hz, dotted line 2800-4000 Hz.

2.4 Room Reverberation and Noise

2.4.1 Basic Room Acoustics

The transmission of sound between two points in a room is affected by the presence of reflections from the various surfaces within the room. The sound field observed at the receiver location consists of a single direct wave and multiple reflected waves. The direct wave is due to the line of sight propagation from source to receiver and thus is the first wavefront to arrive. Subsequent wavefronts are due to reflections from the various surfaces in the room. Since these wavefronts travel by indirect paths, the time of flight to the receiver is longer than that of the direct wave. This is illustrated in Figure 9.

The intensity of the direct wavefront measured at the sound receiver is determined by the source strength and the distance between the source and receiver. As the distance increases, the intensity of the direct wavefront decreases with the square of the distance [34].

In addition to attenuation with distance, the intensity of reflected sound waves is reduced due to acoustic absorption by the reflecting surfaces. The amount of absorption is determined by the acoustic properties of the surface materials. Table 2 lists the pressure reflection coefficient measured at various frequencies for several commonly encountered room surfaces. The pressure reflection coefficient gives the amplitude of the reflected pressure wave with respect to the incident pressure wave. A pressure reflection coefficient of 1 indicates perfect reflection and a coefficient of 0 indicates perfect absorption. As seen in this table, the absorption of most practical materials changes with the frequency of the incident sound wave. Consequently, reverberant sound fields are often complex and difficult to predict accurately.

TABLE 2. Acoustic pressure reflection coefficients of several common building materials measured at different frequencies (obtained from [36]).

Material	Reflection Coefficient					
	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Gypsum wallboard	0.85	0.95	0.97	0.98	0.96	0.96
Linoleum on concrete	0.99	0.98	0.98	0.98	0.98	0.98

TABLE 2. Acoustic pressure reflection coefficients of several common building materials measured at different frequencies (obtained from [36]).

Material	Reflection Coefficient					
	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Typical wood floor	0.92	0.94	0.95	0.97	0.97	0.97
Typical window glass	0.84	0.88	0.91	0.93	0.96	0.98
Carpet with underpadding	0.96	0.85	0.79	0.77	0.72	0.55
Thick drapery	0.92	0.80	0.67	0.55	0.52	0.50
Acoustic ceiling tile	0.75	0.59	0.32	0.28	0.24	0.17

Due to wall reflections, the amount of reflected acoustic energy reaching the receiver can often exceed the amount of direct acoustic energy. Of course, this depends on many factors including the distance from source to receiver and the directivity of the source and receiver.

Wall reflections can create large excursions in the frequency response measured from one point in a room to another. If a reflected wave adds in phase with the direct wave or other reflected waves at a certain frequency, constructive interference occurs producing a frequency response peak. If, however, the reflected waves are out of phase destructive interference occurs producing a frequency response dip. Figure 7 shows a simulated room impulse response for a source/receiver spacing of 1 m in a mildly reverberant room (reverberation time T_{60} of approximately 200ms^{-1}). The sampling frequency used to obtain this impulse response is 22050Hz. The corresponding frequency response magnitude is shown in Figure 7. Even though this room is only mildly reverberant, the frequency response demonstrates significant deviations from an ideal flat response.

It has been shown that the transfer functions observed in typical offices are non-minimum phase [37]. Consequently, direct inversion of a room frequency response by single-channel inverse filtering is not possible since any zeros outside the unit circle would result in an unstable pole. That is, non-minimum phase zeros cannot be inverted. Even if the inverse filter is designed to correct only the minimum phase part of the transfer function, it is not always perceptually effective [37][38].

Miyoshi and Kaneda showed that a non-minimum phase room response can be effectively

Figure 7. Impulse response of a simulated rectangular room with a reverberation time of approximately 200 ms. Spacing between source and receiver is 1m.

Figure 8. Frequency response of the room impulse response shown in Figure 7.

inverted if multiple acoustic channels (and multiple transducers) are available [39]. The invertibility depends on the fact that the different sound transmission channels will not share any common zeros in the transfer function.

2.4.2 Modeling Room Acoustics

Many different methods exist for modeling the acoustics of a room. The choice of a particular method depends upon the intended application. For predicting the sound quality of speech reception in rooms, the most appropriate method is the method of images [40].

The image method is based on the assumption that the wavelengths of the relevant frequencies are much smaller than the dimensions of the enclosure. In addition, it is assumed that all room boundaries are perfectly rigid so that all reflections are specular. That is, all frequencies are reflected equally. Although these assumptions do not strictly hold in realistic rooms, the image method does render a reasonable model of the impact of room boundaries on the transfer function from a point source to a point receiver.

The assumption of specular reflection greatly simplifies the calculation of a room impulse response. Under this assumption, the reflections from each room boundary can be represented by a mirror-image source located at a symmetric position behind the boundary as illustrated in Figure 9. The strength of the image source is equal to the strength of the original source multiplied by the reflection coefficient of the appropriate surface.

The positions of all the relevant image sources can be calculated from a knowledge of the room geometry, reflection coefficients (or conversely, absorption coefficients) and the position of the acoustic source within the room. Theoretically, the total number of image sources (reflections) is infinite; practically however, only those image sources of significant strength need to be included. The time duration at which the impulse response may be truncated with little error can be determined from the reverberation time of the room.

In order to produce a discrete-time impulse response, the times of arrival of the various image wavefronts must be quantized. This quantization introduces an error in the wavefront arrival time. This error can be controlled by using a sufficiently high sampling fre-

Figure 9. Location of image sources representing several horizontal reflections from the walls in a rectangular room. The solid lines indicate the actual paths travelled by the reflected waves; dashed lines indicate the paths travelled by the waves generated using the image sources.

quency for the quantization. The sampling frequency of the final impulse response generated can be reduced by interpolation, as described in [41] or [42].

The image method is widely used for simulating the effects of room acoustics on speech transmission. Using the interpolation procedure described above, the wavefront arrival times can be accurately simulated for the multiple microphones in an array permitting accurate modeling of the effects of array processing.

2.4.3 Relative Levels of Room, Microphone and Preamplifier Noise

Noise is also an important consideration in speech pickup systems. Noise can be introduced at any point along the signal path from the acoustic environment to the electronic equipment used to generate, transmit or store the electrical signals. To understand the relative importance of various noise sources two specific examples will be discussed: speech pickup in rooms and speech pickup in moving vehicles.

Sources of acoustic noise in rooms include office and ventilation equipment and the movements of people within the room. Two other significant noise sources are sensor noise and preamplifier noise. In the case of electret microphones, sensor noise originates in the microphone itself. It is caused by random (thermal) movements of the diaphragm and also by noise contributed by the built-in FET. Preamplifier noise is produced by the voltage and current noise sources associated with electronic amplifiers. In general, all these noise sources are observed over a broad range of frequencies; however, different noise sources dominate in different frequency bands, as shown in Figure 10.

Sources of noise in moving vehicles include the engine, tires, ventilation fans and also wind. Furthermore, each of these components will vary in importance depending on the type of vehicle, its speed and the road surface. Several representative spectra were measured in a typical passenger sedan for different driving conditions [19]. These spectra are shown in Figure 11.

2.4.4 Spatial Correlation of Noise

Consider the problem of array sound pickup from a talker in a reverberant room. In a

Figure 10. Relative levels of room noise, electret microphone noise and preamplifier noise measured in a quiet room. Room noise dominates at frequencies below 1 kHz.

Figure 11. Background noise spectra obtained in a moving vehicle under different driving conditions.

reverberant environment, since the direct wavefront travels the shortest distance to the array it possesses the smallest radius of curvature. Conversely, reflected waves travel a longer path and possess larger radii of curvature. For multiple reflections, as the distance travelled by the wave increases, the reflected waves may be accurately modeled as planar. Of course, the reverberant field is composed of many such waves propagating in different directions. In general, little information is available about the directions of arrival of the various reflected waves. A first-order approximation to describe the noise field, then, is to assume an equal probability for waves arriving from any direction. This leads to the well-known spherically isotropic, or diffuse, noise field. Insofar as the average energy in the reflected wavefronts appear less curved than that of the direct wave, optimizing array response using this model should increase the ratio of direct to reverberant energy at the output of the array.

As another example, there has been considerable interest in the problem of speech pickup inside passenger cars. The acoustic environment inside an automobile is characterized by low levels of reverberation and high levels of background noise when compared to a typical room [6][43]. The main source of speech interference in mobile telephony, therefore, is the background noise due to the engine, tires and wind. The spatial characteristics of this noise field have been studied in the context of enhancing speech pickup. Evidence shows that a reasonable description for these characteristics is also the spherically isotropic noise field [6][43].

Since an isotropic noise field can be thought of as resulting from the superposition of an infinite number of uncorrelated plane waves each with an equiprobable direction of propagation, the spatial correlation can be evaluated analytically [44]. For omnidirectional receivers separated by a distance d , the spatial correlation function $\rho(d)$ is given by the familiar sinc function

$$\rho(d) = \frac{\sin kd}{kd} \quad (1)$$

where k is the wavenumber. The spatial correlation functions for other common types of acoustic receivers can be found in [6] and [45].

For broadband beamforming information about the time behavior of the noise field is also important. The analysis of the spatial correlation in an isotropic field can be extended to include the effects of time delay in one of the receiver signals. The details of such an analysis are provided in Appendix A.

2.5 Objective Measures of Sound Pickup Systems

2.5.1 Room Acoustical Measurements

This section introduces a number of room acoustical measurements which are widely used to concisely describe room characteristics. Due to the complex nature of the sound field within a room, analysis and measurement of such sound fields are facilitated by the assumption of a diffuse sound field.

Perhaps the most fundamental measurement of reverberation is the ratio of the energy in the direct wave to the energy of all reflections. This *direct-to-reverberant energy ratio* is usually expressed in decibels and it gives an indication of the relative amounts of direct and reverberant energy. Since the direct-to-reverberant energy ratio varies with distance from source to receiver, a related measurement called the critical distance, or critical radius, or reverberation distance is often cited. The critical distance is the source-to-receiver spacing at which the direct and reverberant energies are equal (i.e. the direct-to-reverberant ratio is 0dB).

A widely-used measurement of room acoustics is the reverberation time. Reverberation time T_{60} is usually defined as the time required for the steady-state sound intensity in a room to decay by 60dB after the source is extinguished. First proposed by Walter Sabine, the reverberation time can be theoretically related to the average absorption coefficient of the room boundaries, the total area of these boundaries and the room volume [46]. This parameter can also be easily estimated from room impulse response measurements [47].

Another measurement of room acoustics is spectral deviation. Spectral deviation is an indication of the amount of fluctuation in a frequency response compared to a perfectly flat response. It is defined as the square-root of the sum-squared deviations of the fre-

quency response from an ideal flat response. Spectral deviation has been shown to be theoretically related to the direct-to-reverberant energy ratio in rooms [48].

Of the many studies in the area of speech quality, very few have been conducted under monaural or diotic conditions. One notable exception is a study reported by Berkley [49]. Berkley *et al.*, studied the problem of rating the speech quality obtained under various reverberant conditions. The study is particularly relevant to the current problem since listening preferences were recorded for diotic listening.

After collecting listener preference data under a wide variety of simulated reverberant conditions, a relationship was derived between the listener preference scores and two objective measures from room acoustics, reverberation time and spectral deviation. This relationship permits the prediction of the listener preference rating, on a nine-point scale, from objective measurements of a room impulse response. A general conclusion from this work is that users of communications systems tend to prefer speech which has less perceptible reverberation, as indicated by low values of direct-to-reverberant energy.

Chapter 3

INTRODUCTION TO ARRAY PROCESSING

This chapter reviews several key topics in array processing that are required for a thorough understanding of this thesis. First, the basic mathematical formulation of a general array processor is presented. The formulation assumes a narrowband processor for simplicity but it should be noted that the results are applicable to the broadband case, as discussed later. Next, two key array performance indices are introduced. It is shown that, by using matrix notation, convenient descriptions of these performance indices are possible. These descriptions then permit the methods of linear algebra to be applied in solving array optimization problems. In particular, the solutions to three array-gain optimization problems are explored: unconstrained, quadratically constrained and linearly constrained. Finally, the concept of near field is briefly discussed.

3.1 Basic Concepts and Mathematical Formulation

In the discussions and derivations to follow, it is necessary to refer to certain entities both as vectors, when dealing with the entire array, and as scalars, when focusing on a specific sensor. Furthermore, it is also necessary to consider both frequency-domain and time-domain versions. To maintain the connection between the various representations of the same entity, the same alphabetical symbol will be used in all instances. When required, vector and matrix variables are distinguished from their scalar counterparts using subscripts and bold-face type. Similarly, frequency-domain variables are distinguished from their time-domain counterparts using uppercase type and also by explicitly stating the frequency dependence. For instance, the time-domain FIR filter taps for sensor m are $w_m(n)$ and the time-domain stacked weight vector is $\mathbf{w}$. In the frequency domain, the response of the m^{th} sensor filter at frequency ω is $W_m(\omega)$ and the frequency-domain weight vector is $\mathbf{W}$. Similar definitions are used for other quantities.

Consider an array of M sensors immersed in a sound field. For a specific frequency ω , the output $U_m(\omega)$ of sensor m consists of a signal component $S_m(\omega)$ and a noise component

$$N_m(\omega)$$

$$U_m(\omega) = S_m(\omega) + N_m(\omega) ..$$

The array output $Y(\omega)$ is the sum of the weighted sensor outputs

$$Y(\omega) = \sum_{m=0}^{M-1} W_m^*(\omega) \cdot U_m(\omega)$$

where $W_m(\omega)$ is the complex weight for sensor m and $*$ denotes complex conjugation.

This can be written more succinctly by defining a vector of complex sensor weights $\mathbf{W}$

$$\mathbf{W}^T = [W_0(\omega) \ W_1(\omega) \ \dots \ W_{M-1}(\omega)]$$

and a vector of sensor inputs $\mathbf{U}$

$$\mathbf{U}^T = [U_0(\omega) \ U_1(\omega) \ \dots \ U_{M-1}(\omega)] .$$

The array output can then be written as the vector inner product

$$Y(\omega) = \mathbf{W}^H \mathbf{U} .$$

In the above definitions T denotes transpose and H denotes conjugate (Hermitian) transpose.

Typically, an array is designed to receive signals from a given direction while rejecting interference from other directions. The array's ability to perform this task is described by a performance index known as array gain.

3.1.1 Array Gain

Array gain is defined as the ratio of the array output signal-to-noise ratio over that of an individual sensor [50]. For a specific frequency ω the array gain $G(\omega)$ is given by

$$G(f) = \frac{E\{|\mathbf{W}^H \tilde{\mathbf{S}}|^2\} / (E\{|\mathbf{W}^H \tilde{\mathbf{N}}|^2\})}{\sigma_s^2 / \sigma_n^2} = \frac{E\{\mathbf{W}^H \tilde{\mathbf{S}} \cdot \tilde{\mathbf{S}}^H \mathbf{W}\} / \sigma_s^2}{E\{\mathbf{W}^H \tilde{\mathbf{N}} \cdot \tilde{\mathbf{N}}^H \mathbf{W}\} / \sigma_n^2}.$$

In this expression, $\tilde{\mathbf{S}}$ is the vector of signal components, $\tilde{\mathbf{N}}$ is the vector of noise components and σ_s^2 and σ_n^2 are the signal and noise powers observed at a single sensor, respectively. As an example, if the signal is a monochromatic planewave with frequency ω impinging upon the array, the signal vector is

$$\mathbf{S}^T = \begin{bmatrix} e^{j\frac{\omega}{c}\tau_0} & e^{j\frac{\omega}{c}\tau_1} & \dots & e^{j\frac{\omega}{c}\tau_{M-1}} \end{bmatrix} \quad (2)$$

where τ_m are the delays associated with the wave propagation across the array and c is the propagation speed. By defining the signal correlation matrix $\mathbf{R}_{ss}(\omega)$

$$\mathbf{R}_{ss}(\omega) = E\{\tilde{\mathbf{S}} \cdot \tilde{\mathbf{S}}^H\} / \sigma_s^2 = E\{\mathbf{S} \cdot \mathbf{S}^H\} \quad (3)$$

and the noise correlation matrix $\mathbf{R}_{nn}(\omega)$

$$\mathbf{R}_{nn}(\omega) = E\{\tilde{\mathbf{N}} \cdot \tilde{\mathbf{N}}^H\} / \sigma_n^2 = E\{\mathbf{N} \cdot \mathbf{N}^H\} \quad (4)$$

the above expression for array gain becomes

$$G(\omega) = \frac{\mathbf{W}^H \mathbf{R}_{ss}(\omega) \mathbf{W}}{\mathbf{W}^H \mathbf{R}_{nn}(\omega) \mathbf{W}}. \quad (5)$$

The array gain is thus described as the ratio of two quadratic forms (also known as a Rayleigh quotient). This description is convenient since it is in a form that is easily optimized using linear algebra techniques. This fact will be discussed in more detail later.

It should be noted that the expression for array gain includes separate expressions for the signal and noise correlation. In a practical setting, however, the only data available is the combined signal and noise observation. In general, separate estimates of the signal and noise are not possible. Nonetheless, the definition of array gain is useful in the analysis and design of arrays where at least rudimentary information is available about the signal and noise characteristics.

3.1.2 White Noise Gain

Array gain indicates a beamformer's ability to discriminate between signals and noise in the assumed propagation environment. In a practical beamformer implementation, however, there are sources of noise which do not arise from wave propagation. These include sensor and preamplifier noise as well as errors in sensor placement. Since these errors are uncorrelated from one sensor to the next, their effects can be represented by an appropriate amount of white noise added to each sensor signal [51]. A sensitivity analysis shows that an array's response to such errors is governed by a quantity known as white noise gain $G_w(\omega)$ [52], or alternatively its reciprocal noise sensitivity $\Psi(\omega)$ [53][54].

White noise gain can be calculated by assuming that the dominant noise source is uncorrelated from one sensor to the next. The resulting noise correlation matrix $\mathbf{R}_{nn}(\omega)$ reduces to an $M \times M$ identity matrix. White noise gain is obtained by substituting this into the expression for array gain

$$G_w(\omega) = \frac{\mathbf{W}^H \mathbf{R}_{ss}(\omega) \mathbf{W}}{\mathbf{W}^H \mathbf{I} \mathbf{W}}. \quad (6)$$

White noise gain describes the beamformer's gain *against* sensor and preamplifier noise and also against errors such as due microphone misplacement. The higher the value of $G_w(\omega)$, the more robust the beamformer. As an example, the white noise gain for an M -element delay-and-sum beamformer steered for plane waves is M . In this case, array processing reduces uncorrelated noise by a factor of M (improves the signal-to-noise ratio by a factor of M). By comparison, the white noise gain for a single sensor is 1, or 0dB.

3.1.3 Gain Optimization

An interesting problem in array processing is, given an array geometry and specific signal and noise correlation matrices, what choice of sensor weights yields the maximum array gain? Furthermore, what is the maximum array gain for this choice of sensor weights? Fortunately, with the description of array gain provided by (5), the answer to this question is easily found.

Recall that the array gain is the ratio of two quadratic forms. Provided that $\mathbf{R}_{nn}(\omega)$ is non-singular, its value is bounded by the minimum and maximum eigenvalues of the symmetric matrix $\mathbf{R}_{nn}^{-1}(\omega)\mathbf{R}_{ss}(\omega)$ [55]. The array gain is maximized by setting the weight vector $\mathbf{W}$ equal to the eigenvector corresponding to the maximum eigenvalue.

In the special case where $\mathbf{R}_{ss}(\omega)$ is a dyad, that is, it is defined by the outer product

$$\mathbf{R}_{ss}(\omega) = \mathbf{S}\mathbf{S}^H \quad (7)$$

then the weight vector $\mathbf{W}_{\text{opt}}$ that maximizes $G(\omega)$ is given simply by [56][57]

$$\mathbf{W}_{\text{opt}} = \alpha \mathbf{R}_{nn}^{-1}(\omega) \mathbf{S} \quad (8)$$

where α is an arbitrary complex constant. Inclusion of the complex constant α implies that the optimum weight vector is only unique to within a constant scale factor. In fact, it has been shown that the optimum weight solutions for several different optimization strategies can all be expressed as a scalar multiple of the basic solution $\mathbf{R}_{nn}^{-1}(\omega)\mathbf{S}$ [55].

The maximum array gain $G(\omega)_{\text{opt}}$ provided by the weights in (8) is

$$G(\omega)_{\text{opt}} = \mathbf{S}^H \mathbf{R}_{nn}^{-1}(\omega) \mathbf{S}. \quad (9)$$

Specific solutions for $\mathbf{W}_{\text{opt}}$ are determined by the exact values of the signal and noise correlation matrices, $\mathbf{R}_{ss}(\omega)$ and $\mathbf{R}_{nn}(\omega)$. In adaptive beamforming, these correlation matrices are estimated from the sensor signals and the optimum weights are computed in real time. In fixed beamforming applications, these correlation matrices are derived from information assumed or known about the signal and noise fields. For the applications considered in this thesis, simplified models are used to calculate $\mathbf{R}_{ss}(\omega)$ and $\mathbf{R}_{nn}(\omega)$. The choice of models suitable for near-field speech pickup is covered in a later chapter.

3.1.3.1 Quadratically Constrained Gain Optimization

Optimized beamformers have the potential to provide higher gain than available from delay-and-sum beamforming. Unfortunately, in many circumstances the resulting array can be very sensitive to the effects of microphone noise and positioning errors as indicated

by low white noise gain. In extreme cases, white noise gain is reduced to such an extent that the optimum gain is impossible to realize using practical sensors and tolerances.

Fortunately, a portion of the optimized gain can be retained, however, by constraining the loss of white noise gain during the optimization procedure [58]. The design of an optimum beamformer then becomes a trade-off between the array's sensitivity to sensor noise (and other uncorrelated errors) and the desired amount of gain over the spatial noise field. The problem to be solved is

$$\text{maximize } G(\omega) \text{ subject to } G_w(\omega) = \delta^2$$

where the constrained value of white noise δ^2 must be chosen less than or equal to the maximum possible value for the problem to be self-consistent. In the far-field beamforming case, this means that $\delta^2 \leq M$.

The solution to such a constrained optimization problem can be obtained using the method of Lagrange multipliers. Specifically, it is necessary to minimize

$$\frac{1}{G(\omega)} + \kappa \frac{1}{(G_w(\omega) - \delta^2)}$$

where κ is a Lagrange multiplier. This equivalent formulation of the above problem is seen to be easily solved by rearranging the terms in the cost function as follows

$$\frac{\mathbf{W}^H \mathbf{R}_{nn}(\omega) \mathbf{W}}{\mathbf{W}^H \mathbf{R}_{ss}(\omega) \mathbf{W}} + \kappa \frac{\mathbf{W}^H \mathbf{I} \mathbf{W}}{\mathbf{W}^H \mathbf{R}_{ss}(\omega) \mathbf{W}} = \frac{\mathbf{W}^H (\mathbf{R}_{nn}(\omega) + \kappa \mathbf{I}) \mathbf{W}}{\mathbf{W}^H \mathbf{R}_{ss}(\omega) \mathbf{W}}. \quad (10)$$

The cost function is the reciprocal of the array gain expression (5) with $\mathbf{R}_{nn}(\omega)$ replaced by $\mathbf{R}_{nn}(\omega) + \kappa \mathbf{I}$. The additional diagonal component $\kappa \mathbf{I}$ represents, in a simple manner, the effects of noise sources which are uncorrelated from one sensor to the next (i.e. due to preamplifier noise, position errors, etc.). The strength of the constraint is controlled by the magnitude of κ . Setting κ to a large value implies that the dominant noise is uncorrelated from microphone to microphone. When uncorrelated noise is dominant, the optimum weights are those of a conventional delay-and-sum beamformer. Setting $\kappa = 0$, of course,

produces the unconstrained optimum array. Unfortunately, there is no simple relationship between the constraint parameter κ and the constrained value of white noise gain δ^2 . Designing an array for a prescribed value of $G_w(\omega)$ requires an iterative procedure such as the one proposed in [51]. This approach has been used successfully in the design of practical super-directive microphone arrays for hearing-aid applications [59].

The above cost function can thus be minimized by maximizing its reciprocal as given by (10). The optimum weight vector is thus

$$\mathbf{W}_{opt} = (\mathbf{R}_{nn}(\omega) + \kappa \mathbf{I})^{-1} \mathbf{S}$$

where it is assumed that $\mathbf{R}_{ss}(\omega)$ is given by (7).

Of course, a suitable value of $G_w(\omega)$ must be selected. This choice will depend on the exact level of sensor and preamplifier noise present as well as the tolerance on the sensors and their positions. Lower sensor and preamplifier noise and more precise matching between microphones permits more white noise gain to be traded for array gain. As an example, the noise level (in equivalent sound pressure level) provided by modern electret microphones is of the order of 20 – 30dB SPL (that is, dB re: 20×10^{-6} Pa) whereas the acoustic background noise level of typical offices are in the vicinity of 30 – 45dB SPL [60]. Since the uncorrelated sensor noise is about 10 – 15dB lower than the acoustic background noise (due to the assumed noise field) it is possible to trade off some of the sensor SNR for increased rejection of environmental noise and reverberation.

3.1.3.2 Linearly Constrained Gain Optimization

For certain applications, it may be desirable to optimize array gain while constraining the beamformer response for a specific direction. This is typically done to force the array response to zero at some direction or to constrain the response in the look direction; however, derivative and eigenvector constraints can also be used to exert broader control over the response shape [61]. Such linear constraints can be expressed as

$$\mathbf{c}_l^H \mathbf{W} = g_l$$

where c_l is a $M \times 1$ constraint vector and g_l is the constrained gain. Constraint vectors are constructed in an identical manner to signal propagation vectors. Multiple constraints can be enforced by defining a constraint matrix C , whose L columns are the individual constraint vectors c_l , and a gain vector g . In this case, the constraint equation becomes

$$C^H W = g.$$

The optimization problem is now posed as

$$\text{maximize } G(\omega) \text{ subject to } C^H W = g.$$

This problem can be greatly simplified by restating the gain maximization goal in an equivalent form known as the linearly constrained minimum variance (LCMV) method. That is, maximizing array gain is equivalent to minimizing the noise response, as represented by the denominator of (5), with a constraint on the signal response, as represented by the numerator of (5). The signal response constraint can be handled easily as one of the columns in C . The optimization problem is to

$$\text{minimize } W^H R_{nn}(\omega) W + \kappa^H (C^H W - g) + \kappa^T (C^H W - g).$$

Again, the method of Lagrange multipliers can be used to transform the constrained optimization problem into an unconstrained one. The solution to this problem is [62]

$$W_{\text{opt}} = R_{nn}^{-1}(\omega) C [C^H R_{nn}^{-1}(\omega) C]^{-1} g.$$

In the case of a single, signal-response constraint

$$S^H W = g_s$$

then the linearly constrained optimum weights are

$$W_{\text{opt}} = \frac{R_{nn}^{-1}(\omega) S}{S^H R_{nn}^{-1}(\omega) S} g_s \quad (11)$$

which is equal to the unconstrained solution given in (8) with

$$\alpha = \frac{g_s}{S^H R_{nn}^{-1}(\omega) S}.$$

3.2 Near Field/Far Field Distinction

The notation introduced thus far has described signals and noise in general terms only. For specific solutions, it is necessary to consider the propagation environment in which the array is immersed. For the purposes of beamforming in free-space, it is sufficient to consider two wave propagation models: spherical wave and plane wave.

Consider a spherical-wave source (also known as a point source) whose location is given by the position vector $\mathbf{a}_s$, as depicted in Figure 12. Taking the coordinate origin as the phase reference, the acoustic signal $\tilde{S}_m(\omega)$ produced by a sensor located at $\mathbf{a}_m$ is represented by

$$\tilde{S}_m(\omega) = \frac{e^{jk|\mathbf{a}_s - \mathbf{a}_m|}}{|\mathbf{a}_s - \mathbf{a}_m|}$$

where k is the wavenumber. Note that the acoustic field due to a point source depends only on the distance $|\mathbf{a}_s - \mathbf{a}_m|$ between the source and observation point but that both the amplitude and phase are affected. The signal propagation vector $\tilde{\mathbf{S}}$ is then

$$\tilde{\mathbf{S}}^T = \left[\frac{e^{jk|\mathbf{a}_s - \mathbf{a}_0|}}{|\mathbf{a}_s - \mathbf{a}_0|} \quad \frac{e^{jk|\mathbf{a}_s - \mathbf{a}_1|}}{|\mathbf{a}_s - \mathbf{a}_1|} \quad \cdots \quad \frac{e^{jk|\mathbf{a}_s - \mathbf{a}_{M-1}|}}{|\mathbf{a}_s - \mathbf{a}_{M-1}|} \right]. \quad (12)$$

For plane-wave propagation, the distance to the sound source is assumed to be infinite as depicted in Figure 13. In this case, the vector $\mathbf{a}_s$ is a direction vector normalized so that $|\mathbf{a}_s| = 1$. The signal produced by a sensor located at $\mathbf{a}_m$ is now independent of distance and is represented by

$$S_m(\omega) = e^{jk(\mathbf{a}_s^T \cdot \mathbf{a}_m)} \quad (13)$$

where the phase reference is again taken at the coordinate origin. Thus, for plane waves, only the observed phase varies for different sensor positions.

Figure 12. Acoustic source located in the near field of an array.

Figure 13. Acoustic source located in the far field of an array.

In the context of beamforming, a source is said to be located in an array's far field if plane-wave propagation can be assumed without resulting in response errors. If wavefront curvature can be detected within the aperture of the array, however, then some response degradation will result from an assumption of plane-wave propagation [63]. In such cases a spherical-wave point source is a more appropriate signal model. Unfortunately, there is no single demarcation point between the near field and the far field that is suitable to all applications.

For conventional beamforming, one rule of thumb is that the near field extends to a distance of approximately $D^2/2\lambda$ where D is the largest dimension of the aperture and λ is the wavelength [64]. Expressions for the array's direct-wave response have been used to show that the response errors due to neglecting wavefront curvature depends on the angle of incidence [63].

For highly optimized beamformers, however, response distortion due to the incorrect assumption of plane-wave propagation can occur at much larger distances [16]. For optimized far-field beamformers distances of up to $10D^2/\lambda$ may be required to obtain accurate beampatterns [65][66].

Chapter 4

CURRENT APPROACHES TO NEAR-FIELD BEAMFORMING

This section discusses available methods for implementing a near-field beamformer. Beamforming notation is first introduced by describing conventional delay-and-sum beamforming for spherical waves. This is followed by an overview of recent advances in near-field beamformer design. As in the discussion of far-field techniques [62], the available beamforming approaches can be separated into two categories: those based on beam-pattern design and those based on optimizing an array performance index, usually array gain.

Beampattern design approaches attempt to implement an array with a prescribed spatial response pattern, often over a specified bandwidth. Due to the similarities between beam-pattern design and FIR filter design, many of the same techniques can be applied. Optimal beamformers attempt to maximize (or minimize) a beamformer's response based on some performance index. This is usually accomplished using statistical descriptions of the source and interference.

4.1 Delay-and-sum Beamforming for Spherical Waves

Consider an acoustic source in the near field of an M -element receiving array as depicted in Figure 12. The array consists of M omnidirectional sensors and the source location is given by the position vector $\mathbf{a}_s$. The acoustic field observed by sensor m located at $\mathbf{a}_m$ can be represented by

$$\frac{e^{jk|\mathbf{a}_s - \mathbf{a}_m|}}{|\mathbf{a}_s - \mathbf{a}_m|}$$

where k is the wavenumber. The array response $\bar{b}(\mathbf{a}_s, \omega)$ is formed by multiplying the signal from sensor m by the conjugate of a complex weight $W_m(\omega)$ and adding this to the other scaled sensor signals

$$\tilde{b}(\mathbf{a}_s, \omega) = \sum_{m=0}^{M-1} W_m(\omega)^* \cdot \frac{e^{jk|\mathbf{a}_s - \mathbf{a}_m|}}{|\mathbf{a}_s - \mathbf{a}_m|}.$$

Note that this expression uses the conjugate of the weights as denoted by $*$. This is to simplify matrix notation, which is introduced later. The response $\tilde{b}(\mathbf{a}_s, \omega)$ represents the array output as a function of source position and frequency and is referred to as the beam-pattern function. To facilitate comparisons of array performance with that of a single sensor, the array response can be normalized to that of a single sensor located at the coordinate origin. This yields the normalized beampattern $b(\mathbf{a}_s, \omega)$

$$b(\mathbf{a}_s, \omega) = \frac{|\mathbf{a}_s|}{e^{jk|\mathbf{a}_s|}} \cdot \sum_{m=0}^{M-1} W_m(\omega)^* \cdot \frac{e^{jk|\mathbf{a}_s - \mathbf{a}_m|}}{|\mathbf{a}_s - \mathbf{a}_m|}. \quad (14)$$

Using this normalization, as the source distance increases, $|\mathbf{a}_s| \rightarrow \infty$ and $1/|\mathbf{a}_s - \mathbf{a}_m| \approx 1/|\mathbf{a}_s|$ so that the near-field beampattern approaches the far-field beampattern.

For simplicity of notation, it is also convenient to choose the coordinate origin to coincide with the array phase. That is, choose

$$\sum_{m=0}^{M-1} \mathbf{a}_m = \mathbf{0}$$

where $\mathbf{0}$ is the null vector.

Using conventional delay-and-sum beamforming, the array can be focused at a point $\mathbf{a}_f$ by compensating for the wavefront propagation delays. This can be accomplished using the following uniform-magnitude sensor weights

$$W_m(\omega) = e^{jk|\mathbf{a}_f - \mathbf{a}_m|}$$

resulting in the array response

$$b(\mathbf{a}_f, \mathbf{a}_s, \omega) = \frac{|\mathbf{a}_s|}{e^{jk|\mathbf{a}_s|}} \cdot \sum_{m=0}^{M-1} \frac{e^{jk(|\mathbf{a}_s - \mathbf{a}_m| - |\mathbf{a}_f - \mathbf{a}_m|)}}{|\mathbf{a}_s - \mathbf{a}_m|}$$

where the dependence on focal point location $\mathbf{a}_f$ is explicitly shown. The behavior of this beampattern function can be demonstrated using a simple example.

Consider the specific case of a uniformly spaced linear array whose sensors are positioned along the z axis. The location of sensor m is given by z_m and the source is located at (r, θ) . Note that, due to axial symmetry, the array response is invariant to changes in ϕ . The relevant distances are

$$|\mathbf{a}_s - \mathbf{a}_m| = \sqrt{r^2 - 2rz_m \cos \theta + z_m^2} \text{ and} \quad (15)$$

$$|\mathbf{a}_f - \mathbf{a}_m| = \sqrt{r_f^2 - 2r_f z_m \cos \theta_f + z_m^2} \quad (16)$$

where (r_f, θ_f) are the coordinates of the array focal point. The beampattern function becomes

$$b(r_f, \theta_f, r, \theta, \omega) = \frac{r}{e^{jkr}} \cdot \sum_{m=0}^{M-1} \frac{e^{jk(r_{sm} - r_{fm})}}{r_{sm}}$$

where $r_{sm} = |\mathbf{a}_s - \mathbf{a}_m|$ and $r_{fm} = |\mathbf{a}_f - \mathbf{a}_m|$.

To illustrate the near-field behavior, the beampattern of a focused, 15-element, uniformly spaced linear array was evaluated along three circular paths in the near field, as illustrated in Figure 14. The radii of the circular paths are $((D)/2) + d$, D and $10D$ where D is the total array length and d is the sensor spacing, which is equal to one-half the operating wavelength λ . The beampatterns were obtained by maintaining a fixed focal point (r_f, θ_f) and varying the source location (r, θ) . This distinction is important since the beampattern is different from the steered response (which is obtained using fixed (r, θ) and varying (r_f, θ_f)) which is in contrast to the case of delay-and-sum beamformers in the far-field (as stated on page 127 in [50]).

Figure 14. Illustration of three circular paths within the near field of a uniformly spaced linear array. The array length is $D = (M-1)d$ where M is the number of elements and d is the sensor spacing. The radii of the three circular paths are: $r_1 = (D/2) + d$, $r_2 = D$ and $r_3 = 10D$. To maintain small details, the drawing is not to scale.

In the first set of beampattern curves the array is focused at $(r_f, \theta_f) = ((D/2) + d, \pi/2)$ corresponding to point B. The top plot in Figure 15 shows the normalized response evaluated along the path ABCDA where the array axis is along the horizontal axis of the plots. The radius of the circular path is equal to the focal distance of the array r_f . A distinct peak in the response is observed at the focal point $\theta_f = \pi/2$ although the magnitude of the peak is slightly less than M , the far-field array gain. This is caused by the nonuniform contributions of each sensor to the array output due to the varying distances r_{fm} to the focal point. Two other large response peaks are observed at angles 0 and π corresponding to points A and C. These peaks are due to the proximity of the test points to the outermost array sensors. The attenuation provided by the array at these angles is insufficient to overcome the increased amplitude due to source proximity.

The response along paths EFGHE and JKMOJ corresponding to radial distances of D and $10D$ are shown in the middle and bottom plots of Figure 15, respectively. Since the distance from the array exceeds the focal distance, the response magnitude decreases and also becomes less directional. Note that the reduction in magnitude is in addition to inverse square-law attenuation due to the normalization of the beampattern function. Thus, the near-field focused array is providing a certain degree of range discrimination.

For the second set of curves the array is focused at $(r_f, \theta_f) = (10D, \pi/2)$ corresponding to point K in Figure 14. Similarly, the response along paths ABCDA, EFGHE and JKMOJ with the array focused at point K are shown in the top, middle and bottom plots of Figure 16. For the cases where the path radius is smaller than the focal distance (that is, paths ABCDA and EFGHE), the response magnitude is smaller than that at the focal point, demonstrating range discrimination. Along the path JKMOJ, the response is very similar in shape to the far-field response of an identical array steered to an angle of $\pi/2$. This is to be expected since, as the distance from the array increases, the source moves from the near field into the far field and the beampattern approaches the far-field beampattern.

As in the case of far-field behavior, the near-field response shape will also be affected by the ratio of sensor spacing to the operating wavelength. The width of the main beam increases for reduced sensor spacing or longer wavelengths and vice versa.

Figure 15. Delay-and-sum beampatterns for a 15 element linear array with sensor spacing $d = \lambda/2$ focused at $(r_f, \theta_f) = ((D/2) + d, \pi/2)$. The beampattern is evaluated along three different paths as shown in Figure 14: path ABCDA (top), path EFGHE (middle) and path JKMOJ (bottom).

Figure 16. Delay-and-sum beampatterns for a 15 element linear array with sensor spacing $d = \lambda/2$ focused at $(r_f, \theta_f) = (10D, \pi/2)$. The beampattern is evaluated along three different paths as shown in Figure 14: path ABCDA (top), path EFGHE (middle) and path JKMOJ (bottom).

4.2 Near-field Beampattern Design

In many applications, the default beampattern provided by a delay-and-sum array is not adequate. When a specific beampattern shape is required, the methods described in this section can be used to design a beamformer. The assumption is, of course, that a suitable beampattern is known and that it can be implemented by the given array geometry.

4.2.1 Near-field Compensation

One method which can be used to design a beamformer with a prescribed near-field response shape is to employ a gain and phase compensation factor on the signal received from each microphone [8]. The compensation factor attempts to remove the phase delay associated with wavefront curvature and also the amplitude effects due to spherical-wave spreading.

Using the example of a uniform linear array, equation (14) can be rewritten using the substitutions in (15) and (16)

$$b(r_f, \theta_f, r, \theta, \omega) = \frac{r}{e^{jkr}} \cdot \sum_{m=0}^{M-1} W_m(\omega)^* \cdot \frac{e^{jkr_{sm}}}{r_{sm}} \cdot g_m \quad (17)$$

where the compensation factors g_m have been included. The compensation factors are defined to provide correction at a fixed array focal point (r_f, θ_f) . That is

$$g_m = \frac{r_{fm}}{r_f} \cdot e^{jk(r - r_{fm} + z_m \cos \theta_f)}.$$

Substituting this definition into (17) yields the compensated beampattern function $b_c(r_f, \theta_f, r, \theta, \omega)$

$$b_c(r_f, \theta_f, r, \theta, \omega) = \sum_{m=0}^{M-1} W_m(\omega)^* \cdot \frac{r e^{jk(r_{sm} - r)}}{r_{sm}} \cdot \frac{r_{fm} e^{jk(r_f - r_{fm})}}{r_f} \cdot e^{jkz_m \cos \theta}.$$

When the focal point and the actual source location are close together $r_f \approx r$ and $r_{fm} \approx r_{sm}$ and the compensated beampattern simplifies to

$$b_c(r_f, \theta_f, r, \theta, \omega) = \sum_{m=0}^{M-1} W_m(\omega)^* \cdot e^{jkz_m \cos \theta}$$

which resembles a far-field beampattern function. Consequently, conventional far-field beamformer design techniques can be used to derive appropriate sensor weights to provide the desired near-field beampattern.

4.2.2 Radial Beampattern Transformation

This near-field design method is based on the spherical harmonic solution to the wave equation [67][68]. In the method, a desired near-field beampattern is transformed to its equivalent far-field beampattern. Conventional design techniques are then used to design a beamformer with the required far-field response pattern. If the transformed far-field pattern is achieved, then the resulting beamformer will have the desired near-field beampattern at the specified range.

The transformation is derived by noting that a beampattern function $b \equiv b(r, \theta, \phi, t)$ will satisfy the wave equation which, in spherical coordinates, is given by

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial b}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial b}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 b}{\partial \phi^2} = \frac{1}{c^2} \frac{\partial^2 b}{\partial t^2}.$$

The solution is given by

$$b(r, \theta, \phi, k) = \frac{1}{\sqrt{r}} \left(\sum_{n=0}^{\infty} \alpha_n H_{n+1/2}^{(1)}(kr) P_n(\cos \theta) + \sum_{n=1}^{\infty} \sum_{m=1}^n H_{n+1/2}^{(1)}(kr) P_n^m(\cos \theta) \times (\beta_{nm} e^{jm\phi} + \gamma_{nm} e^{-jm\phi}) \right) \quad (18)$$

where $P_n(\cdot)$ is the Legendre function of order n , $P_n^m(\cdot)$ are associated Legendre functions and $H_{n+1/2}^{(1)}(\cdot)$ are half-integer order spherical Hankel functions of the first kind. The coefficients α_n , β_{nm} and γ_{nm} are calculated from the following analysis equations

$$\alpha_n = \frac{n + \frac{1}{2}}{\frac{2\pi H_{n+1/2}^{(1)}(kr)}{\sqrt{r}}} \times \int_0^{2\pi} \int_0^\pi b(r, \theta, \phi, k) P_n(\cos \theta) \sin \theta d\theta d\phi \quad (19)$$

$$\beta_{nm} = \frac{n + \frac{1}{2}}{\frac{2\pi H_{n+1/2}^{(1)}(kr)}{\sqrt{r}}} \cdot \frac{(n-m)!}{(n+m)!} \times \int_0^{2\pi} \int_0^\pi b(r, \theta, \phi, k) P_n^m(\cos \theta) \sin \theta e^{-jm\phi} d\theta d\phi \quad (20)$$

and

$$\gamma_{nm} = \frac{n + \frac{1}{2}}{\frac{2\pi H_{n+1/2}^{(1)}(kr)}{\sqrt{r}}} \cdot \frac{(n-m)!}{(n+m)!} \times \int_0^{2\pi} \int_0^\pi b(r, \theta, \phi, k) P_n^m(\cos \theta) \sin \theta e^{jm\phi} d\theta d\phi. \quad (21)$$

The procedure consists of three steps:

1. Calculate the analysis coefficients using the desired beampattern at the design distance using equations (19)(20) and (21).
2. Transform the near-field beampattern into an equivalent far-field beampattern using $r = \infty$ in equation (18).
3. Design a far-field beamformer to implement the beampattern from Step 2 using conventional techniques.

This technique applies to arrays with arbitrary geometry. Furthermore, a desired beampattern can be achieved over an arbitrary bandwidth by extending the method using the results of [69]. The success of this design directly depends on whether the desired beampattern can be achieved using the chosen array geometry. In practice, of course not all beampatterns will be achievable using all array geometries. This method provides no information on choosing an appropriate beampattern nor on determining if a chosen beampattern can be achieved using a given array geometry.

A further twist on this approach has been used to simultaneously achieve both a desired near-field pattern as well as a desired far-field pattern [70]. Since an exact solution to this problem is, in general, not possible a least-squares solution is used to permit matching both patterns with minimum mean-squared error.

4.2.3 Multidimensional Filter Design

Nordebo *et al.* [71] have used a multidimensional filter design approach to achieve a specified beamformer response in the space-frequency plane. The intent is to achieve a specified physical beampattern (spatial dimension) over a wide bandwidth (frequency dimension). A weighted Chebyshev approximation is used to minimize the maximum value of the approximation error. The method is similar to the minimax design procedure used in the Remez exchange algorithm for FIR filter design. An advantage of this approach is that the beampattern need not be specified in detail. Instead, a loose description in terms of passband/stopband behavior can be employed.

4.2.4 Nonlinear Goal Programming

Pirz combines the concept of a harmonically-nested array with a goal optimization method [9] to optimize the response pattern of a linear array over a near-field region. The goal programming differs from conventional least-squares optimization in that only those errors which reduce the performance are included. That is, deviations from the ideal response which exceed specification are not included in the error calculation. The optimization is applied to the polar response pattern at various distances as well as to the response of the array in a plane parallel to the array axis.

The advantage of this method is that instead of a specific beampattern target, only a beam-pattern goal is required. That is, only a rough description of the desired array response is necessary. The drawback is that the method is somewhat blind to array performance indices. As mentioned in [9] single frequency designs often resulted in superdirective arrays that are of limited practical use unless a proper constraint is employed [58].

4.2.5 Near-field Null Placement

While not strictly a near-field beamformer design, this approach has near-field application potential. Steyskal demonstrated the placement of near-field nulls in such a way as to limit the perturbation produced in a desired far-field response pattern [72]. The method was applied to a case in which there was a significant near-field scattering object. A null was placed at the near-field location of the scattering object resulting in a much smaller pertur-

bation of the desired far-field pattern.

The method is formulated as a constrained least-squared approximation problem. The mean-squared error between the desired $b_d(\theta, \omega)$ and actual $b(\theta, \omega)$ far-field beampatterns are minimized subject to the constraint of zero near-field response at a set of P points (r_p, θ_p) . For a uniformly spaced linear array this can be expressed as minimizing

$$\int_{\theta} |b_d(\theta, \omega) - b(\theta, \omega)|^2 d\theta$$

subject to

$$b(r_p, \theta_p) = 0 \quad p = 1, 2, \dots, P.$$

In reference [72] an example is provided that illustrates minimal perturbation of the far-field beampattern for a single, near-field null.

4.3 Gain Optimization

An alternate approach to beamformer design is based on the optimization of a global performance index. The advantage of this approach is that a suitable array performance index can be devised which has relevance to the particular application. Optimizing the performance index then leads directly to improved performance.

4.3.1 Optimum Array Gain for Spatially Incoherent Noise

In an early application of microphone arrays for speech pickup in cars, Goulding analyzed system performance using a signal-to-noise ratio criterion [6]. This study was restricted to the case of a circular microphone array centered on the mouth of the talker. Furthermore, the noise field was assumed to be incoherent from one microphone to another.

Although the mathematical formulation did not use matrix notation, the expression obtained for signal-to-noise ratio provides equivalent results to that shown in (9) assuming a spatially incoherent noise field. By employing the Schwartz inequality Goulding showed that the real microphone weights that maximize signal-to-noise ratio in incoherent noise

are given by

$$w_m = \frac{\alpha}{r_m}$$

where α is an arbitrary constant.

Following a slightly different strategy, the same result was reported by Nomura *et al.* [73].

4.3.2 Gain Optimization using Sensor Weight and Position Adjustment

Many array design problems are formulated assuming fixed sensor locations. The design problem then reduces to that of finding an optimal set of sensor weights to achieve the desired response. Fortunately, problems formulated in this manner have well-behaved solution subspaces and it is not uncommon that unique, optimum solutions are found. This is not the case, however, if the sensor positions are varied as part of the optimization procedure. In this case, the error performance surface can exhibit multiple minima and procedures must be sought to prevent locally optimum solutions. Berger and Silverman have addressed such a design task [10].

A performance index (the power spectral dispersion function) was devised that restricts noise and signal sources to a line parallel to the linear array's axis. The gains and linear positions of an 8-element array were then optimized using a numerical procedure known as stochastic region contraction. This procedure is specially devised to ensure a globally optimum solution. Comparisons are made with the method of simulated annealing.

Chapter 5

NEAR-FIELD GAIN OPTIMIZATION

As discussed in Chapter 3, array gain $G(\omega)$ at a specific frequency ω is conveniently expressed in matrix notation as

$$G(\omega) = \frac{\mathbf{W}^H \mathbf{R}_{ss}(\omega) \mathbf{W}}{\mathbf{W}^H \mathbf{R}_{nn}(\omega) \mathbf{W}}. \quad (22)$$

To design an optimum fixed-coefficient beamformer for near-field applications, appropriate choices must be made for $\mathbf{R}_{ss}(\omega)$ and $\mathbf{R}_{nn}(\omega)$. In speech-pickup applications, the acoustic source of interest is the talker's mouth. The choice of a suitable beamforming model for speech signals must take into account the physical properties of speech-energy radiation. As discussed in Chapter 2, radiated speech energy is distributed in a uniform manner over the front hemisphere surrounding the talker. Consequently, a reasonable model to use in beamforming is that of an acoustic point source.

For a point source whose position is given by $\mathbf{a}_s$, the signal propagation vector $\tilde{\mathbf{S}}$ is

$$\tilde{\mathbf{S}}^T = \left[\frac{e^{jk|\mathbf{a}_s - \mathbf{a}_0|}}{|\mathbf{a}_s - \mathbf{a}_0|} \quad \frac{e^{jk|\mathbf{a}_s - \mathbf{a}_1|}}{|\mathbf{a}_s - \mathbf{a}_1|} \quad \cdots \quad \frac{e^{jk|\mathbf{a}_s - \mathbf{a}_{M-1}|}}{|\mathbf{a}_s - \mathbf{a}_{M-1}|} \right]. \quad (23)$$

where $\mathbf{a}_m$, $m = 0 \dots M-1$ are the sensor locations. For a point source in the near field of a receiving array, unequal signal amplitudes are observed at each sensor. This complicates the normalization of $\mathbf{R}_{ss}(\omega)$ (see (3)) since the observed signal power σ_s^2 varies between sensors. To provide a consistent normalization, it is necessary to identify a reference location and define σ_s^2 accordingly. The reference location will affect the apparent gain of the array, however, so it must be chosen carefully. The best choice is usually the location of the sensor that is closest to the signal source. That is, choose the reference location $\mathbf{a}_r$, such that

$$|\mathbf{a}_r| = |\mathbf{a}_s - \mathbf{a}_m|_{\min}$$

so the signal power for the unit strength source becomes

$$\sigma_s^2 = \frac{e^{jk|a_r|}}{|a_r|} \cdot \frac{e^{-jk|a_r|}}{|a_r|} = \frac{1}{|a_r|^2}.$$

A normalized signal propagation vector $\mathbf{S}$ can then be defined as

$$\mathbf{S}^T(f) = \frac{|a_r|}{e^{jk|a_r|}} \left[\frac{e^{jk|a_s - a_0|}}{|a_s - a_0|} \frac{e^{jk|a_s - a_1|}}{|a_s - a_1|} \cdots \frac{e^{jk|a_s - a_{M-1}|}}{|a_s - a_{M-1}|} \right]$$

so $\mathbf{R}_{ss}(\omega)$ becomes

$$\mathbf{R}_{ss}(\omega) = \mathbf{S} \cdot \mathbf{S}^H.$$

Unfortunately, the sources of interference for speech pickup do not normally come from a single spatial location, as does the speech. Instead, noise and reverberation can arrive at the array from a large number of *a priori* unknown directions. For this reason, it is most expedient to use an average statistical model to describe the spatial distribution of noise sources. In this way, the average response to noise and interference is minimized relative to the desired signal response. In the optimization procedures to follow, the spherically isotropic noise field is assumed. As discussed in Section 2.4.4 on page 23, there is justification for its use in microphone array applications. Furthermore, in the absence of detailed information about the noise field this assumption will yield a beamformer that has the highest gain in an average sense.

5.1 Unconstrained Optimization

The foregoing signal and noise models can be used to optimize the near-field response of an array. Due to the nature of an isotropic noise field, optimum beamformer designs obtained with this procedure naturally fall into two categories [18]: those obtained for a spatially uncorrelated noise field and for a partially correlated noise field. The former case occurs for a uniformly spaced linear array in an isotropic field when the sensor spacing is equal to a half wavelength. It also occurs in the case when the dominant noise is uncorre-

lated sensor or preamplifier noise. The latter case is more general and occurs in an isotropic field when the above sensor spacing requirement is not met.

5.1.1 Spatially Uncorrelated Noise

When the observed noise signals are uncorrelated between microphones $\mathbf{R}_{nn}(\omega)$ is an $M \times M$ identity matrix. This matrix is readily inverted and the optimum weight solution according to (8) is simply the signal propagation vector [18]. That is

$$\mathbf{W}_{opt} = \alpha \mathbf{S} \quad (24)$$

which for the case of a point source in the near field is given by (23). Thus, for spatially uncorrelated noise, the optimum beamformer weights are real although not uniform in magnitude. Instead, the optimum weights are inversely proportional to the distance to the signal source. To achieve the maximum near-field array gain those microphones that are closest to the focal point must be given the highest weighting. This is intuitively appealing since signals from the closest microphones will have the highest signal-to-interference ratio prior to beamforming. Furthermore, as the source position approaches the location of one of the sensors, the optimum weight amplitude increases for that sensor but decreases for all other sensors. Thus, the performance of the array gradually approaches the performance of a single sensor as the source position approaches the sensor location. At the other extreme, as the source position moves away from the sensors, the optimum weight amplitudes approach a uniform weighting which is the optimum far-field solution for an equivalent array.

The optimum array gain is obtained by substituting the above weight solution into (22)

$$G_{opt}(\omega) = \sum_{m=0}^{M-1} \frac{|\mathbf{a}_s|^2}{|\mathbf{a}_s - \mathbf{a}_m|^2}. \quad (25)$$

As the distance between the array and the source position increases (i.e. as $|\mathbf{a}_s| \rightarrow \infty$) then the optimum near-field gain approaches M the far-field gain obtained using delay-and-sum beamforming. Consequently, this optimum solution can be viewed as the near-field equivalent of delay-and-sum beamforming [18].

This weight solution has been reported previously in [6] and [73]. These authors assumed an uncorrelated noise field, regardless of sensor spacing. In an isotropic noise field, when all sensors are spaced by $\lambda/2$, these solutions are identical. If half-wavelength spacing is not maintained, however, the noise due to an isotropic field is not completely uncorrelated between sensors. The optimum weights in (24) can be used in a broadband beamformer but are not expected to provide optimum gain at frequencies where the microphone spacing is not $\lambda/2$ or an integer multiple of it. This point will be explored further in later discussions.

The optimum weights for uncorrelated noise also provide a very robust beamformer. In fact, the maximum white noise gain possible for an array focused on a near-field point is provided by these optimum weights and not by uniform weights [18]. To see this, consider the expression describing the white noise gain $G_w(\omega)$

$$G_w(\omega) = \frac{\mathbf{W}^H \mathbf{R}_{ss}(\omega) \mathbf{W}}{\mathbf{W}^H \mathbf{I} \mathbf{W}} .$$

This equation is also in the form of a ratio of two quadratic forms. The expression is equivalent to (22) with $\mathbf{R}_{nn}(\omega)$ replaced by an $M \times M$ identity matrix. Consequently, the weight vector $\mathbf{W}_{\max}$ that maximizes this ratio is the same as that which maximizes the array gain when $\mathbf{R}_{nn}(\omega) = \mathbf{I}$. That is

$$\mathbf{W}_{\max} = \mathbf{S} .$$

Again, the weights that maximize $G_w(\omega)$ are not uniform in amplitude as for a delay-and-sum beamformer for far-field sources. The resulting maximum value of $G_w(\omega)$ is equivalent to (25)

$$G_{w \max}(\omega) = G_{\text{opt}}(\omega) = \sum_{m=0}^{M-1} \frac{|\mathbf{a}_s|^2}{|\mathbf{a}_s - \mathbf{a}_m|^2} \leq M . \quad (26)$$

Note that this is less than M , the value of white noise gain obtained for a delay-and-sum beamformer for far-field sources.

5.1.2 Partially Correlated Noise

In cases where the observed noise signals are not completely uncorrelated between microphones $R_{nn}(\omega)$ does not have a convenient analytical form. In this case, a closed-form solution to the weight optimization problem is not readily available and numerical solutions must be sought.

While specific examples will be discussed in the next section, a few general observations about the optimum solutions can be made. For microphone spacing that is smaller than $\lambda/2$, super-directive arrays are produced. As for super-directive beamformers in the far-field, the optimum weights are complex and include a phase adjustment in addition to that required for wavefront alignment. The beamformer has the potential to provide higher gain than that of a delay-and-sum beamformer, however, such gains are difficult to fully realize because the resulting array possesses low white noise gain, indicating a high sensitivity to random errors. A portion of the gain increase can be practically achieved, however, by constraining the loss of white noise gain during the optimization procedure [58][59]. This point is discussed more fully in Section 5.2.

5.1.3 Numerical Example

The effectiveness of the optimization will be demonstrated using a 16-element linear array with a uniform sensor spacing of d and operating wavelength λ . With reference to the classification above, two specific cases will be presented: one for $d = \lambda/2$ and another for $d = \lambda/4$. See Figure 14 for a description of the array position relative to the coordinate axes.

As discussed in Chapter 3, the size of the near field for an array of maximum dimension D can be approximated by $D^2/(2\lambda)$ [64]. For an M -element uniformly spaced array with maximum dimension $D = (M-1)d$, the near field extends to a distance of $(D(M-1)/2)(d/\lambda)$. At a frequency corresponding to $\lambda/2$, the near field extends to $3.75D$ whereas at a frequency corresponding to $\lambda/4$, it extends to only $1.875D$. In the examples to follow, the near-field gain is calculated over a square region $6D$ on a side centered on the array.

For the first example, the optimum gain was calculated over the near-field region defined above [16]. For this example, as the focal point is moved around the near field, the reference point is changed to coincide with the closest microphone location. That is, the calculated array gain is always equal to the gain over the closest omnidirectional microphone. The results of the calculation are shown in Figure 17. For points far from the array, the optimum gain equals that of a far-field delay-and-sum beamformer (i.e. $10\log M$ or 12dB). Closer to the array, however, the optimum gain drops approaching 0dB in the immediate vicinity of the individual sensors. The reduction in array gain extends to the farthest distance for focal points in the endfire direction. This is due to the large change in source-to-sensor distance across the array and the resulting effect on the summation (25). The change in source-to-sensor distance, and hence the change in array gain, is much less pronounced at broadside.

As noted earlier, the white noise gain is at its maximum for this frequency. Since the expression for $G_{w \max}$ (26) is the same as for array gain at this frequency, Figure 17 is identical to the variation of white noise with focal point.

To highlight the effect of optimization, the array gain obtained from a uniformly weighted delay-and-sum beamformer was removed from the optimum gain. The improvement due to optimization is plotted as a function of near-field focal point in Figure 18. As seen in the figure, the gain improvement due to the optimization is small over much of the near-field region. Significant gain enhancement over delay-and-sum beamforming ($> 1\text{ dB}$) is noted only for positions very close to the array ($< D/4$). This reflects the fact that for a microphone spacing of $\lambda/2$, the optimum weights are nearly uniform in amplitude except for focal points located very close to the array. Due to the uncorrelated nature of the noise signals at this frequency, only the varying amplitudes in the received sensor signals are exploited.

The ability of a near-field focused array to discriminate sources at varying distances can be demonstrated using beampattern plots. The beampatterns of both uniformly weighted and optimized beamformers evaluated at varying distances from the array are compared. In both cases, the array is focused at a distance equal to the array length directly broadside

Figure 17. Contour plot of the optimum near-field gain for a 16-element uniformly spaced linear array with a microphone spacing of $\lambda/2$. Contour lines are spaced by 1dB and the x and y coordinates are normalized to the total length of the array. The positions of the array microphones are illustrated by the solid circles along the x axis at $y = 0$.

Figure 18. Contour plot of the net effect of near-field gain optimization for the array shown in Figure 17. The gain of an equivalent delay-and-sum beamformer has been removed to highlight the effect of the optimization.

to the array (this corresponds to point F in Figure 14 and coordinates $(r_f, \theta_f) = (D, \pi/2)$).

Figure 19 shows three array beampatterns calculated at a frequency corresponding to $\lambda/2$ sensor spacing for a uniformly weighted beamformer. The beampatterns were evaluated at three distances: $(D/2) + d$, D and also at ∞ (i.e. a far-field response). At the focal distance, the response pattern exhibits a mainlobe amplitude of almost 16 in the focal direction $\pi/2$, indicating correct focusing. For distances less than or greater than the focal distance the mainlobe amplitude is smaller indicating range discrimination. Figure 20 shows the corresponding beampatterns for the optimized beamformer. The similarity between the uniformly weighted and optimum beampatterns underscores the similar behavior of these two designs for focal points located a distance more than $D/4$ from the array.

For the second example, the gain of an identical array but with a microphone spacing of $\lambda/4$ is considered. The optimum gain was calculated over the same near-field region and the gain of a delay-and-sum beamformer was removed. The net effect of the optimization is shown in Figure 21. As seen in the figure, the optimization provides a higher gain improvement over a wider physical region than in the case of $\lambda/2$ microphone spacing. Unfortunately, the increase in array gain comes at the expense of robustness since the white noise gain is approximately -50dB or less over much of the near-field region. Consequently, there is little hope of realizing a beamformer with this gain.

Beampatterns were evaluated for the same three distances as used in the previous example for both a uniformly weighted and optimized array. The response with uniform weights is shown in Figure 22 and that for the optimum array shown in Figure 23. As in the previous example, the beampatterns both exhibit strong peaks at the focal point, indicating correct focusing, and in the far field, the optimum design has a lower peak response magnitude compared to the uniform weights. The beampatterns obtained at a distance of $(D/2) + d$, however, show significant differences. The optimum beampattern displays significant grating lobes along the endfire directions. These lobes are higher than in the uniformly weighted beamformer because the super-directive nature of the optimum design results in

Figure 19. Delay-and-sum beampatterns for a 16-element uniformly spaced linear array with $\lambda/2$ sensor spacing. The array is focused at $(r_f, \theta_f) = (D, \pi/2)$ and the beampattern is evaluated at three different distances: $r = (D/2) + d$ (top), $r = D$ (middle) and $r = \infty$ (bottom).

$$r = \frac{D}{2} + d$$

$$r = D$$

$$r = \infty$$

Figure 20. Optimum beampatterns for the array in Figure 19.

Figure 21. Contour plot of the optimum near-field gain for the same conditions as in Figure 17 but with a microphone spacing of $\lambda/4$.

large weight amplitudes. At a distance of $(D/2) + d$, points near endfire are in close proximity to the outermost sensors and their responses dominate the array output due to the extremely high weight amplitudes.

5.2 Quadratically Constrained Optimization

As discussed in the previous section, unconstrained optimization can result in a beamformer that is impractical to realize. Realistic optimum designs can be obtained, however, by employing a quadratic constraint in the optimization procedure. As shown in Section 3.1.3.1 on page 33, a quadratic constraint can be enforced by adding a small diagonal component to the noise correlation matrix. That is, define

$$\mathbf{R}_{nn}(\omega) = \mathbf{Q}(\omega) + \kappa \mathbf{I} \quad (27)$$

where $\mathbf{Q}(\omega)$ is the noise correlation matrix associated with the isotropic field and $\kappa \mathbf{I}$ is the diagonal component. The optimum weights are then

$$\mathbf{W}_{opt} = (\mathbf{R}_{nn}(\omega) + \kappa \mathbf{I})^{-1} \mathbf{S} \quad .$$

Unfortunately, there is no simple relationship between the constraint parameter κ and the resulting value of white noise gain $G_w(\omega)$. To design a beamformer for a prescribed value of $G_w(\omega)$ an iterative procedure is used.

As discussed in Section 3.1.3.1 on page 33, the choice of a suitable value of $G_w(\omega)$ will depend on the exact level of sensor and preamplifier noise present as well as the tolerance on the sensors and their positions. Lower sensor and preamplifier noise and more precise matching between microphones permits more white noise gain to be traded for array gain.

For the purposes of this thesis, a white noise gain level of 0dB is used. This choice should yield an optimized beamformer with the same sensitivity to noise as a single microphone. It should also be noted that this limiting value has been used successfully in the design of practical super-directive microphone arrays for hearing-aid applications [59].

Figure 22. Delay-and-sum beampatterns for a 16-element uniformly spaced linear array with $\lambda/4$ sensor spacing. The array is focused at $(r_f, \theta_f) = (D, \pi/2)$ and the beampattern is evaluated at three different distances: $r = (D/2) + d$ (top), $r = D$ (middle) and $r = \infty$ (bottom).

$$r = \frac{D}{2} + d$$

$$r = D$$

$$r = \infty$$

Figure 23. Optimum beampatterns for the array in Figure 22. Note the change in scale for the top graph.

5.2.1 Numerical Example

Since a quadratic constraint is only necessary to limit loss of white noise gain for optimized arrays with $d < \lambda/2$, only the example for $d = \lambda/4$ is continued.

The constrained optimum gain was calculated over the near-field region defined above. As before, when the focal point is moved around the near field, the reference point is changed to coincide with the closest microphone location. The net improvement over delay-and-sum beamforming is plotted in Figure 24. By constraining the optimization, the maximum array gain possible has been reduced. For a distance of $D/2$ in the broadside direction, the gain improvement drops from approximately 4dB with no constraint to approximately 1dB with a white noise gain constrained to 0dB.

Note that the improvement in array gain extends farther for focal points in the broadside direction. In fact, in the endfire direction, there seems to be little dependence of array gain on distance. This can be explained by noting that the wavefront-curvature phase difference between sensors is only observed for directions other than endfire. At endfire, the linear array can only distinguish a plane wave from a spherical wave based on amplitude information. The difference in signal amplitude between sensors does not provide a large potential for gain improvement.

The beampattern of the constrained optimum design is shown in Figure 25. This can be compared with the beampattern of an equivalent uniformly weighted array shown in Figure 22 and of the unconstrained optimum array shown in Figure 23. As before, the optimized array displays correct focusing and a reduction in the far-field response, compared to the uniformly weight array. In this case, however, it is seen that the extremely high grating lobes have been greatly reduced by limiting the minimum white noise gain.

Achievable gains for the broadside direction can be explored further by considering focal points at varying distances along the x -axis for $z = 0$ (see Figure 14). Figure 26 shows the increase in array gain over a delay-and-sum beamformer as a function of white noise gain for several focal distances. As shown in the figure, for focal points far from the array, the improvement due to optimization increases very slowly with reductions in white noise

Figure 24. Contour plot of the net effect of constrained near-field gain optimization for the array shown in Figure 21. White noise gain is constrained to a minimum of 0 dB. The gain of an equivalent delay-and-sum beamformer has been removed to highlight the effect of the optimization.

Figure 25. Quadratically constrained optimum beampatterns for the array in Figure 22. Array gain optimization is constrained to limit the minimum white noise to 0 dB.

gain. To obtain even a 1dB increase in array gain at a distance of $10D$ requires $G_w = -40\text{dB}$. This implies a noise amplification of approximately 40dB, compared to a single sensor. For many applications, this does not represent a reasonable trade-off. For focal points closer to the array, however, larger increases in array gain can be achieved for a given level of white noise gain. At a distance of $D/2$, a 1dB increase in array gain requires only $G_w = -10\text{dB}$. This fact can be exploited to design a broadband beamformer that has a small variation in gain over a wide frequency range [18].

5.3 Linearly Constrained Optimization

For speech pickup in rooms, an interesting problem is whether a null can be placed to eliminate specific strong reflections. This is particularly intriguing when locating a null in the same direction but at a greater distance than the array focal point (for example, nulling the image sources to the right of the talker in Figure 3).

Forcing a response in a specific direction is accomplished using linearly constrained optimization, as discussed in Section 3.1.3.2 on page 35. The weight vector that maximizes array gain subject to the set of constraints $C^H W = g$ is

$$W_{\text{opt}} = R_{nn}^{-1}(\omega) C [C^H R_{nn}^{-1}(\omega) C]^{-1} g. \quad (28)$$

This array optimization is now applied to the specific case of an array optimum subject to dual linear constraints. One constraint is used to control the beamformer response in the target direction while the other is used to control the response in a known interference direction. That is, define the $M \times 2$ constraint matrix

$$C = [S \ N]$$

and the 2×1 gain vector

$$g = \begin{bmatrix} g_s \\ g_n \end{bmatrix}.$$

The effect of this constraint system is to force the beamformer to have a response of g_s to

Figure 26. Increase in array gain as a function of white noise gain for a 16-element uniformly spaced linear array with $\lambda/4$ sensor spacing. The trade-off of white noise gain for array gain varies with the distance to the array focal point. For a given white noise constraint, higher array gain is observed for focal points close to the array.

signal vector S and a response of g_n to (assumed) noise vector N . By substituting the constraints into (28) the following expression can be derived for the optimum weights (for details, please refer to the derivation in Appendix B)

$$W_{\text{opt}} = \frac{1}{1 - \cos(\gamma)^2} \cdot [g_s W_s - g_n W_s S^H W_n + g_n W_n - g_s W_n N^H W_s]$$

where

$$W_s = \frac{R_{nn}^{-1}(\omega)S}{S^H R_{nn}^{-1}(\omega)S}, \quad W_n = \frac{R_{nn}^{-1}(\omega)N}{N^H R_{nn}^{-1}(\omega)N}$$

and $\cos(\gamma)^2$ is the cosine squared of the generalized angle between the constraint vectors S and N

$$\cos(\gamma)^2 = \frac{|S^H R_{nn}^{-1}(\omega)N|^2}{S^H R_{nn}^{-1}(\omega)S \cdot N^H R_{nn}^{-1}(\omega)N}. \quad (29)$$

(For a discussion of the concept of generalized angle see Appendix B or [74].) The array gain for this linearly constrained beamformer is

$$G(\omega) = G_u(\omega) \cdot \frac{|g_s|^2 (1 - \cos(\gamma)^2)}{|g_s|^2 + |g_n|^2 \cdot \frac{S^H R_{nn}^{-1}(\omega)S}{N^H R_{nn}^{-1}(\omega)N} - \frac{2\text{Re}\{g_s g_n^* (N^H R_{nn}^{-1}(\omega)S)\}}{N^H R_{nn}^{-1}(\omega)N}} \quad (30)$$

where $G_u(\omega) = S^H R_{nn}^{-1}(\omega)S$ is the unconstrained optimum gain obtained for a target direction of S . As discussed in Appendix B, the constrained array gain is always less than or equal to the unconstrained array gain. This result applies equally to the white noise gain since no assumption was made about $R_{nn}(\omega)$. Consequently, imposing a linear constraint on an optimized beamformer will degrade both the array gain as well as the beamformer's robustness.

In spite of these drawbacks, sometimes it is necessary to reduce the beamformer response to the direction N beyond what is obtained by unconstrained optimization. In these cases,

the loss in array gain can be limited by appropriate selection of g_n . It is shown in Appendix B that, for a given pair of constraint vectors S and N , the optimal choice for g_n is

$$g_n|_{\text{opt}} = g_s \frac{N^H R_{nn}^{-1}(\omega) S}{S^H R_{nn}^{-1}(\omega) S} = g_s N^H W_s.$$

This value of g_n is the response of an unconstrained optimum array to a source corresponding to N . It is identical to the response which would have been obtained in the absence of a constraint on N . Thus, to maximize array gain, no constraint should be used.

In cases where a discrete interferer exists the above optimal g_n provides no added benefit; however, by selecting g_n as a fraction of the maximizing value, the gain loss can be controlled to any desired level. That is choose

$$g_n = \alpha g_s \frac{N^H R_{nn}^{-1}(\omega) S}{S^H R_{nn}^{-1}(\omega) S} \quad (31)$$

where α is a real constant between 0 and 1. With this choice, (30) reduces to

$$G(\omega) = G_u(\omega) \frac{(1 - \cos(\gamma)^2)}{[1 - (2\alpha - \alpha^2) \cos(\gamma)^2]}. \quad (32)$$

To illustrate the effect of the constraints on gain, the gain reduction ratio $G(\omega)/(G_u(\omega))$ was calculated as a function of $\cos(\gamma)^2$ for several values of α . The results are shown in Figure 27.

The data in this figure show that substantial gain loss occurs as $\cos(\gamma)^2 \rightarrow 1$. The loss in gain decreases as $\cos(\gamma)^2 \rightarrow 0$ and is zero for constraint vectors that are orthogonal $\cos(\gamma)^2 = 0$. Furthermore, for a given value of $\cos(\gamma)^2$ greater loss of array gain occurs when enforcing constraints for small values of α . For example, limiting gain loss to less than 3dB requires that $\cos(\gamma)^2 \leq 0.5$ when $\alpha = -30\text{dB}$. If the constraint is loosened to $\alpha = -3\text{dB}$; however, it is only necessary that $\cos(\gamma)^2 \leq 0.9$. The implications for array focal and constraint points is discussed next with reference to a specific example.

Figure 27. Gain loss due to linear constraint imposition as a function of $\cos(\gamma)^2$ for several values of α .

The primary interest in this thesis is to examine the application of the constrained optimization technique to near-field sources. More specifically, when the signal propagation vector S corresponds to a near-field point source and the noise propagation vector N corresponds to a plane-wave incident from the same angle as S . The feasibility of a far-field response constraint is illustrated by examining the variation of $\cos(\gamma)^2$ for various source positions within the near field of a typical array. The specific example considered is that of a 16-element uniformly spaced linear array focused at a distance equal to the array length. The variation in $\cos(\gamma)^2$ is plotted in Figure 28 as a function of angle of incidence with respect to the array axis. The solid curve was obtained using $R_{nn}(\omega) = I$ and the dashed curve was obtained using the $R_{nn}(\omega)$ that corresponds to a spherically isotropic noise field.

As seen in the figure, the effect of spherically isotropic $R_{nn}(\omega)$ is to increase the generalized angle between the near-field and far-field propagation vectors. Referring to Figure 27, this will lessen the impact of the constraint on the optimization of an array in isotropic noise compared to white noise.

5.3.1 Numerical Example

Linearly constrained optimization is demonstrated by a continuation of the previous example for a linear array with $d = \lambda/4$. The focal point is fixed at a distance of D in the broadside direction. The linear constraints include a constraint on the response magnitude at the focal point and also a null constraint ($\alpha = -\infty$) for plane-wave arrival from the broadside direction. To restrict the discussion only to practical designs, the noise correlation matrix is constructed as in (27) using $\kappa = 0.0018$. This value of κ produced a white noise gain of 0dB in the quadratically constrained example depicted in Figure 25. Since this $R_{nn}(\omega)$ contains both a spherically isotropic portion as well as a spatially incoherent portion, $\cos(\gamma)^2$ for S and N will be somewhere in between the extremes depicted in Figure 28. Nonetheless, these extremes can be used to bracket the predicted gain loss due to the linear constraints (as shown in Figure 27).

The beampattern of the null-constrained optimum design for three different distances is

Figure 28. Variation in $\cos(\gamma)^2$ as a function of angle of incidence with respect to array axis θ . Solid line assumes uncorrelated noise (i.e. $R_{nn} = I$) whereas dashed line assumes isotropic noise.

shown in Figure 29. This can be compared with the beampatterns obtained for the array optimized without linear constraints shown in Figure 25. As seen in Figure 29, the linear constraints provide the desired response magnitude at the focal point and the response pattern at the focal distance remains largely unaffected. In contrast, the far-field beampattern displays a sharp null in the response in the broadside direction. The cost of this null constraint is to reduce the array gain by 1.1 dB from 9.4 dB with no linear constraints down to 8.3 dB. Referring to Figure 28, for an angle of incidence of $\pi/2$, $\cos(\gamma)^2$ lies between 0.1 and 0.4 and the predicted gain loss is between 0.5 dB and 2.2 dB.

This design was repeated using the soft constraint depicted in (31) with $\alpha = -6$ dB. The beampatterns are shown in Figure 30. In this case, the gain loss is only 0.4 dB.

As suggested by the beampatterns for $(D/2) + d$, however, a loss in robustness also results from the additional linear constraints. While the reduction in white noise gain is not explicitly predicted by (32), it is reduced in comparison with the unconstrained array. For the preceding examples, the white noise gain of the null-constrained case is reduced by 4.7 dB where as the soft-constrained case suffers only a 2.6 dB loss.

$$r = \frac{D}{2} + d$$

$$r = D$$

$$r = \infty$$

Figure 29. Null-constrained optimum beampatterns for the array in Figure 22. Note the change in scale for the top graph.

Figure 30. Soft-constrained optimum beampatterns for the array in Figure 22. Array gain is optimized while suppressing the response in the broadside direction by an additional 6 dB compared to that shown in Figure 25. Note the change in scale for the top graph.

Chapter 6

BROADBAND BEAMFORMER DESIGN

The optimum microphone weights described in the previous chapter are valid only within a narrow bandwidth around a single frequency. Since speech signals are broadband, a speech bandwidth array must be optimized separately at many frequencies within the range of interest. A practical implementation can then be realized in the frequency domain (using an FFT) or by designing a FIR filter that provides the correct amplitude and phase at the selected frequencies. This chapter discusses the issues associated with the latter implementation.

6.1 Sensor FIR Filter Design using Complex Approximation

This approach implements broadband beamformers using fixed-coefficient FIR filters. The required complex sensor weights at each frequency provide the frequency-domain design target for the FIR filter calculation. In general, these sensor weights specify an FIR filter that is neither linear phase nor minimum phase. Consequently, a FIR design approach must be chosen that allows the specification of an arbitrary amplitude and phase.

The goal of FIR filter design is to choose a set of filter coefficients $\hat{w}(n)$ such that their frequency response $\hat{W}(\omega)$ approximates a desired response $W_d(\omega)$ [75]. The frequency response $\hat{W}(\omega)$ and the filter coefficients $\hat{w}(n)$ are related by the discrete-time Fourier transform (DTFT)

$$\hat{W}(\omega) = \sum_{n=0}^{N-1} \hat{w}(n)e^{-j\omega n}. \quad (33)$$

By sampling this equation at L arbitrary frequencies the following matrix equation can be written

$$F\hat{w} = \hat{W} \quad (34)$$

where F is an $N \times L$ matrix of Fourier coefficients, $\hat{w}$ is the N -element vector of filter

coefficients and $\hat{\mathbf{W}}$ is the L -element vector of the frequency samples of $\hat{W}(\omega)$. This equation provides the basis for two design approaches considered in this section: the frequency sampling approach and the least-squared error approach.

6.1.1 Frequency Sampling Design

In the frequency sampling approach, the number of frequency response samples L is set equal to the number of FIR filter taps N . In this case, the system of equations (34) can be rewritten using the desired response samples $\mathbf{W}_d$ instead of $\hat{\mathbf{W}}$ and solved directly for the required filter coefficients

$$\hat{\mathbf{w}} = \mathbf{F}^{-1} \mathbf{W}_d \quad (35)$$

provided $\mathbf{F}$ is invertible. This design approach produces a set of FIR filter coefficients that match the desired frequency response exactly at each frequency ω_l . In between these frequencies, however, the response is undefined. Care must be taken to ensure a usable filter response by properly selecting the number of filter taps and the spacing of the sample frequencies ω_l .

If the $L = N$ frequency samples are chosen to be equispaced, then the solution described in (35) can be obtained easily using the inverse FFT. To obtain this solution choose frequency samples ω_l such that

$$\omega_l = \frac{2\pi}{N}l, \text{ where } l = 0, 1 \dots N-1. \quad (36)$$

The filter coefficients $\hat{\mathbf{w}}$ are then obtained as the inverse FFT of the samples of the desired frequency response $\mathbf{W}_d$.

In the application to beamforming, however, this simple approach has several drawbacks. First, it is necessary to specify the desired filter response at all frequencies in the range of 0 to π . This means that even at frequencies where no beamforming is required (for frequencies below the speech band, for instance) it is still necessary to specify a filter response. This specification can use up filter-design degrees of freedom if not chosen correctly. The result of this is that more filter taps are needed for a given beamformer design.

Secondly, since the response at frequencies in between the sample points ω_l is unspecified (and unconstrained), there can be significant phase and amplitude errors in the resulting filters. These errors will lead to array missteering at the intermediate frequencies. The remedy, of course, is to make the frequency samples closer by increasing the number of filter taps, with a resulting increase in implementation cost. As a result of these drawbacks, the direct frequency sampling design is not considered further.

By choosing a nonuniform spacing of the N frequency samples, greater flexibility in the filter design process is achieved. A significant advantage is obtained through the use of transition bands. Specifying a don't care region for the filter design allows the approximation error to be larger in the out-of-band regions and typically results in lower error for the in-band regions. This alleviates the first drawback associated with the uniform frequency sampling approach described above. A transition band is specified simply by not specifying the response over that frequency band. For instance, instead of specifying uniformly spaced frequency samples between 0 to π , it is possible to use limited frequency range say, ω_a to ω_b . That is, choose

$$\omega_l = \frac{(\omega_b - \omega_a)}{N}l, \text{ where } l = 0, 1 \dots N-1. \quad (37)$$

Unfortunately, the response at frequencies in between the sample points remains undefined. Once transition bands are specified, improvements in filter response can only be obtained by increasing the number of filter taps. The drawbacks here are twofold: First, if there are sharp discontinuities in the desired response, then increasing the number of taps will have no effect on decreasing the amplitude of the resulting Gibbs oscillations. The response oscillations due to the Gibbs phenomenon produce sensor filter errors that can also lead to array missteering. Second, as the number of frequency sample points increases, adjacent frequencies become closer and the matrix $\mathbf{F}$ becomes ill-conditioned. An ill conditioned $\mathbf{F}$ will lead to large errors in filter design upon its inversion.

For the various drawbacks cited above, the simple frequency-sampling filter design approach is not considered for sensor filter design.

6.1.2 Least squares complex approximation

In order to overcome the filter-design limitations discussed in the previous section, the approach used must allow the specification of non-uniform frequency samples and also a large number of closely spaced points at which the response is desired. The least-squares design approach permits this flexibility.

The least-squares approach uses a number of frequency sample points that is larger than the number of filter taps, that is $L > N$. In this case, the matrix F is not square and, therefore, not invertible so equation (34) is overdetermined and cannot be exactly solved for $\hat{\mathbf{w}}$. In such cases, a least-squared error solution can be found that minimizes the approximation error for this equation.

The error measure is defined as the weighted sum of the squared differences between the desired $W_d(\omega)$ and actual $\hat{W}(\omega)$ frequency responses

$$E = \sum_{l=0}^{L-1} V(\omega_l) |W_d(\omega_l) - \hat{W}(\omega_l)|^2. \quad (38)$$

In this equation, $V(\omega_l)$ is an arbitrary positive weight function and ω_l are the discrete frequency points. In matrix terms, the error can be written as the quadratic form

$$E = \mathbf{e}^H \mathbf{V} \mathbf{e} \quad (39)$$

where $\mathbf{V}$ is an $L \times L$ diagonal matrix of error weights $V(\omega_l)$ and $\mathbf{e}$ is the L -element vector of error terms given by

$$\mathbf{e} = \mathbf{F} \hat{\mathbf{w}} - \mathbf{W}_d \quad (40)$$

where $\mathbf{W}_d$ is the L -sample vector of desired frequency response samples. The filter coefficient vector that minimizes the squared error is given by the well-known least-squares solution [75]

$$\hat{\mathbf{w}} = (\mathbf{F}^H \mathbf{V} \mathbf{F})^{-1} \mathbf{F}^H \mathbf{V} \mathbf{W}_d. \quad (41)$$

In the broadband beamformer design problem, a separate filter response function must be implemented for each sensor. The desired response points W_d are given by the optimum sensor coefficients evaluated at the frequencies ω_l .

This approach allows the use of non-uniformly spaced frequency samples, including transition bands, and also permits the weighting of error in different frequency bands. Furthermore, since F does not need to be inverted, it permits the use of a large number of closely spaced frequency sample points within the band of interest. Although the desired response is not achieved exactly at the sample points, the approximation error is guaranteed to be minimum for a given number of taps. Consequently, this approach is used for beamformer design.

6.1.3 Complex Chebyshev Error Approximation

As a further elaboration of the FIR filter design problem, it is possible to formulate a design problem so as to minimize the maximum weighted error over a frequency band rather than simply the mean squared error. The problem is to find filter coefficients $\hat{w}(n)$ that minimize

$$\max_{\omega \in \omega_l} V(\omega) \left| W_d(\omega) - \sum_{n=0}^{N-1} \hat{w}(n) e^{-j\omega n} \right|. \quad (42)$$

When applied to the design of conventional filters, the Chebyshev error approach can lead to reductions in the number of taps required to implement a given filter response. The design process is much more involved, however, particularly when implementing a filter with arbitrary phase response. Furthermore, a more precise realization of the individual sensor filters does not necessarily produce an improved beamformer response as discussed later. For those readers that are interested in Chebyshev approximation more information can be found in [76][77][78][75].

6.2 Direct Synthesis using Time-domain Problem Formulation

The problem of broadband beamformer design can be addressed directly at the array

design stage by reformulating the array-gain equation in the time domain. The problem formulation used here is the same as that in [61]. The broadband beamformer structure used is an FIR filter at the output of each sensor, as depicted in Figure 31.

At a given time instant n the output of the M sensors can be expressed as the vector $\mathbf{v}(n)$ where

$$\mathbf{v}(n) = [u_0(n) \ u_1(n) \ \dots \ u_{M-1}(n)]^T.$$

This vector is referred to as the array snapshot vector. The MN dimensional stacked snapshot vector $\mathbf{u}(n)$ is formed from N delayed array outputs contained in the delay line

$$\mathbf{u}(n) = [\mathbf{v}^T(n) \ \mathbf{v}^T(n-1) \ \dots \ \mathbf{v}^T(n-N+1)]^T.$$

The beamformer output $y(n)$ can then be expressed as the inner product of the stacked snapshot vector and the (real-valued) stacked weight vector $\mathbf{w}$

$$y(n) = \mathbf{w}^T \mathbf{u}(n). \quad (43)$$

where

$$\mathbf{w} = [w_0(0) \ w_1(0) \ w_2(0) \ \dots \ w_0(1) \ \dots \ w_0(N-1) \ w_1(N-1) \ \dots \ w_{M-1}(N-1)]^T$$

and $w_m(n)$ is the impulse response of the m^{th} sensor filter. The output $u_m(n)$ from sensor m consists of both a signal $s_m(n)$ and noise $n_m(n)$ component

$$u_m(n) = s_m(n) + n_m(n).$$

The array response to these components is, similarly,

$$y_s(n) = \mathbf{w}^T \mathbf{s}(n) \quad (44)$$

and

$$y_n(n) = \mathbf{w}^T \mathbf{n}(n) \quad (45)$$

Figure 31. Broadband FIR beamformer structure.

where $s(n)$ is the stacked signal snapshot vector and $n(n)$ is the stacked noise snapshot vector.

The expression for array gain can thus be written as

$$G = \frac{E\{|w^T s(n)|^2\} / (E\{|w^T n(n)|^2\})}{\sigma_s^2 / \sigma_n^2} = \frac{E\{w^T s(n) \cdot s^T(n) w\} / \sigma_s^2}{E\{w^T n(n) \cdot n^T(n) w\} / \sigma_n^2} . \quad (46)$$

By defining the signal spatiotemporal correlation matrix R_{ss}

$$R_{ss} = E\{s(n) \cdot s^T(n)\} \quad (47)$$

and the noise spatiotemporal correlation matrix R_{nn}

$$R_{nn} = E\{n(n) \cdot n^H(n)\} \quad (48)$$

the above expression for array gain becomes

$$G = \frac{w^T R_{ss} w}{w^T R_{nn} w} \cdot \frac{\sigma_n^2}{\sigma_s^2} \quad (49)$$

which, within a constant scale factor, is in the convenient ratio of two quadratic forms.

As discussed in Chapter 5, for a general matrix R_{ss} this ratio is maximized by setting the weight vector equal to the eigenvector corresponding to the maximum eigenvalue of the symmetric matrix $R_{nn}^{-1} R_{ss}$. The solution, therefore, implies an eigenanalysis of this matrix. The apparent reward for this added complexity, of course, is the fact that the weight solution is provided directly in the time domain. The optimum stacked weight vector would then directly provide the FIR tap weights required to implement the broadband beamformer. In practice, however, the situation is more involved.

As discussed in the description of the frequency-domain problem, the optimum weights for a given frequency are unique only within a constant scale factor. That is, the optimum solution can be scaled by multiplying each sensor weight by the same complex constant. When separately optimizing array gain at each frequency, this fact was exploited to force

the sensor weights to provide a flat frequency response across all frequencies. In the time-domain formulation, this normalization corresponds to a frequency-dependent equalizer to flatten the overall beamformer response. This equalizer must be applied equally to all sensor filters to prevent distortion of the beamformer response.

The simplest method to calculate this equalizer is to compute the beamformer response to the target signal. The individual filter responses can then be equalized using this target response either by deconvolving the time responses or by dividing the frequency responses. The main difficulty with this approach is due to the numerical problems encountered if the overall beamformer response approaches zero at a given frequency. At such a frequency, it is not possible to design a proper equalizer to correct the sensor filters responses. This problem can be overcome by modifying the problem formulation.

The linearly constrained minimum variance (LCMV) beamforming problem formulation will be used for direct broadband beamformer design [61]. The beamformer structure is identical to that depicted in Figure 31.

The idea behind LCMV beamforming is to minimize the beamformer output power while constraining the array response to target signals from a given spatial region. Using matrix notation, the beamformer output $y(n)$ is

$$y(n) = \mathbf{w}^T \mathbf{u}(n) \quad (50)$$

and the output variance to be minimized is given by

$$\sigma_y^2 = E\{|y(n)|^2\} = E\{\mathbf{w}^T \mathbf{u}(n) \cdot \mathbf{u}^T(n) \mathbf{w}\} = \mathbf{w}^T \mathbf{R}_{uu} \mathbf{w}. \quad (51)$$

In the absence of any constraints, the output power is minimized by the trivial solution $\mathbf{w} = \mathbf{0}$. The beamformer response is constrained by requiring that

$$\mathbf{c}_l^T \mathbf{w} = g_l \quad (52)$$

where $\mathbf{c}_l$ is a stacked constraint vector and g_l is the desired response level. A single constraint (point constraint) is required to control the beamformer response at a specific loca-

tion for a specific frequency. In practice, multiple constraints are used to exert more control over the response at multiple locations and/or frequencies. Multiple linear constraints are more concisely expressed as

$$\mathbf{C}^T \mathbf{w} = \mathbf{g} \quad (53)$$

where $\mathbf{C}$ is now an $MN \times L$ matrix whose columns are the individual constraint vectors $\mathbf{c}_l$ and $\mathbf{g}$ is the $L \times 1$ response vector. Thus, the problem is to minimize

$$\mathbf{w}^T \mathbf{R}_{uu} \mathbf{w} \quad (54)$$

subject to the set of linear constraints

$$\mathbf{C}^T \mathbf{w} = \mathbf{g}. \quad (55)$$

The solution to this constrained optimization problem is found using the method of Lagrange multipliers. The vector of beamformer coefficients which satisfies the above conditions is given by [61]

$$\mathbf{w}_{\text{opt}} \approx \mathbf{R}_{uu}^{-1} \mathbf{C} (\mathbf{C} \mathbf{R}_{uu}^{-1} \mathbf{C})^{-1} \mathbf{g}. \quad (56)$$

Although the problem formulation is different, in the narrowband case, the solution to the LCMV beamformer problem is identical within a scale factor of the maximum signal-to-noise ratio solution [55]. This fact is exploited here to design a broadband beamformer to achieve the array performance described in Chapter 5. Using this approach, it is necessary to specify suitable expressions for the data correlation and constraint matrices.

The array data correlation matrix is defined using the stacked sensor output vector $\mathbf{u}(n)$

$$\mathbf{R}_{uu} = E\{\mathbf{u}(n) \cdot \mathbf{u}^T(n)\}. \quad (57)$$

The sensor output signal consists of both signal $s(n)$ and noise $\mathbf{n}(n)$ components so that $\mathbf{R}_{uu}$ can be expressed as

$$\mathbf{R}_{uu} = \mathbf{R}_{ss} + \mathbf{R}_{nn} + \sigma_w^2 \mathbf{I}. \quad (58)$$

In this equation, it is assumed that the sensor outputs are contaminated by a small amount of spatially incoherent noise represented by the $MN \times MN$ identity matrix I and white noise power σ_w^2 . This small amount of noise ensures that R_{uu} is invertible and also serves to control the beamformer's robustness.

Note that R_{ss} and R_{nn} are broadband spatiotemporal correlation matrices, distinct from the narrowband spatial correlation matrices introduced in (3) and (4). The mathematical formulation of these matrices includes the effects of temporal variations in the signal as well as spatial variations due to acoustic propagation. Expressions describing the elements of R_{ss} and R_{nn} can be derived from the defining equations, (47) and (48). The details of this derivation are provided in Appendix A for the case of a broadband point source and also broadband isotropic noise.

The crosscorrelation $\rho_{mp}(\tau)$ between the m^{th} and p^{th} sensor signals as a function of time lag τ due to a broadband point source is given by

$$\rho_{mp}(\tau) = \sigma_s^2 \cdot \frac{r^2}{r_m r_p} \cdot \frac{(\sin \omega_b(\tau + (r_p - r_m)/c) - \sin \omega_a(\tau + (r_p - r_m)/c))}{(\omega_b - \omega_a)(\tau + (r_p - r_m)/c)}, \quad (59)$$

where r_m and r_p are the distances from the source to each sensor. The source signal is assumed to have a uniform power spectral density over a frequency range from ω_a to ω_b and a total power σ_s^2 when observed at a distance r from the point source. The elements of R_{ss} for this case are

$$R_{ss} = \begin{bmatrix} \rho_{00}(0) & \rho_{01}(0) & \dots & \rho_{00}(1) & \dots & \rho_{0,M-1}(N-1) \\ \rho_{10}(0) & \rho_{11}(0) & \dots & \rho_{10}(1) & \dots & \rho_{1,M-1}(N-1) \\ \rho_{20}(0) & \rho_{21}(0) & \dots & \rho_{20}(1) & \dots & \rho_{2,M-1}(N-1) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{00}(1) & \rho_{01}(1) & \dots & \rho_{00}(0) & \dots & \rho_{0,M-1}(N-2) \\ \rho_{10}(1) & \rho_{11}(1) & \dots & \rho_{10}(0) & \dots & \rho_{1,M-1}(N-2) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{M-1,0}(N-1) & \rho_{M-1,1}(N-1) & \dots & \rho_{M-1,0}(N-2) & \dots & \rho_{M-1,M-1}(0) \end{bmatrix}. \quad (60)$$

In the case of broadband isotropic noise, the crosscorrelation $\rho_{mp}(\tau)$ between the m^{th} and

p^{th} sensor signals is given by

$$\rho_{mp}(\tau) = \sigma_n^2 \cdot \frac{c}{2d_{mp}(\omega_b - \omega_a)} \cdot [\text{si}[\omega_b u_1] - \text{si}[\omega_a u_1] - \text{si}[\omega_b u_2] + \text{si}[\omega_a u_2]] \quad (61)$$

where $u_1 = d_{mp}/c + \tau$, $u_2 = -d_{mp}/c + \tau$, d_{mp} is the distance between the sensors, σ_n^2 is the noise power observed at a single sensor and $\text{si}(x)$ and $\text{ci}(x)$ are the sine and cosine integrals, respectively (see Appendix A). The elements of $\mathbf{R}_{nn}$ are distributed in the same manner as illustrated in (60) for $\mathbf{R}_{ss}$.

As discussed earlier, beamformer response is controlled at a specific spatial location and frequency using a single point constraint. The constraint matrix $\mathbf{C}$ is composed of a number of such point constraints distributed over a spatial region and frequency band of interest. The columns of $\mathbf{C}$ for a real-data beamformer are derived from hypothetical stacked array signal vectors that correspond to a source at a desired spatial location. For instance, ensuring a beamformer response amplitude of A and phase ϕ for a planewave impinging from an angle θ with a frequency ω , requires the following constraints [61]

$$\begin{bmatrix} \mathbf{c}_\theta^T(\omega) \\ \mathbf{s}_\theta^T(\omega) \end{bmatrix} \mathbf{w} = \begin{bmatrix} A \cos \phi \\ A \sin \phi \end{bmatrix} \quad (62)$$

where $\mathbf{c}_\theta(\omega)$ is

$$\mathbf{c}_\theta(\omega) = [\cos \omega \tau_{1,\theta} \cos \omega \tau_{2,\theta} \dots \cos \omega (\tau_{1,\theta} - 1) \dots \cos \omega (\tau_{M-1,\theta} - N + 1)]^T$$

and

$$\mathbf{s}_\theta(\omega) = [\sin \omega \tau_{1,\theta} \sin \omega \tau_{2,\theta} \dots \sin \omega (\tau_{1,\theta} - 1) \dots \sin \omega (\tau_{M-1,\theta} - N + 1)]^T.$$

In these equations, $\tau_{m,\theta}$ is the propagation delay in time samples (with respect to the array phase centre) associated with wave arrival from angle θ at sensor m . The phase ϕ typically corresponds to the expected time delay through the tapped delay line at frequency ω .

For a spherical-wave source, the above constraints must be modified to account for the

varying amplitude across the array. Referring to Figure 12, the amplitude observed at sensor m relative to that at the origin is $|a_s|/|a_s - a_m|$. Thus, the above constraints become

$$c_{a_s}(\omega) = |a_s| \left[\frac{\cos \omega \tau_{0,a_s}}{|a_s - a_0|} \frac{\cos \omega \tau_{1,a_s}}{|a_s - a_1|} \cdots \frac{\cos \omega (\tau_{0,a_s} - 1)}{|a_s - a_0|} \cdots \frac{\cos \omega (\tau_{M-1,a_s} - N + 1)}{|a_s - a_{M-1}|} \right]^T$$

and

$$s_{a_s}(\omega) = |a_s| \left[\frac{\sin \omega \tau_{0,a_s}}{|a_s - a_0|} \frac{\sin \omega \tau_{1,a_s}}{|a_s - a_1|} \cdots \frac{\sin \omega (\tau_{0,a_s} - 1)}{|a_s - a_0|} \cdots \frac{\sin \omega (\tau_{M-1,a_s} - N + 1)}{|a_s - a_{M-1}|} \right]^T.$$

6.3 Design Example

The preceding methods for broadband beamformer design can be compared using a specific example. For the present purposes, the application to computer-telephony integration is suitable. The physical setup is depicted in Figure 44.

An array of seven omnidirectional microphones is placed along the top edge of a computer monitor. The array is to be used to pickup speech from the operator so it is focused at a point directly in front of the monitor at a distance of 0.55m. The spacing between microphones is 0.07m and the sampling frequency to be used is 12kHz. The beamformer's response is to be optimized over a frequency range of 300Hz to 3.6kHz.

To permit practical implementation, it is necessary to limit the beamformer's loss of robustness due to response optimization. While no precise limits can be defined, for the present purposes a reasonable white noise gain target is approximately 0dB. For both design approaches, this target white noise gain is obtained through trial-and-error. In the case of the time-domain formulation, however, less control can be exercised over this parameter and the target level is not obtained at all frequencies.

In a broadband beamformer design, two aspects of performance are important: the array gain achieved and the beamformer frequency response. For a broadband beamformer design to be useful it must achieve the optimum array gain and it must pass a desired sig-

nal from a target location with little distortion. Unfortunately, as will be seen, a given level of error in achieving array gain does not directly relate to the level of error in the beamformer response. Consequently, comparing different broadband beamformer designs requires two separate error measurements: deviation from the ideal array gain and deviation from an ideal beamformer response.

The ideal array gain $G(\omega)_{\text{opt}}$ at a given frequency is the theoretical optimum array gain for the stipulated level of white noise gain. For a given beamformer implementation, the actual array gain $G(\omega)_{\text{act}}$ may deviate from the ideal value due to practical implementation of the sensor filters. The array gain error $\epsilon_G(\omega)$ is defined as the difference between the ideal and actual array gain, when these two quantities are expressed in decibels. That is

$$\epsilon_G(\omega) = 10\log|G(\omega)_{\text{act}}| - 10\log|G(\omega)_{\text{opt}}|. \quad (63)$$

The ideal beamformer response is obtained for a source located exactly at the array focal point. For a distortion-free representation of the source signal, the ideal beamformer response should be flat over the frequency range of interest. For a broadband beamformer with sensor weights $w_m(n)$, the actual array impulse response $w_A(n)$ is given by

$$w_A(n) = \sum_{m=0}^{M-1} w_m(n) \otimes d_m(n) \quad (64)$$

where $d_m(n)$ is the response produced in the m^{th} sensor by a unit pulse emitted from a source located at the array focal point and $\otimes$ denotes the convolution operation. The actual beamformer frequency response $W_A(\omega)$ is the DTFT of $w_A(n)$

$$W_A(\omega) = \sum_{n=0}^{N-1} w_A(n) e^{-j\omega n}.$$

The beamformer response error $\epsilon_W(\omega)$ is the difference between the ideal response W_I and the actual response, when both are expressed in decibels

$$\epsilon_W(\omega) = 20\log|W_A(\omega)| - 20\log|W_I|. \quad (65)$$

Typically, the ideal frequency response at the target location is perfectly flat. That is, W_I is a constant for all frequencies.

The performance of a specific design is quantified by calculating the statistics of $\epsilon_G(\omega)$ and $\epsilon_W(\omega)$ over the frequency range of interest. Specifically, the mean, standard deviation and maximum absolute values are used. These measures are used in the following sections to compare different design methods and also to compare similar designs with varying numbers of taps.

6.3.1 Complex Approximation

To design the individual sensor FIR filters, the complex microphone weights for each beamformer were calculated at a total of 200 points over the frequency range of 300Hz to 3.6kHz. The response was unspecified for frequencies less than 300Hz to allow a transition band for filter design. At 0Hz and 6kHz, the target response is set equal to a real constant proportional to the inverse of the distance from the sound source. This provides for a well-behaved filter response at extreme frequencies. The FIR filter coefficients are constrained to be real by forcing conjugate symmetry in the frequency domain. Furthermore, the impulse responses were centered within the N FIR taps by adding a linear phase delay equal to $(N - 1)/2$ taps to the complex weights. Centering the impulse response in this way reduces the error associated with approximating fractional sample delays [79].

Beamformer designs were implemented for a varying number of filter taps per channel from 8 to 48. For each design, the error statistics described above were computed. The gain errors obtained are plotted in Figure 32 and the response errors in Figure 33, both as a function of the number of FIR taps.

As seen in Figure 32, for beamformers with 24 taps or more per channel, the ideal array gain is achieved with an accuracy of ± 0.5 dB. Array gain error decreases further as the number of taps is increased until a level of 40 taps per channel is attained. Increasing the number of FIR taps beyond 40 does not change the error level significantly.

Reducing the beamformer response error to a level of ± 0.5 dB also requires approximately

Figure 32. Gain error statistics as a function of the number of FIR taps for the least-squared error design method: standard deviation (solid), peak error (dash-dot) and mean (dotted).

Figure 33. Response error statistics as a function of the number of FIR taps for the least-squared error design method: standard deviation (solid), peak error (dash-dot) and mean (dotted).

24 taps. Further reductions are possible using a larger number of taps.

6.3.2 Time-domain Formulation

Applying the time-domain method involves a determination of the number of constraints as well as the frequency spacing. As mentioned earlier, the LCMV problem formulation is equivalent to the maximum SNR problem when a simple gain constraint is used. Consequently, for the current example, the linear constraints are limited to a single spatial location and multiple frequencies. For the following examples, all frequency constraints refer to the beamformer response at the desired focal point.

For a single spatial location the maximum number of constraints of the form (62) is limited to $N/2$ (recall that constraining amplitude and phase at a single frequency requires both a sine and cosine term). For the present example, the frequency points are equispaced over the design frequency range of 300Hz to 3.6kHz. This choice allows for a transition band below 300Hz and above 3.6kHz. In the constraint calculation, the impulse responses were centered within the N FIR taps by setting the phase portion of the gain constraint to a linear phase delay equal to $(N - 1)/2$ taps.

The array data correlation matrix is constructed as described in (58). The signal and noise correlation matrix entries are given by (59) and (61) and both σ_s^2 and σ_n^2 are set to unity. Note that due to the constraint at the array focal point, the chosen signal level σ_s^2 will have no effect on the optimum solution. It could equally well have been set to zero. The level of σ_w^2 is adjusted to limit the minimum white noise gain to 0dB. Since the appropriate value varies with the number of FIR taps chosen, this adjustment is performed (through trial and error) for each design case. Despite this adjustment, the resulting white noise gain target is not achieved at all frequencies. This is a consequence of the design process: only indirect control over white noise gain is possible.

Beamformer designs were implemented for a number of filter taps per channel from 8 to 36. For each design, the array gain and response error statistics were computed. The gain errors obtained are plotted in Figure 34 and the response errors in Figure 35, both as a function of the number of FIR taps.

Figure 34. Gain error statistics as a function of the number of FIR taps for the time-domain design method: standard deviation (solid), peak error (dash-dot) and mean (dotted).

Figure 35. Response error statistics as a function of the number of FIR taps for the time-domain design method: standard deviation (solid), peak error (dash-dot) and mean (dotted).

For the time-domain method, a slightly larger number of taps is required to achieve the same level of gain error compared to the least-squared method. In both cases, however, using a number of taps much larger than 30 provides diminishing returns.

The behavior of the beamformer response error, however, is very different. As seen in Figure 35, the response error level is approximately ± 0.1 dB for a beamformer with only 10 taps per channel. This is due to the fact that the beamformer response is exactly constrained in the design process. For the least-squared method, the beamformer response is not directly constrained.

6.3.3 Sensor Filters

As a final point of comparison, it is illustrative to examine the sensor filters responses resulting from each of the design methods. The specific case of a beamformer with 28 tap FIR filters on each sensor is considered. The ideal response is plotted together with the actual frequency responses of the sensor filters for each design method in Figure 36. It can be seen that the accuracy of implementing the sensor FIR filter does not necessarily relate to better overall array behavior.

Figure 36. Response of 28-tap sensor filter 3 for near-field beamformer: ideal response (solid), least-squares design (dash-dot) and time-domain design (dashed).

Chapter 7

APPLICATIONS OF NEAR-FIELD BEAMFORMING

This section discusses two practical microphone array applications where near-field beamforming is of benefit: mobile telephony and computer telephony. The application to mobile telephony demonstrates beamforming improvements in an environment where the dominant noise is not coherent with the desired signal. The example provided for computer telephony demonstrates performance in reverberation where the interference is correlated with the desired signal.

7.1 Mobile Telephony

In mobile telephony, due to the small size of the passenger compartment in a typical automobile, the talker's mouth is located within the array's near field. This fact has been recognized and several methods for handling near-field beamformer design have been proposed for mobile telephony [6][71].

This section examines the theoretical and practical performance of fixed, optimum near-field beamformers for application to mobile telephony [19]. The near-field optimization approach described in the preceding chapters is used to design two broadband beamformers for speech pickup in a moving car. The first beamformer has frequency independent weights and provides optimum gain in spatially incoherent noise. As mentioned earlier, this is equivalent to a delay-and-sum beamformer with simple amplitude weights and so it will be referred to as a weighted, delay-and-sum (WDS) beamformer (even though it is implemented as an FIR filter). The second beamformer provides optimum gain in spherically isotropic noise. In this case, the microphone weights are separately optimized for each frequency and must be implemented using a digital filter. Consequently, this design will be referred to as the filter-and-sum (FS) beamformer.

Both beamformers are evaluated using the specific example of an 8-element array with a microphone spacing of 7cm. This spacing is equivalent to approximately 0.75λ at the highest frequency of interest, 3500Hz. Choosing the microphone spacing higher than

$\lambda/2$ yields a slightly higher gain at the upper end of the design frequency range. Typically, microphone spacing above $\lambda/2$ is avoided since it will result in spatial aliasing when the main beam is steered away from broadside. In the present example, however, the main beam direction is fixed so the added gain can be exploited. The physical layout showing relative positions of the microphones and source is shown in Figure 37. The acoustic source of interest is located 33.8cm from the array axis and 37.5cm along the array axis, measured from microphone 8.

7.1.1 Broadband Beamformer Design

Broadband beamformers were designed using the least-squares method of complex approximation, outlined in Section 6.1 on page 82. The complex microphone weights for each beamformer were calculated at a total of 400 points over the frequency range of 300Hz to 5kHz (half the sampling frequency). The microphone weights were unspecified for frequencies between 0Hz and 300Hz to allow a transition band for filter design. At 0Hz, the microphone weights for both beamformers are real constants proportional to the inverse of the distance from the sound source.

The robustness of the FS beamformer was maintained by constraining the minimum allowable white noise gain. For this experiment, the minimum white noise gain was fixed at 0dB for all frequencies in the range of interest, 300Hz to 4000Hz. The constraint was enforced through an iterative, trial-and-error design process. This process was repeated until the white noise gain target was met within a small fraction of a decibel.

Of course, the number of FIR taps chosen has a significant impact on the accuracy of the sensor weight implementation. For the present purposes, an appropriate number of taps was chosen by trial and error. Deciding the number of taps that are appropriate entails an examination of how closely the target array gain is achieved and also how faithfully the desired signal is received. Using FIR filter length of 60 taps allows the target array gain to be achieved within a fraction of a decibel and also provides a source frequency response that is flat to within $\pm 0.5\text{dB}$.

In the design process, real FIR coefficients were ensured by forcing conjugate symmetry

Figure 37. Physical layout showing relative positions of the microphone array and mannequin. Microphone number 3 is located closest to the mannequin's mouth and is taken as the reference.

in the frequency domain. Furthermore, the impulse responses were centered within the L FIR taps by adding a linear phase delay equal to $(L - 1)/2$ taps to the complex weights. Centering the impulse response in this way reduces the error associated with approximating fractional sample delays [79].

As an example, the frequency response obtained for microphone filter number 3 for the FS beamformer is shown in Figure 38 together with the design target. As seen in this figure, using a 60-tap filter allows the amplitude target to be matched very closely over the design frequency range of 300Hz to 4000Hz.

7.1.2 Test Conditions

In order to evaluate fixed beamformer strategies under realistic conditions, a series of recordings were made inside a typical, full-size passenger sedan. A uniformly spaced, linear microphone array was mounted inside the passenger compartment and used to record speech and noise while the vehicle was driven under a variety of conditions. The layout inside the passenger compartment, illustrating the relative positions of the array and mannequin, is shown in Figure 37. The array was mounted horizontally near the ceiling inside the passenger compartment to avoid visual obstruction for the driver. The microphones are mounted, with a spacing of 7cm, on an aluminum boom to ensure positional accuracy. Rubber vibration isolators are used on each capsule to reduce the pickup of mechanical vibration. Absolute sensitivity of the microphones was measured and the data used to ensure a matching of better than 0.25dB. The microphones in the array are omnidirectional electret condenser microphones possessing a flat frequency response from 20Hz to 16kHz.

Anechoic speech was recorded on digital audio tape (DAT) and played back through the artificial mouth of a head and torso simulator (mannequin). This device was designed to simulate the acoustical properties of the human voice during speech. Its use allows the precise control over experimental conditions and permits the objective measurement of the effects of speech pickup enhancements. The mannequin was seated in the right front seat for all test conditions. The perpendicular distance from the mannequin's mouth to the

Figure 38. Target (circles) and actual (solid line) frequency response for microphone filter number 3 in the FS beamformer.

array axis is 33.8cm .

Multichannel audio recordings are used to evaluate the effectiveness of the beamformers on real microphone signals. To preserve all necessary phase information, a multichannel simultaneous recording was necessary. This was accomplished using a high-speed data acquisition card, a personal computer and custom software. A block diagram of the system is shown in Figure 39. Data are recorded at a sampling rate of 20,000 samples per second per channel.

Recordings were made under a number of different test conditions. These are summarized in Table 3. All recordings were 52s in duration and consisted of a brief period (3 – 6s) of the noise field only followed by a 40s segment of pre-recorded speech. The balance of the 52s recordings consisted of the noise field only.

To characterize the recorded noise signals, power spectra were estimated from the noise-only segments of the audio recordings using the Welch method. The estimates were obtained by averaging 64 successive 1024-point FFTs, overlapped by 512 points. Representative power spectra obtained using the signal from microphone 3 are shown in Figure 40. The general shape of the spectra indicates the dominant low-frequency energy consistent with car-noise spectra reported by others.

7.1.3 Measured Gain

The array gain within the speech bandwidth of 300Hz to 4000Hz was measured for both beamformers under all recording conditions. Figure 41 shows the predicted and measured array gain averaged over all recording conditions. As seen in this figure, the average performance of the WDS beamformer agrees well with theoretical predicted performance in isotropic noise. With the possible exception of the 1kHz to 2kHz range, the average performance for the FS beamformer also displays reasonable agreement with predictions. It is also seen that the FS beamformer provides higher array gain than the WDS beamformer at frequencies below 1kHz . In fact, at 300Hz , the FS design provides an added 5dB of noise reduction.

Figure 39. Simplified block diagram of the multichannel recording system.

Figure 40. Vehicle noise-power spectra obtained under various conditions.

Figure 41. Predicted (dashed line) and measured (solid line) array gain for both the WDS beamformer (top graph) and the FS beamformer (bottom graph) as a function of frequency. These data are the average gains obtained for all conditions tested.

In the 1kHz to 2kHz range, a significant discrepancy between predicted and measured performance for the FS beamformer can be observed. This is caused by an electrical interference problem in the recording equipment. This can be seen in Figure 40 as the tonal components in the ambient noise spectra that protrude approximately 10 – 15dB above the instrument noise floor. These components are observed at 600Hz, 900Hz and in the 1kHz to 2kHz range. In examining the noise recorded at the other microphones, similar components are observed with varying strengths. Due to the dominance of the low frequency energy in the acoustic noise spectra, the electrical interference in the low frequencies is largely overshadowed. This is not always the case in the 1kHz to 2kHz range, however, particularly for those conditions with low cabin noise.

The two fixed beamformers have quite different responses to such noise signals. The reason for this is that the electrical interference appears to the beamformer as a spatially incoherent noise source. The WDS beamformer provides an optimum response under these conditions. On the other hand, the FS beamformer expects a spatially isotropic noise field. In the design process, part of the robustness to uncorrelated noise has been traded off in exchange for improved gain in isotropic noise. The reduced robustness is illustrated by the white noise gain which is shown in Figure 42. For the FS beamformer the white noise gain is close to its target value of 0dB up to a frequency of about 1.8kHz. At this point, the FS white noise gain approaches that of the WDS beamformer. As seen in the array gain plots, the problems coincide with the frequency region of reduced white noise gain.

With hindsight, of course, the FS beamformer could be designed with a different constraint in the troublesome region of 1kHz to 2kHz. Identical performance with the WDS beamformer would then be obtained under these conditions.

Individual data obtained for the specific cases listed in Table 3 can be found in [19]. General agreement between theory and measurement is also observed in the individual noise reduction plots; however, individual curves display a greater variation due to the lack of averaging.

The total noise power reduction within the speech bandwidth of 300Hz to 4000Hz was calculated for both beamformers under all recording conditions. In all cases, noise reduc-

Figure 42. White noise gain as a function of frequency for both broadband beamformers.

tion is measured in comparison to the noise observed by the reference microphone. The noise reduction figures are summarized in Table 3.

TABLE 3. Total speech-bandwidth noise power reduction in decibels provided by the two broadband beamformers.

Condition	WDS Beamformer	FS Beamformer
Ambient noise (parked inside building)	3.7	3.7
Fan on high-speed setting	1.0	4.1
Engine revving at 2000rpm	1.1	4.3
Moving at 60km/h	1.5	4.6
Moving at 60km/h fan on high speed	1.6	4.7
Moving at 60km/h , front windows open	4.5	4.6
Moving at 60km/h , rear windows open	2.0	4.6
Moving at 100km/h	1.9	4.8
Moving at 100km/h , fan on high speed	2.4	5.2

As seen in this table, the noise reduction provided by the FS beamformer is greater than that for the WDS beamformer in all but two cases where it is identical. Furthermore, the added noise reduction is comparable to the increased low-frequency gain provided by the FS beamformer (see Figure 41) due to the dominant low frequency energy in the vehicle noise spectra.

7.1.4 Effect of Beamformers on Speech Signal

To verify the beamformers' responses to the target speech signal, an anechoic response measurement was made. The mannequin and microphone array were placed in an anechoic chamber in the same relative positions as used in the car. Beamformer responses were calculated using the measured anechoic responses from the mannequin to each microphone. The differences in anechoic response between the beamformers and microphone 3, the closest microphone to the mannequin, are shown in Figure 43. As seen in this figure, the response of both beamformers is essentially flat over the frequency range of interest. A slight reduction of approximately 1 dB in frequencies above 1 kHz is due to the

fact that the mannequin is not omnidirectional at these frequencies as assumed by the source model. Informal listening indicated that these response deviations do not cause significant perceptual distortion so the effect of the beamformers on the speech signal will be neglected. This greatly simplifies subsequent analysis since it permits array gain to be measured directly by estimating the noise reduction.

7.1.5 Discussion

Using a fixed beamformer, noise reductions on the order of 4dB to 5dB are obtained for the conditions described in this chapter. These reductions are in line with predicted gains assuming an acoustic point source in an isotropic field. Affes and Grenier [80] measured the noise reduction performance and speech distortion levels obtained using several adaptive beamforming algorithms for another eight-element (approximately linear) microphone array. Noise reductions in the range of -6.7 dB (i.e. noise enhancement) to 10.8dB were observed. Unfortunately, in some instances, severe speech distortion was also measured making direct comparisons with the results of this paper difficult (since there is no speech distortion). In one case where speech distortion was minimum, noise reductions of 4.6dB to 5.1dB were recorded. These levels of noise reduction, though obtain under different conditions, are certainly comparable to those obtained in this study.

7.2 Computer Telephony Integration

Computer telephony (CT) is another application area where the source of interest is located within the array's near field. The application of an optimized, fixed beamformer is particularly well suited to the CT application since the talker is typically located directly in front of the computer screen and a movable beam is not necessary.

Figure 44 shows a linear microphone array positioned along the top edge of a video monitor. The array consists of 7 omnidirectional microphones equally spaced at intervals of 7cm, as in the application to mobile telephony. Care was taken to place the microphones at the vertex of the upper edge of the monitor to minimize the effects of sound diffraction [81]. A Cartesian coordinate system is defined such that the origin coincides with the cen-

Figure 43. Anechoic frequency response deviations produced by the broadband beamformers. The curves show the difference between the response provided by the beamformers compared to that of the reference microphone for speech signals from the mannequin.

ter of the array. The array lies along the x axis and the y axis is perpendicular to the computer screen. The assumed location of the desired talker is on the y axis at a distance of 55cm. At the highest frequency of interest, 3500Hz, this distance is within the array's near field which extends to a distance of approximately 0.91m.

The computer monitor was located on a small table inside a rectangular room measuring $4.6\text{m} \times 6.7\text{m} \times 2.4\text{m}$. A human head and torso simulator (mannequin) was seated in a chair directly in front of the computer monitor. This mannequin was designed to simulate the acoustical properties of the human voice during speech. Its use allows the precise control over experimental conditions and permits the objective measurement of the effects of speech pickup enhancements. The distance from the mannequin's artificial mouth opening to the array centre was 55cm. The physical layout of equipment within the room is illustrated in Figure 45.

An interesting feature of the chosen room location is the presence of a strong "flutter echo" which is clearly audible at the computer monitor's location. This echo presumably occurs when sound bounces back and forth between the two walls on the left and right side of the diagram in Figure 45 (these walls have higher reflection coefficients than the top and bottom walls, see Section 2.4.1 on page 17). It will be shown later that this strong flutter echo affects the isotropy of the reverberant sound field, a fact that can be exploited to further increase array performance.

The background noise level in the room is relatively low compared to a typical office environment. Consequently, beamformer effectiveness is examined solely by demonstrating reverberation reduction.

7.2.1 Room Impulse Response Measurements

To simulate the effect of beamforming on reverberation reduction, impulse response measurements were made from three separate sound sources to each microphone in the array. The sound sources are the mannequin's artificial mouth, to represent a near-end talker, and the left and right speakers built into the multimedia monitor, to represent a far-end speech source. A single microphone was used to take measurements for all array microphone

Figure 44. Microphone array intended for computer telephony application. Array consists of ten microphones equispaced at intervals of 7cm. The array focal point is at a distance of 50cm along the y axis.

Figure 45. Physical layout of computer monitor and mannequin within the room.

locations. The sampling frequency was 12kHz and the length of the measured impulse responses was 8192 samples corresponding to a time duration of 0.68s. This measurement length is more than adequate to cover the room reverberation time of approximately 0.3s.

The beamformer response to a given acoustic source is simulated by filtering the individual microphone impulse responses using the appropriate sensor FIR filter and summing the result from all seven microphones. To permit the direct comparison of single microphone responses with those from the array, the array responses are normalized to provide the same target signal level as a single microphone.

In the following section, three broadband beamformer designs will be studied: a maximum gain beamformer similar to that used in the mobile telephony example, and two maximum gain beamformers subject to a single far-field response constraint in the same angular direction as the focal point. The difference between the latter two designs is in the constrained response magnitude. In reference to (31), the beamformers are labelled according to the value of the constant α . In all cases, the time-domain formulation described in Section 6.2 on page 86 is used to implement the beamformer.

7.2.2 Beamformer Designs

The first beamformer is designed assuming that the reverberant sound field is isotropic. No additional constraint is placed on the array response pattern so $\alpha = 1$. Consequently, the designs are identical to those considered in the example shown in Section 6.3.2 on page 99. The performance data, discussed shortly, were obtained using a 30-tap beamformer. For this case $\sigma_w^2 = 0.04$, $\sigma_s^2 = 1$ and $\sigma_n^2 = 1$ resulting in a peak response error of 0.01dB and a peak gain error of 0.22dB, compared to an ideal optimum beamformer with white noise gain equal to 0dB. The actual array gain is shown as the solid curve in Figure 46 together with the gains of corresponding constrained beamformers (to be discussed later). Similarly, the white noise gain is shown in Figure 47.

The second beamformer provides the maximum array gain subject to the constraint that the far-field response in the broadside direction is exactly zero, that is using $\alpha = 0$. In

Figure 46. Array gain provided by broadband beamformers designed for CTI. The curve labelled $\alpha = 1$ is the maximum gain solution, whereas the curves labelled $\alpha = 0.707$ and $\alpha = 0$ were obtained for constrained designs.

Figure 47. White noise gain provided by CTI beamformers.

this case $\sigma_w^2 = 0.1$, $\sigma_s^2 = 1$ and $\sigma_n^2 = 1$ resulting in a peak response error of 0.65dB and a peak gain error of 9.9dB , compared to an ideal optimum beamformer with white noise gain equal to 0dB . Furthermore, achieving this performance in a broadband design requires 60 FIR taps.

The performance cost of imposing the far-field null is seen by referring to the array gain plotted in Figure 46 (dash-dot curve labelled $\alpha = 0$). The array gain has been drastically reduced by imposing the null constraint. In fact, for frequencies below approximately 1.6kHz, the null-constrained beamformer has a negative gain (i.e. the performance is worse than a single microphone). Interestingly, the performance penalty is not constant with frequency. Low frequencies are penalized much more than high frequencies. This is due to the variation in $\cos(S, N, R_{nn}^{-1}(\omega))^2$ (see (29)) when S corresponds to the array focal point and N corresponds to far-field point at which the response is constrained. For a given array geometry, the apparent angle between S and N is smaller ($\cos(S, N, R_{nn}^{-1}(\omega))^2 \rightarrow 1$) for lower frequencies (longer wavelengths).

Not surprisingly, imposing the null constraint also affects the white noise gain, as shown in Figure 47. In fact, the reduction in white noise gain is greater than that observed for the array gain: a consequence of the effect of the matrix $R_{nn}^{-1}(\omega)$ on the generalized angle between S and N .

The gain reductions can be mitigated by “softening” the constraint as discussed in Section 5.3 on page 72 and illustrated in Figure 27. Thus, the third beamformer replaces the far-field null constraint with a response constraint corresponding to $\alpha = -3\text{dB}$ or 0.707 . As in the null-constraint case, $\sigma_w^2 = 0.1$, $\sigma_s^2 = 1$ and $\sigma_n^2 = 1$; however, the resulting peak response error is only 0.06dB and the peak gain error drops to 2.3dB . Again, achieving this performance in a broadband design requires 60 FIR taps.

Again, the array gain is plotted in Figure 46 (dashed curve labelled $\alpha = 0.707$). Note that the array gain loss is limited to approximately 2dB at its worst, near a frequency of 500Hz . Similarly, the impact on white noise gain has also been reduced as shown in Figure 47.

7.2.3 Performance in Reverberant Environment

The performance of the above beamformer designs is now measured under realistic reverberant conditions. Each broadband beamformer is applied to the individual impulse responses described earlier. The response to both the near-end speech as well as the far-end speech are considered.

7.2.3.1 Response to Near-end Signal

In the mobile telephony example, it was possible to estimate array gain by measuring the noise power reduction obtained during segments of the audio recordings where speech was absent. This technique worked in this case since the dominant interference was the cabin noise. In cases where the dominant interference is reverberation, the corresponding technique would be to separately measure the effect of the beamformer on the direct and reverberant parts of the signal. In a simulated environment, this is easy to do. In a practical setting, however, separating these components is very difficult. Consequently, a different method must be adopted for measuring beamformer effectiveness in reverberation.

Two measurements are used to demonstrate the effect of the beamformers on near-end speech: the near-end frequency response (measured from the mannequin's mouth to the array output) and the reverberant energy decay observed at the array output. Frequency response curves are used to ensure that the beamformer passes the desired signal relatively undistorted. Decay curves are used to illustrate the suppression of reverberant energy.

Ideally, only the direct part of the reverberant impulse response would be used to generate the beamformer response; however, separating the direct and reflected parts of an impulse response is very difficult under realistic conditions. As a compromise, frequency response curves are generated by taking the DFT of the first 1024 impulse response samples. This corresponds to a time window of 85.3ms. The near-end frequency responses to be shown thus include the effects of the direct wavefront, including frequency response of artificial mouth, as well as a large number of reflections. Consequently, even though the beamformer response at the focal point is designed to be flat, when combined with the effects of room reflections the overall measured response will not be. To highlight the differences

between the beamformers, only the magnitude differences will be displayed. The reference curve for both cases is the frequency response obtained by the $\alpha = 1$ beamformer. The results are shown in Figure 48 and Figure 49.

As seen in Figure 48, the response of the $\alpha = 0$ beamformer, exhibits deviations of $\pm 8\text{dB}$ or more over much of the speech frequency range. Such large deviations have a negative impact on audio quality. The response of the $\alpha = 0.707$ beamformer is shown in Figure 49. In this case, the frequency response deviations are maintained at a much lower level. The impact on audio quality should be much less pronounced.

The effect of beamformers on reverberation is most readily demonstrated using energy decay curves. An energy decay curve is a graphical method for showing how rapidly the reverberant energy in a room impulse response dies out. For a given time instant, t , the energy decay is obtained by subtracting the energy in the impulse response tail from t to ∞ from the total impulse response energy, normalized to the total impulse power (see the method of integrated impulse response in [46]). Generally speaking, a more rapid decay in reverberant energy is considered beneficial for speech pickup.

The energy decay curves corresponding to speech emitted by the near-end talker (the mannequin) are illustrated in Figure 50 for each of the beamformer designs. The impulse response data were bandlimited to the range 300Hz to 4kHz before generating the decay curves. From these curves it is seen that the beamformer with $\alpha = 0.707$ provides the most rapid decay of reverberant energy. This suggests that the reverberant field in the room contains a significant non-isotropic component whose direction coincides with that of the desired talker. Interestingly, the beamformer with an exact null ($\alpha = 0$) provides a slower decay although the penalty is not as severe as might be expected from the array gain shown in Figure 46. Determining an optimum value of α requires an investigation of the detailed characteristics of the reverberant field. This non-trivial task [82] is beyond the scope of this thesis.

7.2.3.2 Response to Far-end Signal

In this example, the far-end signal is assumed to originate from the two, multimedia loud-

Figure 48. Target frequency response provided by the CTI beamformer optimized with $\alpha = 0$.

Figure 49. Target frequency response provided by the CTI beamformer optimized with $\alpha = 0.707$.

Figure 50. Energy decay curves for talker position provided by CTI beamformers.

speakers built into the computer monitor. As part of a complete handsfree system, it is interesting to demonstrate the beamformers' effect of the amount of far-end signal picked up by the array. This effect can be summarized by the echo return loss enhancement (ERLE). This figure of merit indicates the additional amount of far-end signal rejected by the beamformer in comparison to a single microphone. That is

$$\text{ERLE} = 10 \log \frac{\sum h_r^2(n)}{\sum_n h_t^2(n)} \quad (66)$$

where $h_r(n)$ is the impulse response to the (single) reference microphone only and $h_t(n)$ is the total impulse response of the beamformer. The summation is over all points in the impulse response.

The data illustrate that adding a response constraint also causes a reduction in ERLE. The ERLE provided by the $\alpha = 1$ beamformer is the highest at 8.1dB. The $\alpha = 0.707$ design provides 7.8dB whereas the null-constrained design amplifies the far-end signal with an ELRE of -2.3dB.

Chapter 8

CONCLUSIONS

This thesis explores array optimization applied to acoustic targets located in the near field of a microphone array. The results of this work illustrate the performance improvements that result when differences in curvature between the desired and interfering wavefronts are exploited.

For gain optimization alone, incorporation of a spherical-wave target model into the optimization criterion produces gain improvements for focal points (target locations) closer to the array. Furthermore, in the near-field, the maximum white noise gain is achieved by weighting the microphone signals by an amount inversely proportional to the distance to the sound source. This is in contrast to the far-field case where uniform weights provide the maximum white noise gain.

It is also shown that near-field gain optimization may be combined with linear constraints to force a given response level in an arbitrary far-field direction. In the specific case where the far-field response is constrained along the same angular direction as the near-field target, care must be taken in choosing the constraint response level so as to not degrade overall performance too drastically. A suitable “soft” constraint level is derived that permits a trade off between constraint level and array gain. The “soft” constraint is shown to provide added reverberation reduction in a non-isotropic field.

Two methods are used to generate broadband beamformer designs based on the above gain optimization techniques. The first method is based on narrowband array optimization and FIR filter design using complex approximation. The second method uses concepts from linearly constrained minimum variance (LCMV) beamforming to simultaneously optimize the array and design the sensor filters. By examining a specific example, the method based on LCMV beamforming is shown to provide comparable array gain but a flatter frequency response for a given number of FIR taps. Furthermore, the example also shows that improved accuracy in implementing the ideal sensor-filter frequency response does not necessarily translate into better array performance.

The theoretical results of the thesis are validated through two applications which require near-field beamforming: speech pickup in mobile telephony and computer telephony.

In mobile telephony, quadratically constrained near-field gain optimization provides 3dB of additional speech-bandwidth noise reduction compared to an existing near-field optimum solution. These gains are comparable to those obtained by other researchers using adaptive beamforming methods under similar conditions.

In a computer telephony application, linearly constrained optimization is used to reduce the beamformer response to wall reflections originating behind the desired talker. When applied to real measured room responses, the linearly constrained beamformer provides additional reduction in reverberation compared to an optimized-gain array.

BIBLIOGRAPHY

- [1] J.L. Flanagan, D.A. Berkley, G.W. Elko, J.E. West and M.M. Sondhi, "Autodirective microphone systems", *Acustica*, vol. 73, pp. 58-71, 1991.
- [2] R.W. Stadler, and W.M. Rabinowitz, "On the Potential of fixed arrays for hearing aids", *J. Acoust. Soc. Am.*, Vol. 94, pp. 1332-1342, Sep. 1993.
- [3] J.M. Kates, "Superdirective arrays for hearing aids", *J. Acoust. Soc. Am.*, Vol. 94, pp. 1930-1933, Oct. 1993.
- [4] W. Soede, F.A. Bilsen and A.J. Berkhout, "Development of a directional hearing instrument based on array technology", *J. Acoust. Soc. Am.*, Vol. 94, Pt. 1, pp. 785-798, Aug. 1993.
- [5] M.W. Hoffman, T.D. Trine, K.M. Buckley and D.J. Van Tasell, "Robust adaptive microphone array processing for hearing aids: realistic speech enhancement", *J. Acoust. Soc. Am.*, Vol. 96, pp. 759-770, Aug. 1994.
- [6] M.M. Gouling and J.S. Bird, "Speech enhancement for mobile telephony", *IEEE Trans. on Veh. Tech.*, vol. 39, pp. 316-326, Nov. 1990.
- [7] Nordholm, S., Claesson, I. and Bengtsson, B., "Adaptive Array Noise Suppression of Handsfree Speaker Input in Cars", *IEEE Trans. on Veh. Tech.*, vol. 42, Nov. 1993, pp. 514-518.
- [8] F. Khalil, J.P. Jullien and A. Gilloire, "Microphone Array for Sound Pickup in Teleconference Systems", *J. Audio Eng. Soc.*, vol. 42, pp. 691-700, 1994.
- [9] F. Pirz, "Design of a wideband, constant beamwidth, array microphone for use in the near field", *Bell Syst. Tech. J.*, vol. 58, pp. 1839-1850, Oct. 1979.
- [10] M. F. Berger and H.F. Silverman, "Microphone array optimization by stochastic region contraction", *IEEE Trans. on Signal Processing*, vol. 39, pp. 2377-2386, Dec. 1991.
- [11] J.E. Greenberg and P.M. Zurek, "Adaptive beamformer performance in reverberation", in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 1991*, Oct. 1991.
- [12] P.M. Peterson, "Using linearly-constrained adaptive beamforming to reduce interference in hearing aids from competing talkers in reverberant rooms", in *Proc. IEEE ICASSP 1987*, pp. 2364-2367.
- [13] T. Shan and T. Kailath, "Adaptive beamforming for coherent signals and interference", *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 33, pp. 527-536, June 1985.

- [14] P.M. Peterson, S. Wei, W.M. Rabinowitz and P.M. Zurek, "Robustness of an adaptive beamforming method for hearing aids", *Acta Otolaryngol*, Suppl. 469, pp. 85-90, 1990.
- [15] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction", *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. 34, pp. 1391-1400, Dec. 1986.
- [16] J.G. Ryan, and R.A. Goubran, "Nearfield beamforming for microphone arrays", in *Proc. IEEE ICASSP 1997*, pp. 363-366.
- [17] J.G. Ryan, and R.A. Goubran, "Optimum nearfield response for microphone arrays", in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 1997*, Oct. 1997.
- [18] J.G. Ryan, and R.A. Goubran, "Array optimization applied in the near field of a microphone array", *IEEE Trans. on Speech and Audio Processing*, accepted for publication.
- [19] J.G. Ryan, and R.A. Goubran, "Application of near-field optimum microphone arrays to handsfree mobile telephony", *IEEE Trans. on Vehicular Technology*, submitted for publication, 1999.
- [20] W. Koenig, "Subjective effects in binaural hearing", *J. Acoust. Soc. Am.*, vol. 22, pp. 61-62, Jan. 1950.
- [21] A.H. Koenig, J.B. Allen, D.A. Berkley and T.H. Curtis, "Determination of masking-level differences in a reverberant environment", *J. Acoust. Soc. Am.*, Vol. 61, pp. 1374-1376, 1977.
- [22] F.A. Bilsen, "Pitch of noise signals: evidence for a central spectrum", *J. Acoust. Soc. Am.*, Vol. 61, pp. 150-161, 1977.
- [23] A.H. Koenig, D.A. Berkley, T.H. Curtis and J.B. Allen, "Magnitude of JNDs for diotic and dichotic perception of spectrally colored noise", *J. Acoust. Soc. Am.*, vol. 58(S1), p. S55, 1975.
- [24] J.P. Moncur and D. Dirks, "Binaural and monaural speech intelligibility in reverberation", *J. Speech and Hearing Research*, vol. 10, pp. 186-195, 1967.
- [25] B.S. Atal, M.R. Schroeder and K.H. Kuttruff, "Perception of coloration in filtered Gaussian noise- short time spectral analysis by the ear", *4th Intl. Cong. Acoust.*, paper H31, 1962.
- [26] P.M. Zurek, "Measurements of binaural echo suppression", *J. Acoust. Soc. Am.*, vol. 66, pp. 1750-1757, 1979.

- [27] H.K. Dunn and D.W. Farnsworth, "Exploration of pressure field around the human head during speech", in *J. Acoust. Soc. Am.*, vol. 10, pp. 184-199, 1939.
- [28] H.F. Olson, *Acoustical Engineering*, Princeton, NJ: D. Van Nostrand Company Inc., 1957, p. 293.
- [29] C.W. Reedyk, "Noise-cancelling electret microphone for lightweight head telephone sets", *J. Acoust. Soc. Am.*, vol. 53, pp. 1609-1615, 1973.
- [30] G.M. Sessler and J.E. West, "First-order gradient microphones based on the foil-electret principle: discrimination against air-borne and solid-borne noises", *J. Acoust. Soc. Am.*, vol. 46, pp. 1081-1086, 1969.
- [31] G.M. Sessler, J.E. West and R.A. Kubli, "Unidirectional second-order gradient microphone", *J. Acoust. Soc. Am.*, vol. 86, pp. 2063-2066, 1989.
- [32] J.L. Flanagan, J.D. Johnston, R. Zahn and G.W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms", *J. Acoust. Soc. Am.*, Vol. 78, pp. 1508-1518, Nov. 1985.
- [33] T. Nishi and T. Inoue, "Development of a multi-beam array microphone for multi-channel pickup of sound fields", *Acustica*, vol. 76, pp. 163-172, 1992.
- [34] L.E. Kinsler, A.R. Frey, A.B. Coppens J.V. and Sanders, *Fundamentals of Acoustics*, 3rd ed., New York: John Wiley and Sons, 1982.
- [35] J. L. Flanagan, "Analog measurements of sound radiation from the mouth", *J. Acoust. Soc. Am.*, Vol. 32, pp. 1613-1620, 1960.
- [36] R.G. Cann and K.A. Hoover, "Architectural Acoustic Principles and Design Techniques" in *Audio Engineering Handbook* K.B. Benson, Ed. New York: McGraw-Hill, 1988, p. 3.7.
- [37] S.T. Neely and J.B. Allen, "Invertibility of a room impulse response", *J. Acoust. Soc. Am.*, vol. 66, pp. 165-169, 1979.
- [38] B. Gover and J.G. Ryan, "Least-squares inverse filtering of speech", *Canadian Acoustics*, vol. 26, pp. 36-37, September 1998.
- [39] M. Miyoshi and Y. Kaneda, "Inverse Filtering of Room Acoustics", *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 36, pp. 145-152, Dec. 1988.
- [40] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small-room acoustics", *J. Acoust. Soc. Am.*, vol. 65, pp. 943-950, 1979.

- [41] P.M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room", *J. Acoust. Soc. Am.*, vol. 80, pp. 1527-1529, 1986.
- [42] B. Champagne, "Simulation of the response of multiple microphones to a moving point source", *Applied Acoustics*, vol. 42(4), pp. 313-332, 1994.
- [43] N. Dal Degan and C. Prati, "Acoustic noise analysis and speech enhancement techniques for mobile radio applications", *Signal Processing*, vol. 15, pp. 43-56, July 1988.
- [44] C.T. Morrow, "Point-to-point correlation of sound pressures in reverberation chambers", *J. Sound and Vibration*, vol. 16(1), pp. 29-42, 1971.
- [45] G.W. Elko, "Adaptive noise cancellation with directional microphones", in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 1997*, Oct. 1997.
- [46] H. Kuttruff, *Room Acoustics*, 3rd ed. Essex, England: Elsevier Applied Science, 1991, pp. 116-117.
- [47] M.R. Schroeder, "New method of measuring reverberation time", *J. Acoust. Soc. Am.*, vol. 37, pp. 409-412, 1965.
- [48] J.J. Jetzt, "Critical distance measurement of rooms from the sound energy spectral response", *J. Acoust. Soc. Am.*, vol. 65, pp. 1204-1211, 1979.
- [49] D.A. Berkley, "Normal Listeners in Typical Rooms", in *Acoustical factors Affecting Hearing Aid Performance*, G.A. Studebaker and I. Hochberg Ed. University Park Press, 1980.
- [50] D.H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, NJ: P T R Prentice Hall, 1993.
- [51] H. Cox, R.M. Zeskind and M.M. Owen, "Robust adaptive beamforming", *IEEE Trans. on A.S.S.P.*, vol. 35, pp. 1365-1375, October 1987.
- [52] H. Cox, "Sensitivity considerations in adaptive beamforming", in *Signal Processing*, J.W.R. Griffiths, P.L. Stocklin and C. Van Schooneveld, Eds., New York: Academic Press, 1973.
- [53] P.M. Peterson, *Adaptive Array Processing for Multiple Microphone Hearing Aids*, Ph.D. Thesis, 1989.
- [54] E.N. Gilbert and S.P. Morgan, "Optimum design of directive antennas subject to random variations", *Bell System Technical Journal*, vol. 34, pp. 637-663, May 1955.

- [55] R.A. Monzingo and T.A. Miller, *Introduction to Adaptive Arrays*, New York: John Wiley & Sons, 1980.
- [56] D.K. Cheng, "Optimization Techniques for Antenna Arrays", *Proc. of the IEEE*, vol. 59, pp. 1664-1674, Dec. 1971.
- [57] D.K. Cheng and F.I. Tseng, "Gain optimization for arbitrary antenna arrays", *IEEE Trans. on Antennas and Propagation*, vol. 13, pp. 973-974, Nov. 1965.
- [58] H. Cox, R.M. Zeskind, and T. Kooij, "Practical supergain", *IEEE Trans. on A.S.S.P.*, vol. 34, pp. 393-398, June 1986.
- [59] J.M. Kates and M.R. Weiss, "A comparison of hearing-aid array-processing techniques", *J. Acoust. Soc. Am.*, Vol. 99, pp. 3139-3148, May 1996.
- [60] W.E. Blazier, "Noise control criteria for heating, ventilating, and air-conditioning systems", chap. 43 in *Handbook of Acoustical Measurements and Noise Control*, C.M. Harris, Ed., 3rd ed. New York: McGraw Hill, 1991.
- [61] K.M. Buckley, "Spatial/spectral filtering with linearly constrained minimum variance beamformers", *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 35, pp. 249-266, Mar. 1987.
- [62] B.D. Van Veen and K.M. Buckley, "Beamforming: a versatile approach to spatial filtering", *IEEE Signal Processing Magazine*, pp.4-24, Apr. 1988.
- [63] J.G. Ryan, "Criterion for the minimum source distance at which plane-wave beamforming can be applied", in *J. Acoust. Soc. Am.*, vol. 104, No. 1, pp. 595-598, 1998.
- [64] B.D. Steinberg, *Principles of Aperture and Array System Design*, New York: John Wiley & Sons, 1976, p. 12.
- [65] P.S. Hacker and H.E. Schrank, "Range distance requirements for measuring low and ultralow sidelobe antenna patterns", *IEEE Trans. on Antennas and Propagation*, vol. 30, pp. 956-966, Sep. 1982.
- [66] R.C. Hansen, "Measurement distance effects on low sidelobe patterns", *IEEE Trans. on Antennas and Propagation*, vol. 32, pp. 591-594, June 1984.
- [67] R.A. Kennedy, T. Abhayapala and D.B. Ward, "Broadband nearfield beamforming using a radial beam pattern transformation", *IEEE Trans. on Signal Processing*, vol. 46, pp. 2147-2156, Aug. 1998.
- [68] R.A. Kennedy, T. Abhayapala, D.B. Ward, "Nearfield Broadband Frequency Invariant Beamforming", *Proc. of I.C.A.S.S.P.*, 1996.

- [69] D.B. Ward, R.A. Kennedy, and R.C. Williamson, "Theory and design of broadband sensor arrays with frequency-invariant far-field beampatterns", *J. Acoust. Soc. Am.*, vol. 97, pp. 1023-1034, Feb. 1995.
- [70] D.B. Ward and G.W. Elko, "Mixed near-field/far-field beamforming: A new technique for speech acquisition in a reverberant environment", *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 1997*, October 1997.
- [71] S. Nordebo, I. Claesson and S. Nordholm, "Weighted Chebyshev approximation for the design of broadband beamformers using quadratic programming", *IEEE Signal Processing Letters*, vol. 1, pp. 103-105, July 1994.
- [72] H. Steyskal, "Synthesis of antenna patterns with imposed near-field nulls", *Electronics Letters*, vol. 30, pp. 2000-2001, Nov. 1994.
- [73] H. Nomura, Y. Kaneda and J. Kojima, "Optimum gains of a delay-and-sum microphone array for near sound field", *J. Acoust. Soc. Am.*, vol. 100, No. 4, Pt. 2, p. 2697, Oct. 1996.
- [74] H. Cox, "Resolving power and sensitivity to mismatch of optimum array processors", *J. Acoust. Soc. Am.*, vol. 54, pp. 771-785, 1973.
- [75] T.W. Parks and C.S. Burrus, *Digital Filter Design*. New York, NY: John Wiley & Sons, Inc, 1987.
- [76] R.L. Streit and A.H. Nutall, "A general Chebyshev complex function approximation procedure and an application to beamforming", *J. Acoust. Soc. Am.*, vol. 72(1), pp. 181-190, July 1982.
- [77] A.S. Alkhairy, K.G. Christian and J.S. Lim, "Design and characterization of optimal FIR filters with arbitrary phase", *IEEE Trans. on Signal Processing*, vol. 41(2), pp. 559-572, Feb. 1993.
- [78] L.J. Karam and J.H. McClellan, "Complex Chebyshev approximation for FIR filter design", *IEEE Trans. on Circuits and System II: Analog and Digital Signal Processing*, vol. 42(3), pp. 207-216, Mar. 1995.
- [79] T.I. Laakso, V. Valimaki, M. Karjalainen and U. K. Laine, "Splitting the Unit Delay", *IEEE SP Magazine*, vol. 13, pp. 30-60, Jan. 1996.
- [80] S. Affes and Y. Grenier, "Test of adaptive beamformers for speech acquisition in cars", *Proc. 5th International Conference on Signal Processing Applications and Technology*, pp. 154-159, October 1994.
- [81] M.R. Stinson, J.G. Ryan and G. Daigle, "Sound pickup in the presence of diffraction effects", *Canadian Acoustics*, vol. 26, pp. 34-35, September 1998.

- [82] T.J. Schultz, "Diffusion in Reverberation Rooms", *J. Sound and Vibration*, vol. 16(1), pp. 17-28, May 1971.
- [83] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York, NY: McGraw-Hill Book Company, 1984.
- [84] A. B. Carlson, *Communication Systems: An Introduction to Signals and Noise in Electrical Communication*. New York, NY: McGraw-Hill Book Company, 1975.
- [85] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*. New York, NY: Academic Press Inc., 1965, p. 187.
- [86] A. Fenwick, "An array processor subject to a constraint", *J. Acoust. Soc. Am.*, vol. 49(4), pp. 1108-1110, April 1971.
- [87] C. Drane and J. McIlvenna, "Gain Maximization and Controlled Null Placement Simultaneously Achieved in Aerial Array Patterns", *The Radio and Electronic Engineer*, Vol. 39, pp. 49-57, Jan. 1970.

Appendix A

BROADBAND CORRELATION FUNCTIONS

This Appendix derives expressions for the point-to-point broadband correlation functions required to describe the LCMV beamformer design equations.

A.1 Broadband Acoustic Point Source

Consider an acoustic point source in the vicinity of two omnidirectional point receivers as depicted in Figure 51. The signal power spectral density is η^2 and uniform over the frequency range f_a to f_b . Assuming ideal spherical-wave spreading, the power spectral density $S_s(\omega)$ observed at a distance r from the source location is

$$S_s(\omega) = \begin{cases} \frac{\eta^2}{r^2}, & (2\pi f_a \leq \omega \leq 2\pi f_b) \\ \frac{\eta^2}{r^2}, & (-2\pi f_b \leq \omega \leq -2\pi f_a) \\ 0, & (\text{otherwise}) \end{cases} \quad (67)$$

The corresponding autocorrelation function $\tilde{\rho}_s(\tau)$ is obtained by taking the inverse Fourier transform of (67)

$$\tilde{\rho}_s(\tau) = \frac{\eta^2}{\pi\tau} \cdot \frac{1}{r^2} \cdot (\sin \omega_b \tau - \sin \omega_a \tau) \quad (68)$$

where τ is continuous time lag.

The crosscorrelation between the signals observed by two different sensors can be obtained using the methods of linear system theory. The relative impulse response $h_{21}(t)$ from sensor 2 to sensor 1 can be written as

Figure 51. Acoustic point source located in the vicinity of two ideal point receivers.

$$h_{21}(t) = \frac{r_2}{r_1} \cdot \delta\left(t - \frac{r_1 - r_2}{c}\right) \quad (69)$$

where $\delta(t)$ is a continuous-time Dirac delta function. The corresponding transfer function $H_{21}(\omega)$ is

$$H_{21}(\omega) = \frac{r_2}{r_1} \cdot e^{-j\omega\left(\frac{r_1 - r_2}{c}\right)} \quad (70)$$

For a linear system with input power spectral density $S_{ii}(\omega)$ and transfer function $H(\omega)$, the cross-power spectrum between the input and output $S_{io}(\omega)$ is given by [83]

$$S_{io}(\omega) = S_{ii}(\omega) \cdot H(\omega)^* \quad (71)$$

where $*$ denotes complex conjugation. Therefore, the crosspower spectrum between sensor 2 and sensor 1 is

$$S_{21}(\omega) = S_{22}(\omega) \cdot H_{21}(\omega)^* \quad (72)$$

The crosscorrelation function $\tilde{\rho}_{21}(\tau)$ is the inverse Fourier transform of (72)

$$\tilde{\rho}_{21}(\tau) = \frac{1}{2\pi} \cdot \int_{-\pi}^{\pi} S_{22}(\omega) \cdot \left(\frac{r_2}{r_1} \cdot e^{j\omega\left(\frac{r_1 - r_2}{c}\right)} \right) e^{j\omega\tau} d\omega \quad (73)$$

Similarly, the cross power spectrum between sensor 1 and sensor 2 is

$$S_{12}(\omega) = S_{11}(\omega) \cdot H_{12}(\omega)^* \quad (74)$$

and the crosscorrelation function $\tilde{\rho}_{12}(\tau)$ is the inverse Fourier transform of (74)

$$\tilde{\rho}_{12}(\tau) = \frac{1}{2\pi} \cdot \int_{-\pi}^{\pi} S_{11}(\omega) \cdot \left(\frac{r_1}{r_2} \cdot e^{j\omega\left(\frac{r_2 - r_1}{c}\right)} \right) e^{j\omega\tau} d\omega \quad (75)$$

By the scaling and time-shifting properties of the Fourier transform [84], $\tilde{\rho}_{21}(\tau)$ can be deduced from (68) as

$$\tilde{\rho}_{21}(\tau) = \frac{r_2}{r_1} \cdot \tilde{\rho}_{22}\left(\tau + \left(\frac{r_1 - r_2}{c}\right)\right) \quad (76)$$

where $\tilde{\rho}_{22}(\tau)$ is given by (68) with $r = r_2$, so that

$$\tilde{\rho}_{21}(\tau) = \frac{\eta^2}{\pi} \cdot \frac{1}{r_1 r_2} \cdot \frac{(\sin \omega_b(\tau + (r_1 - r_2)/c) - \sin \omega_a(\tau + (r_1 - r_2)/c))}{(\tau + (r_1 - r_2)/c)}. \quad (77)$$

Similarly, $\tilde{\rho}_{12}(\tau)$ is given by

$$\tilde{\rho}_{12}(\tau) = \frac{r_1}{r_2} \cdot \tilde{\rho}_{11}\left(\tau + \left(\frac{r_2 - r_1}{c}\right)\right), \quad (78)$$

and

$$\tilde{\rho}_{12}(\tau) = \frac{\eta^2}{\pi} \cdot \frac{1}{r_1 r_2} \cdot \frac{(\sin \omega_b(\tau + (r_2 - r_1)/c) - \sin \omega_a(\tau + (r_2 - r_1)/c))}{(\tau + (r_2 - r_1)/c)}. \quad (79)$$

It many instances it is useful normalize the crosscorrelation functions to the signal power observed at a reference sensor location. For a reference sensor located a distance r from the source, the total power $\tilde{\rho}(0)$ is obtained by integrating (67) over all frequencies. This gives

$$\tilde{\rho}(0) = \frac{\eta^2}{\pi r^2} \cdot (\omega_b - \omega_a). \quad (80)$$

The normalized versions of (77) and (79) are

$$\rho_{21}(\tau) = \frac{r^2}{r_1 r_2} \cdot \frac{(\sin \omega_b(\tau + (r_1 - r_2)/c) - \sin \omega_a(\tau + (r_1 - r_2)/c))}{(\omega_b - \omega_a)(\tau + (r_1 - r_2)/c)} \quad (81)$$

and

$$\rho_{12}(\tau) = \frac{r^2}{r_1 r_2} \cdot \frac{(\sin \omega_b(\tau + (r_2 - r_1)/c) - \sin \omega_a(\tau + (r_2 - r_1)/c))}{(\omega_b - \omega_a)(\tau + (r_2 - r_1)/c)}, \quad (82)$$

respectively. These expressions are similar to the autocorrelation function for a broadband planewave source presented in [61].

As an example, the normalized crosscorrelation function according to (81) was calculated as a function of τ for different separation distances $\Delta r = r_1 - r_2$. To illustrate the behavior of this function in the context of discrete-time systems, time delays τ and distances $r_1 - r_2$, are normalized to the sampling interval $1/F_s$ and the signal frequency range is 0 to $F_s/2$. The reference sensor is chosen as sensor 2 so that $r = r_2$. The plots shown in Figure 52 were generated using $r_2 = 10$ samples and varying r_1 , as shown.

A.2 Broadband Isotropic Noise

An isotropic field can be thought of as resulting from the superposition of an infinite number of uncorrelated plane waves each with an equiprobable direction of propagation. In the case of a broadband isotropic field, crosscorrelation between two points must take into account spatial variations due to planewave propagation as well as temporal variations due to finite bandwidth.

Figure 53 shows a planewave propagating past two omnidirectional point receivers. The signal carried by the planewave is assumed to be broadband with uniform power spectral density η^2 over a finite bandwidth from f_a to f_b . That is,

$$S_p(\omega) = \begin{cases} \eta^2, & (2\pi f_a \leq \omega \leq 2\pi f_b) \\ \eta^2, & (-2\pi f_b \leq \omega \leq -2\pi f_a) \\ 0, & (\text{otherwise}) \end{cases} \quad (83)$$

The corresponding autocorrelation function $\tilde{\rho}_p(\tau)$ is obtained via inverse Fourier transform

$$\tilde{\rho}_p(\tau) = \frac{\eta^2}{\pi\tau} \cdot (\sin \omega_b \tau - \sin \omega_a \tau). \quad (84)$$

For planewave propagation, the signal received by sensor 1 is simply a time-shifted version of that received by sensor 2. The time delay between sensors is given by the propagation delay of the acoustic wavefront $d \cos \theta / c$. Following an approach similar to that used to derive (76), the crosscorrelation $\tilde{\rho}_{21}(\tau)$ between sensor 2 and sensor 1 in a planewave

Figure 52. Crosscorrelation function according to (81) for the physical arrangement shown in Figure 51. Both τ and $\Delta r = r_1 - r_2$ are normalized to the sampling interval $1/F_s$. The signal frequency range is 0 to $F_s/2$.

Figure 53. Acoustic planewave propagating across two ideal point receivers.

field is

$$\tilde{\rho}_{21}(\tau) = \tilde{\rho}_{22}(\tau + d\cos\theta/c). \quad (85)$$

The total crosscorrelation in an isotropic field $\tilde{\rho}_n(\tau)$ is obtained by taking the expected value of (85) over all arrival directions θ . That is

$$\tilde{\rho}_n(\tau) = \frac{1}{2} \int_0^\pi \tilde{\rho}_{22}(\tau + d\cos\theta/c) \sin\theta d\theta. \quad (86)$$

$$(87)$$

Substituting (84) for $\tilde{\rho}_{22}(\tau)$ yields

$$\tilde{\rho}_n(\tau) = \frac{\eta^2}{2\pi} \int_0^\pi \left(\frac{\sin\omega_b(\tau + d\cos\theta/c) - \sin\omega_a(\tau + d\cos\theta/c)}{\tau + d\cos\theta/c} \right) \sin\theta d\theta. \quad (88)$$

By introducing the change of variables

$$u = \frac{d\cos\theta}{c} + \tau$$

and

$$du = -\frac{d\sin\theta}{c} d\theta$$

the above can be expressed as

$$\tilde{\rho}_n(\tau) = -\frac{\eta^2}{2\pi} \cdot \frac{c}{d} \cdot \int_{\frac{d}{c} + \tau}^{-\frac{d}{c} + \tau} \left(\frac{\sin\omega_b u - \sin\omega_a u}{u} \right) du. \quad (89)$$

From [85] it is found that

$$\int \frac{\sin kx}{(a + bx)} dx = \frac{1}{b} \left[\cos \frac{ka}{b} \text{si}(u) - \sin \frac{ka}{b} \text{ci}(u) \right] \text{ where } u = \frac{k}{b}(a + bx).$$

Making the appropriate substitutions in (89) and simplifying yields

$$\tilde{\rho}_n(\tau) = \frac{\eta^2}{2\pi} \cdot \frac{c}{d} \cdot [\text{si}[\omega_b u_1] - \text{si}[\omega_a u_1] - \text{si}[\omega_b u_2] + \text{si}[\omega_a u_2]] \quad (90)$$

where $u_1 = d/c + \tau$, $u_2 = -d/c + \tau$ and $\text{si}(x)$ and $\text{ci}(x)$ are the sine and cosine integrals respectively defined as [85]

$$\text{si}(x) = -\int_x^{\infty} \frac{\sin u}{u} du.$$

and

$$\text{ci}(x) = -\int_x^{\infty} \frac{\cos u}{u} du.$$

The normalized crosscorrelation $\rho_n(\tau)$ is obtained by dividing (90) by

$$\tilde{\rho}_p(0) = \frac{\eta^2}{\pi} \cdot (\omega_b - \omega_a) \quad (91)$$

which is the power received by a single sensor. This produces

$$\rho_n(\tau) = \frac{c}{2d(\omega_b - \omega_a)} \cdot [\text{si}[\omega_b u_1] - \text{si}[\omega_a u_1] - \text{si}[\omega_b u_2] + \text{si}[\omega_a u_2]]. \quad (92)$$

As an example, the normalized crosscorrelation function according to (92) was calculated as a function of τ for different sensor separations d . As in the previous example, time delays τ and distance d , are both normalized to the sampling interval $1/F_s$ and the signal frequency range is 0 to $F_s/2$.

Figure 54. Crosscorrelation function according to (92) for the physical arrangement shown in Figure 53. Both τ and d are normalized to the sampling interval $1/F_s$. The signal frequency range is 0 to $F_s/2$.

Appendix B

OPTIMIZATION SUBJECT TO LINEAR CONSTRAINTS

This Appendix presents a derivation of the sensor weights, array gain and white noise gain for a beamformer optimized subject to dual linear constraints. The derivation generalizes that found in [86] incorporating constraints that are not necessarily null and making use of the concept of generalized angle as in [74]. For the specific case of a single look-direction gain constraint and a single null constraint, the result is shown to be identical to that in [86]. The results of this derivation are used to demonstrate near-field gain optimization subject to a far-field constraint.

B.1 Optimum Weights

The problem is that of maximizing array gain G subject to the set of linear constraints $C^H W = g$. As discussed in Section 3.1.3.2 on page 35 this can be accomplished by using the linearly constrained minimum variance (LCMV) approach. Omitting the explicit frequency dependence of the quantities, the weight vector that simultaneously maximizes array gain and satisfies the constraints is

$$W_{\text{opt}} = R_{nn}^{-1} C [C^H R_{nn}^{-1} C]^{-1} g . \quad (93)$$

Consider the specific case of two linear constraints. That is, for

$$C = \begin{bmatrix} S & N \end{bmatrix} \text{ and } g = \begin{bmatrix} g_s \\ g_n \end{bmatrix} .$$

In this derivation, the only assumption made about S and N is that they are not identical (otherwise the matrix represented by $C^H R_{nn}^{-1} C$ would not be invertible). In this thesis, results are presented assuming a near-field S and a far-field N ; however, either S or N could be in the near field or far field.

Substituting these expressions into that for W_{opt} yields

$$\mathbf{W}_{\text{opt}} = \mathbf{R}_{nn}^{-1} [\mathbf{S} \ \mathbf{N}] \left[\begin{bmatrix} \mathbf{S}^H \\ \mathbf{N}^H \end{bmatrix} \mathbf{R}_{nn}^{-1} [\mathbf{S} \ \mathbf{N}] \right]^{-1} \begin{bmatrix} g_s \\ g_n \end{bmatrix}$$

which becomes

$$\mathbf{W}_{\text{opt}} = \begin{bmatrix} \mathbf{R}_{nn}^{-1} \mathbf{S} & \mathbf{R}_{nn}^{-1} \mathbf{N} \end{bmatrix} \begin{bmatrix} \mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{S} & \mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{N} \\ \mathbf{N}^H \mathbf{R}_{nn}^{-1} \mathbf{S} & \mathbf{N}^H \mathbf{R}_{nn}^{-1} \mathbf{N} \end{bmatrix}^{-1} \begin{bmatrix} g_s \\ g_n \end{bmatrix}$$

after multiplication. The inverse of the 2×2 matrix is

$$\begin{bmatrix} \mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{S} & \mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{N} \\ \mathbf{N}^H \mathbf{R}_{nn}^{-1} \mathbf{S} & \mathbf{N}^H \mathbf{R}_{nn}^{-1} \mathbf{N} \end{bmatrix}^{-1} = \frac{1}{\Delta} \cdot \begin{bmatrix} \mathbf{N}^H \mathbf{R}_{nn}^{-1} \mathbf{N} & -\mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{N} \\ -\mathbf{N}^H \mathbf{R}_{nn}^{-1} \mathbf{S} & \mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{S} \end{bmatrix}$$

where Δ is the matrix determinant

$$\Delta = [\mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{S}][\mathbf{N}^H \mathbf{R}_{nn}^{-1} \mathbf{N}] - |\mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{N}|^2.$$

Substituting the expression for the inverse matrix, multiplying through and rearranging terms yields

$$\begin{aligned} \mathbf{W}_{\text{opt}} = \frac{1}{1 - \cos(\gamma)^2} \cdot & \left[g_s \frac{\mathbf{R}_{nn}^{-1} \mathbf{S}}{\mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{S}} - g_n \frac{\mathbf{R}_{nn}^{-1} \mathbf{S}}{\mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{S}} \cdot \frac{\mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{N}}{\mathbf{N}^H \mathbf{R}_{nn}^{-1} \mathbf{N}} \right. \\ & \left. + g_n \frac{\mathbf{R}_{nn}^{-1} \mathbf{N}}{\mathbf{N}^H \mathbf{R}_{nn}^{-1} \mathbf{N}} - g_s \frac{\mathbf{R}_{nn}^{-1} \mathbf{N}}{\mathbf{N}^H \mathbf{R}_{nn}^{-1} \mathbf{N}} \cdot \frac{\mathbf{N}^H \mathbf{R}_{nn}^{-1} \mathbf{S}}{\mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{S}} \right] \end{aligned}$$

where $\cos(\gamma)^2$ is the cosine squared of the generalized angle between the constraint vectors $\mathbf{S}$ and $\mathbf{N}$

$$\cos(\gamma)^2 = \frac{|\mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{N}|^2}{\mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{S} \cdot \mathbf{N}^H \mathbf{R}_{nn}^{-1} \mathbf{N}}.$$

Note that due to the Schwarz inequality,

$$0 \leq \cos(\gamma)^2 \leq 1.$$

When $\cos(\gamma)^2 = 0$, the vectors $\mathbf{S}$ and $\mathbf{N}$ are orthogonal in the generalized sense. Conversely, when $\cos(\gamma)^2 = 1$, the vectors are colinear. When $\mathbf{R}_{nn} = \mathbf{I}$, the generalized angle reduces to a more familiar form. A more complete discussion of the concept of generalized angle can be found in [74].

The apparently complicated expression for $\mathbf{W}_{\text{opt}}$ can be simplified by defining

$$\mathbf{W}_s = \frac{\mathbf{R}_{nn}^{-1} \mathbf{S}}{\mathbf{S}^H \mathbf{R}_{nn}^{-1} \mathbf{S}} \text{ and } \mathbf{W}_n = \frac{\mathbf{R}_{nn}^{-1} \mathbf{N}}{\mathbf{N}^H \mathbf{R}_{nn}^{-1} \mathbf{N}}.$$

Referring to (11), $\mathbf{W}_s$ and $\mathbf{W}_n$ can be recognized as the weight vectors that provide optimum gain for signal directions given by $\mathbf{S}$ and $\mathbf{N}$, respectively. Substituting the expressions for $\mathbf{W}_s$ and $\mathbf{W}_n$ yields

$$\mathbf{W}_{\text{opt}} = \frac{1}{1 - \cos(\gamma)^2} \cdot [g_s \mathbf{W}_s - g_n \mathbf{W}_s \mathbf{S}^H \mathbf{W}_n + g_n \mathbf{W}_n - g_s \mathbf{W}_n \mathbf{N}^H \mathbf{W}_s]. \quad (94)$$

Thus, the optimum linearly constrained weights are obtained directly from the constraint vectors $\mathbf{S}$ and $\mathbf{N}$ and the individual optimum weight sets $\mathbf{W}_s$ and $\mathbf{W}_n$.

Note the denominator term involving $\cos(\gamma)^2$. If the vectors $\mathbf{S}$ and $\mathbf{N}$ are orthogonal, then $\cos(\gamma)^2 = 0$ and the denominator term has no effect. As $\mathbf{S}$ approaches $\mathbf{N}$, however, $\cos(\gamma)^2 \rightarrow 1$, the denominator approaches 0 and the weight magnitudes increase drastically. Under this condition, the resulting array has very low robustness.

B.2 Array Gain

The array gain provided by the weight vector $\mathbf{W}$ for a signal with propagation vector $\mathbf{S}$ immersed in a noise field described by $\mathbf{R}_{nn}$ is given by

$$G = \frac{\mathbf{W}^H \mathbf{S} \mathbf{S}^H \mathbf{W}}{\mathbf{W}^H \mathbf{R}_{nn} \mathbf{W}}.$$

Substituting the expression for W_{opt} in (93) and simplifying yields the following expression for the expected gain of a beamformer optimized subject to two linear constraints

$$G = S^H R_{nn}^{-1} S \cdot \frac{|g_s|^2 (1 - \cos(\gamma)^2)}{|g_s|^2 + |g_n|^2 \cdot \frac{S^H R_{nn}^{-1} S}{N^H R_{nn}^{-1} N} - \frac{2 \text{Re}\{g_s g_n^* (N^H R_{nn}^{-1} S)\}}{N^H R_{nn}^{-1} N}}. \quad (95)$$

The quantity $S^H R_{nn}^{-1} S$ is the optimum gain provided by the W_s beamformer. That is, the gain of an array optimized for reception of S subject to a single gain constraint in the look direction and no second constraint in the direction of N . This quantity is referred to as the unconstrained gain G_u so that

$$G = G_u \cdot \frac{|g_s|^2 (1 - \cos(\gamma)^2)}{|g_s|^2 + |g_n|^2 \cdot \frac{S^H R_{nn}^{-1} S}{N^H R_{nn}^{-1} N} - \frac{2 \text{Re}\{g_s g_n^* (N^H R_{nn}^{-1} S)\}}{N^H R_{nn}^{-1} N}}. \quad (96)$$

Although it is difficult to determine from inspection of (96), it is shown in [87] that the constrained gain is always less than or equal to the unconstrained gain. For the specific case where the linear constraints consist of a single gain constraint $S^H W_s = g_s = 1$ and a single null constraint $N^H W_n = g_n = 0$, W_{opt} simplifies to

$$W_{\text{opt}} = \frac{1}{1 - \cos(\gamma)^2} \cdot [W_s - W_n N^H W_s].$$

The quantity $N^H W_s$ is the response of a beamformer optimized for reception of S to the direction N . This is precisely the amount of component N that must be cancelled in the response of the W_s beamformer. Thus, the linearly constrained optimum solution is composed of the sum of an optimum beamformer W_s and a nulling beamformer $-W_n N^H W_s$, in agreement with [86]. Furthermore, the array gain simplifies to

$$G = G_u \cdot (1 - \cos(\gamma)^2).$$

Thus, in the case of a single null constraint, the resulting array gain is always smaller than the gain of an equivalent unconstrained array since S and N cannot be identical.

It should be noted that this result applies equally well to array gain and white noise gain (obtained by setting $R_{nn} = I$) since no assumption was made about R_{nn} . Consequently, imposing a linear constraint on an optimized beamformer will degrade both the array gain as well as the beamformer's robustness, as measured by the white noise gain.

In spite of these drawbacks, sometimes it is necessary to reduce the beamformer response to the direction N beyond what is obtained by unconstrained optimization. In these cases, the loss in array gain can be limited by choosing a compromise value of g_n that minimizes the impact on array gain while allowing suppression of discrete interferers. This is discussed in the following section.

B.3 Choice of Soft Response Constraint

For a given pair of constraint vectors S and N , it is possible to select the gain parameter g_n such that the array gain is maximized. Referring to the above expression, it can be seen that only the denominator is a function of g_n . Consequently, array gain can be maximized by choosing the value of g_n that minimizes the denominator. This value can be found by taking the partial derivative of the denominator expression with respect to g_n^* and setting the result equal to 0. The value of g_n that minimizes the denominator is $g_n|_{\text{opt}}$ and it is given by

$$g_n|_{\text{opt}} = g_s \frac{N^H R_{nn}^{-1} S}{S^H R_{nn}^{-1} S} = g_s N^H W_s. \quad (97)$$

Thus, the choice of g_n that minimizes impact on the array gain is the response of the unconstrained beamformer to the null direction, N : the same response that would be obtained in the absence of the constraint on N . All other values of g_n will result in a lower value of G .

By selecting g_n as a fraction of the maximizing value (97), then the gain loss can be controlled to any desired level. That is choose

$$g_n = \alpha \cdot g_s \frac{N^H R_{nn}^{-1} S}{S^H R_{nn}^{-1} S}$$

where α is an arbitrary constant. Then the array gain is

$$G = G_u \cdot \frac{[1 - \cos(\gamma)^2]}{[1 + (\alpha^2 - 2\text{Re}\{\alpha\})\cos(\gamma)^2]}.$$

The effect of linear constraints is demonstrated in Figure 27 where the ratio $G(\omega)/G_u(\omega)$ is plotted as a function of $\cos(\gamma)^2$.