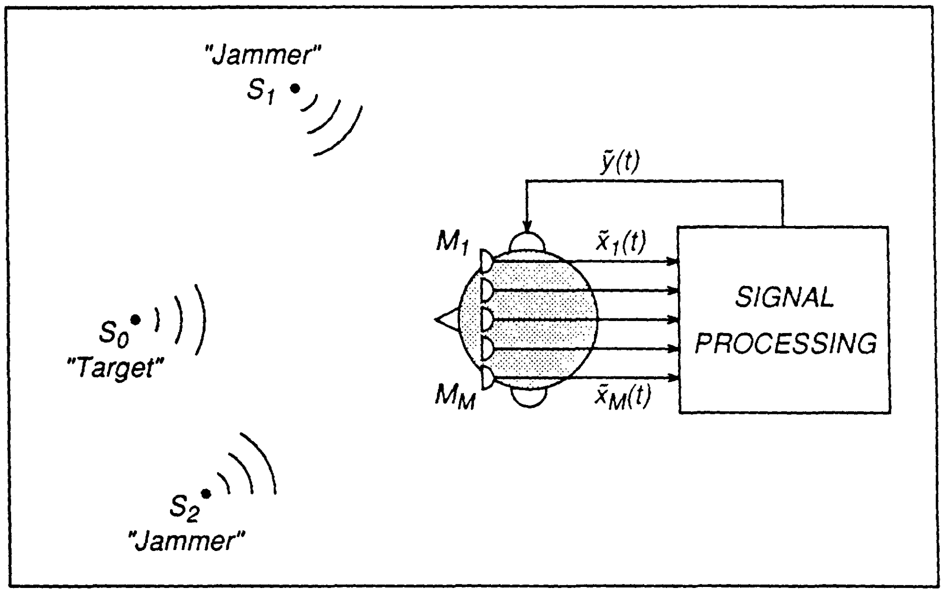
Adaptive Array Processing for Multiple Microphone Hearing Aids

RLE Technical Report No. 541

February 1989

Patrick M. Peterson



Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

Adaptive Array Processing for Multiple Microphone Hearing Aids

RLE Technical Report No. 541

February 1989

Patrick M. Peterson

**Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139 USA**

This work has been supported in part by a grant from the National Institute of Neurological and Communicative Disorders and Stroke of the National Institutes of Health (Grant RO1-NS21322).

Adaptive Array Processing for Multiple Microphone Hearing Aids

by

Patrick M. Peterson

Submitted to the Department of Electrical Engineering and
Computer Science on February 2, 1989 in partial fulfillment of the
requirements for the degree of Doctor of Science

Abstract

Hearing-aid users often have difficulty functioning in acoustic environments with many sound sources and/or substantial reverberation. It may be possible to improve hearing aids (and other sensory aids, such as cochlear implants or tactile aids for the deaf) by using multiple microphones to distinguish between spatially-separate sources of sound in a room. This thesis examines adaptive beamforming as one method for combining the signals from an array of head-mounted microphones to form one signal in which a particular sound source is emphasized relative to all other sources.

In theoretical work, bounds on the performance of adaptive beamformers are calculated for head-sized line arrays in stationary, anechoic environments with isotropic and multiple-discrete-source interference. Substantial performance gains relative to a single microphone or to conventional, non-adaptive beamforming are possible, depending on the interference, allowable sensitivity to sensor noise, array orientation, and number of microphones. Endfire orientations often outperform broadside orientations and using more that about 5 microphones in a line array does not improve performance.

In experimental work, the intelligibility of target speech is measured for a two-microphone beamformer operating in simulated environments with one interference source and different amounts of reverberation. Compared to a single microphone, beamforming improves the effective target-to-interference ratio by 30, 14, and 0 dB in anechoic, moderate, and severe reverberation. In no case does beamforming lead to worse performance than human binaural listening.

Thesis Supervisor: Nathaniel Durlach
Title: Senior Scientist

ACKNOWLEDGEMENTS

Although I must accept full responsibility for the contents of this thesis, it is only part of a larger "Multi-Mike" collaboration among members of the MIT Sensory Communication Group, in particular, Pat Zurek, Bill Rabinowitz, Nat Durlach, Roel Hammerschlag, Su-Min Wei, and Julie Greenberg. I use the first-person plural throughout the thesis to acknowledge the innumerable contributions of these coworkers. Nat Durlach supervised the thesis and William Siebert, Bruce Musicus, and Arthur Baggeroer served as readers. Most of the work was supported under grant R01-NS21322 from the National Institute of Neurological and Communicative Disorders and Stroke. On a personal level, I would also like to acknowledge the help and encouragement of my family and all my friends in the Sensory Communication Group and elsewhere (You know who you are!).

Contents

Chapter 1

Introduction

1.1 A Deficiency in Monaural Hearing-Aids

Hearing-impaired listeners using monaural hearing aids often have difficulty understanding speech in noisy and/or reverberant environments (Gelfand and Hochberg, 1976; Nabelek, 1982). In these situations, the fact that normal listeners have less difficulty, due to a phenomenon of two-eared listening known as the "cocktail-party effect" (Koenig, 1950; Kock, 1950; Moncur and Dirks, 1967; MacKeith and Coles, 1971; Plomp, 1976), indicates that impaired listeners might do better with aids on both ears (Hirsh, 1950). Unfortunately, this strategy doesn't always work, possibly because hearing impairments can degrade binaural as well as monaural abilities (Jerger and Dirks, 1961; Markides, 1977; Siegenthaler, 1979). Furthermore, it is often impossible to provide binaural aid, as in the case of a person with no hearing at all in one ear, or in the case of persons with tactile aids or cochlear implants, where sensory limitations, cost, or risk preclude binaural application. All of these effectively-monaural listeners find themselves at a disadvantage in understanding speech in poor acoustic environments. A single output hearing aid that enhanced "desired" signals in such environments would be quite useful to these impaired listeners.

Such an aid could be built with a single microphone input if a method were available for recovering a desired speech signal from a composite signal containing interfering speech or noise. Although much research has been devoted to such single-channel speech enhancement systems (Frazier, Samsam, Braida and Oppenheim, 1976; Boll, 1979; Lim and Oppenheim, 1979), no system has been found effective in increasing speech intelligibility (Lim, 1983) [1]. Fundamentally, single-microphone

[1] Recent adaptive systems proposed for single-channel hearing aids apply adaptive linear bandpass

systems cannot provide the direction-of-arrival information which two-eared listeners use to discriminate among multiple talkers in noisy environments (Dirks and Moncur, 1967; Dirks and Wilson, 1969; Blauert, 1983).

1.2 A Strategy for Improvement

This thesis will explore the strategy of using multiple microphones and adaptive combination methods to construct adaptive multiple-microphone monaural aids (AMMAs). For spatially-separated sound sources in rooms, multiple microphones in a spatially-extended array will often receive signals with intermicrophone differences that can be exploited to enhance a desired signal for monaural presentation. The effectiveness of this "spatial-diversity" strategy is indicated by the fact that it is employed by almost all natural sensory systems. In human binaural hearing, the previously mentioned cocktail-party effect is among the advantages afforded by spatial diversity (Durlach and Colburn, 1978).

To truly duplicate the abilities of the normal binaural hearing system, a monaural aid should enable the listener to concentrate on a selected source while monitoring, more or less independently, sources from other spatial positions (Durlach and Colburn, 1978). In principle, these abilities could be provided by first resolving spatially-separate signal sources and then appropriately coding the separated information into one monaural signal. While other researchers (Corbett, 1986; Durlach, Corbett, McConnell, et al., 1987) investigate the coding problem, this thesis will concentrate on the signal separation problem. The immediate goal is a processor that enhances a signal from one particular direction (straight-ahead, for example).

filtering to the composite signal by either modifying the relative levels of a few different frequency bands (Graupe, Grosspietsch and Basseas, 1987), or by modifying the cutoff-frequency of a high-pass filter (Ono, Kanzaki and Mizoi, 1983). As we will discuss in more detail later, speech intelligibility depends primarily on speech-to-noise ratio in third-octave-wide bands, with slight degradations due to "masking" when noise in one band is much louder than speech in an adjacent band. Since the proposed adaptive filtering systems cannot alter the within-band speech-to-noise ratio, we would expect no intelligibility improvement except in the case of noises with pronounced spectral peaks. Careful evaluations of these systems (Van Tassell, Larsen and Fabry, 1988; Neuman and Schwander, 1987) confirm our expectations about intelligibility. Of course, hearing-aid users also consider factors beyond intelligibility, such as comfort, that may well be improved with adaptive filtering.

Many of these processors, each enhancing signals from different directions and all operating simultaneously, would form a complete signal separation system for the ultimate AMMA. However, even a single, straight-ahead directional processor could provide useful monaural aid in many difficult listening situations.

The existence of the cocktail party effect indicates that information from multiple, spatially-separated acoustic receivers can increase a selected source's intelligibility. Unfortunately, we do not understand human binaural processing well enough to duplicate its methods of enhancing desired speech signals. Certain phenomena, such as the precedence effect (Zurek, 1980), indicate that this enhancement involves non-linear processing, which can be difficult to analyze and may not be easy to synthesize.

Linear processing, on the other hand, which simply weights and combines signals from a receiver array, can be easily synthesized to optimize a mathematical performance criterion, such as signal-to-noise ratio (SNR). Although linear processing may ultimately prove inferior to some as-yet-unknown non-linear scheme, and although improving SNR does not necessarily improve intelligibility (Lim and Oppenheim, 1979), the existence of a well-defined mathematical framework has encouraged research and generated substantial insight into linear array processing techniques. Techniques based on antenna theory (Elliott, 1981) can be used to design fixed weightings that have unity gain in a desired direction with minimum average gain in all other directions. However, if we restrict microphone placement (for cosmetic or practical reasons) tɔ locations on a human head, then the array size will be small relative to acoustic wavelengths and overall improvements will be limited. On the other hand, adaptive techniques developed in radar, sonar, and geophysics can provide much better performance by modifying array weights in response to the actual interference environment (Monzingo and Miller, 1980). The existence of a substantial literature and the success of adaptive arrays in other applications make them an attractive approach to developing AMMAs.

To date, only a few attempts have been made to apply adaptive array techniques to the hearing aid problem (Christiansen, Chabries and Lynn, 1982; Brey and Robinette, 1982; Foss, 1984). These attempts have generally proven inconclusive, either because they used unrealistic microphone placements (i.e., near the sound sources) or because they used real-time hardware that severely limited performance. In no case has the potential problem of reverberation been addressed and there has been no effort to compare alternative methods.

This thesis will focus on determining the applicability of adaptive array processing methods to hearing aids and to the signal separation problem in particular. We will not look at fixed array weighting systems or non-linear processing schemes, although other researchers should not overlook these alternate approaches. Furthermore, we will not directly address the issue of practicality. Our primary goal is to determine the potential benefits of adaptive array processing in the hearing aid environment, independent of the practicality of realizing these benefits with current technology.

We will determine the potential of array processing both theoretically and experimentally. Theoretical limits on signal-to-noise ratio (SNR) improvement can be calculated for particular environments and array geometries independent of the processing algorithm. Experimentally, specific algorithms can be implemented and actual improvements in SNR and intelligibility can then be measured and related to the theoretical limits.

The effects of reverberation will be investigated empirically by measuring performance in simulated reverberant environments with precisely known and modifiable characteristics. We will not include the effects of head-mounting on our microphone arrays. This makes the theoretical analysis tractable and reflects our intuition that, while the amplitude and phase effects introduced by the head are substantial and may change the magnitude of our calculated limits, they will not alter the general pattern of results.

Research areas which might benefit from this work include: hearing aids and sensory substitution aids, human binaural hearing, human speech perception in reverberant environments, and adaptive array signal processing.

We believe that the combination of theoretical and experimental approaches to the AMMA problem is especially significant, and that the experimental work reported in this thesis is at least equal in importance to the theoretical work. The simulation methods and experiments of Chapters 5 and 6 may have been presented in less depth only because they are the subjects of previous publications (Peterson, 1986; Peterson, 1987; Peterson, Durlach, Rabinowitz and Zurek, 1987).

Chapter 2

Background

Before describing adaptive multi-microphone hearing aids (AMMAs), we will discuss three separate background topics: (1) performance of non-impaired listeners in noisy, reverberant environments, (2) principles of operation and capabilities of currently-available multi-microphone aids, and (3) principles governing multi-microphone aids based on linear combination, whether those aids are adaptive or not.

2.1 Human Speech Reception in Rooms

The speech reception abilities of human binaural listeners provide at least three valuable perspectives on AMMAs. Firstly, the fact that listening with two ears helps humans to ignore interference demonstrates that multiple acoustic inputs can be useful in interference reduction. Secondly, the degree to which humans reject interference provides a point of comparison in evaluating AMMA performance. Finally, knowing something about how the human system works may be useful in designing AMMAs.

Considerable data are available on the ability of human listeners to understand a target speaker located straight-ahead in the presence of an interference source, or jammer, at various azimuth angles in anechoic environments (Dirks and Wilson, 1969; Tonning, 1971; Plomp, 1976; Plomp and Mimpen, 1981). Zurek has constructed a model of human performance in such situations that is consistent with much of the available experimental data (Zurek, 1988 (in revision)). The model is based primarily on two phenomena: head shadow and binaural unmasking (Durlach and Colburn, 1978).

Figure 2.1 shows the predicted target-to-jammer power ratio (TJR) required to

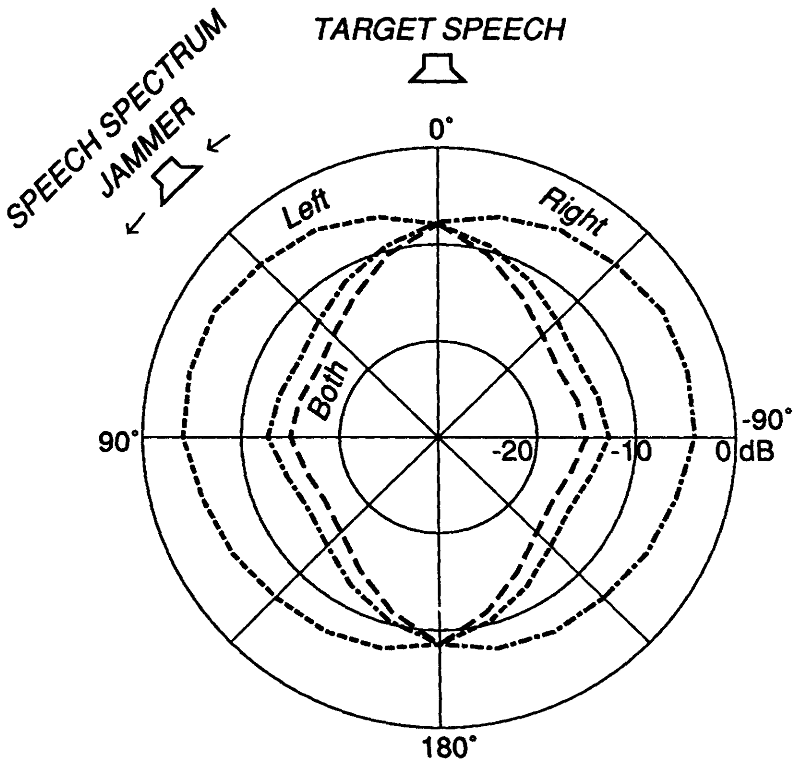


Figure 2.1: Human sensitivity to interference in an anechoic environment as predicted by Zurek's model. Target-to-Jammer ratio (TJR) needed for constant intelligibility is plotted as a function of interference angle and listening condition (left, right, or both ears).

maintain constant target intelligibility as a function of interference azimuth for three different listening conditions: right ear only, left ear only, and two-eared listening. (Better performance corresponds to smaller target-to-jammer ratio.) At 0° interference azimuth, the interference and target signal coincide and monaural and binaural listening are equivalent[1]. At any other interference angle, one ear will give better monaural performance than the other due to the head's shadowing of the jammer. At 90° for instance, the right ear picks up less interference than the left ear and thus performs better. Simply choosing the better ear would enable a listener to perform

[1] Although this equivalence may seem necessary, there is some experimental evidence (MacKeith and Coles, 1971; Gelfand and Hochberg, 1976; Plomp, 1976) that binaural listening can be advantageous for coincident signal and interference.

at the minimum of the left and right ear curves. For a particular interference angle, the difference between these curves, which Zurek calls the ***head-shadow advantage***, can be as great as 10 dB. The additional performance improvement represented by the binaural curve, called the ***binaural-interaction advantage***, comes from binaural unmasking and amounts to 3 dB at most in this particular situation. The maximum interference rejection occurs for a jammer at 120° and amounts to about 9 db relative to the rejection of a jammer at 0°. It should be emphasized that even a 3 dB increase in effective TJR can dramatically improve speech reception since the relationship between speech intelligibility and TJR can be very steep (Kalikow, Stevens and Elliott, 1977).

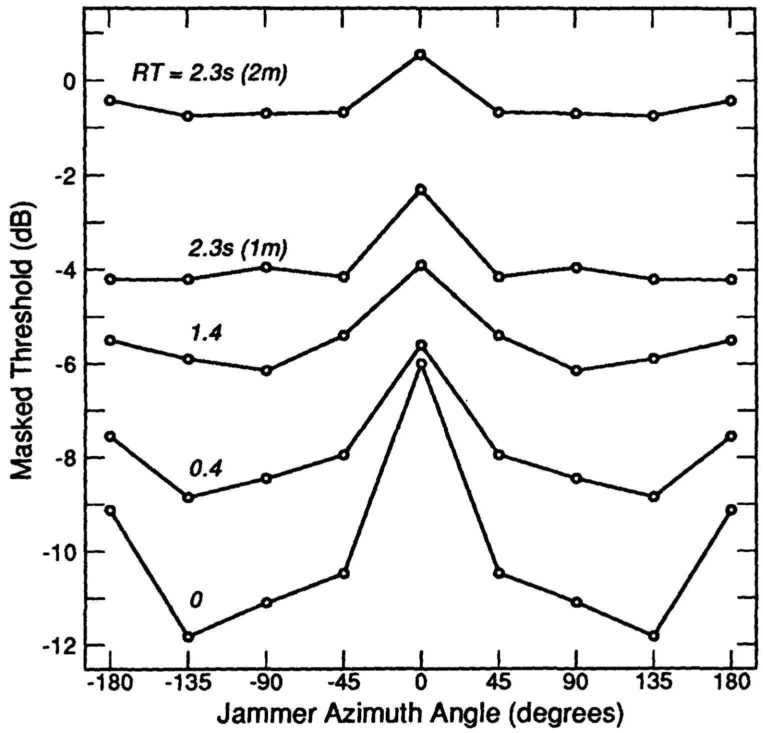


Figure 2.2: Plomp's measurements of human sensitivity to a single jammer as a function of jammer azimuth and reverberation time RT.

For reverberant environments, there is no comparable intelligibility model and experimental measurements are fewer and more complex (Nabelek and Pickett, 1974; Gelfand and Hochberg, 1976; Plomp, 1976). Plomp measured the intelligibility threshold of a target source at 0° azimuth as a function of reverberation time and azimuth of a single competing jammer. Figure 2.2 summarizes his data on TJR at the threshold of intelligibility for binaural listening and shows that the maximum interference rejection relative to coincident target and jammer drops to less than 2 dB for long reverberation times.

Plomp's data also indicate that, even for coincident target and jammer, intelligibility decreases as reverberation time increases. This effect could be explained if target signal arriving via reverberant paths acted as interference. Lochner and Burger (Lochner and Burger, 1961) and Santon (Santon, 1976) have studied this phenomenon in detail and conclude that target echoes arriving within a critical time of about 20 milliseconds may actually enhance intelligibility while late-arriving echoes do act as interference.

2.2 Present Multi-microphone Aids

Although it seems desirable that the performance of an AMMA approach that of a human binaural listener (or, perhaps, even exceed it), to be significant such an aid need only exceed the performance of presently available multiple-input hearing aids. These devices fall into two categories: true multimicrophone monaural aids (often called CROS-type aids) and directional-microphone aids.

The many CROS-type aids (Harford and Dodds, 1974) are all loosely based on the idea of sampling the acoustic environment at multiple locations (usually at the two ears) and presenting this information to the user's one good ear. In particular, the aid called BICROS combines two microphone signals (by addition) into one monaural signal. Unfortunately, there are no normal performance data on BICROS or any other CROS-type aid. Studies with impaired listeners have shown that CROS aids can improve speech reception for some interference configurations but may also decrease performance in other situations (Lotterman, Kasten and Revoile, 1968; Markides, 1977). Overall, performance has not improved. For this reason, commercial aids are sometimes equipped with a user-operated switch to control the combination of microphone signals (MULTI-CROS). An AMMA which achieved any automatic interference reduction would represent an improvement over all but the user-operated CROS-type aids.

Directional microphones also sample the sound field at multiple points, but the sample points are very closely spaced and the signals are combined acoustically rather than electrically. These microphones can be considered non-adaptive arrays and can be analyzed using the same principles from antenna theory (Baggeroer, 1976; Elliott, 1981) that are covered more fully in the next section. In particular, they depend on "superdirective" weighting schemes to achieve directivity superior to simple omnidirectional microphones. Ultimately, the sensitivity of superdirective arrays to weighting errors and electronic noise limit the extent to which directional microphones can emphasize on-axis relative to off-axis signals (Newman and Shrote, 1982). This emphasis is perhaps 10 dB at most for particular angles and about 3 dB averaged over all angles (Knowles Electronics, 1980). Nonetheless, directional-microphones seem to be successful additions to hearing aids (Madison and Hawkins, 1983; Mueller, Grimes and Erdman, 1983) and an AMMA should perform at least as well to be considered an improvement.

Recent work on higher-order directional microphones that use more sensing elements indicates that overall gains of 8.5 dB may be practical without excessive noise sensitivity (Rabinowitz, Frost and Peterson, 1985). Clearly, as improvements are made to directional microphones, the minimum acceptable AMMA performance will increase.

2.3 Multi-microphone Arrays

To describe the design and operation of multiple-microphone arrays, whether adaptive or not, we will need some mathematical notation and a few basic concepts. Figure 2.3 shows a generic multi-microphone monaural hearing aid in a typical listening situation. The listener is in a room with one *target* or desired sound source, labelled S_0, and J *jammers* or interfering sound sources, S_1 through S_J. For the example in the figure, $J = 2$. The sound from each source travels through the

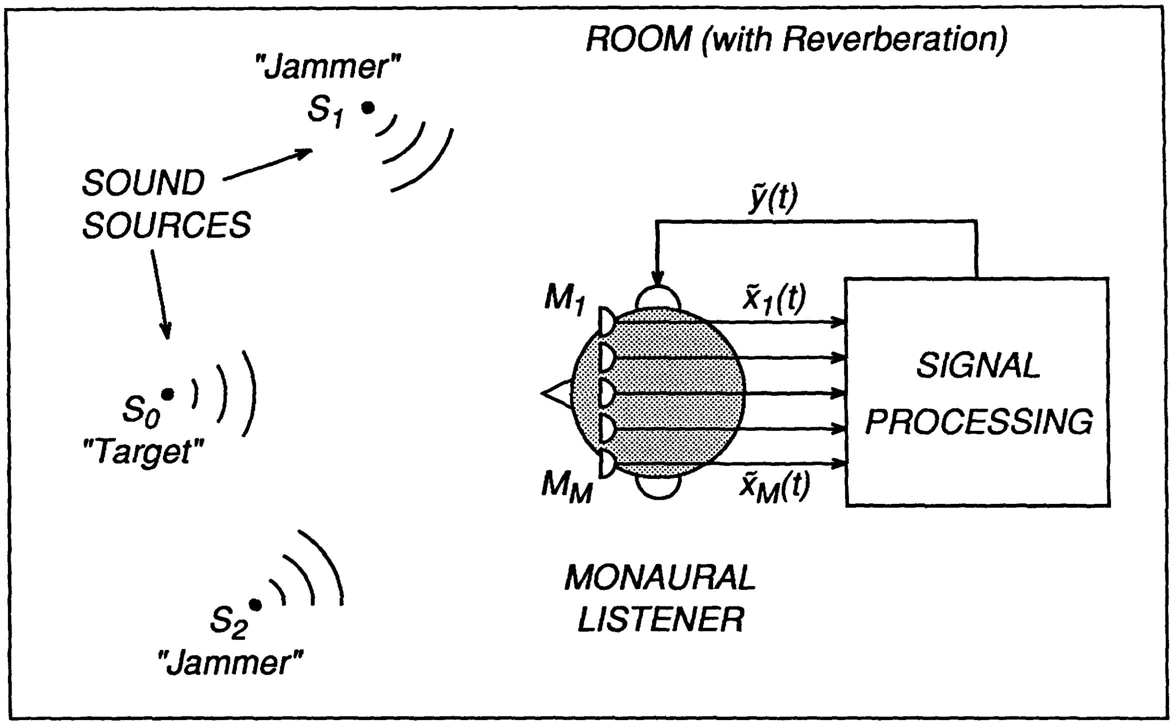


Figure 2.3: The generic multiple microphone hearing aid.

room to each of M microphones, M_1 through M_M, that are mounted somewhere on the listener's head, not necessarily in a straight line. The multiple microphone signals are then processed to form one output signal that is presented to the listener.

2.3.1 Received Signals

Let the continuous-time signal from S_j be $\tilde{s}_j(t)$, the room impulse response from S_j to microphone M_m be $\tilde{h}_{mj}(t)$, and the sensor noise at microphone M_m be $\tilde{u}_m(t)$. Then the received signal, $\tilde{x}_m(t)$, at microphone M_m is

$$\begin{aligned} \tilde{x}_m(t) &= \sum_{j=0}^{J} \tilde{h}_{mj}(t) \otimes \tilde{s}_j(t) + \tilde{u}_m(t) \qquad (2.1) \\ &= \sum_{j=0}^{J} \int_{-\infty}^{\infty} \tilde{h}_{mj}(\tau)\, \tilde{s}_j(t-\tau)\, d\tau + \tilde{u}_m(t) \end{aligned}$$

where $\otimes$ denotes continuous-time convolution. If we define the Fourier transform of a continuous-time signal $\tilde{s}(t)$ as

$$\tilde{\mathcal{S}}(f) = \int_{-\infty}^{\infty} \tilde{s}(t)\, e^{-j2\pi f t}\, dt$$

then the frequency domain equivalent of Equation (2.1) can be written

$$\tilde{\mathcal{X}}_m(f) = \sum_{j=0}^{J} \tilde{\mathcal{H}}_{mj}(f)\, \tilde{\mathcal{S}}_j(f) + \tilde{\mathcal{U}}_m(f) \tag{2.2}$$

The signal processing schemes that we will consider are all ***sampled-data, digital, linear*** systems. Thus, the microphone signals will always be passed through anti-aliasing low-pass filters and sampled periodically with sampling period T_s. The resulting discrete-time signal, or sequence of samples, from microphone m will be defined by

$$x_m[n] \triangleq T_s\, \tilde{x}_m(nT_s)\ , \tag{2.3}$$

where we use brackets for the index of a discrete-time sequence and parentheses for the argument of a continuous-time signal. The scaling factor, T_s, is necessary to preserve the correspondence between continuous- and discrete-time convolution (Oppenheim and Johnson, 1972). That is, if

$$\tilde{f}(t) = \int_{-\infty}^{\infty} \tilde{g}(\tau)\, \tilde{h}(t-\tau)\, d\tau\ ,$$

if $\tilde{f}()$, $\tilde{g}()$, and $\tilde{h}()$ are all bandlimited to frequencies less than $1/2T_s$, and if we define the corresponding discrete-time sequences as in equation (2.3), then we can write

$$f[n] = \sum_{l=-\infty}^{\infty} g[l]\, h[n-l]\ .$$

This makes it possible to place all of our derivations in the discrete-time context with the knowledge that, if necessary, we can always determine the appropriate correspondence with continuous-time (physical) signals.

In particular, we can view the sampled input signal, $x_m[n]$, as arising from a multiple-input discrete-time system with impulse responses $h_{mj}[n]$ operating on discrete-time source signals, $s_j[n]$, and corrupted by sensor noise $u_m[n]$. That is, if we use $*$ to denote discrete-time convolution,

$$\begin{aligned} x_m[n] &= \sum_{j=0}^{J} h_{mj}[n] * s_j[n] + u_m[n] \\ &= \sum_{j=0}^{J} \sum_{l=-\infty}^{\infty} h_{mj}[l]\, s_j[n-l] + u_m[n]\,. \end{aligned} \tag{2.4}$$

If the corresponding continuous-time signals in (2.1) are bandlimited to frequencies less than $1/2T_s$, then

$$\begin{aligned} h_{mj}[n] &= T_s\, \tilde{h}_{mj}(nT_s)\,, \\ s_j[n] &= T_s\, \tilde{s}_j(nT_s)\,, \\ \text{and} \quad u_m[n] &= T_s\, \tilde{u}_m(nT_s)\,. \end{aligned}$$

Defining the Fourier transform of a discrete-time signal $s[n]$ as

$$\mathcal{S}(f) = \sum_{n=-\infty}^{\infty} s[n]\, e^{-j2\pi f n T_s}\,,$$

we can express equation (2.4) in the frequency domain as

$$\mathcal{X}_m(f) = \sum_{j=0}^{J} \mathcal{H}_{mj}(f)\, \mathcal{S}_j(f) + \mathcal{U}_m(f) \tag{2.5}$$

and, as long as the continuous-time signals are bandlimited to half the sampling rate,

$$\begin{aligned} \mathcal{X}_m(f) &= \tilde{\mathcal{X}}_m(f) \\ \mathcal{H}_{mj}(f) &= \tilde{\mathcal{H}}_{mj}(f) \\ \mathcal{S}_j(f) &= \tilde{\mathcal{S}}_j(f) \\ \mathcal{U}_m(f) &= \tilde{\mathcal{U}}_m(f)\,. \end{aligned}$$

Note that the bandlimited signal assumption is not very restrictive. Since the source-room-microphone system is linear and time-invariant (LTI), the order of the room and anti-aliasing filters can be reversed without altering the sampled microphone signals. Thus, low-pass filtering of the microphone signals prior to sampling is functionally equivalent to using low-pass filtered source signals. Since the room response at frequencies above $1/2T_s$ either cannot be excited or cannot be observed, we are free to assume that it is zero.

Since we will be concerned exclusively with estimating the target signal, $s_0[n]$, we can simplify notation by defining:

$$\begin{aligned} s[n] &\triangleq s_0[n] && (2.6) \\ v_m[n] &\triangleq \sum_{j=1}^{J} h_{mj}[n] * s_j[n] && (2.7) \\ \text{and } z_m[n] &\triangleq v_m[n] + u_m[n] \,, && (2.8) \end{aligned}$$

so that $s[n]$ is the target signal, $v_m[n]$ is the total received interference at microphone m, and $z_m[n]$ is the total noise from external and internal sources at microphone m. The received-signal equations can now be rephrased as

$$x_m[n] = h_{m0}[n] * s[n] + v_m[n] + u_m[n] = h_{m0}[n] * s[n] + z_m[n] \qquad (2.9)$$

in the time-domain or, in the frequency domain,

$$\mathcal{X}_m(f) = \mathcal{H}_{m0}(f)\,\mathcal{S}(f) + \mathcal{V}_m(f) + \mathcal{U}_m(f) = \mathcal{H}_{m0}(f)\,\mathcal{S}(f) + \mathcal{Z}_m(f) \,. \qquad (2.10)$$

The received microphone signals can be combined into one observation vector and the time-domain convolution can be expressed as a matrix multiplication, giving rise to the following matrix equation for the observations:

$$\begin{bmatrix} x_1[n] \\ x_2[n] \\ \vdots \\ x_M[n] \end{bmatrix} = \begin{bmatrix} h_{10}[0] & h_{10}[1] & h_{10}[2] & \cdots \\ h_{20}[0] & h_{20}[1] & h_{20}[2] & \cdots \\ \vdots & \vdots & \vdots & \\ h_{M0}[0] & h_{M0}[1] & h_{M0}[2] & \cdots \end{bmatrix} \begin{bmatrix} s[n] \\ s[n-1] \\ s[n-2] \\ \vdots \end{bmatrix} + \begin{bmatrix} z_1[n] \\ z_2[n] \\ \vdots \\ z_M[n] \end{bmatrix} \qquad (2.11)$$

If we use boldface to denote vectors, we can write the time-domain equation more compactly as

$$\boldsymbol{x}[n] = \begin{bmatrix} \boldsymbol{h}[0] & \boldsymbol{h}[1] & \boldsymbol{h}[2] & \cdots \end{bmatrix} \begin{bmatrix} s[n] \\ s[n-1] \\ s[n-2] \\ \vdots \end{bmatrix} + \boldsymbol{z}[n] = \boldsymbol{H}\boldsymbol{s}[n] + \boldsymbol{z}[n] \,, \qquad (2.12)$$

where $\boldsymbol{x}[n]$ is the vector of M microphone samples at sampling instant n and $\boldsymbol{z}[n]$ is the vector of M total noise values. Each $\boldsymbol{h}[i]$ is the vector of microphone responses i sample times after emission of an impulse from the desired source, and $\boldsymbol{s}[n]$ is a vector of (possibly many) past samples of the desired source signal.

The frequency-domain version of the matrix observation equation is simpler because convolution can be expressed as a multiplication of scalar functions:

$$\begin{bmatrix} \mathcal{X}_1(f) \\ \vdots \\ \mathcal{X}_M(f) \end{bmatrix} = \begin{bmatrix} \mathcal{H}_1(f) \\ \vdots \\ \mathcal{H}_M(f) \end{bmatrix} \mathcal{S}(f) + \begin{bmatrix} \mathcal{Z}_1(f) \\ \vdots \\ \mathcal{Z}_M(f) \end{bmatrix} \tag{2.13}$$

where the elements of each vector are identical to the elements of equation (2.10). Using an underscore to denote vectors in the frequency-domain, this equation can be condensed to

$$\underline{\mathcal{X}}(f) = \underline{\mathcal{H}}(f)\,\mathcal{S}(f) + \underline{\mathcal{Z}}(f)\,. \tag{2.14}$$

In developing the target estimation equations, the signals $s[n]$ (target), $v_m[n]$ (received interference), and $u_m[n]$ (sensor noise) will usually be treated as zero-mean random processes, so that signals derived from them by linear filtering, such as the received-target signal,

$$r_m[n] = \sum_{k=-\infty}^{\infty} h_{m0}[k]\,s[n-k]\,, \tag{2.15}$$

will also be zero-mean random processes. We will also usually assume that these random processes are wide-sense stationary so that, for any two such processes, p and q, we can define the *correlation function*

$$R_{pq}[k] \triangleq E\left\{p[n]\,q[n-k]\right\}\;, \tag{2.16}$$

and its Fourier transform, the *cross-spectral-density function*

$$S_{pq}(f) \triangleq \sum_{n=-\infty}^{\infty} R_{pq}[n]\,e^{-j2\pi f n T_s}\,. \tag{2.17}$$

If $p = q$, of course, these functions become the *autocorrelation* and *spectral-density* functions, respectively.

In the most general case, $\boldsymbol{p}$ and $\boldsymbol{q}$ may be complex, vector-valued random processes and have a *correlation matrix*

$$R_{\boldsymbol{pq}}[k] \triangleq E\left\{\boldsymbol{p}[n]\,\boldsymbol{q}^{\dagger}[n-k]\right\} , \tag{2.18}$$

where † indicates the complex-conjugate transpose. The related *cross-spectral-density matrix* is then

$$\mathcal{S}_{\boldsymbol{pq}}(f) \triangleq \sum_{k=-\infty}^{\infty} R_{\boldsymbol{pq}}[k]\, e^{-j2\pi fkT_s} , \tag{2.19}$$

where the elements of $\mathcal{S}_{\boldsymbol{pq}}$ are the Fourier transforms of the elements of $R_{\boldsymbol{pq}}$.

If the $\boldsymbol{p}$ and $\boldsymbol{q}$ processes are derived from a common process, say r, the correlation and spectral-density matrices can be expressed in terms of the corresponding matrices for r. If $\boldsymbol{p}$, $\boldsymbol{q}$, and r are related by the convolutions

$$\begin{aligned} \boldsymbol{p}[n] &= \boldsymbol{a}[n] * r[n] && (2.20) \\ \boldsymbol{q}[n] &= \boldsymbol{b}[n] * r[n] , && (2.21) \end{aligned}$$

then

$$\begin{aligned} R_{\boldsymbol{pq}}[k] &= E\left\{\boldsymbol{p}[n]\,\boldsymbol{q}^{T}[n-k]\right\} \\ &= E\left\{\sum_{l=-\infty}^{\infty} \boldsymbol{a}[l]\,r[n-l] \sum_{m=-\infty}^{\infty} r[n-k+m]\,\boldsymbol{b}^{T}[-m]\right\} \\ &= \sum_{m=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} \boldsymbol{a}[l]\,R_{rr}[k-l-m]\,\boldsymbol{b}^{T}[-m] \\ &= \sum_{m=-\infty}^{\infty} (\boldsymbol{a}[k-m] * R_{rr}[k-m])\,\boldsymbol{b}^{T}[-m] \\ &= \boldsymbol{a}[k] * R_{rr}[k] * \boldsymbol{b}^{T}[-k] , && (2.22) \end{aligned}$$

and

$$\mathcal{S}_{\boldsymbol{pq}}(f) = \underline{\mathcal{A}}(f)\,\mathcal{S}_{rr}(f)\,\underline{\mathcal{B}}^{\dagger}(f) . \tag{2.23}$$

An alternative form of the correlation matrix can be derived if we express convolutions (2.20) and (2.21) as matrix multiplications, in the style of (2.11) and (2.12),

$$\begin{aligned} \boldsymbol{p}[n] &= \boldsymbol{A}\,\boldsymbol{r}[n] = \begin{bmatrix} \boldsymbol{a}[0] & \boldsymbol{a}[1] & \cdots \end{bmatrix} \begin{bmatrix} r[n] \\ r[n-1] \\ \vdots \end{bmatrix} & (2.24) \\ \boldsymbol{q}[n] &= \boldsymbol{B}\,\boldsymbol{r}[n]\,. & (2.25) \end{aligned}$$

In this case,

$$R_{\boldsymbol{pq}}[k] = E\left\{\boldsymbol{A}\,\boldsymbol{r}[n]\,\boldsymbol{r}^T[n-k]\,\boldsymbol{B}^T\right\} = \boldsymbol{A}\,R_{\boldsymbol{rr}}[k]\,\boldsymbol{B}^T\ , \qquad (2.26)$$

where it should be noted that $R_{\boldsymbol{rr}}$ is a matrix function of k:

$$R_{\boldsymbol{rr}}[k] = E\left\{\begin{bmatrix} r[n] \\ r[n-1] \\ \vdots \end{bmatrix}\begin{bmatrix} r[n-k] \\ r[n-k-1] \\ \vdots \end{bmatrix}^T\right\} = \begin{bmatrix} R_{rr}[k] & R_{rr}[k+1] & \cdots \\ R_{rr}[k-1] & R_{rr}[k] & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \qquad (2.27)$$

When a derivation depends only on the value of $R_{\boldsymbol{pq}}[0]$, we will use the shortened notation $R_{\boldsymbol{pq}}$ to denote this value.

In our application we will always assume that $s[n]$, $v_m[n]$, and $u_m[n]$ are mutually uncorrelated so that

$$R_{s\boldsymbol{v}}[k] = R_{s\boldsymbol{u}}[k] = \begin{bmatrix} \vdots \\ 0 \\ \vdots \end{bmatrix} \qquad (2.28)$$

and

$$R_{\boldsymbol{vu}}[k] = \begin{bmatrix} & \vdots & \\ \cdots & 0 & \cdots \\ & \vdots & \end{bmatrix} \qquad (2.29)$$

We will also assume that the sensor noise is white, uncorrelated between microphones, with an energy per sample of σ_u^2 at each microphone. That is,

$$R_{\boldsymbol{uu}}[k] = \begin{bmatrix} \sigma_u^2 & 0 & \cdots & 0 \\ 0 & \sigma_u^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_u^2 \end{bmatrix} \delta[k] = \sigma_u^2\, I_M\, \delta[k]\ , \qquad (2.30)$$

where I_M is the $M \times M$ identity matrix and $\delta[n]$ is the discrete-time delta function,

$$\delta[n] = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.31}$$

We will usually not make any assumptions about the structure of the received interference, $\boldsymbol{v}[n]$, so that $R_{\boldsymbol{vv}}[k]$ can be arbitrary.

The target autocorrelation function, $R_{ss}[k]$, if needed, must be determined from properties of the particular source signal, either using *a priori* knowledge or by estimation. Similarly, the received-interference autocorrelation matrix, $R_{\boldsymbol{vv}}[k]$, will depend on the properties of the particular interference signals and on the propagation (room) configuration.

It is now a simple matter to specify the statistical properties of the received microphone signals. Using a vector form of (2.9), the convolutional definition of $x_m[n]$,

$$\boldsymbol{x}[n] = \boldsymbol{h}[n] * s[n] + \boldsymbol{z}[n]\,, \tag{2.32}$$

and following (2.22) and (2.23), we can determine that

$$\begin{aligned} \mathcal{S}_{\boldsymbol{xx}}(f) &= \underline{\mathcal{H}}(f)\,\mathcal{S}_{ss}(f)\,\underline{\mathcal{H}}^{\dagger}(f) + \mathcal{S}_{\boldsymbol{zz}}(f) & (2.33)\\ \mathcal{S}_{s\boldsymbol{x}}(f) &= \mathcal{S}_{ss}(f)\,\underline{\mathcal{H}}^{\dagger}(f) & (2.34)\\ \mathcal{S}_{\boldsymbol{x}s}(f) &= \underline{\mathcal{H}}(f)\,\mathcal{S}_{ss}(f)\ . & (2.35) \end{aligned}$$

Using the multiplicative definition of $\boldsymbol{x}[n]$ in (2.12),

$$\boldsymbol{x}[n] = \boldsymbol{H}\boldsymbol{s}[n] + \boldsymbol{z}[n]\,,$$

and using (2.26), it also follows that

$$\begin{aligned} R_{\boldsymbol{xx}}[k] &= \boldsymbol{H}R_{\boldsymbol{ss}}[k]\,\boldsymbol{H}^T + R_{\boldsymbol{zz}}[k] & (2.36)\\ R_{s\boldsymbol{x}}[k] &= R_{\boldsymbol{s}s}[k]\,\boldsymbol{H}^T & (2.37)\\ R_{\boldsymbol{x}s}[k] &= \boldsymbol{H}R_{\boldsymbol{s}s}[k]\,. & (2.38) \end{aligned}$$

When necessary, we can express the total noise statistics in terms of received interference and sensor noise:

$$\begin{aligned} R_{\boldsymbol{zz}}[k] &= R_{\boldsymbol{vv}}[k] + \sigma_u^2\, I_M\,\delta[k] & (2.39)\\ \mathcal{S}_{\boldsymbol{zz}}(f) &= \mathcal{S}_{\boldsymbol{vv}}(f) + \sigma_u^2\, I_M\ . & (2.40) \end{aligned}$$

2.3.2 Signal Processing

After the received signals have been sampled, they are converted from analog to digital form for subsequent digital processing[2]. The processing schemes under consideration form output samples, $y[n]$, by weighting and combining a finite number of present and past input samples. If the weights are fixed, the processing amounts to LTI FIR (linear-time-invariant finite-impulse-response) filtering. In our case, however, the weights are adaptive and depend on the input and/or output signals. Strictly speaking, then, the processing will be neither linear, time-invariant, nor even finite-impulse-response (when the weights depend on the output samples). If the adaptation is slow enough, however, the system will be *almost* LTI FIR over short intervals. After the output samples are computed, they are converted from digital to analog form and passed through a low-pass reconstruction filter that produces $\tilde{y}(t)$ for presentation to the listener.

There are at least two ways, shown in Figures 2.4 and 2.5, to view the operation of the digital processing section. Figure 2.4 shows the processing in full detail. Each discrete-time microphone signal passes through a string of $L-1$ unit delays, making the L most recent input values available for processing. The complete set of ML values are multiplied by individual weights, $w_m[l]$ (where $m = 1 \ldots M$ and $l = 0 \ldots L-1$), and added together to form the output $y[n]$. This processing can be expressed in algebraic terms by the equation

$$y[n] = \mathbf{w}^T\mathbf{x}[n] = \begin{bmatrix} \boldsymbol{w}^T[0] & \boldsymbol{w}^T[1] & \cdots & \boldsymbol{w}^T[L-1] \end{bmatrix} \begin{bmatrix} \boldsymbol{x}[n] \\ \boldsymbol{x}[n-1] \\ \vdots \\ \boldsymbol{x}[n-L+1] \end{bmatrix}, \qquad (2.41)$$

where $\boldsymbol{w}[l]$ is the vector of weights at delay l, and $\boldsymbol{x}[n]$, as defined in the previous section (equation (2.12)), is the vector of sampled microphone signals at time index

[2] Although this conversion process introduces quantization errors, we will usually assume that the errors are small enough to ignore and use Equation (2.3) to describe both digital and analog samples.

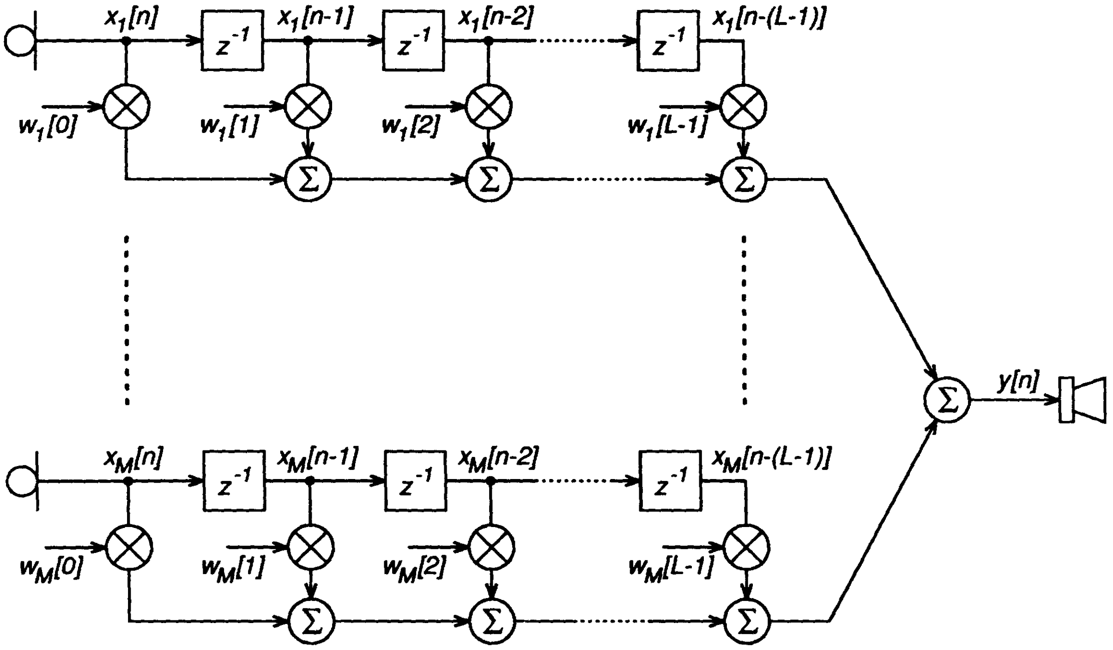


Figure 2.4: Detailed view of signal processing operations.

n. Specifically,

$$\boldsymbol{w}[l] = \begin{bmatrix} w_1[l] \\ w_2[l] \\ \vdots \\ w_M[l] \end{bmatrix} \quad \text{and} \quad \boldsymbol{x}[n] = \begin{bmatrix} x_1[n] \\ x_2[n] \\ \vdots \\ x_M[n] \end{bmatrix} .$$

If the weights, **w**, are adaptive, they will, of course, depend on the time index, n. We have not expressed this dependence in our notation because changes in **w** are normally orders of magnitude slower than changes in **x** and, therefore, the weights comprise a quasi-LTI system over short intervals. The notation was chosen to emphasize the interpretation of the weighting vector as a filter.

If we consider the weights for each channel as a filter, then we can view the processing more abstractly, as shown in Figure 2.5, where each microphone signal passes through a filter with impulse response $w_m[n]$. In this view, the output is

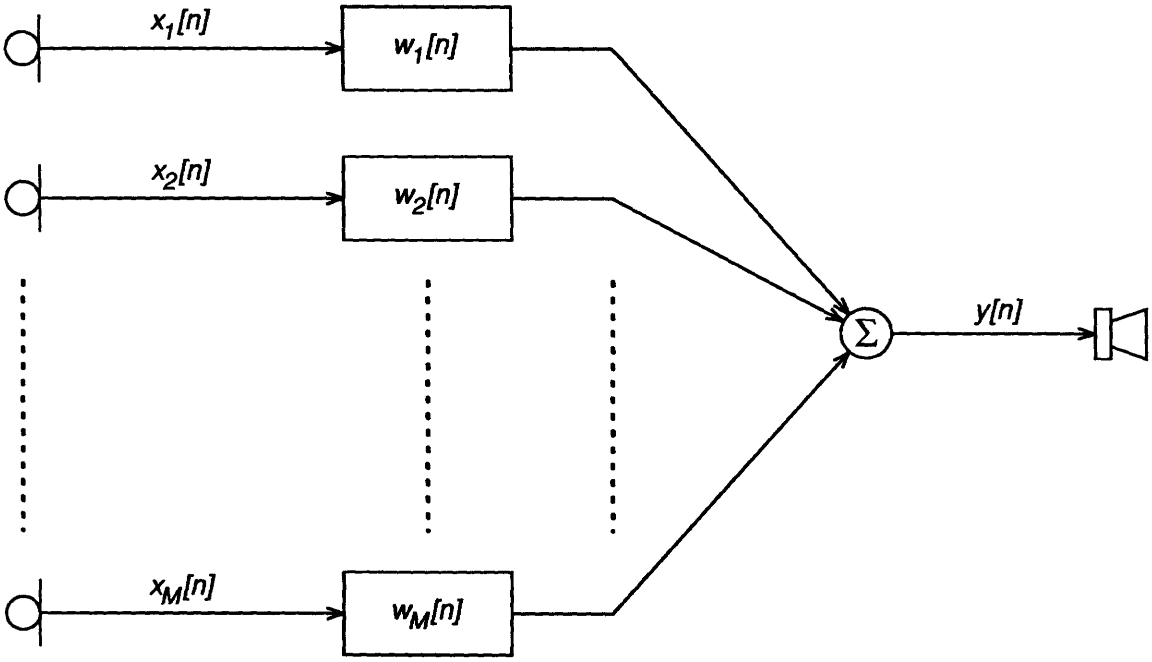


Figure 2.5: Filtering view of signal processing operations.

simply the sum of M filtered input signals,

$$\begin{aligned} y[n] &= \sum_{m=1}^{M} w_m[n] * x_m[n] & (2.42) \\ &= \sum_{m=1}^{M} \sum_{l=0}^{L-1} w_m[l]\, x_m[n-l] \\ &= \sum_{l=0}^{L-1} \sum_{m=1}^{M} w_m[l]\, x_m[n-l] \\ &= \sum_{l=0}^{L-1} \boldsymbol{w}^T[l]\, \boldsymbol{x}[n-l] \\ &= \boldsymbol{w}^T[n] * \boldsymbol{x}[n] & (2.43) \end{aligned}$$

or, in the frequency domain,

$$\mathcal{Y}(f) = \sum_{m=1}^{M} \mathcal{W}_m(f)\, \mathcal{X}_m(f) = \underline{\mathcal{W}}^T(f)\, \underline{\mathcal{X}}(f)\,. \qquad (2.44)$$

If the $x_m[n]$ are stationary random processes, we can use (2.43) and, following (2.22) and (2.23), determine the output autocorrelation and spectral-density functions:

$$\begin{aligned} R_{yy}[k] &= \boldsymbol{w}^T[k] * R_{\boldsymbol{xx}}[k] * \boldsymbol{w}[-k] && (2.45) \\ \mathcal{S}_{yy}(f) &= \underline{\mathcal{W}}^T(f)\,\mathcal{S}_{\boldsymbol{xx}}(f)\,\underline{\mathcal{W}}^*(f)\,. && (2.46) \end{aligned}$$

We can also use (2.41) and (2.26) to derive the multiplicative form of the output autocorrelation function:

$$\begin{aligned} R_{yy}[k] &= E\left\{y[n]\,y[n-k]\right\} \\ &= E\left\{\mathbf{w}^T\,\mathbf{x}[n]\,\mathbf{x}^T[n-k]\,\mathbf{w}\right\} \\ &= \mathbf{w}^T R_{\mathbf{xx}}[k]\,\mathbf{w}\,, && (2.47) \end{aligned}$$

where $R_{\mathbf{xx}}[k]$ is an $ML \times ML$ matrix of correlations among all the delayed microphone samples in the array, which can be expressed in terms of the $M \times M$ correlation matrix $R_{\boldsymbol{xx}}[k]$.

$$R_{\mathbf{xx}}[k] = E\left\{\begin{bmatrix} \boldsymbol{x}[n] \\ \boldsymbol{x}[n-1] \\ \vdots \end{bmatrix}\begin{bmatrix} \boldsymbol{x}[n-k] \\ \boldsymbol{x}[n-k-1] \\ \vdots \end{bmatrix}^T\right\} = \begin{bmatrix} R_{\boldsymbol{xx}}[k] & R_{\boldsymbol{xx}}[k+1] & \cdots \\ R_{\boldsymbol{xx}}[k-1] & R_{\boldsymbol{xx}}[k] & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{2.48}$$

$R_{\mathbf{xx}}[k]$ can also be expressed in terms of target and interference statistics. The vector of ML array observations, $\mathbf{x}$, can be modelled by extending (2.12):

$$\begin{aligned} \mathbf{x}[n] &= \begin{bmatrix} \boldsymbol{x}[n] \\ \boldsymbol{x}[n-1] \\ \vdots \\ \boldsymbol{x}[n-L+1] \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{h}[0] & \boldsymbol{h}[1] & \cdots & \cdots & \cdots \\ 0 & \boldsymbol{h}[0] & \cdots & \cdots & \cdots \\ \vdots & & \ddots & & \\ 0 & 0 & \cdots & \boldsymbol{h}[0] & \cdots \end{bmatrix}\begin{bmatrix} s[n] \\ s[n-1] \\ \vdots \\ s[n-L+1] \\ \vdots \end{bmatrix} + \begin{bmatrix} \boldsymbol{z}[n] \\ \boldsymbol{z}[n-1] \\ \vdots \\ \boldsymbol{z}[n-L+1] \end{bmatrix} \\ &= \mathbf{H}\,\mathbf{s}[n] + \mathbf{z}[n]\,. && (2.49) \end{aligned}$$

Using this model,

$$R_{\mathbf{xx}}[k] = E\left\{\mathbf{x}[n]\mathbf{x}^T[n-k]\right\} = \mathbf{H}\,R_{\mathbf{ss}}[k]\,\mathbf{H}^T + R_{\mathbf{zz}}[k]\,, \tag{2.50}$$

where $R_{\mathbf{ss}}[k]$ and $R_{\mathbf{zz}}[k]$ are extended versions of $R_{\boldsymbol{ss}}[k]$ and $R_{\boldsymbol{zz}}[k]$.

2.3.3 Response Measures

Once an array processor (a set of microphone locations and weights) has been specified, we can evaluate the response of that processor in at least two ways. The array ***directional response***, or sensitivity to plane-waves as a function of arrival direction, can be determined from the specification of the array processor alone. When we know, in addition, the statistics of a particular signal or noise field, we can determine the overall ***signal-*** or ***noise-field response*** of the array for that specific field.

Directional Response

An array's directional response can be defined as the ratio of the array processor's output to that of a nearby reference microphone as a function of the direction of a distant test source that generates the equivalent of a plane wave in the vicinity of the array. We will assume that our arrays are mounted in free space with no head present and that the microphones are omnidirectional, and small enough not to disturb the sound field[3]. We will also assume that the microphones have poor enough coupling to the field (due to small size and high acoustic impedance) that inter-microphone loading effects are negligible.

Let the location of microphone m be $\vec{r}_m$, its three-dimensional coordinate vector relative to a common array origin; let $\vec{\alpha}$ be a unit vector in the direction of signal propagation; let c be the velocity of propagation; and let $\mathcal{S}_T(f)$ represent the test signal as measured by a reference microphone at the array origin. Then the

[3] The presence of a head or of microphone scattering will introduce direction- and frequency-dependent amplitude and phase differences from the simplified plane-wave field that we have assumed. The directional response is then harder to calculate and dependent on the specific head and/or microphone configuration.

amplitude and phase of the signal at microphone m will be given by

$$\mathcal{X}_m(f,\vec{\alpha}) = \mathcal{S}_T(f)e^{-j2\pi f\tau_m(\vec{\alpha})} = \mathcal{S}_T(f)e^{-j2\pi f\frac{\vec{\alpha}\cdot\vec{r}_m}{c}} , \tag{2.51}$$

where $\tau_m(\vec{\alpha}) = \vec{\alpha}\cdot\vec{r}_m/c$ represents the relative delay in signal arrival at microphone m. The array output for the test signal is then

$$\mathcal{Y}_T(f,\vec{\alpha}) = \sum_{m=1}^{M} \mathcal{W}_m(f)\mathcal{X}_m(f,\vec{\alpha}) = \sum_{m=1}^{M} \mathcal{W}_m(f)\mathcal{S}_T(f)e^{-j2\pi f\tau_m(\vec{\alpha})} \tag{2.52}$$

and the array's directional response (sometimes called the *array factor*) is given by

$$\mathcal{G}(f,\vec{\alpha}) = \frac{\mathcal{Y}_T(f,\vec{\alpha})}{\mathcal{S}_T(f)} = \sum_{m=1}^{M} \mathcal{W}_m(f)e^{-j2\pi f\tau_m(\vec{\alpha})} . \tag{2.53}$$

Since $\vec{\alpha}$ can be expressed in terms of azimuth angle, θ, and elevation angle, ϕ, we can also write the directional response as $\mathcal{G}(f,\theta,\phi)$.

An array's directional response is often described by considering only sources in the horizontal plane of the array and plotting the magnitude of $\mathcal{G}(f,\theta,0)$ at a particular frequency f as a function of arrival angle θ. To illustrate the utility of such *beam patterns*, Figure 2.6 shows patterns for an endfire array (whose elements are lined up in the target direction, 0°) of 21 elements spaced 3 cm apart for a total length of 60 cm, or about 2 feet. The processor that gave rise to these patterns, a delay-and-sum beamformer, delayed the microphone signals to make target waveforms coincident in time and then summed all microphones with identical weights. That is, for a delay-and-sum beamformer,

$$\mathcal{W}_m(f) = \frac{1}{M}e^{j2\pi f\tau_m(0^\circ)} . \tag{2.54}$$

The single-frequency beam patterns of Figure 2.6 (a), (b), and (c) illustrate the fact that delay-and-sum beam patterns become more "directive" (preferentially sensitive to arrivals from 0°) at higher frequencies. In quantitative terms, the 3 dB response beamwidth (Elliott, 1981, page 150) varies from about 160° at 250 Hz to 76° at 1 KHz to 38° at 4 KHz. Alternatively, directivity can be characterized by the *directivity factor* or *directivity index*, D, defined as the ratio of the response power

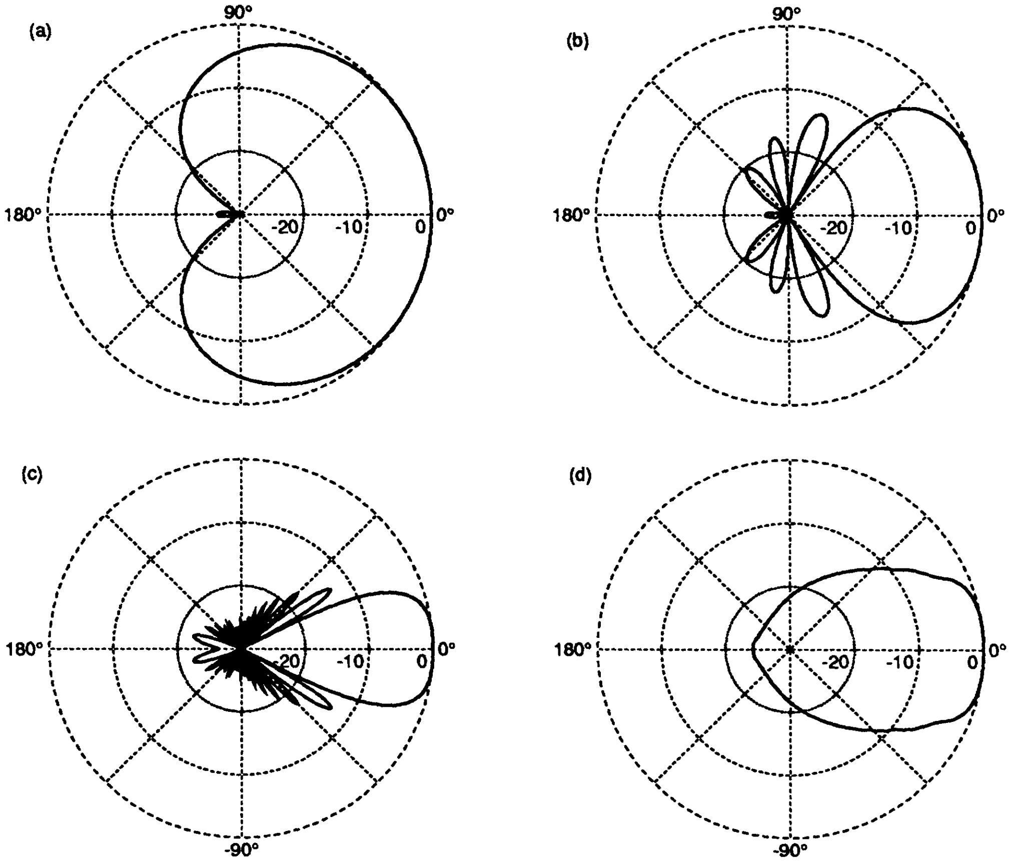


Figure 2.6: Beam patterns for a 21-element, 60 cm endfire array of equispaced microphones with delay-and-sum beamforming. Patterns are shown for (a) 250 Hz, (b) 1000 Hz, (c) 4000 Hz, and (d) the "intelligibility-weighted" average of the response at 257 frequencies spaced uniformly from 0 through 5000 Hz. Radial scale is in decibels.

at 0° to the average response power over all spherical angles (Schelkunoff, 1943; Elliott, 1981):

$$D(f) = \frac{|\mathcal{G}(f,0,0)|^2}{\frac{1}{4\pi}\iint |\mathcal{G}(f,\theta,\phi)|^2 \, d\theta \, d\phi} . \tag{2.55}$$

For our 21-element array, we can use an equation for the directivity of a uniformly-weighted, evenly-spaced, endfire array (Schelkunoff, 1943, page 107), to calculate the directivities of patterns (a), (b), and (c) as 3.7, 9.3, and 16 dB, respectively.

The final pattern in Figure 2.6 presents a measure of the array's broadband directional response, the "intelligibility-average" across frequency of the array's directional response function for sources in the horizontal plane. This average is designed to reflect the net effect of a particular frequency response on speech intelligibility and can be calculated as

$$\langle \mathcal{G}(\theta) \rangle_I = \int_0^{1/2T_s} W_{AI}(f)\, 20 \log_{10} \mathrm{rms}_{1/3}(|\mathcal{G}(f,\theta)|)\, df \, . \tag{2.56}$$

The function $\mathrm{rms}_{1/3}()$ smooths a magnitude spectrum by averaging the power in a third-octave band around each frequency and is defined as

$$\mathrm{rms}_{1/3}(|H(f)|) = \sqrt{\frac{\int_{2^{-1/6}f}^{2^{1/6}f} |H(\nu)|^2 \, d\nu}{(2^{1/6} - 2^{-1/6})f}} \, . \tag{2.57}$$

This smoothing reflects the fact that, in human hearing, sound seems to be analyzed in one-third-octave-wide frequency bands, within which individual components are averaged together[4]. The smoothed magnitude response is then converted to decibels to reflect the ear's logarithmic sensitivity to the sound level in a band. Next, the smoothed frequency-response in decibels is multiplied by a weighting function, $W_{AI}(f)$, that reflects the relative importance of different frequencies to speech intelligibility. The weighting function is normalized to have an integral of 1.0 and is based on results from Articulation Theory (French and Steinberg, 1947; Kryter, 1962a; Kryter, 1962b; ANSI, 1969), which was developed to predict the intelligibility of filtered speech by estimating the audibility of speech sounds. Finally, the integral of the weighted, logarithmic, smoothed frequency-response gives the intelligibility-averaged gain, $\langle \mathcal{G} \rangle_I$, of the system. In the special case of frequency-independent directional response, i.e. $\mathcal{G}(f,\theta) = K(\theta)$, smoothing and weighting will have no effect and $\langle \mathcal{G}(\theta) \rangle_I = 20 \log_{10} K(\theta)$.

Intelligibility-averaged gain can be described as the relative level required for a signal in the unprocessed condition to be equal in intelligibility to the processed

[4] Of course, the presumed smoothing of human audition must operate on the array output signal, and smoothing the magnitude response function (which is only a transfer function), as in (2.56) will be exactly equivalent only when the input spectrum is flat. When the input spectrum is known, we could calculate $\langle \mathcal{G} \rangle_I$ more precisely by comparing smoothed input and output spectra. However, the simplified formula of (2.56) gives very similar results as long as either the input spectrum or the response magnitude is relatively smooth, and can be used to compare array responses independent of input spectrum.

signal. For the broadband beam pattern in Figure 2.6(d), the gain at 0° is 0 dB because signals from 0° are passed without modification and intelligibility is not changed. At 45°, the intelligibility-averaged broadband gain of -12 dB implies that processing has reduced the ability of the jammer to affect intelligibility to that of an unprocessed jammer of 12 dB less power.

The absolute effect of a given jammer on intelligibility will depend on the characteristics of the target. As an example, consider first a "reference" condition with target and jammer coincident at 0°, equal in level, and with identical spectra. In this situation, Articulation Theory would predict an Articulation Index (the fraction of target speech elements that are audible) of 0.4, which is sufficient for 50% to 95% correct on speech intelligibility tests of varying difficulty. Now, if that same jammer moves to 45°, its level would have to be increased by 12 dB to produce the same Articulation Index and target intelligibility as the reference condition[5]. Alternatively, the target could be reduced in power by 12 dB and still be as intelligible as it was in the reference condition. This implies one last interpretation of $\langle \mathcal{G} \rangle_I$ as that target-to-jammer ratio necessary to maintain constant target intelligibility (similar to the predictions of Zurek's binaural intelligibility model in section 2.1).

Based on intelligibility-averaged broadband gain, the four broadband beam patterns in Figure 2.7 can then be used to illustrate the rationale for adaptive beamforming. Pattern (a) is, once again, the average directional response of a 21-element 60-cm (2-foot) delay-and-sum endfire array. Although its directivity might be satisfactory for a hearing aid, its size is excessive. Pattern (b) is the result of reducing the delay-and-sum beamformer array to six elements over 15-cm (0.5 foot). Now the size is acceptable but directivity has decreased substantially. Patterns (c) and (d) show the results of applying "optimum" beamforming (to be discussed in

[5] Strictly speaking, $\langle \mathcal{G} \rangle_I$ only approximates the result of a search for the input Target-to-Jammer-Ratio that would give an A.I. of 0.4 if the A.I. calculation were performed in full (non-linear) detail. However, for a number of cases in which full calculations were made, the approximation error was less than 0.5 dB if the range of the frequency response was less than 40 dB. For frequency responses with ranges greater than 40 dB, the approximation was always conservative, underestimating the effective jammer reduction.

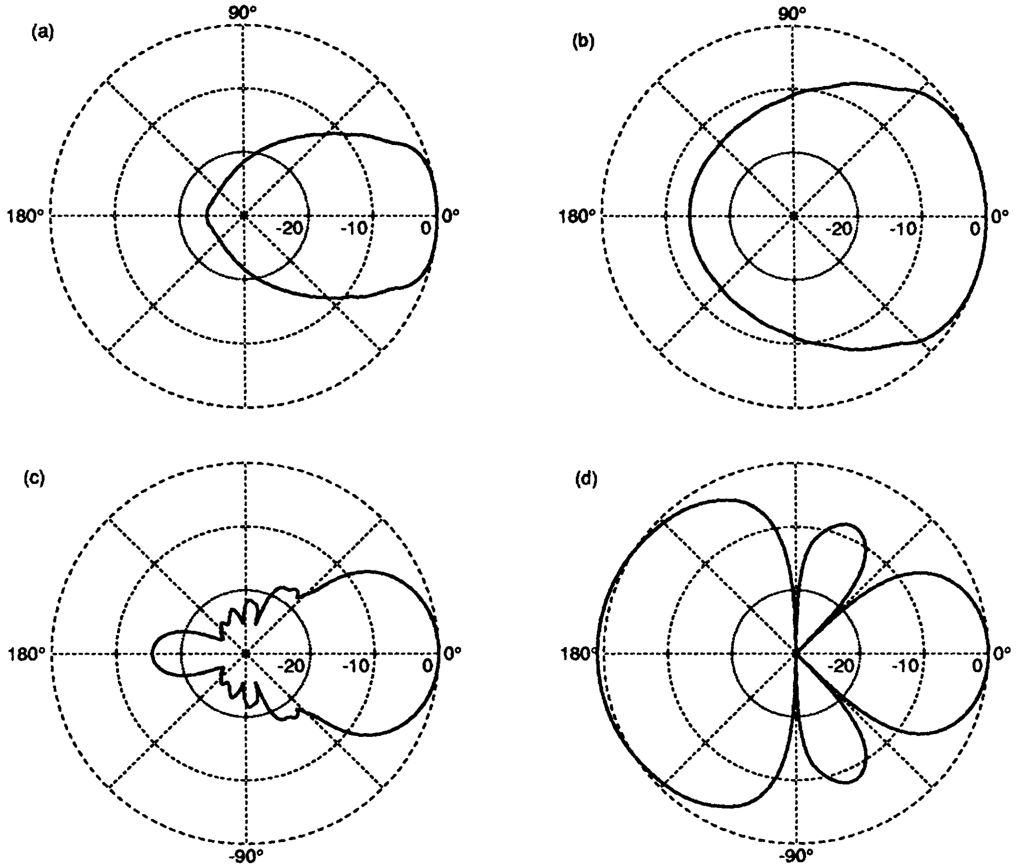


Figure 2.7: Broadband beam patterns for four equispaced, endfire arrays: (a) 21 elements, 60 cm, delay-and-sum beamforming; (b) 6 elements, 15 cm, delay-and-sum beamforming; (c) 6 elements, 15 cm, weights chosen to maximize directivity; (d) 6 elements, 15 cm, weights chosen to minimize jammers at 45° and −90°.

the next chapter) to the same 6-element, half-foot endfire array. In pattern (c), the processing weights have been optimized to maintain the target signal but minimize the response to isotropic noise or, equivalently, to maximize the directivity index (Duhamel, 1953; Bloch, Medhurst and Pool, 1953; Weston, 1986). This processing scheme provides directivity similar to that in pattern (a) with an array four times smaller. It should be noted, however, that endfire arrays designed to maximize directivity (so called "superdirective" arrays) are often quite sensitive to sensor noise and processing inaccuracies (Chu, 1948; Taylor, 1948; Cox, 1973a; Hansen, 1981). For hearing aid applications, this sensitivity may be reduced while significant directivity is retained by using "suboptimum" design methods (Cox, Zeskind and Kooij, 1985; Rabinowitz, Frost and Peterson, 1985; Cox, Zeskind and Kooij, 1986). In pattern (d), the processing weights have been optimized to minimize the array output power for the case of jammers at 45° and −90° in an anechoic environment with a small amount of sensor noise. Although the beam pattern hardly seems directional and even shows excess response for angles around 180°, only the responses at angles of 0°, 45°, and −90° are relevant because there are no signals present at any other angles. Pattern (d) is functionally the most directive of all for this particular interference configuration because it has the smallest response in the jammer directions. If the interference environment changes, however, the processor that produced pattern (d) must adapt its weights to maintain minimum interference response. This is precisely the goal of adaptive beamformers.

Signal- and Noise-Field Response

When we know the characteristics of a specific sound field, such as the field generated by the two directional sources in the last example, we can define the array response to that particular field as the ratio of array output power to the average power received by the individual microphones. This response measure takes into account all the complexities of the sound field, such as the presence of multiple sources or correlated reverberant echoes from multiple directions.

We will use $K_n(f)$ to denote an array's noise-field response at frequency f to noise with an inter-microphone cross-spectral-density matrix of $\mathcal{S}_{\boldsymbol{nn}}(f)$. The noise-field response will depend on $\mathcal{S}_{\boldsymbol{nn}}(f)$ and on the processor weights, $\underline{\mathcal{W}}(f)$, as follows. The average microphone power is the average of the diagonal elements of $\mathcal{S}_{\boldsymbol{nn}}(f)$ or trace$(\mathcal{S}_{\boldsymbol{nn}}(f))/M$. The array output power, given by equation (2.46), is simply $\underline{\mathcal{W}}^T(f)\,\mathcal{S}_{\boldsymbol{nn}}(f)\,\underline{\mathcal{W}}^*(f)$. The array's noise-field response is then

$$K_n(f) = \frac{\underline{\mathcal{W}}^T(f)\,\mathcal{S}_{\boldsymbol{nn}}(f)\,\underline{\mathcal{W}}^*(f)}{\frac{1}{M}\,\mathrm{trace}(\mathcal{S}_{\boldsymbol{nn}}(f))}\,. \tag{2.58}$$

A similar array response can be defined for any signal or noise field. In particular, we will be most interested in the response to the total noise signal, z:

$$K_z(f) = \frac{\underline{\mathcal{W}}^T(f)\,\mathcal{S}_{\boldsymbol{zz}}(f)\,\underline{\mathcal{W}}^*(f)}{\frac{1}{M}\,\mathrm{trace}(\mathcal{S}_{\boldsymbol{zz}}(f))}\;; \tag{2.59}$$

the response to sensor noise, u, whose cross-correlation matrix is $\sigma_u^2\, I_M$:

$$K_u(f) = \frac{\underline{\mathcal{W}}^T(f)\,\sigma_u^2\, I_M\,\underline{\mathcal{W}}^*(f)}{\frac{1}{M}\,\mathrm{trace}(\sigma_u^2\, I_M)} = \underline{\mathcal{W}}^T(f)\,\underline{\mathcal{W}}^*(f) = |\underline{\mathcal{W}}(f)|^2\;; \tag{2.60}$$

and the response to the received target signal, $\boldsymbol{r}[n] = \boldsymbol{h}[n] * s[n]$, from (2.32):

$$\begin{aligned} K_r(f) &= \frac{\underline{\mathcal{W}}^T(f)\,\mathcal{S}_{\boldsymbol{rr}}(f)\,\underline{\mathcal{W}}^*(f)}{\frac{1}{M}\,\mathrm{trace}(\mathcal{S}_{\boldsymbol{rr}}(f))} = \frac{\underline{\mathcal{W}}^T(f)\,\underline{\mathcal{H}}(f)\,\mathcal{S}_{ss}(f)\,\underline{\mathcal{H}}^\dagger(f)\,\underline{\mathcal{W}}^*(f)}{\frac{1}{M}\,\mathrm{trace}(\underline{\mathcal{H}}(f)\,\mathcal{S}_{ss}(f)\,\underline{\mathcal{H}}^\dagger(f))} \\ &= \frac{\underline{\mathcal{W}}^T(f)\,\underline{\mathcal{H}}(f)\,\underline{\mathcal{H}}^\dagger(f)\,\underline{\mathcal{W}}^*(f)}{\frac{1}{M}\,\mathrm{trace}(\underline{\mathcal{H}}^\dagger(f)\,\underline{\mathcal{H}}(f))} = \frac{|\underline{\mathcal{W}}^T(f)\,\underline{\mathcal{H}}(f)|^2}{\frac{1}{M}\,|\underline{\mathcal{H}}(f)|^2}\,, \end{aligned} \tag{2.61}$$

where we have factored out the scalar signal power, $\mathcal{S}_{ss}(f)$ and used the identity $\mathrm{trace}(AB) = \mathrm{trace}(BA)$.

A measure of array performance that often appears in the literature, ***array gain*** G_A, is the ratio of output to input signal-to-noise ratios (Bryn, 1962; Owsley, 1985; Cox, Zeskind and Kooij, 1986) and is easily shown to be

$$G_A(f) = \frac{K_r(f)}{K_z(f)}\,. \tag{2.62}$$

Note that array gain could be described as the gain ***against*** the total noise field and is opposite in sense to the total-noise response, $K_z(f)$, but has the intuitive appeal that higher gains are better. We will extend the array gain notion by defining similar gains for particular noise-fields of interest. Specifically, if we use $G_n(f)$ to denote the ratio of output to input signal-to-noise ratios for noise n, then we can define a total-noise gain,

$$G_z(f) = \frac{K_r(f)}{K_z(f)} = G_A(f)\;, \tag{2.63}$$

which is identical to array gain; an isotropic-noise gain, or array gain against isotropic noise,

$$G_i(f) = \frac{K_r(f)}{K_i(f)} = D(f)\;, \tag{2.64}$$

which is identical to the array directivity defined in (2.55); and a jammer-noise gain, or gain against received directional-jammer signals,

$$G_j(f) = \frac{K_r(f)}{K_j(f)}\,, \tag{2.65}$$

where $K_i(f)$ and $K_j(f)$ are the array responses, defined as above, to isotropic and directional-jammer noise fields, respectively[6].

[6] This family of gain measures is missing one member that we will not use. A commonly-used measure of array insensitivity to errors, ***white noise gain***, or gain ***against*** spatially- and temporally-uncorrelated noise is defined as

$$G_W(f) = G_u(f) = \frac{K_r(f)}{K_u(f)}\,. \tag{2.66}$$

This measures the degree to which the signal is amplified preferentially to white noise and random errors (Cox, Zeskind and Kooij, 1986). Thus, larger values of G_W are better, although the significance of a small G_W will depend on the amount of white noise or the magnitude of error actually present. In fact, G_W predicts the ratio by which white noise would have to exceed the signal to produce equal power in the output. Note that white noise gain, G_W, is ***inversely*** proportional to the sensor-noise response, K_u. In the common special case where the signal gain, K_r, is unity,

$$G_W(f) = \frac{1}{K_u(f)} = \frac{1}{|\underline{W}(f)|^2} \qquad (K_r = 1)\,. \tag{2.67}$$

We prefer to use $K_u(f)$ directly as a measure of the ***sensitivity*** of a processor to sensor-noise.

Chapter 3

Optimum Array Processing

In the last chapter we described the signal-processing structure of our proposed multi-microphone monaural hearing aid and used response patterns to illustrate the potential benefit of processing that is matched to the received interference. In this chapter we derive specific processing methods that are, in various senses, "optimum" for removing stationary interference. In subsequent chapters we will analyze the performance of these optimum processors and describe adaptive processing methods that can approach optimum performance in non-stationary hearing-aid environments.

Our investigation of optimum processing will proceed in three steps. First, we will consider various optimization criteria for processing based on unlimited observations (i.e., processing that uses data from all time) and show that the various criteria lead to similar frequency-domain processors. Second, we will consider a few of the same criteria for processing based on limited observations, which will lead to optimum time-domain processors. Third, we will try to relate the frequency- and time-domain results and discuss ways in which the different methods can be used.

3.1 Frequency-Domain Optimum Processors

Although our ultimate goal is an AMMA based on a limited number of microphone signal samples, as shown in Figure 2.4, we can gain considerable insight with relatively simple calculations by first considering the case in which samples from all time are available. If the sampled signal is stationary, it will have a *spectral representation*, similar to the Fourier transform of a deterministic signal, that depends on the signal samples over all time. Because the components of the signal's spectral representation at different frequencies will be uncorrelated, the derivation and application of optimum processors in the frequency domain will be greatly simplified. The results of frequency-domain processing can then be used to bound the performance of realizable processors based on a limited number of samples.

Spectral Representation of a Random Process. To present a rigorously correct definition of the spectral representation of a stationary random process would involve mathematical issues beyond the scope of this thesis (Wiener, 1930; Doob, 1953; Van Trees, 1968; Gardner, 1986). We will use an approximation that is essentially correct but requires a bit of justification.

Over a finite interval, a function $x[n]$ can be represented as a sum of orthonormal basis functions;

$$x[n] = \sum_{k=0}^{N-1} X_k \, \phi_k[n] \qquad (-N/2 \leq n < N/2) \tag{3.1}$$

where the basis functions, $\phi_k[n]$, satisfy

$$\sum_{n=-N/2}^{N/2-1} \phi_j^*[n] \, \phi_k[n] = \delta[j-k] \tag{3.2}$$

and the X_ks can be determined by

$$X_k = \sum_{n=-N/2}^{N/2-1} x[n] \, \phi_k^*[n] \, . \tag{3.3}$$

If the basis functions are known, then the set of X_ks, $\{X_k \,|\, 0 \leq k < N\}$, and the values of $x[n]$, $\{x[n] \,|\, -N/2 \leq n < N/2\}$, are equivalent representations of the same function.

When $x[n]$ is a random process, its values will be random variables and the X_ks will be linearly related random variables. Karhunen and Loève have shown that it is possible to choose a set of basis functions such that the X_ks are uncorrelated (Van Trees, 1968), i.e.

$$E\left\{X_j^* \, X_k\right\} = \lambda_j \, \delta[j-k] \, . \tag{3.4}$$

Assuming that $x[n]$ is stationary, this special set of basis functions will satisfy

$$\sum_{m=-N/2}^{N/2-1} R_{xx}[n-m]\,\phi_k[m] = \lambda_k\,\phi_k[n]\,, \tag{3.5}$$

in which ϕ_k is an eigenfunction and λ_k is the corresponding eigenvalue. As $N \to \infty$, this equation approaches the form of a convolution of ϕ_k with R_{xx}, which can be viewed as the impulse response of an LTI system, whose eigenfunctions must be complex exponentials.

In fact, it can be shown (Davenport and Root, 1958; Van Trees, 1968; Gray, 1972) that for large N,

$$\begin{aligned} \phi_k[n] &\simeq \frac{1}{\sqrt{N}} e^{j2\pi kn/N} = \frac{1}{\sqrt{N}} e^{j2\pi f_k nT_s} & (3.6)\\ \lambda_k &\simeq \mathcal{S}_{xx}\left(\frac{k}{NT_s}\right) = \mathcal{S}_{xx}(f_k)\,, & (3.7) \end{aligned}$$

where $f_k = \frac{k}{NT_s}$. (The previously mentioned mathematical issues arise in rigorously taking the limit of these expressions as $N \to \infty$.) This leads to our approximate (for large N) spectral representation,

$$\breve{\mathcal{X}}_N(f) = \frac{1}{\sqrt{N}} \sum_{n=-N/2}^{N/2-1} x[n]\, e^{j2\pi f nT_s}\,, \tag{3.8}$$

for which

$$E\left\{\breve{\mathcal{X}}_N(f_j)\,\breve{\mathcal{X}}_N(f_k)\right\} \simeq \mathcal{S}_{xx}(f_k)\,\delta[j-k]\,. \tag{3.9}$$

The validity of this approximation will depend on N being much greater than the non-zero extent of $R_{xx}[n]$ or, equivalently, greater than some function of the "sharpness" of features in $\mathcal{S}_{xx}(f)$.

We can now proceed to consider various optimizing criteria in the derivation of frequency-domain optimum processors.[1] These derivations will all be based on a model of the received signal, generalized from Section 2.3.1, as

$$\underline{\breve{\mathcal{X}}}_N(f) = \underline{\mathcal{H}}(f)\,\breve{\mathcal{S}}_N(f) + \underline{\breve{\mathcal{Z}}}_N(f)\,. \tag{3.10}$$

[1] The basic concept and many of the results of this section were originally presented by Cox in an excellent paper (Cox, 1968) and later expanded slightly by Monzingo and Miller (Monzingo and Miller, 1980).

In all cases we will assume that $\underline{\mathcal{H}}(f)$, the vector of transfer functions from the target source to each of the microphones, is known. To simplify notation, we will drop both the explicit argument f in frequency-domain functions and the subscript N that denotes the extent of approximate spectral representations. Thus, equation (3.10) can be compressed to

$$\underline{\breve{\mathcal{X}}} = \underline{\mathcal{H}}\breve{\mathcal{S}} + \underline{\breve{\mathcal{Z}}} . \tag{3.11}$$

3.1.1 Maximum A Posteriori Probability

If we assume that the target, jammers, and receiver noise are all zero-mean real Gaussian random processes, then, in the frequency domain, the target and total received noise will both have zero-mean complex Gaussian distributions (Reed, 1962; Goodman, 1963) given by

$$p\left(\breve{\mathcal{S}}\right) = \frac{1}{\pi \det\left(\mathcal{S}_{ss}\right)} \exp\left(-\breve{\mathcal{S}}^{*} \mathcal{S}_{ss}^{-1} \breve{\mathcal{S}}\right) \tag{3.12}$$

$$p\left(\underline{\breve{\mathcal{Z}}}\right) = \frac{1}{\pi^{M} \det\left(\mathcal{S}_{zz}\right)} \exp\left(-\underline{\breve{\mathcal{Z}}}^{\dagger} \mathcal{S}_{zz}^{-1} \underline{\breve{\mathcal{Z}}}\right) , \tag{3.13}$$

which we can denote by

$$\breve{\mathcal{S}} \sim N\left(0, \mathcal{S}_{ss}\right) \tag{3.14}$$

$$\underline{\breve{\mathcal{Z}}} \sim N\left(\underline{0}, \mathcal{S}_{zz}\right) . \tag{3.15}$$

From these distributions and the received-signal model, it follows that

$$\underline{\breve{\mathcal{X}}} \sim N\left(\underline{0}, \underline{\mathcal{H}}\mathcal{S}_{ss}\underline{\mathcal{H}}^{\dagger} + \mathcal{S}_{zz}\right) , \tag{3.16}$$

and the *a posteriori* probability of $\breve{\mathcal{S}}$ given the observation $\underline{\breve{\mathcal{X}}}$ is

$$\begin{aligned} p\left(\breve{\mathcal{S}}|\underline{\breve{\mathcal{X}}}\right) &= \frac{p\left(\underline{\breve{\mathcal{X}}}|\breve{\mathcal{S}}\right) p\left(\breve{\mathcal{S}}\right)}{p\left(\underline{\breve{\mathcal{X}}}\right)} = \frac{p_{\underline{\breve{\mathcal{Z}}}}\left(\underline{\breve{\mathcal{X}}} - \underline{\mathcal{H}}\breve{\mathcal{S}}\right) p\left(\breve{\mathcal{S}}\right)}{p\left(\underline{\breve{\mathcal{X}}}\right)} \\ &= k \exp\left(-\left(\underline{\breve{\mathcal{X}}} - \underline{\mathcal{H}}\breve{\mathcal{S}}\right)^{\dagger} \mathcal{S}_{zz}^{-1} \left(\underline{\breve{\mathcal{X}}} - \underline{\mathcal{H}}\breve{\mathcal{S}}\right) - \left(\breve{\mathcal{S}}^{\dagger} \mathcal{S}_{ss}^{-1} \breve{\mathcal{S}}\right)\right) . \end{aligned} \tag{3.17}$$

The MAP target estimate, $\hat{\mathcal{S}}_{\mathrm{MAP}}$, is that value of $\breve{\mathcal{S}}$ for which the *a posteriori* distribution is maximum. Since the exponential function is monotonic, the maximum of (3.17) occurs where the exponent itself is maximum. That is,

$$\frac{\partial}{\partial\breve{\mathcal{S}}}\left[-\left(\underline{\breve{\mathcal{X}}}-\underline{\mathcal{H}}\breve{\mathcal{S}}\right)^{\dagger}\mathcal{S}_{zz}^{-1}\left(\underline{\breve{\mathcal{X}}}-\underline{\mathcal{H}}\breve{\mathcal{S}}\right)-\left(\breve{\mathcal{S}}^{\dagger}\mathcal{S}_{ss}^{-1}\breve{\mathcal{S}}\right)\right]\Bigg|_{\breve{\mathcal{S}}=\hat{\mathcal{S}}_{\mathrm{MAP}}}=0\,, \tag{3.18}$$

which implies[2]

$$\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{zz}^{-1}\left(\underline{\breve{\mathcal{X}}}-\underline{\mathcal{H}}\hat{\mathcal{S}}_{\mathrm{MAP}}\right)-\mathcal{S}_{ss}^{-1}\,\hat{\mathcal{S}}_{\mathrm{MAP}}=0 \tag{3.19}$$

and, therefore,

$$\hat{\mathcal{S}}_{\mathrm{MAP}}=\left(\mathcal{S}_{ss}^{-1}+\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{zz}^{-1}\,\underline{\breve{\mathcal{X}}}=\underline{\mathcal{W}}_{\mathrm{MAP}}^{T}\,\underline{\breve{\mathcal{X}}}\,. \tag{3.20}$$

Thus, the optimum processor for the MAP criterion combines microphone signals with the weighting function

$$\underline{\mathcal{W}}_{\mathrm{MAP}}^{T}=\left(\mathcal{S}_{ss}^{-1}+\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{zz}^{-1}\,, \tag{3.21}$$

or, after applying the matrix identity (A.2),

$$\begin{aligned}\underline{\mathcal{W}}_{\mathrm{MAP}}^{T} &= \mathcal{S}_{ss}\,\underline{\mathcal{H}}^{\dagger}\left(\underline{\mathcal{H}}\,\mathcal{S}_{ss}\,\underline{\mathcal{H}}^{\dagger}+\mathcal{S}_{zz}\right)^{-1} && (3.22)\\ &= \mathcal{S}_{sx}\,\mathcal{S}_{xx}^{-1}\,. && (3.23)\end{aligned}$$

Note that this processor is based on knowledge of $\underline{\mathcal{H}}$, $\mathcal{S}_{ss}$ and $\mathcal{S}_{zz}$, and on the assumption that s and z are Gaussian.

[2] The notation $\frac{\partial}{\partial\mathbf{v}}$, where $\mathbf{v}$ is complex, stands for the derivatives with respect to the real and imaginary parts of $\mathbf{v}$. We use the following notation and rules for differentiation with respect to complex vectors (or scalars). If s is a real scalar, $\mathbf{v}$ and $\mathbf{w}$ are complex vectors, and $\mathbf{M}$ and $\mathbf{H}$ are complex matrices, then $\frac{\partial}{\partial\mathbf{v}}s$ is a vector of partial derivatives of s with respect to the real and imaginary parts of each element of $\mathbf{v}$, and, in particular,

$$\begin{aligned}\frac{\partial}{\partial\mathbf{v}}\left(\mathbf{v}^{\dagger}\,\mathbf{M}\,\mathbf{v}\right) &= 2\,\mathbf{M}\,\mathbf{v}\\ \frac{\partial}{\partial\mathbf{v}}\left(\mathbf{v}^{\dagger}\,\mathbf{w}+\mathbf{w}^{\dagger}\,\mathbf{v}\right) &= 2\,\mathbf{w}\\ \frac{\partial}{\partial\mathbf{v}}\left[(\mathbf{w}-\mathbf{H}\,\mathbf{v})^{\dagger}\,\mathbf{M}\,(\mathbf{w}-\mathbf{H}\,\mathbf{v})\right] &= -2\,\mathbf{H}^{\dagger}\,\mathbf{M}\,(\mathbf{w}-\mathbf{H}\,\mathbf{v})\,.\end{aligned}$$

These relationships can be derived using rules for real-vector differentiation (Selby, 1975; Monzingo and Miller, 1980) and considering separately the derivatives with respect to real and imaginary parts.

We can characterize the performance of the MAP processor by calculating the expected squared error (or variance) of the signal estimate.

$$\begin{align}
\varepsilon^2_{\text{MAP}} &= E\left\{\left(\hat{\mathcal{S}}_{\text{MAP}} - \check{\mathcal{S}}\right)^2\right\} = E\left\{\left(\underline{\mathcal{W}}^T_{\text{MAP}}\,\check{\underline{\mathcal{X}}} - \check{\mathcal{S}}\right)\left(\check{\underline{\mathcal{X}}}^\dagger\,\underline{\mathcal{W}}^*_{\text{MAP}} - \check{\mathcal{S}}^*\right)\right\} \nonumber \\
&= \underline{\mathcal{W}}^T_{\text{MAP}}\,\mathcal{S}_{\boldsymbol{xx}}\,\underline{\mathcal{W}}^*_{\text{MAP}} - \mathcal{S}_{s\boldsymbol{x}}\,\underline{\mathcal{W}}^*_{\text{MAP}} - \underline{\mathcal{W}}^T_{\text{MAP}}\,\mathcal{S}_{\boldsymbol{x}s} + \mathcal{S}_{ss} \nonumber \\
&= \mathcal{S}_{ss} - \mathcal{S}_{s\boldsymbol{x}}\,\mathcal{S}^{-1}_{\boldsymbol{xx}}\,\mathcal{S}_{\boldsymbol{x}s} \tag{3.24} \\
&= \mathcal{S}_{ss} - \mathcal{S}_{ss}\,\underline{\mathcal{H}}^\dagger\left(\underline{\mathcal{H}}\,\mathcal{S}_{ss}\,\underline{\mathcal{H}}^\dagger + \mathcal{S}_{\boldsymbol{zz}}\right)^{-1}\underline{\mathcal{H}}\,\mathcal{S}_{ss} \tag{3.25} \\
&= \left(\mathcal{S}^{-1}_{ss} + \underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{\boldsymbol{zz}}\,\underline{\mathcal{H}}\right)^{-1} . \tag{3.26}
\end{align}$$

3.1.2 Minimum Mean Squared Error

The MMSE target estimate, $\hat{\mathcal{S}}_{\text{MMSE}}$, minimizes the expected squared estimation error

$$\begin{align}
\varepsilon^2 &= E\left\{\left(\hat{\mathcal{S}} - \check{\mathcal{S}}\right)^2\right\} = E\left\{\left(\underline{\mathcal{W}}^T\,\underline{\mathcal{X}} - \check{\mathcal{S}}\right)\left(\underline{\mathcal{X}}^\dagger\,\underline{\mathcal{W}}^* - \check{\mathcal{S}}^*\right)\right\} \nonumber \\
&= \underline{\mathcal{W}}^T\,\mathcal{S}_{\boldsymbol{xx}}\,\underline{\mathcal{W}}^* - \underline{\mathcal{W}}^T\,\mathcal{S}_{\boldsymbol{x}s} - \mathcal{S}_{s\boldsymbol{x}}\,\underline{\mathcal{W}}^* + \mathcal{S}_{ss} . \tag{3.27}
\end{align}$$

The minimum of this quadratic form will occur where the gradient with respect to $\underline{\mathcal{W}}$ is zero.

$$\left.\frac{\partial}{\partial\underline{\mathcal{W}}}\left(\varepsilon^2\right)\right|_{\underline{\mathcal{W}}=\underline{\mathcal{W}}_{\text{MMSE}}} = 2\,\mathcal{S}_{\boldsymbol{xx}}\,\underline{\mathcal{W}}^*_{\text{MMSE}} - 2\,\mathcal{S}_{\boldsymbol{x}s} = \underline{0} \tag{3.28}$$

then implies

$$\begin{align}
\underline{\mathcal{W}}^*_{\text{MMSE}} &= \mathcal{S}^{-1}_{\boldsymbol{xx}}\,\mathcal{S}_{\boldsymbol{x}s} \nonumber \\
\underline{\mathcal{W}}^T_{\text{MMSE}} &= \mathcal{S}_{s\boldsymbol{x}}\,\mathcal{S}^{-1}_{\boldsymbol{xx}} \tag{3.29} \\
&= \mathcal{S}_{ss}\,\underline{\mathcal{H}}^\dagger\left(\underline{\mathcal{H}}\,\mathcal{S}_{ss}\,\underline{\mathcal{H}}^\dagger + \mathcal{S}_{\boldsymbol{zz}}\right)^{-1} \tag{3.30} \\
&= \left(\mathcal{S}^{-1}_{ss} + \underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{\boldsymbol{zz}}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{\boldsymbol{zz}} . \tag{3.31}
\end{align}$$

The MMSE processor is identical to the MAP processor and also depends on knowledge of $\underline{\mathcal{H}}$, $\mathcal{S}_{ss}$ and $\mathcal{S}_{\boldsymbol{zz}}$, but the derivation does not depend on the Gaussian assumption. This is consistent with the generally known result that, when $\underline{\mathcal{S}}$ and $\underline{\mathcal{X}}$ *are* jointly Gaussian, then the MMSE and MAP estimates of $\underline{\mathcal{S}}$ given $\underline{\mathcal{X}}$ are identical (Van Trees, 1968).

Since the MMSE and MAP processors are identical, the mean squared error for the two processors will also be equal.

$$\begin{aligned}
\varepsilon^2_{\text{MMSE}} &= E\left\{\left(\hat{\mathcal{S}}_{\text{MMSE}} - \breve{\mathcal{S}}\right)^2\right\} = \varepsilon^2_{\text{MAP}} \\
&= \mathcal{S}_{ss} - \mathcal{S}_{sx}\,\mathcal{S}_{xx}^{-1}\,\mathcal{S}_{xs} && (3.32) \\
&= \mathcal{S}_{ss} - \mathcal{S}_{ss}\,\underline{\mathcal{H}}^{\dagger}\left(\underline{\mathcal{H}}\,\mathcal{S}_{ss}\,\underline{\mathcal{H}}^{\dagger} + \mathcal{S}_{zz}\right)^{-1}\underline{\mathcal{H}}\,\mathcal{S}_{ss} && (3.33) \\
&= \left(\mathcal{S}_{ss}^{-1} + \underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\right)^{-1} . && (3.34)
\end{aligned}$$

3.1.3 Maximum Signal-to-Noise Ratio

The SNR estimator attempts to maximize the ratio of target power to interference power in the processor output, which can be decomposed into target and interference components, $\mathcal{V}$ and $\mathcal{U}$:

$$\mathcal{Y} = \underline{\mathcal{W}}^T\,\underline{\mathcal{X}} = \underline{\mathcal{W}}^T\left(\underline{\mathcal{H}}\,\breve{\mathcal{S}} + \breve{\underline{\mathcal{Z}}}\right) = \underline{\mathcal{W}}^T\,\underline{\mathcal{H}}\,\breve{\mathcal{S}} + \underline{\mathcal{W}}^T\,\breve{\underline{\mathcal{Z}}} = \mathcal{V} + \mathcal{U} \qquad (3.35)$$

Assuming that s and z are uncorrelated, the output power is simply the sum

$$\mathcal{S}_{yy} = \mathcal{S}_{vv} + \mathcal{S}_{uu} = \underline{\mathcal{W}}^T\,\underline{\mathcal{H}}\,\mathcal{S}_{ss}\,\underline{\mathcal{H}}^{\dagger}\,\underline{\mathcal{W}}^* + \underline{\mathcal{W}}^T\,\mathcal{S}_{zz}\,\underline{\mathcal{W}}^* . \qquad (3.36)$$

We want to find $\underline{\mathcal{W}}_{\text{SNR}}$, that $\underline{\mathcal{W}}$ which maximizes

$$\left(\frac{\text{S}}{\text{N}}\right) = \frac{\underline{\mathcal{W}}^T\,\underline{\mathcal{H}}\,\mathcal{S}_{ss}\,\underline{\mathcal{H}}^{\dagger}\,\underline{\mathcal{W}}^*}{\underline{\mathcal{W}}^T\,\mathcal{S}_{zz}\,\underline{\mathcal{W}}^*} . \qquad (3.37)$$

By defining[3] $\underline{\mathcal{P}} = \mathcal{S}_{zz}^{1/2}\,\underline{\mathcal{W}}^*$ and $\underline{\mathcal{R}} = \mathcal{S}_{zz}^{-1/2}\,\underline{\mathcal{H}}\,\mathcal{S}_{ss}^{1/2}$, this expression can be rewritten

$$\left(\frac{\text{S}}{\text{N}}\right) = \frac{\underline{\mathcal{P}}^{\dagger}\,\underline{\mathcal{R}}\,\underline{\mathcal{R}}^{\dagger}\,\underline{\mathcal{P}}}{\underline{\mathcal{P}}^{\dagger}\,\underline{\mathcal{P}}} . \qquad (3.38)$$

According to Rayleigh's principle (Strang, 1976), this quadratic form will be maximized by setting $\underline{\mathcal{P}}$ equal to the eigenvector of $\underline{\mathcal{R}}\,\underline{\mathcal{R}}^{\dagger}$ with the largest eigenvalue.

[3] The factorization $\mathcal{S}_{zz} = \mathcal{S}_{zz}^{1/2}\,\mathcal{S}_{zz}^{1/2}$ exists if and only if $\mathcal{S}_{zz}$ is positive definite (Strang, 1976), a reasonable assumption in our application.

Luckily, the rank 1 matrix $\underline{\mathcal{R}}\,\underline{\mathcal{R}}^{\dagger}$ has only *one* non-zero eigenvalue and the corresponding eigenvector must be of the form $\alpha\,\underline{\mathcal{R}}$. Thus,

$$\begin{aligned}
\underline{\mathcal{P}}_{\mathrm{MAX}} &= \alpha\,\underline{\mathcal{R}} && (3.39)\\
\mathcal{S}_{zz}^{1/2}\,\underline{\mathcal{W}}^{*}_{\mathrm{SNR}} &= \alpha\,\mathcal{S}_{zz}^{-1/2}\,\underline{\mathcal{H}}\,\mathcal{S}_{ss}^{1/2} && (3.40)\\
\underline{\mathcal{W}}^{T}_{\mathrm{SNR}} &= \beta\,\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{zz}^{-1}\,. && (3.41)
\end{aligned}$$

The arbitrary constant, β, determines the overall level of the output but does not affect signal-to-noise ratio.

The performance of the SNR estimator can be evaluated by calculating the maximized output signal-to-noise ratio:

$$\begin{aligned}
\left(\frac{\mathrm{S}}{\mathrm{N}}\right)_{\mathrm{MAX}} &= \frac{\underline{\mathcal{W}}^{T}_{\mathrm{SNR}}\,\underline{\mathcal{H}}\,\mathcal{S}_{ss}\,\underline{\mathcal{H}}^{\dagger}\,\underline{\mathcal{W}}^{*}_{\mathrm{SNR}}}{\underline{\mathcal{W}}^{T}_{\mathrm{SNR}}\,\mathcal{S}_{zz}\,\underline{\mathcal{W}}^{*}_{\mathrm{SNR}}}\\
&= \frac{\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\,\mathcal{S}_{ss}\,\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}}{\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}}\\
&= \mathcal{S}_{ss}\,\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\,. && (3.42)
\end{aligned}$$

Note that, although the performance calculation requires knowledge of $\underline{\mathcal{H}}$, $\mathcal{S}_{zz}$ and $\mathcal{S}_{ss}$, the SNR processor is based on $\underline{\mathcal{H}}$ and $\mathcal{S}_{zz}$ only.

3.1.4 Maximum-Likelihood

The preceeding processors were all based on the assumption that the target signal could be modelled as a random process characterized by, at least, its spectral density function and, in the MAP case, by Gaussian statistics. It is possible, however, to derive processors based on the less restrictive assumption that the target is simply an unknown, deterministic signal for which *a priori* information, in the form of target statistics, is unavailable.

If we assume that the interference alone is Gaussian, i.e.,

$$\check{\underline{\mathcal{Z}}} \sim N(\underline{0}, \mathcal{S}_{zz}) \;, \qquad (3.43)$$

then the probability density function for the observations $\underline{\check{\mathcal{X}}}$ will depend on the unknown $\mathcal{S}$,

$$\begin{aligned} p\left(\underline{\check{\mathcal{X}}}\right) &= p_{\underline{\check{\mathcal{Z}}}}\left(\underline{\check{\mathcal{X}}} - \underline{\mathcal{H}}\mathcal{S}\right) \\ &= k \exp\left(-\left(\underline{\check{\mathcal{X}}} - \underline{\mathcal{H}}\mathcal{S}\right)^{\dagger} \mathcal{S}_{\boldsymbol{z}\boldsymbol{z}}^{-1} \left(\underline{\check{\mathcal{X}}} - \underline{\mathcal{H}}\mathcal{S}\right)\right) . \end{aligned} \tag{3.44}$$

The ML target estimate, $\hat{\mathcal{S}}_{\text{ML}}$, is that value of $\mathcal{S}$ which maximizes (3.44) or, in other words, the target signal for which the observation $\underline{\check{\mathcal{X}}}$ is most likely. Once again, because the exponential function is monotonic and the exponent is quadratic, the maximum of (3.44) can be found where

$$\frac{\partial}{\partial \mathcal{S}}\left[-\left(\underline{\check{\mathcal{X}}} - \underline{\mathcal{H}}\mathcal{S}\right)^{\dagger} \mathcal{S}_{\boldsymbol{z}\boldsymbol{z}}^{-1} \left(\underline{\check{\mathcal{X}}} - \underline{\mathcal{H}}\mathcal{S}\right)\right]\Bigg|_{\mathcal{S}=\hat{\mathcal{S}}_{\text{ML}}} = 0\,, \tag{3.45}$$

which implies that

$$\underline{\mathcal{H}}^{\dagger} \mathcal{S}_{\boldsymbol{z}\boldsymbol{z}}^{-1} \left(\underline{\check{\mathcal{X}}} - \underline{\mathcal{H}}\hat{\mathcal{S}}_{\text{ML}}\right) = 0 \tag{3.46}$$

$$\hat{\mathcal{S}}_{\text{ML}} = \left(\underline{\mathcal{H}}^{\dagger} \mathcal{S}_{\boldsymbol{z}\boldsymbol{z}}^{-1} \underline{\mathcal{H}}\right)^{-1} \underline{\mathcal{H}}^{\dagger} \mathcal{S}_{\boldsymbol{z}\boldsymbol{z}}^{-1} \underline{\check{\mathcal{X}}} = \underline{\mathcal{W}}_{\text{ML}}^{T} \underline{\check{\mathcal{X}}}\,. \tag{3.47}$$

Thus, the ML processor is

$$\underline{\mathcal{W}}_{\text{ML}}^{T} = \left(\underline{\mathcal{H}}^{\dagger} \mathcal{S}_{\boldsymbol{z}\boldsymbol{z}}^{-1} \underline{\mathcal{H}}\right)^{-1} \underline{\mathcal{H}}^{\dagger} \mathcal{S}_{\boldsymbol{z}\boldsymbol{z}}^{-1}\,, \tag{3.48}$$

which depends only on $\underline{\mathcal{H}}$ and $\mathcal{S}_{\boldsymbol{z}\boldsymbol{z}}$ and is based on the assumption of Gaussian interference. When the target signal actually is a stationary process, it can be shown (see Appendix B) that

$$\underline{\mathcal{W}}_{\text{ML}}^{T} = \left(\underline{\mathcal{H}}^{\dagger} \mathcal{S}_{\boldsymbol{x}\boldsymbol{x}}^{-1} \underline{\mathcal{H}}\right)^{-1} \underline{\mathcal{H}}^{\dagger} \mathcal{S}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\,. \tag{3.49}$$

In other words, for an unknown but stationary target, knowledge of $\mathcal{S}_{\boldsymbol{x}\boldsymbol{x}}$, which can be estimated from the observations, is equivalent to knowledge of $\mathcal{S}_{\boldsymbol{z}\boldsymbol{z}}$.

The performance of the ML processor can be characterized by the expected squared error (or variance) of the signal estimate, where the signal is now fixed.

$$\begin{aligned}
\varepsilon^2_{\mathrm{ML}} &= E\left\{\left(\hat{\mathcal{S}}_{\mathrm{ML}} - \mathcal{S}\right)^2\right\} = E\left\{\left(\underline{\mathcal{W}}^T_{\mathrm{ML}}\,\check{\underline{\mathcal{X}}} - \mathcal{S}\right)^2\right\} \\
&= E\left\{\left(\left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\left(\underline{\mathcal{H}}\,\mathcal{S} + \check{\underline{\mathcal{Z}}}\right) - \mathcal{S}\right)^2\right\} \\
&= E\left\{\left(\left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\,\check{\underline{\mathcal{Z}}}\right)^2\right\} \\
&= \left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\,\mathcal{S}_{zz}\,\mathcal{S}^{-1}_{zz}\,\underline{\mathcal{H}}\left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\,\underline{\mathcal{H}}\right)^{-1} \\
&= \left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\,\underline{\mathcal{H}}\right)^{-1} \qquad (3.50)
\end{aligned}$$

3.1.5 Minimum-Variance Unbiased

The MV processor produces a signal estimate, $\hat{\mathcal{S}}_{\mathrm{MV}}$, that is unbiased and has the lowest variance of all unbiased estimates. The zero bias requirement can be stated as

$$E\left\{\hat{\mathcal{S}}_{\mathrm{MV}}\right\} = E\left\{\underline{\mathcal{W}}^T_{\mathrm{MV}}\,\check{\underline{\mathcal{X}}}\right\} = E\left\{\underline{\mathcal{W}}^T_{\mathrm{MV}}\left(\underline{\mathcal{H}}\,\mathcal{S} + \check{\underline{\mathcal{Z}}}\right)\right\} = \underline{\mathcal{W}}^T_{\mathrm{MV}}\,\underline{\mathcal{H}}\,\mathcal{S} = \mathcal{S}\ , \qquad (3.51)$$

which implies that

$$\underline{\mathcal{W}}^T_{\mathrm{MV}}\,\underline{\mathcal{H}} = 1\ . \qquad (3.52)$$

The variance to be minimized is

$$\begin{aligned}
\varepsilon^2_{\mathrm{MV}} &= E\left\{\left(\hat{\mathcal{S}}_{\mathrm{MV}} - \mathcal{S}\right)^2\right\} \quad = \quad E\left\{\hat{\mathcal{S}}^2_{\mathrm{MV}} - 2\,\hat{\mathcal{S}}_{\mathrm{MV}}\,\mathcal{S} + \mathcal{S}^2\right\} \\
&= E\left\{\hat{\mathcal{S}}^2_{\mathrm{MV}}\right\} - \mathcal{S}^2 \\
&= E\left\{\underline{\mathcal{W}}^T_{\mathrm{MV}}\left(\underline{\mathcal{H}}\,\mathcal{S} + \check{\underline{\mathcal{Z}}}\right)\left(\underline{\mathcal{H}}\,\mathcal{S} + \check{\underline{\mathcal{Z}}}\right)^\dagger\underline{\mathcal{W}}^*_{\mathrm{MV}}\right\} - \mathcal{S}^2 \\
&= \underline{\mathcal{W}}^T_{\mathrm{MV}}\mathcal{S}_{zz}\underline{\mathcal{W}}^*_{\mathrm{MV}}\ , \qquad (3.53)
\end{aligned}$$

where the last step depends on $\check{\underline{\mathcal{Z}}}$ being zero-mean and on the assumption of zero bias, i.e., that $\underline{\mathcal{W}}^T_{\mathrm{MV}}\,\underline{\mathcal{H}} = 1$. The minimization of (3.53) with respect to $\underline{\mathcal{W}}$ must be constrained to produce a $\underline{\mathcal{W}}$ that satisfies the zero-bias condition $\underline{\mathcal{W}}^T_{\mathrm{MV}}\,\underline{\mathcal{H}} = 1$. To do this, we introduce the constraint with a Lagrange multiplier, λ, and minimize

$$\underline{\mathcal{W}}^T_{\mathrm{MV}}\mathcal{S}_{zz}\underline{\mathcal{W}}^*_{\mathrm{MV}} + 2\,\lambda\left(\underline{\mathcal{W}}^T_{\mathrm{MV}}\,\underline{\mathcal{H}} - 1\right)\ , \qquad (3.54)$$

which will be independent of λ as long as the constraint is satisfied. Differentiating with respect to $\underline{\mathcal{W}}^*$,

$$\begin{aligned} 2\,\mathcal{S}_{zz}\,\underline{\mathcal{W}}^*_{\text{MV}} + 2\,\lambda\,\underline{\mathcal{H}} &= 0 \\ \underline{\mathcal{W}}^*_{\text{MV}} &= -\lambda\,\mathcal{S}^{-1}_{zz}\,\underline{\mathcal{H}} \\ \underline{\mathcal{W}}^T_{\text{MV}} &= -\lambda\,\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\,. \end{aligned} \tag{3.55}$$

Since this $\underline{\mathcal{W}}$ must satisfy the zero-bias constraint,

$$\begin{aligned} \underline{\mathcal{W}}^T_{\text{MV}}\,\underline{\mathcal{H}} \;=\; -\lambda\,\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\,\underline{\mathcal{H}} &= 1 \\ \lambda &= -\left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\,\underline{\mathcal{H}}\right)^{-1}\,, \end{aligned} \tag{3.56}$$

which allows us to eliminate λ in (3.55),

$$\underline{\mathcal{W}}^T_{\text{MV}} = \left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\,. \tag{3.57}$$

Thus, the MV processor is identical to the ML processor, depending on $\underline{\mathcal{H}}$ and $\mathcal{S}_{zz}$, but can be derived without making the Gaussian assumption. Since the processors are identical, it will also be true for the MV processor that, if the target signal is stationary,

$$\underline{\mathcal{W}}^T_{\text{MV}} = \left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{xx}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{xx}\,. \tag{3.58}$$

And, finally, the performance of the MV processor must equal that of the ML processor:

$$\varepsilon^2_{\text{MV}} = \varepsilon^2_{\text{ML}} = \left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}^{-1}_{zz}\,\underline{\mathcal{H}}\right)^{-1}\,. \tag{3.59}$$

3.1.6 Summary

Table 3.1 summarizes the results of this section. The most important result is that, for all criteria, the processors are identical to within a scalar function of frequency (the denominator expressions are all scalars) that depends on our *a priori* knowledge of the signal and interference spectra. This frequency-dependent weighting function controls the contribution of energy at different frequencies to the overall target-to-jammer ratio (TJR). As discussed in Section 2.3.3, however, human speech reception

Criterion	Assumptions	Processor $\left(\underline{\mathcal{W}}^T\right)$	Performance(ε^2)
MAP	$\underline{\mathcal{H}}$, $\breve{\mathcal{S}} \sim N(0\,, \mathcal{S}_{ss})$ $\underline{\breve{\mathcal{Z}}} \sim N(\underline{0}\,, \mathcal{S}_{zz})$	$\frac{\underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1}}{\mathcal{S}_{ss}^{-1} + \underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1} \underline{\mathcal{H}}}$	$\frac{1}{\mathcal{S}_{ss}^{-1} + \underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1} \underline{\mathcal{H}}}$
MMSE	$\underline{\mathcal{H}}, \mathcal{S}_{ss}\,, \mathcal{S}_{zz}$	$\frac{\underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1}}{\mathcal{S}_{ss}^{-1} + \underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1} \underline{\mathcal{H}}}$	$\frac{1}{\mathcal{S}_{ss}^{-1} + \underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1} \underline{\mathcal{H}}}$
SNR	$\underline{\mathcal{H}}, \mathcal{S}_{ss}\,, \mathcal{S}_{zz}$	$\beta\, \underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1}$	$\left(\frac{\mathrm{S}}{\mathrm{N}}\right) = \underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1} \underline{\mathcal{H}} \mathcal{S}_{ss}$
ML	$\underline{\mathcal{H}}, \underline{\breve{\mathcal{Z}}} \sim N(\underline{0}\,, \mathcal{S}_{zz})$	$\frac{\underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1}}{\underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1} \underline{\mathcal{H}}}$	$\frac{1}{\underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1} \underline{\mathcal{H}}}$
MV	$\underline{\mathcal{H}}, \mathcal{S}_{zz}$	$\frac{\underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1}}{\underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1} \underline{\mathcal{H}}}$	$\frac{1}{\underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1} \underline{\mathcal{H}}}$
ML or MV	also, s is stationary	$\frac{\underline{\mathcal{H}}^\dagger \mathcal{S}_{xx}^{-1}}{\underline{\mathcal{H}}^\dagger \mathcal{S}_{xx}^{-1} \underline{\mathcal{H}}} = \frac{\underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1}}{\underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1} \underline{\mathcal{H}}}$	$\frac{1}{\underline{\mathcal{H}}^\dagger \mathcal{S}_{zz}^{-1} \underline{\mathcal{H}}}$

Table 3.1: Summary of Frequency Domain Optimum Processors. In addition to the listed assumptions, we always assume that target and interference random processes are zero-mean and independent.

does not depend on overall TJR but on the TJR within narrow frequency bands, which can only be changed if both the TJR and the weighting function involve substantial variation within these bands. In other words, only if the target and/or jammer spectra are "peaky" and the processor is capable of a "peaky" response can extra information about target and jammer lead to better performance. Whenever either the TJR or the processor response is "smooth", any of the processors derived in this section ought to produce equally-intelligible output.

Thus, except for targets or jammers with unusual spectra, the design of "optimum" processors (i.e., linear processors with squared-error type criteria) for improving intelligibility is relatively insensitive to many initial assumptions, such as target stationarity or Gaussian distribution of target and jammers. We need only assume zero-mean stationary interference and knowledge of $\underline{\mathcal{H}}$ and $\mathcal{S}_{zz}$. When $\mathcal{S}_{zz}$ is not known, $\mathcal{S}_{xx}$, which can be estimated from observations, will work as well, but at the cost of assuming a zero-mean, stationary target that is independent of the interference. The principle of processing with an estimated $\mathcal{S}_{xx}$ lies at the heart of many practical implementations.

3.2 Time-Domain Optimum Processors

Having derived many different optimum processors for observations over all time, we must now relate those results to our proposed processing architecture (described in Section 2.3), which uses observations over a limited time. In this section, we will derive optimum MMSE and MV time-domain processors for limited observations[4]. Echoing our results for frequency-domain processors, these time-domain processors will be quite similar, indicating that the exact choice of optimization criterion and *a priori* information may not be critical.

[4] The derivation of frequency-domain processors for limited observations is impractical because, for limited observations, the basis functions of the Karhunen-Loève transformation are signal dependent and often intractable (Van Trees, 1968). Thus, practicality (rather than fundamental principle) dictates the use of time-domain processing for limited observations and frequency-domain processing for unlimited observations. This dichotomy also appears in the optimum-filtering literature in the use of Kalman (time-domain) processing for limited observations and Wiener (frequency-domain) processing for unlimited observations (Anderson and Moore, 1979; Wiener, 1949).

Both MMSE and MV time-domain estimators will be based on the microphone samples available to the processor, which are described by the model in (2.49),

$$\mathbf{x}[n] = \mathbf{H}\mathbf{s}[n] + \mathbf{z}[n]\ ,$$

where $\mathbf{x}[n]$ is the vector of ML most recent microphone samples at sampling instant n (L past values for each of M microphones), $\mathbf{H}$ is a transfer function matrix, defined in (2.49), $\mathbf{s}[n]$ is a vector containing as much of the past target signal as can be observed, through $\mathbf{H}$, by the processor, and $\mathbf{z}[n]$ is the vector of ML most recent total (internal plus received) noise values.

The time-domain processors will produce a desired-signal estimate, $\hat{d}[n]$, by combining the observations, $\mathbf{x}[n]$, as described in equation (2.41),

$$\hat{d}[n] = y[n] = \mathbf{w}^T\mathbf{x}[n]\ .$$

The desired signal itself, $d[n]$, can be defined in terms of the target signal as

$$d[n] = \mathbf{f}^T\mathbf{s}[n] = \begin{bmatrix} f[0] & f[1] & \cdots \end{bmatrix} \begin{bmatrix} s[n] \\ s[n-1] \\ \vdots \end{bmatrix}\ . \tag{3.60}$$

In most cases the desired signal "filter", $\mathbf{f}$, will be no more than a delay, which allows us to compensate for delay in the transfer function $\mathbf{H}$ or even, by adding extra delay, to estimate a target sample based on observations of both past and future samples.

3.2.1 Minimum Mean-Square Error

The MMSE estimator is based on the assumption that both the target and interference are stationary, independent, zero-mean random processes with covariance matrices $R_{\mathbf{ss}}$ and $R_{\mathbf{zz}}$. The processor is designed to minimize the squared error

$$\begin{aligned}
\varepsilon^2 &= E\left\{\left(\hat{d}[n] - d[n]\right)^2\right\} \\
&= E\left\{\left(\mathbf{w}^T\mathbf{x}[n] - \mathbf{f}^T\mathbf{s}[n]\right)\left(\mathbf{x}^T[n]\,\mathbf{w} - \mathbf{s}^T[n]\,\mathbf{f}\right)\right\} \\
&= \mathbf{w}^T R_{\mathbf{xx}}\,\mathbf{w} - 2\,\mathbf{w}^T R_{\mathbf{xs}}\,\mathbf{f} + \mathbf{f}^T R_{\mathbf{ss}}\,\mathbf{f}\,. \qquad (3.61)
\end{aligned}$$

This quadratic form in $\mathbf{w}$ will be minimized when

$$\frac{\partial}{\partial \mathbf{w}}\left(\varepsilon^2\right)\bigg|_{\mathbf{w}=\mathbf{w}_{\mathrm{MMSE}}} \;=\; 2\,R_{\mathbf{xx}}\,\mathbf{w}_{\mathrm{MMSE}} - 2\,R_{\mathbf{xs}}\,\mathbf{f} \;=\; 0$$

which implies

$$\begin{aligned}
\mathbf{w}_{\mathrm{MMSE}} &= R_{\mathbf{xx}}^{-1}\,R_{\mathbf{xs}}\,\mathbf{f} \\
\mathbf{w}_{\mathrm{MMSE}}^T &= \mathbf{f}^T R_{\mathbf{sx}}\,R_{\mathbf{xx}}^{-1} && (3.62)\\
\mathbf{w}_{\mathrm{MMSE}}^T &= \mathbf{f}^T R_{\mathbf{ss}}\,\mathbf{H}^T\left(\mathbf{H}\,R_{\mathbf{ss}}\,\mathbf{H}^T + R_{\mathbf{zz}}\right)^{-1} && (3.63)\\
\mathbf{w}_{\mathrm{MMSE}}^T &= \mathbf{f}^T\left(R_{\mathbf{ss}}^{-1} + \mathbf{H}^T R_{\mathbf{zz}}^{-1}\,\mathbf{H}\right)^{-1}\mathbf{H}^T R_{\mathbf{zz}}^{-1}\,, && (3.64)
\end{aligned}$$

where (3.64) is derived from (3.63) by applying matrix identity (A.2). Thus, the MMSE processor depends on knowledge of $\mathbf{H}$, $R_{\mathbf{ss}}$, and $R_{\mathbf{zz}}$ (or $R_{\mathbf{xx}}$).

The performance of the MMSE processor can be evaluated by using $\mathbf{w}_{\mathrm{MMSE}}$ in the squared error equation (3.61):

$$\begin{aligned}
\varepsilon_{\mathrm{MMSE}}^2 &= \left(\mathbf{f}^T R_{\mathbf{sx}}\,R_{\mathbf{xx}}^{-1}\right) R_{\mathbf{xx}}\left(R_{\mathbf{xx}}^{-1}\,R_{\mathbf{xs}}\,\mathbf{f}\right) - 2\,\left(\mathbf{f}^T R_{\mathbf{sx}}\,R_{\mathbf{xx}}^{-1}\right) R_{\mathbf{xs}}\,\mathbf{f} + \mathbf{f}^T R_{\mathbf{ss}}\,\mathbf{f}\\
&= \mathbf{f}^T R_{\mathbf{ss}}\,\mathbf{f} - \mathbf{f}^T R_{\mathbf{sx}}\,R_{\mathbf{xx}}^{-1}\,R_{\mathbf{xs}}\,\mathbf{f}\\
&= \mathbf{f}^T\left(R_{\mathbf{ss}} - R_{\mathbf{ss}}\,\mathbf{H}^T\left(\mathbf{H}\,R_{\mathbf{ss}}\,\mathbf{H}^T + R_{\mathbf{zz}}\right)^{-1}\mathbf{H}\,R_{\mathbf{ss}}\right)\mathbf{f}\\
&= \mathbf{f}^T\left(R_{\mathbf{ss}}^{-1} + \mathbf{H}^T R_{\mathbf{zz}}^{-1}\,\mathbf{H}\right)^{-1}\mathbf{f}\,, && (3.65)
\end{aligned}$$

where the last expression is obtained by using matrix identity (A.1).

3.2.2 Minimum Variance Unbiased

The time-domain MV processor is based on the assumptions that the target signal, $\mathbf{s}[n]$, is completely unknown (as opposed to random), that the total noise, $\mathbf{z}[n]$, is a stationary zero-mean process, and that $\mathbf{H}$ and $R_{\mathbf{zz}}$ are known. The MV processor is designed to produce an unbiased desired-signal estimate, $\hat{d}_{\mathrm{MV}}[n]$, which must satisfy

$$E\left\{\hat{d}_{\mathrm{MV}}[n]\right\} = E\left\{\mathbf{w}_{\mathrm{MV}}^T\mathbf{x}[n]\right\} = E\left\{\mathbf{w}_{\mathrm{MV}}^T\left(\mathbf{H}\,\mathbf{s}[n] + \mathbf{z}[n]\right)\right\} = \mathbf{w}_{\mathrm{MV}}^T\mathbf{H}\,\mathbf{s}[n] = d[n]\,. \qquad (3.66)$$

For this to be true for any $\mathbf{s}[n]$, $\mathbf{w}_{\mathrm{MV}}^T$ must satisfy the constraint

$$\mathbf{w}_{\mathrm{MV}}^T \mathbf{H} = \mathbf{f}^T . \tag{3.67}$$

Subject to this constraint, the MV estimate must also minimize the variance

$$\begin{aligned} \varepsilon_{\mathrm{MV}}^2 &= E\left\{\left(\hat{d}_{\mathrm{MV}}[n] - d[n]\right)^2\right\} \quad = \quad E\left\{\hat{d}_{\mathrm{MV}}^2[n] - 2\,\hat{d}_{\mathrm{MV}}[n]\,d[n] + d^2[n]\right\} \\ &= E\left\{\hat{d}_{\mathrm{MV}}^2[n]\right\} - d^2[n] \\ &= E\left\{\mathbf{w}_{\mathrm{MV}}^T (\mathbf{H}\,\mathbf{s}[n] + \mathbf{z}[n]) \left(\mathbf{s}^T[n]\,\mathbf{H}^T + \mathbf{z}^T[n]\right) \mathbf{w}_{\mathrm{MV}}\right\} - d^2[n] \\ &= \mathbf{w}_{\mathrm{MV}}^T\, R_{\mathbf{zz}}\, \mathbf{w}_{\mathrm{MV}} \ , \end{aligned} \tag{3.68}$$

where we have used (3.66) and the fact that $\mathbf{z}[n]$ is zero-mean. The constrained minimization is performed by minimizing

$$\mathbf{w}_{\mathrm{MV}}^T\, R_{\mathbf{zz}}\, \mathbf{w}_{\mathrm{MV}} - 2\, \left(\mathbf{w}_{\mathrm{MV}}^T\, \mathbf{H} - \mathbf{f}^T\right) \lambda$$

which will be equivalent to (3.68) as long as the constraint is satisfied. Differentiating with respect to $\mathbf{w}_{\mathrm{MV}}$,

$$\begin{aligned} 2\, R_{\mathbf{zz}}\, \mathbf{w}_{\mathrm{MV}} - 2\,\mathbf{H}\,\lambda &= 0 \\ \mathbf{w}_{\mathrm{MV}} &= R_{\mathbf{zz}}^{-1}\,\mathbf{H}\,\lambda \\ \mathbf{w}_{\mathrm{MV}}^T &= \lambda^T\,\mathbf{H}^T R_{\mathbf{zz}}^{-1} . \end{aligned} \tag{3.69}$$

We can use the zero-bias constraint to solve for the Lagrange multiplier λ,

$$\begin{aligned} \mathbf{w}_{\mathrm{MV}}^T\,\mathbf{H} \quad = \quad \lambda^T\,\mathbf{H}^T R_{\mathbf{zz}}^{-1}\,\mathbf{H} &= \mathbf{f}^T \\ \lambda^T &= \mathbf{f}^T \left(\mathbf{H}^T R_{\mathbf{zz}}^{-1}\,\mathbf{H}\right)^{-1} , \end{aligned} \tag{3.70}$$

and substitute into (3.69),

$$\mathbf{w}_{\mathrm{MV}}^T \ = \ \mathbf{f}^T \left(\mathbf{H}^T R_{\mathbf{zz}}^{-1}\,\mathbf{H}\right)^{-1} \mathbf{H}^T R_{\mathbf{zz}}^{-1} , \tag{3.71}$$

which specifies the MV limited-observation time-domain processor. With a derivation similar to Appendix B it can also be shown that, if $\mathbf{s}[n]$ is a stationary random process,

$$\left(\mathbf{H}^T R_{\mathbf{zz}}^{-1}\,\mathbf{H}\right)^{-1} \mathbf{H}^T R_{\mathbf{zz}}^{-1} = \left(\mathbf{H}^T R_{\mathbf{xx}}^{-1}\,\mathbf{H}\right)^{-1} \mathbf{H}^T R_{\mathbf{xx}}^{-1} ,$$

and

$$\mathbf{w}_{\mathrm{MV}}^T = \mathbf{f}^T\left(\mathbf{H}^T R_{\mathbf{xx}}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T R_{\mathbf{xx}}^{-1}. \tag{3.72}$$

The importance of this alternative form is that $R_{\mathbf{xx}}$ can be estimated from the observed microphone signals. Thus, the MV processor depends on $\mathbf{H}$ and $R_{\mathbf{zz}}$ or, if $\mathbf{s}[n]$ is stationary, on $\mathbf{H}$ and $R_{\mathbf{xx}}$.

Finally, the performance of the MV processor is given by the squared error

$$\begin{aligned}\varepsilon_{\mathrm{MV}}^2 &= \mathbf{w}_{\mathrm{MV}}^T R_{\mathbf{zz}}\,\mathbf{w}_{\mathrm{MV}}\\ &= \mathbf{f}^T\left(\mathbf{H}^T R_{\mathbf{zz}}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T R_{\mathbf{zz}}^{-1}R_{\mathbf{zz}}R_{\mathbf{zz}}^{-1}\mathbf{H}\left(\mathbf{H}^T R_{\mathbf{zz}}^{-1}\mathbf{H}\right)^{-1}\mathbf{f}\\ &= \mathbf{f}^T\left(\mathbf{H}^T R_{\mathbf{zz}}^{-1}\mathbf{H}\right)^{-1}\mathbf{f}. \end{aligned} \tag{3.73}$$

Note that this error is equivalent to $\varepsilon_{\mathrm{MMSE}}^2$ with $R_{\mathbf{ss}} = \infty$, another way to express ignorance of the signal.

3.2.3 Summary

Criterion Assumptions	Processor $\left(\mathbf{w}^T\right)$	Performance (ε^2)
MMSE $\mathbf{H}$, $R_{\mathbf{ss}}$, $R_{\mathbf{zz}}$	$\mathbf{f}^T\left(R_{\mathbf{ss}}^{-1}+\mathbf{H}^T R_{\mathbf{zz}}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T R_{\mathbf{zz}}^{-1}$	$\mathbf{f}^T\left(R_{\mathbf{ss}}^{-1}+\mathbf{H}^T R_{\mathbf{zz}}^{-1}\mathbf{H}\right)^{-1}\mathbf{f}$
MV $\mathbf{H}$, $R_{\mathbf{zz}}$	$\mathbf{f}^T\left(\mathbf{H}^T R_{\mathbf{zz}}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T R_{\mathbf{zz}}^{-1}$	$\mathbf{f}^T\left(\mathbf{H}^T R_{\mathbf{zz}}^{-1}\mathbf{H}\right)^{-1}\mathbf{f}$
in addition, stationary $\mathbf{s}$	$\mathbf{f}^T\left(\mathbf{H}^T R_{\mathbf{xx}}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T R_{\mathbf{xx}}^{-1}$ $\equiv \mathbf{f}^T\left(\mathbf{H}^T R_{\mathbf{zz}}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T R_{\mathbf{zz}}^{-1}$	$\mathbf{f}^T\left(\mathbf{H}^T R_{\mathbf{zz}}^{-1}\mathbf{H}\right)^{-1}\mathbf{f}$

Table 3.2: Summary of Time-Domain Optimum Processors. In addition to the listed assumptions, we always assume that target and interference random processes are zero-mean and independent.

Table 3.2 summarizes the processors derived in this section[5]. The processors differ only in the $R_{\mathbf{ss}}$ term that incorporates *a priori* target information. This matrix factor could significantly modify the processing but, since the corresponding factor in the frequency-domain processor only added a frequency-dependent weighting, we will presume that this time-domain factor has a similar effect. Certainly this must be true in the limit of long observations since the time-domain processor must approach frequency-domain processing (Gray, 1972).

If our presumption is correct, time-domain and frequency-domain processors to minimize interference are equally insensitive to initial assumptions, such as target stationarity or Gaussian distribution of target and jammers. The minimum necessary assumptions are zero-mean stationary interference and, in the time-domain, knowledge of $\mathbf{H}$ and $R_{\mathbf{zz}}$. When $R_{\mathbf{zz}}$ is not known, $R_{\mathbf{xx}}$, which can be estimated from observations, will work as well, but at the cost of assuming a zero-mean, stationary target that is independent of the interference. The adaptive beamformer described in Chapter 5 is based on this final scheme for processing with an estimated $R_{\mathbf{xx}}$.

3.3 Comparison of Frequency- and Time-Domain Processing

Clearly, a multi-microphone hearing-aid must be based on limited observations and time-domain optimum processing. However, frequency-domain results are easier to derive and interpret and, since processing in the two domains must be asymptotically equivalent as the observation time becomes long, it would be convenient to use frequency-domain techniques to derive bounds on limited-observation processors. These bounds would be more meaningful if we knew how many observations were necessary for a time-domain processor to approach the performance of the corresponding frequency-domain processor. It is likely that the answer to this question is situation-dependent, but a simple example may shed some light on the different

[5] MAP, SNR (perhaps), and ML processors could also be derived but, as for frequency-domain processors, they would not be unique. In the interest of brevity the derivations are not included.

processing methods and on their asymptotic equivalence.

Consider an anechoic room with a target and one off-axis jammer. A two-microphone array is mounted in free space oriented broadside to the target. The jammer is one sample-time closer to microphone 1 than to microphone 2 or, in other words, a jammer impulse will arrive one sample-time sooner in microphone 1 than in microphone 2. Sensor noise is present in both microphones and its spectral-density function is β times that of the received-jammer. Finally, the two microphone signals are processed by a Minimum-Variance processor that assumes a target transfer function of unity.

When this example is worked out in detail, the asymptotic (i.e., frequency-domain) ratio of output to input noise power is approximately β, which is what one might expect if the jammer were cancelled completely and only sensor noise remained. For time-domain processors limited to L samples per microphone, the ratio of output to input noise power is approximately $1/L$. For sensor-to-directional noise ratios, β, of -10, -20, and -30 dB, the number of time-domain samples, L, necessary to attain the sensor-noise performance limit would then be 10, 100, and 1000, respectively.

Chapter 4

Optimum Performance

In this chapter we analyze the frequency-domain performance of the MV processor for a number of special cases where we can determine $\mathcal{S}_{zz}(f)$ and $\underline{\mathcal{H}}(f)$. Since adaptive beamformers approach optimum MV performance (as will be shown), the results of this chapter represent the best possible performance that an adaptive beamforming hearing aid could achieve in the cases that we study. These cases cannot be exhaustive, however, and some very significant factors, such as head-shadow and complex array configurations, have not been considered in order to make the analysis feasible. Consequently, the performance bounds that we develop will be most valuable not in any absolute sense, but in evaluating the relative effects of various environmental factors and processing configurations. At the end of this chapter we consider, in a non-rigorous fashion, the possible effects of head-shadow and other microphone arrangements.

In most configurations that we analyze, the target source is located in anechoic space "straight-ahead" of the array[1] and in its far-field, so that received target signals have equal magnitudes but phases that differ according to each microphone's displacement in the target direction relative to the array center. We choose to estimate the target signal as it would be measured at the array center, which makes

[1] We will always assume that the array microphones are omnidirectional, are coupled poorly enough to the field that inter-microphone loading effects are negligible (Beranek, 1988), and are small enough that scattering can be ignored. Scattering at 5 KHz would cause field perturbations 3 cm away from a 1-cm microphone of only -13 dB (Morse, 1976). At 1.6 KHz the perturbations would be -32 dB. In either case, the presence of a nearby head would introduce much more dramatic effects.

$\underline{\mathcal{H}}(f)$ especially simple:

$$\underline{\mathcal{H}}(f) = \begin{bmatrix} \vdots \\ e^{j2\pi f r_{mx}/c} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ e^{j\phi_m(f)} \\ \vdots \end{bmatrix}, \tag{4.1}$$

where c is the velocity of sound and r_{mx} is the displacement of the mth microphone in the positive-x direction (defined to be the target direction). This formulation does not include head-shadow, which would introduce additional amplitude and phase factors in $\underline{\mathcal{H}}(f)$.

To characterize performance we use response measures introduced in section 2.3.3, such as: $G_z(f)$, $G_i(f)$ and $G_j(f)$, array gains against total noise, isotropic noise, and directional-jammer noise, respectively; $K_z(f)$ and $K_j(f)$, the array responses to total noise and to directional jammers; and $K_u(f)$, the array response to uncorrelated sensor noise, which we will simply call the array *noise sensitivity.* The various gain measures predict the benefit from processing when the actual $\mathcal{S}_{\mathbf{zz}}(f)$ and $\underline{\mathcal{H}}(f)$ match our assumptions, while noise sensitivity indicates not only the sensitivity of the processor to sensor noise but also, in some sense, its sensitivity to deviations of $\mathcal{S}_{\mathbf{zz}}(f)$ and $\underline{\mathcal{H}}(f)$ from our assumptions[2].

We can expand the relevant array response and array gain definitions in terms

[2] Random perturbations of $\underline{\mathcal{H}}(f)$ can be thought of as adding random errors to the received target signal or, equivalently, adding an extra received-noise component (Cox, 1973b; Cox, Zeskind and Kooij, 1986). Similarly, errors in $\mathcal{S}_{\mathbf{zz}}(f)$ can be thought of as caused by an extra received-noise component. To the extent that such virtual noise signals are white (which will depend on the perturbation mechanisms), sensitivity to perturbations is predicted by noise sensitivity $K_u(f)$.

of $\mathcal{S}_{zz}$ and $\underline{\mathcal{H}}$ as follows.

$$\begin{aligned}
K_r(f) &= \frac{\left|\underline{\mathcal{W}}^T(f)\,\underline{\mathcal{H}}(f)\right|^2}{\left|\underline{\mathcal{H}}(f)\right|^2/M} = \frac{\left|\left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^\dagger\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\right|^2}{\underline{\mathcal{H}}^\dagger\,\underline{\mathcal{H}}/M} = 1 && (4.2)\\
K_z(f) &= \frac{\underline{\mathcal{W}}^T(f)\,\mathcal{S}_{zz}(f)\,\underline{\mathcal{W}}^*(f)}{\mathrm{trace}(\mathcal{S}_{zz}(f))/M}\\
&= \frac{\left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^\dagger\,\mathcal{S}_{zz}^{-1}\,\mathcal{S}_{zz}\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\right)^{-1}}{\mathrm{trace}(\mathcal{S}_{zz})/M}\\
&= \frac{\left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\right)^{-1}}{\mathrm{trace}(\mathcal{S}_{zz})/M} && (4.3)\\
G_z(f) &= \frac{K_r(f)}{K_z(f)} = \frac{\left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\right)\mathrm{trace}(\mathcal{S}_{zz})}{M} && (4.4)\\
K_u(f) &= \underline{\mathcal{W}}^T(f)\,\underline{\mathcal{W}}^*(f) = \left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^\dagger\,\mathcal{S}_{zz}^{-1}\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\left(\underline{\mathcal{H}}^\dagger\,\mathcal{S}_{zz}^{-1}\,\underline{\mathcal{H}}\right)^{-1} && (4.5)
\end{aligned}$$

Analogous expressions for $G_i(f)$, $G_j(f)$, and $K_j(f)$ will be determined as needed.

4.1 Uncorrelated Noise

The total noise signals are uncorrelated between microphones in two situations: when there are no jamming sources and only receiver noise is present, and when the number of jammers and/or the amount of reverberation is great enough to create an isotropic noise field[3] and, in addition, the ratio of microphone spacing to sound wavelength is large. As an example of the second situation, measurements in a large reverberant room with two microphones located near the ears of a human head have shown inter-microphone correlations of zero above 1 KHz (Lindevald and Benade, 1986).

To capture the essential characteristics of these two situations, we can look at the simple case of independent white noise of power σ_u^2 per sample at each microphone,

[3] To be discussed in greater detail in the next section.

for which

$$R_{zz}[k] = \begin{bmatrix} \sigma_u^2 & 0 & \cdots & 0 \\ 0 & \sigma_u^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_u^2 \end{bmatrix} \delta[k] = \sigma_u^2 \boldsymbol{I}_M \, \delta[k] \tag{4.6}$$

$$\mathcal{S}_{zz}(f) = \mathcal{P}_u(f)\, \boldsymbol{I}_M = \sigma_u^2 \boldsymbol{I}_M \,. \tag{4.7}$$

Substituting into optimum weight equation (3.57),

$$\begin{aligned} \underline{\mathcal{W}}^T_{\mathrm{MV}}(f) &= \left(\underline{\mathcal{H}}^\dagger \frac{1}{\sigma_u^2} \boldsymbol{I}_M \, \underline{\mathcal{H}}\right)^{-1} \underline{\mathcal{H}}^\dagger \frac{1}{\sigma_u^2} \boldsymbol{I}_M \\ &= \left(\underline{\mathcal{H}}^\dagger \underline{\mathcal{H}}\right)^{-1} \underline{\mathcal{H}}^\dagger \\ &= \left(\begin{bmatrix} \cdots & e^{-j\phi_m(f)} & \cdots \end{bmatrix} \begin{bmatrix} \vdots \\ e^{j\phi_m(f)} \\ \vdots \end{bmatrix} \right)^{-1} \begin{bmatrix} \cdots & e^{-j\phi_m(f)} & \cdots \end{bmatrix} \\ &= \frac{1}{M} \begin{bmatrix} \cdots & e^{-j\phi_m(f)} & \cdots \end{bmatrix} . \end{aligned} \tag{4.8}$$

The essence of this processor (which should not be surprising) is to compensate for the arrival phase of the target component in each microphone signal and then average the phase-aligned signals.

To characterize MV performance against uncorrelated noise, we evaluate G_z and K_u:

$$G_z = \frac{1}{M} \left(\underline{\mathcal{H}}^\dagger \frac{1}{\sigma_u^2} \boldsymbol{I}_M \, \underline{\mathcal{H}}\right) \mathrm{trace}(\sigma_u^2 \boldsymbol{I}_M) = M \tag{4.9}$$

$$K_u = \frac{1}{M} \begin{bmatrix} \cdots & e^{-j\phi_m(f)} & \cdots \end{bmatrix} \begin{bmatrix} \vdots \\ e^{j\phi_m(f)} \\ \vdots \end{bmatrix} \frac{1}{M} = \frac{1}{M} . \tag{4.10}$$

Once again, these results should not be surprising. When M microphones are averaged together, white noise power is reduced by a factor of M and, if only uncorrelated noise is present, the output signal-to-noise ratio is increased by a factor of M.

4.2 Isotropic Noise

Isotropic noise, which can be defined as the superposition of independent plane waves with identical spectra and uniformly distributed incident angles, is an interesting environment for at least two reasons. First, the ***diffuse sound field*** in reverberant environments, which is composed of all reflected sounds, can be quite isotropic (Beranek, 1954; Cook, Waterhouse, Berendt, et al., 1955; Lindevald and Benade, 1986). Consequently, performance against isotropic noise represents the limiting performance of adaptive beamformers in environments where diffuse, reverberant energy dominates the received signal. Second, an array designed to perform optimally against isotropic noise has maximum directivity (as defined in section 2.3.3), a common requirement for non-adaptive multimicrophone receivers. Therefore, the methods and results of this section can be applied to the design and analysis of fixed-weight arrays and, in particular, can be used to determine the best possible non-adaptive system for comparison with our adaptive systems.

In an isotropic noise field the cross-spectral-density function for two points separated in space by distance d is (Cook, Waterhouse, Berendt, et al., 1955; Cron and Sherman, 1962; Baggeroer, 1976):

$$\mathcal{S}(f,d) = \mathcal{P}_i(f)\,\mathrm{sinc}(2\pi f d/c)\ , \tag{4.11}$$

where $\mathcal{P}_i(f)$ is the common source spectral-density function. The spectral density matrix for isotropic noise incident on an array is then

$$\mathcal{S}_{ii}(f) = \mathcal{P}_i(f)\ \mathrm{sinc}\left(\frac{2\pi f}{c}\begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1M} \\ d_{21} & d_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ d_{M1} & \cdots & \cdots & d_{MM}\end{bmatrix}\right) = \mathcal{P}_i(f)\ \mathrm{sinc}\left(\frac{2\pi f}{c}\boldsymbol{D}\right)\ , \tag{4.12}$$

where d_{ij} is the distance between microphones i and j. Note that when this matrix is used in (4.4) and (4.5) to calculate $G_z(f)$ and $K_u(f)$, the source spectrum $\mathcal{P}_i(f)$ cancels out and we can disregard it (or, equivalently, assume that it is unity). The remaining structure of $\mathcal{S}_{ii}(f)$ is determined by array geometry alone.

4.2.1 Fundamental Performance Limits

Table 4.1 summarizes known limits on optimum performance in isotropic noise for microphone arrays with different geometries. The linear and ring arrays are composed of equispaced sensors and performance depends on the number of microphones, M, on the ratio of inter-microphone spacing to sound wavelength, $d/\lambda = df/c$, and on the orientation of the array relative to the target source, which can vary between *broadside* and *endfire* (or *edgefire* for the ring). The spherically symmetric array can be analyzed without assuming regularity of sensor spacing beyond that required for symmetry. In this case d is not well defined, but it turns out that performance does not depend on microphone spacing when the noise is

Geometry	Element Spacing	
	$d \ll \lambda/2$	large d
Linear Endfire	$G_z = M^2$ $\lim_{d/\lambda \to 0} K_u = \infty$	$G_z = M$ $K_u = 1/M$
Linear Broadside	$G_z = \prod_{m=1}^{\lfloor \frac{M-1}{2} \rfloor} \left(\frac{2m+1}{2m}\right)^2 \simeq \frac{4\lfloor \frac{M-1}{2} \rfloor + 3}{\pi}$ $\lim_{d/\lambda \to 0} K_u = \infty$	$G_z = M$ $K_u = 1/M$
Ring Edgefire	$G_z \simeq 0.53\,(M+1)^{3/2}$	$G_z = M$ $K_u = 1/M$
Spherically Symmetric	$G_z = M$	$G_z = M$ $K_u = 1/M$

Table 4.1: Fundamental limits on the performance of M-element microphone arrays in isotropic noise for various geometries, orientations, and element spacings. By "large" d we mean $d \gg \lambda/2$ or d equal to an integer multiple of $\lambda/2$.

isotropic[4].

When microphones are spaced more than a few wavelengths apart (or at exact multiples of $\lambda/2$), the isotropic noise is uncorrelated between sensors, a situation equivalent to the case analyzed in the previous section, and gain is equal to M with noise sensitivity of $1/M$, regardless of geometry or orientation. When microphone spacing approaches zero, performance does *not*, as one might expect, approach that of a single microphone but, rather, becomes *superdirective*, with gain greater than unity and high sensitivity to noise (Hansen, 1981). Small endfire arrays are theoretically capable of M^2 gain (Uzkov, 1946; Weston, 1986), but, as spacing approaches zero, their noise sensitivity (i.e. K_u or the squared magnitude of $\underline{W}(f)$) grows without bound (Chu, 1948), which imposes practical limits on performance (Taylor, 1948; Newman and Shrote, 1982). Broadside arrays of more than two microphones can also be superdirective, with similar sensitivity problems, but their gain is only proportional to M (Pritchard, 1954; Vanderkulk, 1963). Rings of equispaced microphones exhibit gain proportional to $M^{3/2}$ at small separations (Vanderkulk, 1963), a dependence between that of broadside and endfire linear arrays, and, although noise sensitivity has not been calculated directly, other performance measures indicate sensitivity similar to that of linear arrays. The spherically symmetric array can be shown to have a gain of M averaged over all orientations of the array, regardless of array size (Vanderkulk, 1963). For large microphone spacings this result is consistent with the fact that gain is M for any orientation. For small spacings, the average gain of M implies that an orientation *must* exist with *at least* a gain of M.

Comparing fundamental limits across geometries, all the arrays in Table 4.1 have identical performance when spacing is large (or when frequency is high for a given inter-microphone distance). When spacing is small (or frequency low for a given inter-microphone distance), equispaced linear endfire and broadside arrays represent two extremes of superdirective performance, although the practical significance of these performance limits is not clear because noise sensitivity can be so high.

[4] When sensor noise is considered, perfomance does depend on microphone spacing, but can be analyzed using average sensor density instead of sensor spacing (Vanderkulk, 1963).

4.2.2 Performance Limits for Hearing-Aid Arrays

To develop performance limits that are more relevant to the hearing aid application, we can analyze the performance of "head-sized", equispaced, linear, endfire and broadside arrays in the presence of both isotropic and sensor noise. Because the endfire and broadside orientations in some sense bound the performance of other geometries with equal numbers of microphones, this analysis is interesting in a fairly broad context. In addition, linear geometries are interesting because they are easy to analyze and construct, and non-trivial multimicrophone hearing-aids can be composed of linear, equispaced sub-arrays[5]. The inclusion of sensor noise is extremely important because its presence reduces the superdirectivity of the optimum processor. In fact, by using various levels of assumed sensor noise, we can generate "sub-optimum" array processors that trade superdirective gain for reduced noise sensitivity. We would hope that, over some range of assumed sensor noise levels, we could achieve moderate amounts of supergain with acceptably low noise sensitivity.

Figure 4.1 shows a detailed analysis of the performance of 4-microphone optimum linear endfire and broadside arrays in a fixed isotropic noise field with different amounts of assumed sensor noise. The total noise spectral matrix is given by

$$\mathcal{S}_{\boldsymbol{zz}}(f) = \mathcal{S}_{\boldsymbol{ii}}(f) + \sigma_u^2 \boldsymbol{I}_M = \mathcal{P}_i(f) \left(\operatorname{sinc}\left(\frac{2\pi f}{c} \boldsymbol{D} \right) + \beta(f) \boldsymbol{I}_M \right) , \qquad (4.13)$$

where $\beta(f) = \sigma_u^2 / \mathcal{P}_i(f)$ is the ratio of assumed sensor noise to isotropic noise. We will always use frequency-independent constants for β, thereby making the implicit assumption that both noises have the same spectral shape. Fortunately, none of our results are sensitive to this assumption. The plotted performance measures are noise sensitivity, K_u, and array gain against isotropic noise, G_i (equivalent to the directivity D defined in Section 2.3.3), which can be expressed as

$$G_i(f) \;=\; \frac{K_r(f)}{K_i(f)} = \frac{1}{K_i(f)} = \frac{\operatorname{trace}(\mathcal{S}_{\boldsymbol{ii}}(f))/M}{\underline{\mathcal{W}}^T(f)\,\mathcal{S}_{\boldsymbol{ii}}(f)\,\underline{\mathcal{W}}^*(f)} = \frac{1}{\underline{\mathcal{W}}^T\,\mathcal{S}_{\boldsymbol{ii}}\,\underline{\mathcal{W}}^*} . \qquad (4.14)$$

[5] The optimum weights for a composite array are not, in general, the composite of the optimum sub-array weights (Baggeroer, 1976), but many general insights that we develop in studying linear arrays will still apply to composite arrays and simple, sub-optimum combination rules can be used to establish lower bounds on optimum composite performance.

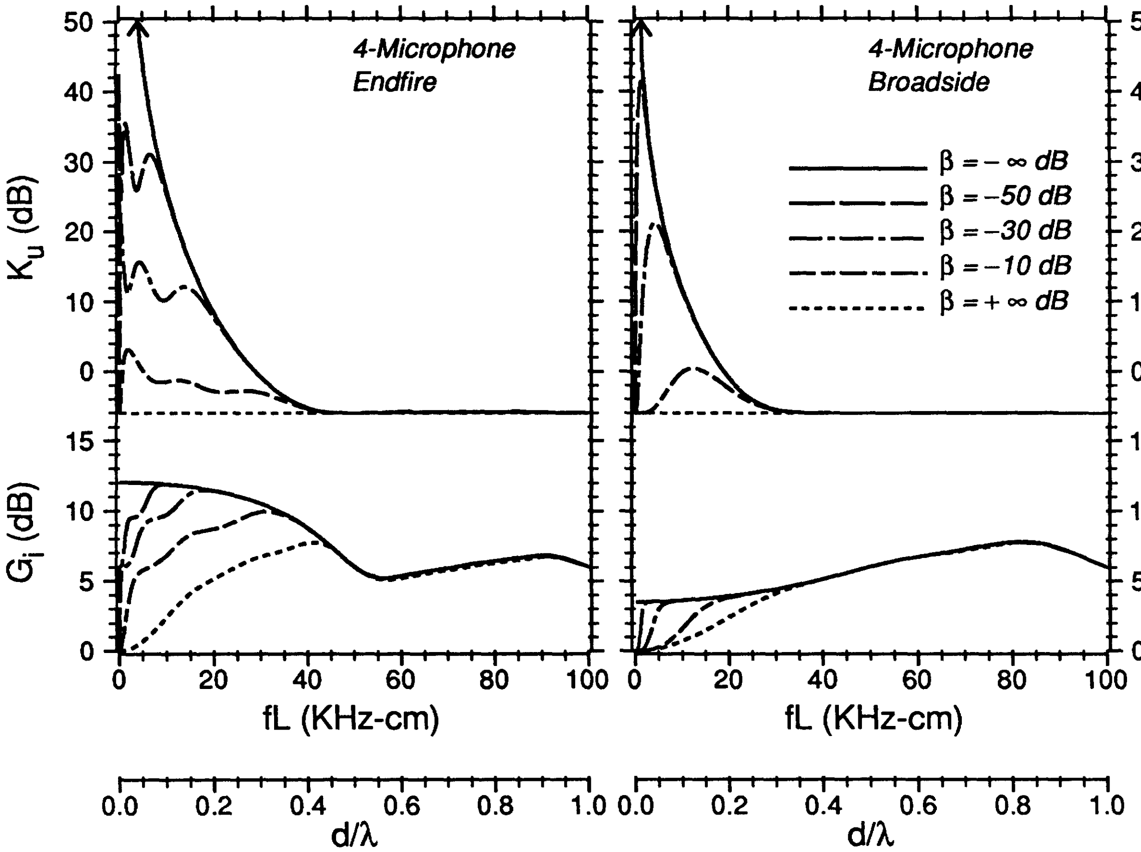


Figure 4.1: Gain against isotropic noise, G_i, and noise sensitivity, K_u, as a function of the frequency array-length product, fL, for linear arrays of 4 equispaced microphones optimized for various sensor-to-isotropic noise ratios, β. Sensor noise is assumed temporally and spatially white, i.e., uncorrelated between samples and between sensors. The two horizontal scales are equivalent, but note the difference in vertical scales.

For MV processing, $\underline{\mathcal{W}}^T = \left(\underline{\mathcal{H}}^\dagger \left(\mathcal{S}_{ii}(f) + \beta\, \boldsymbol{I}_M\right)^{-1} \underline{\mathcal{H}}\right)^{-1} \underline{\mathcal{H}}^\dagger \left(\mathcal{S}_{ii}(f) + \beta\, \boldsymbol{I}_M\right)^{-1}$.

When actual and assumed sensor noises are equal, the array gain against the total (isotropic plus sensor) noise can be expressed in terms of G_i, K_u, and β as

$$G_z(f) \;=\; \frac{K_r(f)}{K_z(f)} = \frac{1}{K_z} = \frac{1+\beta}{\underline{\mathcal{W}}^T \mathcal{S}_{ii} \underline{\mathcal{W}}^* + \beta\, \underline{\mathcal{W}}^T \underline{\mathcal{W}}^*} = \frac{1+\beta}{1/G_i + \beta\, K_u} \,. \tag{4.15}$$

Often, $\beta \ll 1$, $G_i\, \beta\, K_u < 1$, and $G_z \simeq G_i$. When actual and assumed sensor noise differ, the actual G_z can be determined by using G_i and K_u to compute separately the output noise components due to isotropic- and sensor-noise.

Since $\mathcal{S}_{ii}$, G_i, and K_u depend on the product $fd = cd/\lambda = fL/(M-1)$, performance in Figure 4.1 can be shown as a function of the ratio d/λ or as a function of the frequency array-length product, fL. These equivalent variables are both shown along the abscissa. When the length of a particular array is known, we can also interpret the curves as functions of frequency, where the frequency scale is determined by dividing fL by the actual array length. For example, the 0 to 5 KHz response of a 5-cm array is given by the segment of the response curve from $fL = 0$ to 25 KHz-cm.

Looking at the plots in detail, we first note that the limits of performance with no sensor noise ($\beta = -\infty$) correspond to the results in Table 4.1. Specifically, at $d/\lambda = 0.5$ and $d/\lambda = 1.0$ (and presumably at $d/\lambda \gg 1$), and in both orientations, gain approaches $M = 4$, or 6 dB, and noise sensitivity approaches $1/M$, or -6 dB. As $d/\lambda \rightarrow 0$, sensitivity becomes extremely high and gain approaches 16 (12 dB) for the endfire and 2.25 (3.5 dB) for the broadside orientation.

When sensor noise is much greater than isotropic noise (e.g., $\beta = \infty$), the total noise will be uncorrelated and the optimum weights will be uniform. Noise sensitivity is then -6 dB, independent of frequency, while G_i (i.e. directivity) exhibits the frequency dependence of a conventional, uniformly-weighted array. Below $d/\lambda = 0.4$, this dependence is approximately $4L/\lambda$ for the endfire and $2L/\lambda$ for the broadside array. Above $d/\lambda = 0.4$, G_i eventually stops rising due to the effects of spatial-undersampling, which is more detrimental in the endfire configuration.

The area at low frequencies where the curves diverge is the region of superdirective effects. Note that these effects are only observed when $d/\lambda < 0.5$, which is a general result (Vanderkulk, 1963) that is also observed on similar plots (not shown) for 2, 8, 12, and 16 microphones. As the level of assumed sensor noise, β, is increased, the optimum processor's noise sensitivity, K_u, and isotropic gain, G_i, both decrease, as expected. The decreases in G_i are larger at lower frequencies, with the result that, as sensor noise increases, superdirective gain disappears first at low frequencies. More importantly, the decreases in G_i are not proportional to the decreases in K_u and for some values of β, such as $\beta = -10$ dB, superdirective gain is substantial while noise sensitivity is not excessive.

At low frequencies it is also apparent that $K_u < 1/\beta$, or K_u (dB) $< -\beta$ (dB) . In other words $1/\beta$ functions as a "noise sensitivity limit". To see that this must be true in general, consider the case in which the received interference differs only slightly from received target (e.g., at very low fL both target and isotropic noise have intermicrophone correlations of about 1.0 with only slight phase differences). In this situation, the optimum processor will have to use very large weights because any interference cancellation will also cause some target cancellation and, to maintain a target gain of 1.0, the output target level can only be restored by increasing the magnitude of the weights. As the weights are increased, however, sensor noise is amplified in proportion. The optimum weighting is reached when the decrease in output interference noise is matched by the increase in output sensor noise. This is a complicated tradeoff but it would certainly never be advantageous to increase sensor noise by more than $1/\beta$ because then the output sensor noise would be greater than the non-processed interference and processing would be making matters worse. To a first approximation, then, assuming a relative sensor noise level of β in the design of an optimum processor is equivalent to setting a noise-sensitivity limit or weight-magnitude limit of $1/\beta$.

We can gain some insight into the mechanisms of superdirectivity and noise sensitivity by examining beam patterns for a few of the weightings represented in Figure 4.1. Figure 4.2 illustrates optimum endfire and broadside beam patterns for $d/\lambda = 0.1$ at four levels of sensor noise. The lowest value of β (-50 dB) is equivalent to no sensor noise at this value of d/λ and produces a maximally directive pattern. The highest value of β ($+\infty$ dB), as explained earlier, gives rise to uniform

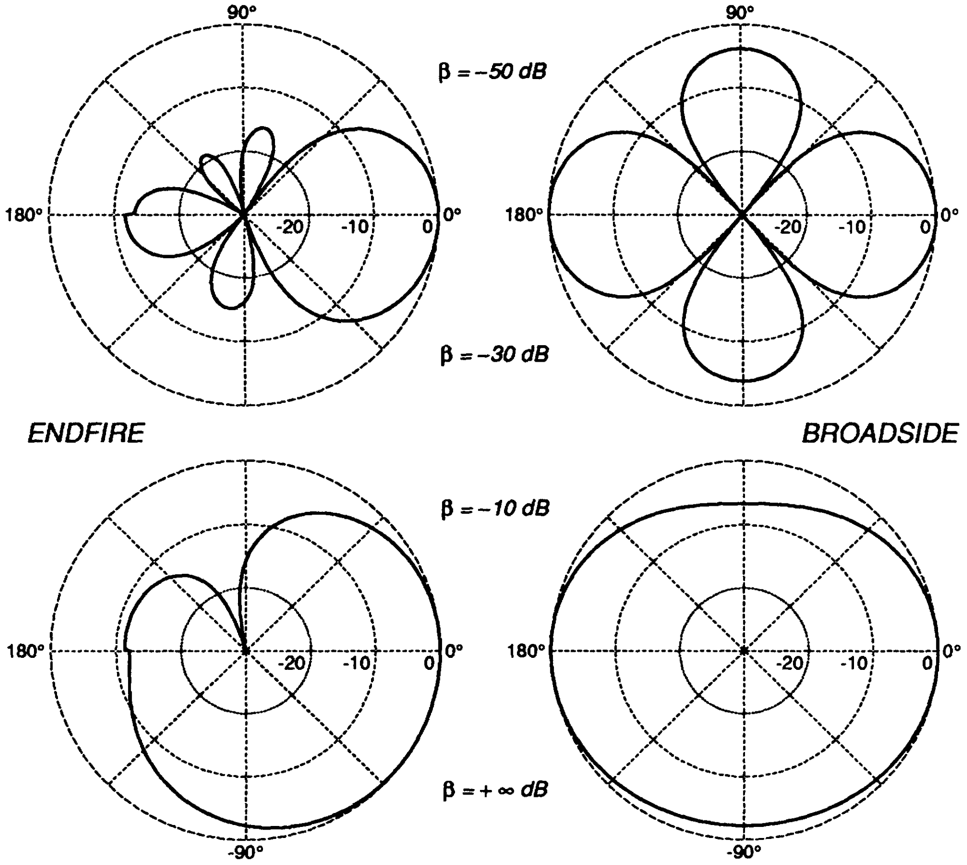


Figure 4.2: Beam patterns for 4-microphone endfire and broadside linear arrays at $d/\lambda = 0.1$ for various sensor-to-isotropic noise ratios, β. Only half of each pattern is shown (from 0° to 180° or from 180° to 0°) since the full patterns are always symmetric about the 0° axis.

weights and the beam pattern of a conventional, uniformly-weighted array. As sensor noise decreases and superdirective gain increases, the beam patterns exhibit more nulls, a behavior similar to that of more conventional arrays that have been "oversteered" or "steered past endfire" (Cox, Zeskind and Kooij, 1986). Steering refers to the process of compensating for relative propagation delays before adding the microphone signals in conventional beamformers. An array that has been steered to endfire has unity response in the target direction and off-axis response that generally falls with increasing angle from endfire. Oversteering is a technique for exagerrating the dependence of gain on angle by using greater than necessary delays at endfire. The result is a more directive beam pattern but less sensitivity in the target direction, which requires compensation with extra overall gain (i.e., larger weights). The larger weights increase sensitivity to uncorrelated noise.

The frequency-gain data in Figure 4.1 can be used to calculate intelligibility-averaged isotropic gain, $\langle G_i \rangle_I$, and thereby predict the performance of array processors in hearing-aid applications. To illustrate such a calculation, consider a broadside array with $L = 5$ cm, a frequency limit of 5 KHz, and processing optimized for $\beta = -30$ dB. For this array processor, the isotropic gain function, $G_i(f)$, from 0 to 5 KHz would correspond to the ($\beta = -30$ dB) G_i curve in Figure 4.1 from $fL = 0$ to 25 KHz-cm. Using this frequency response and equation (2.56) to calculate $\langle G_i \rangle_I$ provides a measure of the intelligibility benefit (about 2.3 dB) that such an array processor could provide in isotropic noise.

Figure 4.3 shows intelligibility-averaged isotropic gain for linear arrays of 2, 4, 8, 12, and 16 microphones in endfire and broadside orientations as a function of array length. Each panel contains gain functions for a different sensor-to-isotropic noise ratio, β. When $\beta = \infty$ dB (i.e. sensor noise dominates), the optimum processor corresponds to a conventional delay-and-sum beamformer, for which isotropic gain (i.e. directivity) will be small at these array sizes. For all other values of β, the optimum processor weightings are, to a greater or lesser extent, superdirective. When $\beta = -\infty$ dB (i.e. no sensor noise), the optimum processor approaches the fundamental limits of Table 4.1, and gains are highest, approaching M^2 for short endfire arrays. The left three panels attempt to show how isotropic gain is affected

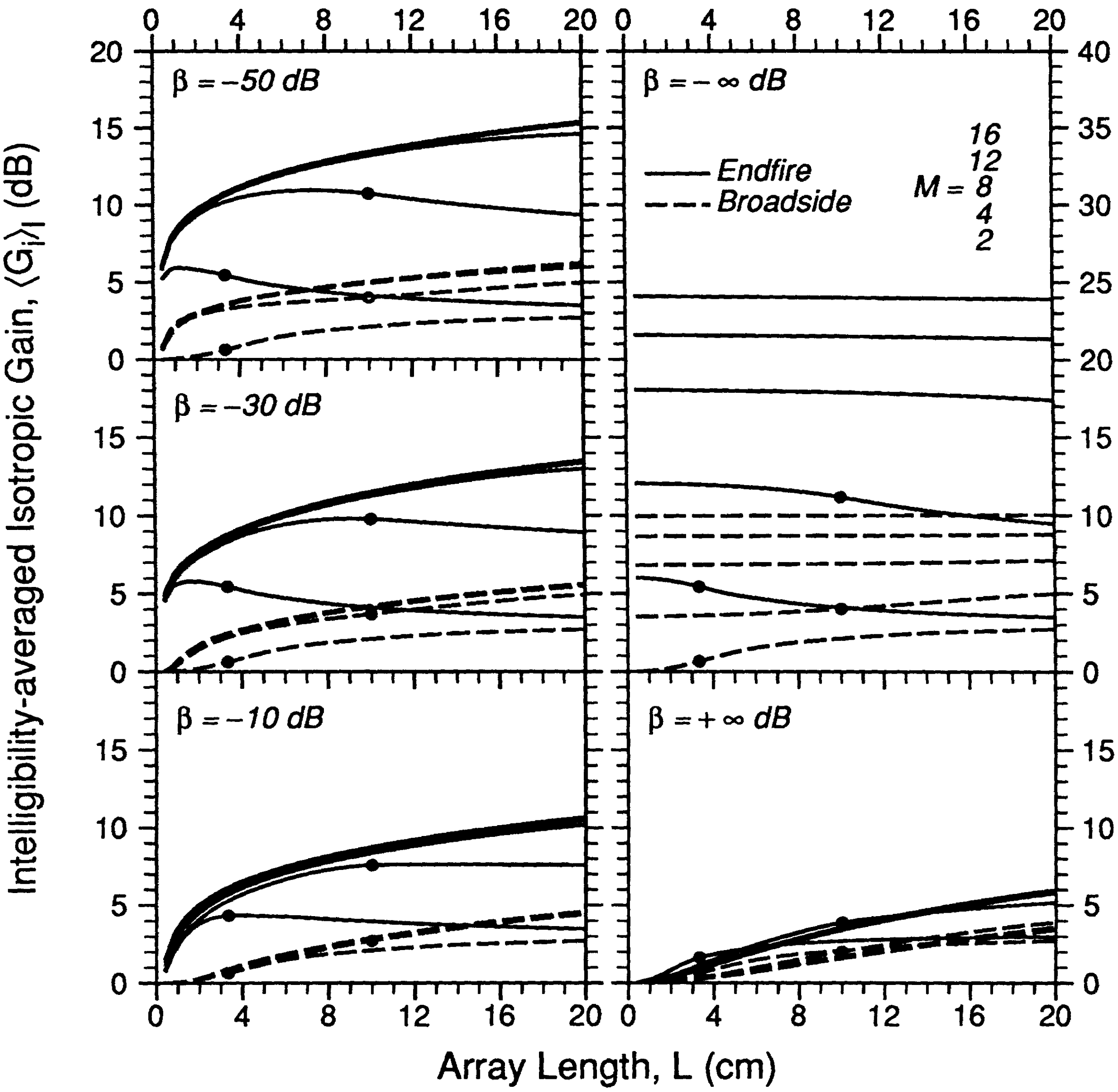


Figure 4.3: Intelligibility-averaged isotropic gain, $\langle G_i \rangle_I$, for endfire and broadside linear arrays limited to $f_{\text{MAX}} = 5$ KHz, as a function of array length, L. Each panel presents the data for a different sensor-to-isotropic noise ratio, β. Each family of curves represents the performance for M, the number of microphones, equal to 2, 4, 8, 12 and 16. The higher curves (better performance) always correspond to larger M (except for $\beta = \infty$ dB). The circles indicate points on the 2- and 4-microphone curves at which the inter-microphone spacing equals $\lambda_{\text{MAX}}/2$. For larger spacings the sound field is spatially undersampled.

by differing levels of assumed sensor noise, β (which serves to limit noise sensitivity, K_u, to less than $1/\beta$). From this figure we can make the following observations concerning linear arrays operating to improve target speech intelligibility in isotropic noise.

- Superdirective weightings, even with β as large as -10 dB, significantly outperform conventional weightings.
- Except for short arrays optimized for $\beta = \infty$ dB (i.e., with conventional weightings), performance always increases with number of microphones.
- Whenever sensor noise is present (i.e. $\beta > -\infty$ dB), the incremental improvement in performance resulting from additional microphones becomes insignificant beyond roughly 4 to 8 microphones for "head-sized" arrays.
- For short arrays with M held constant, endfire configurations significantly outperform broadside configurations.
- For long arrays with M held constant and $d \gg \lambda_{\text{MAX}}/2$, endfire and broadside performance tend to be roughly comparable. ($\lambda_{\text{MAX}} = c/f_{\text{MAX}}$)
- Long arrays generally outperform short arrays except for sparse endfire configurations where $d > \lambda_{\text{MAX}}/2$. In other words, for the given range of L, spatial undersampling ($d > \lambda_{\text{MAX}}/2$) is detrimental to endfire but not to broadside arrays.

In general, these results suggest that optimum array processing can provide significant benefit for head-sized arrays, with acceptable noise sensitivity, even in isotropic noise. Of course, very few interference environments are completely isotropic and it may be necessary to modify some of our conclusions after considering other interference processes.

4.3 Directional Noise

Hearing-aid users encounter many sources of interference (e.g., noisy appliances or competing speech from talkers, televisions, or radios) that are spatially localized and generate direct signals that propagate across the array from one direction. As we have already demonstrated in section 2.3.3, the potential benefit of adaptive array processing is especially great in reducing such directional interference. In this section we will analyze optimum performance in the presence of direct signals from localized sources of interference. In the next section we will consider the effects of signal reflections.

The direct signal from a spatially-localized source is completely correlated from microphone to microphone (i.e., knowledge of the received signal at one microphone and of the inter-microphone transfer function is sufficient to predict the signal at the other microphone). This can be seen by calculating the cross-spectral density matrix for an M-microphone array in the directional-noise field generated by J jammers in an anechoic (i.e., direct signal only) environment. Extending the notation of section 2.3.1, let $\underline{\mathcal{H}}_j(f)$ be the M-vector of transfer functions from jammer j, as observed at the array center, to each of the M microphones, analogous to our use of $\underline{\mathcal{H}}(f)$ in equations (2.14) and (4.1) to represent target transfer functions. If we assume that jammer j is in the array's far field, each transfer function depends on the jammer propagation vector, $\vec{\alpha}_j$ [as defined for equation (2.51)], and on microphone location, $\vec{r}_m$[6]:

$$\underline{\mathcal{H}}_j(f) = \begin{bmatrix} \vdots \\ e^{-j2\pi f \tau_{mj}} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ e^{-j\phi_{mj}(\vec{r}_m, \vec{\alpha}_j, f)} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ e^{-j2\pi f \vec{r}_m \cdot \vec{\alpha}_j / c} \\ \vdots \end{bmatrix} . \qquad (4.16)$$

The received signal from the J jammers is then

$$\underline{\mathcal{V}}(f) = \sum_{j=1}^{J} \underline{\mathcal{H}}_j(f)\, \mathcal{S}_j(f) \qquad (4.17)$$

[6] For jammers in the array's near field, the plane-wave assumption does not hold and the transfer functions will be influenced by differences in source distances and directions. Such details can be taken into account when necessary and do not alter our basic conclusions.

and, assuming statistically independent jammers, the cross-spectral-density matrix is

$$\mathcal{S}_{vv}(f) = E\left\{\underline{\mathcal{V}}(f)\,\underline{\mathcal{V}}^{\dagger}(f)\right\} = \sum_{j=1}^{J} \mathcal{S}_{jj}(f)\,\underline{\mathcal{H}}_j(f)\,\underline{\mathcal{H}}_j^{\dagger}(f)\,. \tag{4.18}$$

If the only noise present is this directional interference, and we define $\underline{\mathcal{H}}_0(f) \triangleq \underline{\mathcal{H}}(f)$ and $\mathcal{S}_0(f) \triangleq \mathcal{S}(f)$, the observation equation reduces to

$$\underline{\mathcal{X}}(f) = \sum_{j=0}^{J} \underline{\mathcal{H}}_j(f)\,\mathcal{S}_j(f)\,. \tag{4.19}$$

When the transfer functions are known, this matrix equation is simply a system of M equations in $J+1$ unknowns, which can be solved for all of the source signals if $J+1 \leq M$. In other words, if we have as many microphones as independent jammer and target sources, it should be possible to separate all of the sources perfectly.

In practice, of course, no noise field is perfectly directional because, even in an anechoic environment, there is always at least some receiver noise. Our analysis will focus on the optimum performance of an array in the presence of both directional interference and uncorrelated sensor noise. The array's total-noise cross-spectral-density matrix is then

$$\mathcal{S}_{zz}(f) = \boldsymbol{\mathcal{S}_{vv}} + \boldsymbol{\mathcal{S}_{uu}} = \sum_{j=1}^{J} \underline{\mathcal{H}}_j\,\underline{\mathcal{H}}_j^{\dagger}\mathcal{S}_{jj} + \sigma_u^2 \boldsymbol{I}_M\,. \tag{4.20}$$

To simplify the analysis, we consider J jammers with white spectra (similar to the sensor noise) and equal powers that sum to a constant, $\mathcal{P}_j$, that is independent of the number of jammers. If we continue to use β to denote the ratio of sensor-to-received noise,

$$\beta = \sigma_u^2 / \mathcal{P}_j\,, \tag{4.21}$$

and

$$\mathcal{S}_{zz}(f) = \mathcal{P}_j\left(\frac{1}{J}\sum_{j=1}^{J} \underline{\mathcal{H}}_j\,\underline{\mathcal{H}}_j^{\dagger} + \beta \boldsymbol{I}_M\right)\,. \tag{4.22}$$

We can then write specific expressions for the total noise response K_z, the gain against total noise G_z, and the gain against directional jammers G_j.

$$\begin{aligned}
K_z(f) &= \frac{\underline{\mathcal{W}}^T \mathcal{S}_{zz} \underline{\mathcal{W}}^*}{\mathrm{trace}\mathcal{S}_{zz}/M} = \frac{M}{\left(\underline{\mathcal{H}}^\dagger \, \mathcal{S}_{zz}^{-1} \, \underline{\mathcal{H}}\right) \mathrm{trace}\mathcal{S}_{zz}} \\
&= \frac{M}{\left(\underline{\mathcal{H}}^\dagger \left(\frac{1}{J}\sum_{j=1}^{J} \underline{\mathcal{H}}_j \, \underline{\mathcal{H}}_j^\dagger + \beta \boldsymbol{I}_M\right)^{-1} \underline{\mathcal{H}}\right) \mathrm{trace}\left(\frac{1}{J}\sum_{j=1}^{J} \underline{\mathcal{H}}_j \, \underline{\mathcal{H}}_j^\dagger + \beta \boldsymbol{I}_M\right)} \\
&= \frac{1}{\left(\underline{\mathcal{H}}^\dagger \left(\frac{1}{J}\sum_{j=1}^{J} \underline{\mathcal{H}}_j \, \underline{\mathcal{H}}_j^\dagger + \beta \boldsymbol{I}_M\right)^{-1} \underline{\mathcal{H}}\right) (1+\beta)} \qquad (4.23)
\end{aligned}$$

$$G_z(f) = \frac{1}{K_z(f)} = (1+\beta) \left(\underline{\mathcal{H}}^\dagger \left(\frac{1}{J}\sum_{j=1}^{J} \underline{\mathcal{H}}_j \, \underline{\mathcal{H}}_j^\dagger + \beta \boldsymbol{I}_M\right)^{-1} \underline{\mathcal{H}}\right) \qquad (4.24)$$

$$\begin{aligned}
G_j(f) &= \frac{\mathrm{trace}\left(\frac{1}{J}\sum_{j=1}^{J} \underline{\mathcal{H}}_j \, \underline{\mathcal{H}}_j^\dagger\right)/M}{\underline{\mathcal{W}}^T \left(\frac{1}{J}\sum_{j=1}^{J} \underline{\mathcal{H}}_j \, \underline{\mathcal{H}}_j^\dagger\right) \underline{\mathcal{W}}^*} \\
&= \frac{(1+\beta)}{\left(\underline{\mathcal{H}}^\dagger \, \mathcal{S}_{zz}^{-1} \, \underline{\mathcal{H}}\right)^{-1} \underline{\mathcal{H}}^\dagger \, \mathcal{S}_{zz}^{-1} \left(\frac{1}{J}\sum_{j=1}^{J} \underline{\mathcal{H}}_j \, \underline{\mathcal{H}}_j^\dagger\right) \mathcal{S}_{zz}^{-1} \, \underline{\mathcal{H}} \left(\underline{\mathcal{H}}^\dagger \, \mathcal{S}_{zz}^{-1} \, \underline{\mathcal{H}}\right)^{-1}} \qquad (4.25)
\end{aligned}$$

Note that the scale factor $\mathcal{P}_j$, the total jammer power, always cancels out. This would not be the case if we were using MAP or MMSE estimators, which make use of information about relative target and jammer levels.

4.3.1 Performance against One Directional Jammer

Before considering the problem of multiple jammers, we will analyze the simpler single-jammer case. Figure 4.4 shows an analysis of the performance of endfire and broadside 2-microphone arrays in the presence of sensor noise and one directional jammer incident from an angle of 45°. The figure presents G_j, G_z, and K_u as a function of fL for three different values of β.

At $fL = 0$ (and at any value of fL where the jammer arrives with multiples of 360° phase shift between microphones), the jammer is indistinguishable from target and cannot be cancelled. The processor adds the microphone signals with equal weights, the jammer is not attenuated ($G_j = 0$), and the sensor noise is reduced by 3 dB ($K_u = -3$ dB). As fL increases, the jammer becomes more distinguishable

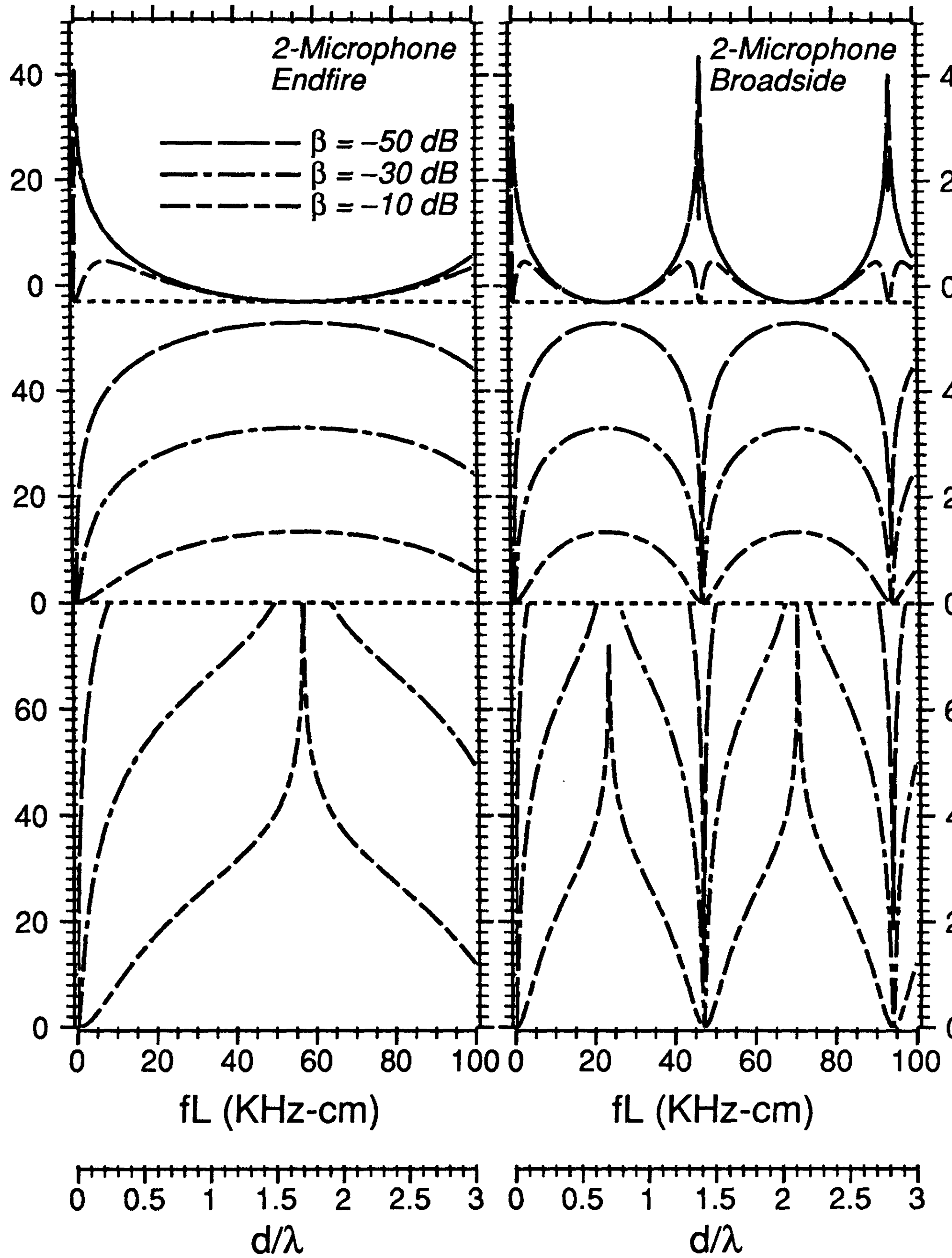


Figure 4.4: Array Performance as a function of fL for 2-microphone endfire and broadside arrays in the presence of sensor noise and one directional jammer at 45°. The plotted performance measures are array gain against the jammer G_j, gain against total noise G_z, and noise sensitivity K_u. Performance is shown for three different values of β, the sensor-to-directional noise power ratio.

from target and can be reduced, but at the cost of increasing the uncorrelated output noise (G_j rises, K_u rises even faster). These effects occur at lower frequencies for broadside than for endfire arrays because broadside intermicrophone phase is more sensitive to deviations of source angle from straight-ahead. As fL increases still further, the jammer becomes even easier to distinguish from target and G_j continues to rise while K_u falls. At some values of fL, the jammer arrives with 180° of phase shift and can be cancelled completely while the sensor-noise is attenuated by 3 dB. Over a great range of frequencies, the jammer can be reduced to well below the level of sensor noise.

Figure 4.5 provides another perspective on directional-noise reduction by showing the directional response at various values of fL for 2-microphone endfire and broadside arrays with the same 45° jammer and $\beta = -10$ dB. The response to the jammer can be measured along the 45° radius. (In general, response to sensor noise cannot be inferred from beam patterns). These patterns illustrate how attempts to null out the jammer generally improve with increasing fL. They also illustrate how performance can be extremely sensitive to the exact location of a null, an important consideration when one tries to realize adaptive null placement against jammers that are moving or strongly time-varying. Figure 4.6 shows the beam patterns for $L = 5$ cm and $L = 20$ cm averaged over frequency using the "intelligibility-averaging" technique of section 2.3.3. Clearly, substantial nulling advantages are obtained even over the broad bandwidth of speech signals.

To show the dependence of performance on jammer angle, we can measure the intelligibility-averaged broadband jammer response, $\langle K_j \rangle_I$, at the jammer angle *only* and plot this response versus jammer angle, as shown in Figures 4.7 and 4.8. These figures also show total noise response, $\langle K_z \rangle_I$, as a function of jammer angle. The arrays described in these figures are 2-microphone 5- and 20-cm endfire and broadside arrays with sensor-to-directional noise ratios, β, of -10 dB.

The response patterns in figures 4.7 and 4.8 have a few properties in common. Although the jammer can be attenuated by more than 20 dB over a wide range of angles, the -10 dB sensor noise limits total-noise reduction to about 10 dB for both array orientations. The average noise reduction is also similar for both orientations.

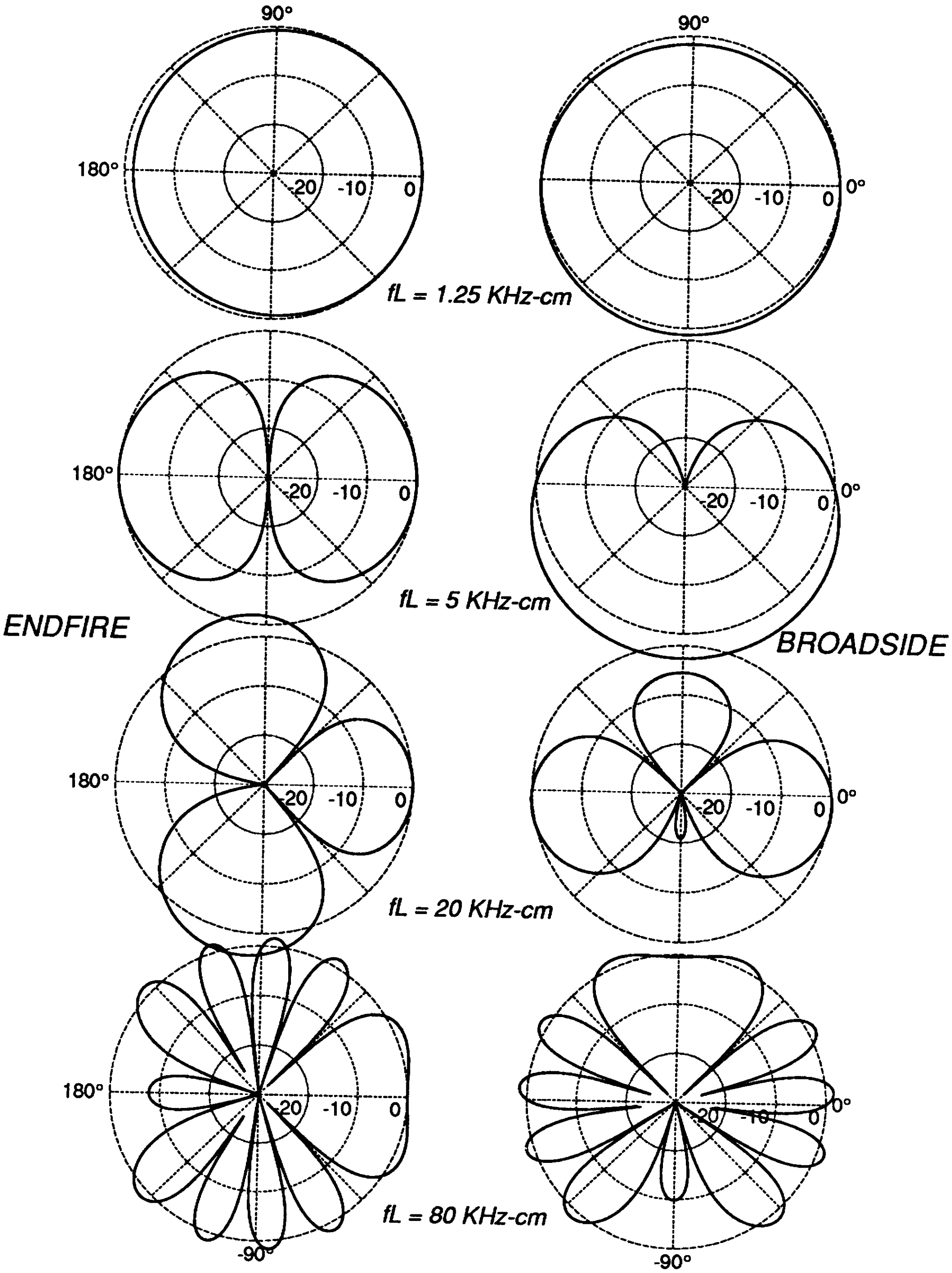


Figure 4.5: Beam patterns for 2-microphone endfire and broadside arrays in the presence of a single jammer at 45° with relative sensor noise at $\beta = -10$ dB. Each row represents the directional response for a different value of fL.

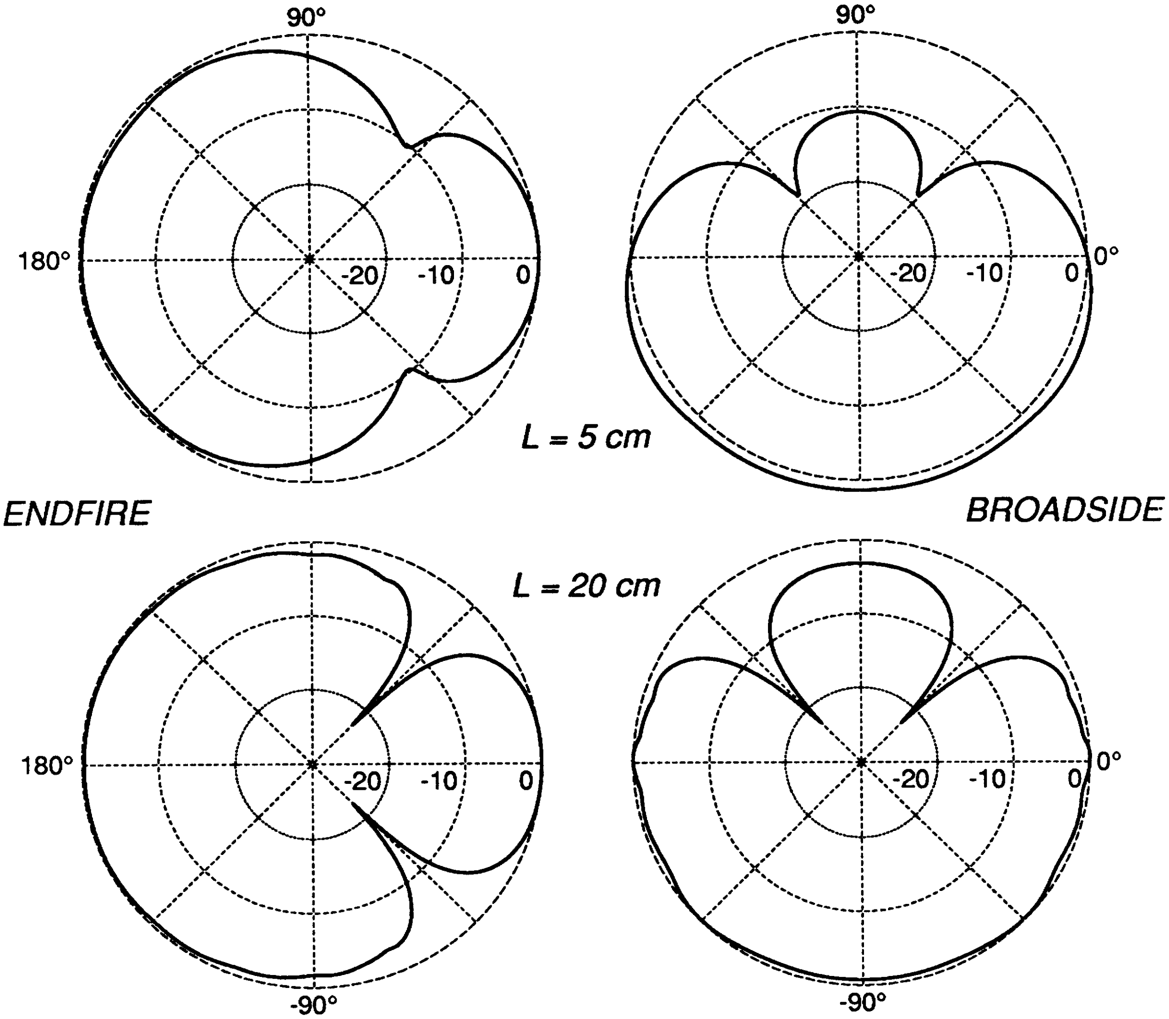


Figure 4.6: Intelligibility-averaged broadband beam patterns, $\langle \mathcal{G}(f,\theta) \rangle_I$, corresponding to the arrays and noise environment of Figure 4.5 for the cases $L = 5$ cm and $L = 20$ cm.

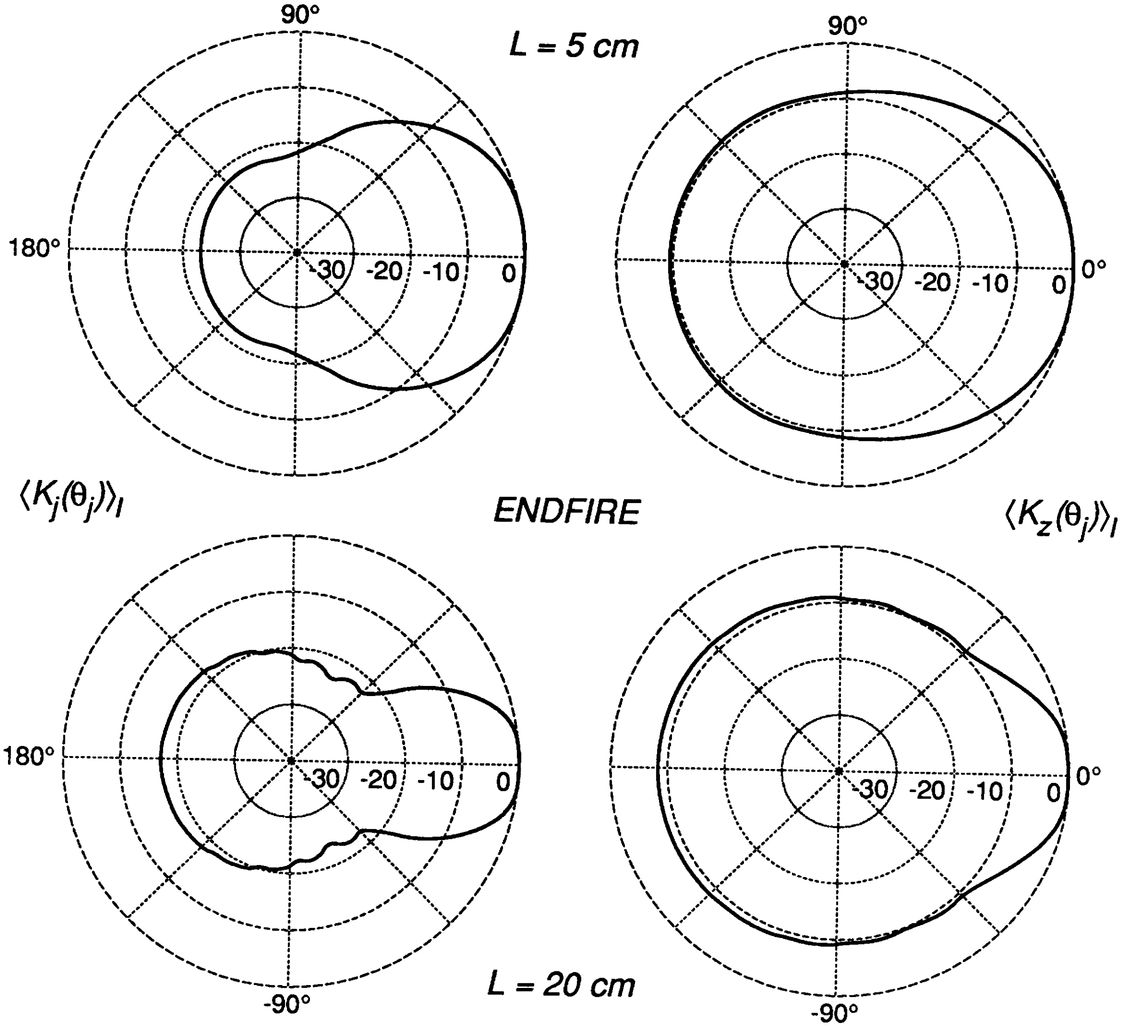


Figure 4.7: Broadband Response to one jammer and to total noise as a function of the jammer's angle for 2-microphone endfire arrays with β = -10 dB.

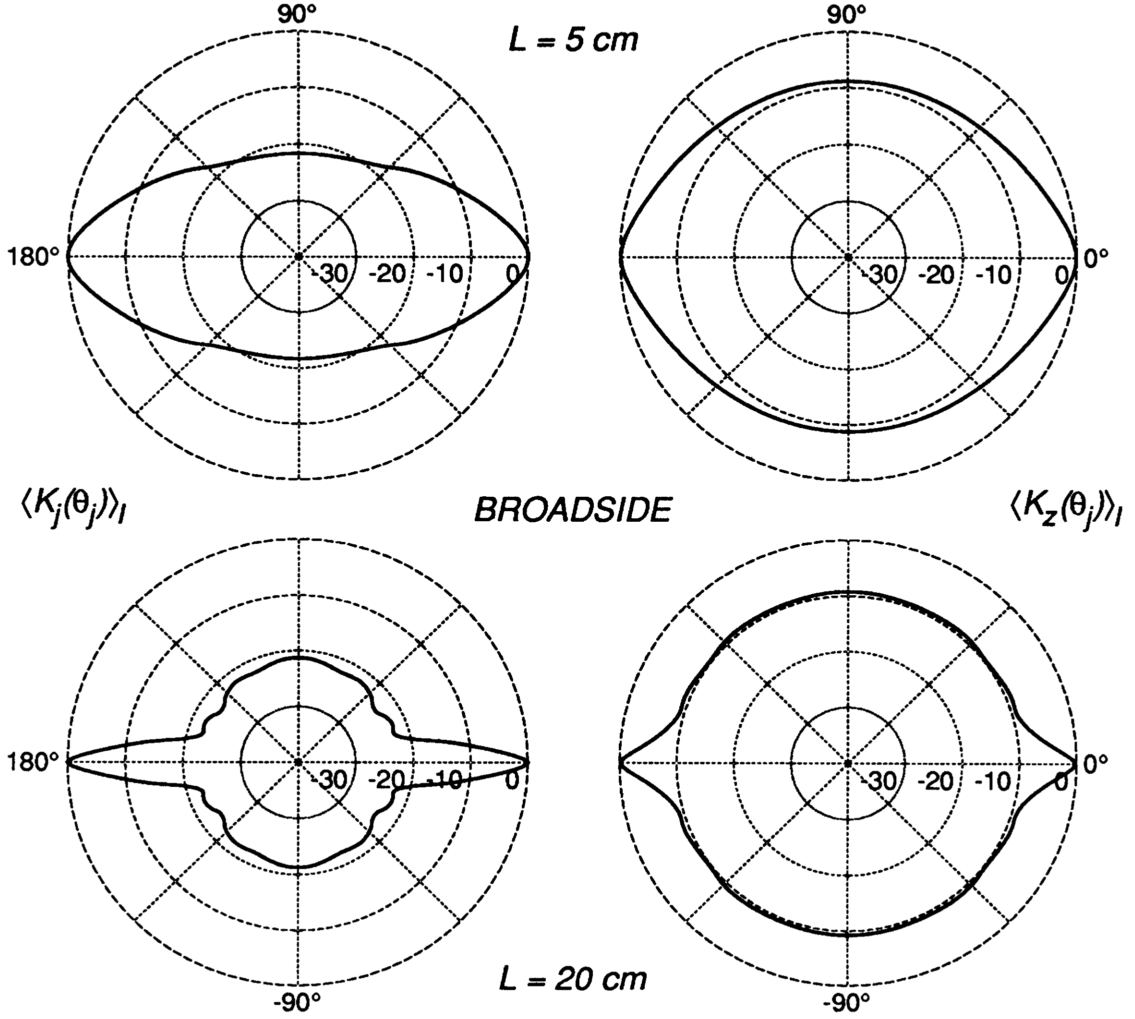


Figure 4.8: Broadband Response to one jammer and to total noise as a function of the jammer's angle for 2-microphone broadside arrays with β = -10 dB. The patterns for L = 20 cm can be compared with Figure 2.1.

The main difference between endfire and broadside response patterns lies in their shape. Due to symmetry, broadside arrays cannot distinguish between sources directly ahead-of and directly behind the array, although we do not consider this an insurmountable shortcoming because arrays can be mounted on the head to make use of head-shadow or they can be composed of both endfire and broadside subarrays. More significantly, the broadside response is much more sensitive to jammer angle near 0° and broadside arrays will be more successful in reducing jammers at small jammer angles. However, this behavior may have drawbacks that appear when the target itself does not appear at exactly 0° (i.e., when the target is *misaligned*). In this case we would expect a broadside array to cancel more of the target than an endfire array, a characteristic that may be undesirable in some applications[7]. The primary point of this discussion is that the shape of the response function can be important, in ways that may not be obvious.

Note that the broadside 20-cm array can be compared directly to the human binaural system modelled by Zurek in Figure 2.1 and Zurek's measure of sensitivity can be compared directly to the total-noise response shown in Figure 4.8. It is interesting that an optimum 2-microphone receiver with $\beta = -10$ dB matches the performance of Zurek's model reasonably well.

Figures 4.9 and 4.10 show total noise response only, again as a function of jammer angle for 2-microphone endfire and broadside arrays with lengths, L, of 5 and 20 cm, but now at two different values of β. Clearly, the maximum and average gains increase and the "beamwidth" decreases as sensor noise decreases.

4.3.2 Performance against Multiple Jammers

Finally, we must consider configurations with more than one jammer. Since multi-jammer directional sensitivity patterns would be difficult to visualize and computationally prohibitive to construct, we will use a Monte Carlo technique to evaluate

[7] Unfortunately, a proper discussion of target misalignment is beyond the scope of this thesis. However, we should note that target misalignment can be quite detrimental to some algorithms (Peterson, Wei, Rabinowitz and Zurek, 1989) and algorithm modifications can reduce misalignment effects (Griffiths and Jim, 1982; Greenberg, Zurek and Peterson, 1989).

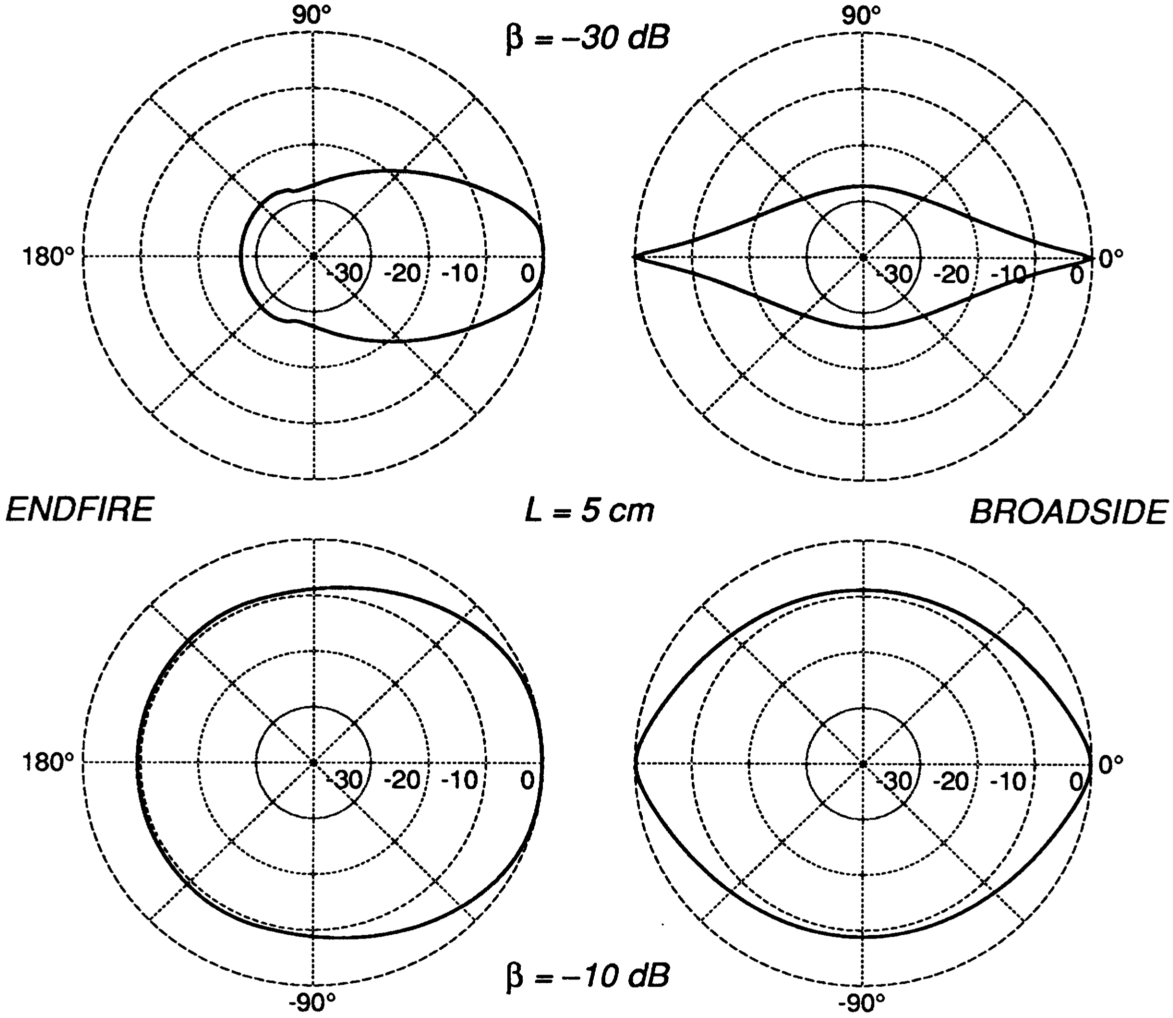


Figure 4.9: Broadband response of short 2-microphone arrays to one jammer as a function of jammer angle and sensor noise level.

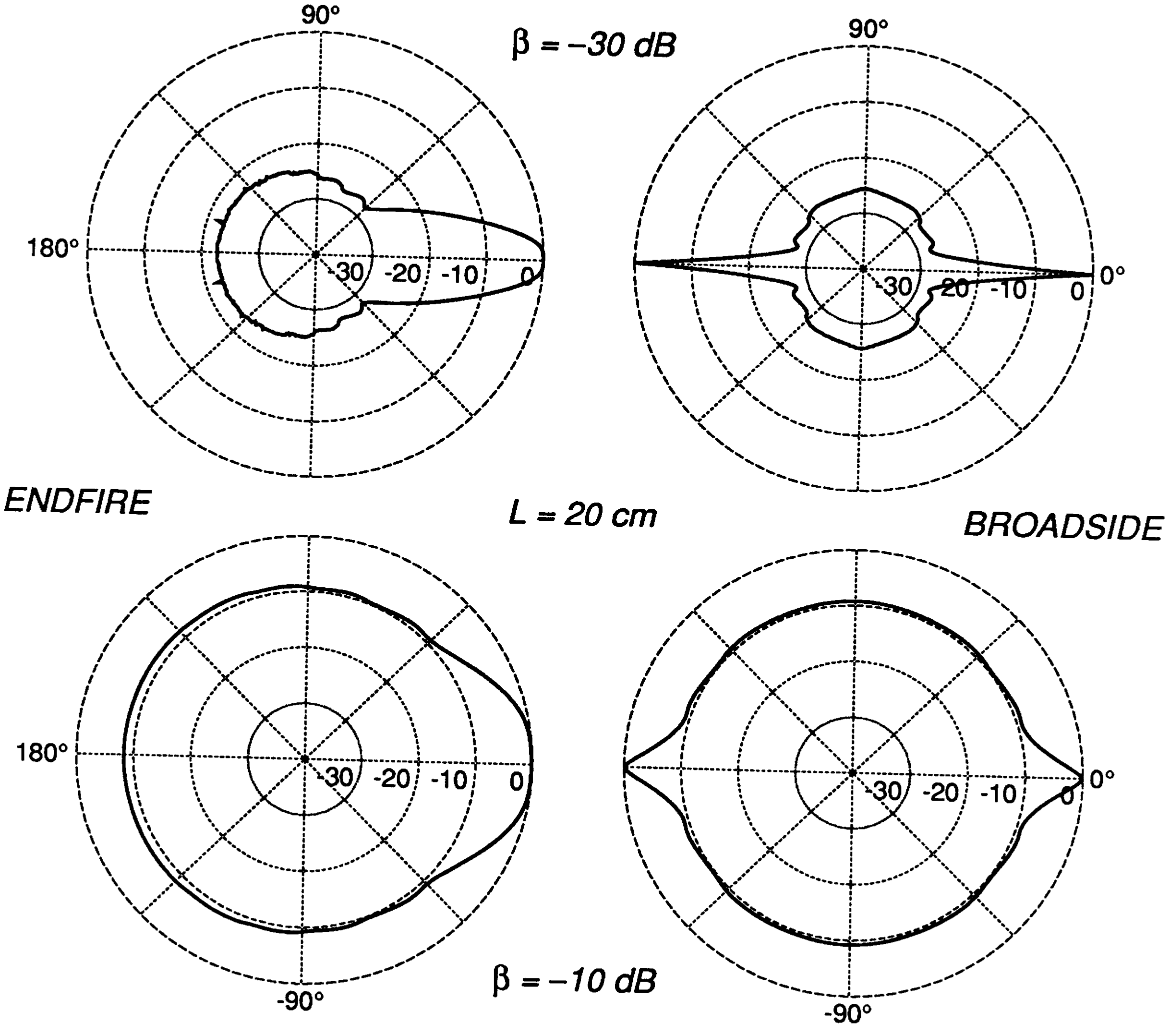


Figure 4.10: Broadband response of long 2-microphone arrays to one jammer as a function of jammer angle and sensor noise level.

multi-jammer environments. Figure 4.11 illustrates the method for one-jammer configurations of the type discussed in the previous subsection.

For each combination of array and sensor noise, we choose 1000 jammer angles from a uniform random distribution, determine the optimum processor and total-noise response K_z for each jammer angle, and accumulate the distribution of response magnitudes. Figure 4.11 shows this distribution plotted in a form that emphasizes its relationship to the directional-noise sensitivity plots in the previous figures. As an example, consider the 20-cm endfire array with -30 dB sensor noise. The plot tells us that about 93% of the randomly-chosen jammer angles resulted in total responses below -10 dB or, equivalently, 7% of the responses were above -10 dB. In Figure 4.10, the -10 dB beamwidth for the same configuration is 26°, or about 7% of 360°. In general, of course, these distributions tell us very little about the shape of the response functions, especially for multiple jammers, where the directional response will have many local maxima and minima.

We will use the mean of each distribution as our primary performance measure. For 1000 random jammer configurations, the 3σ confidence limit for the sample mean is 0.5 dB[8]. We will also use standard deviation of response to convey some information about the shape of the response function. For example, in Figure 4.11, the gain distributions for $L = 2.5$ cm endfire and broadside configurations at $\beta = -30$ dB have almost identical average values but the endfire array has a larger standard deviation (meaning that its gain is more likely to be either very high or very low). This is another way to characterize the shapes of the endfire and broadside response patterns shown in Figures 4.9 and 4.10.

For multiple jammers, the Monte Carlo technique is virtually identical, except that each random jammer configuration is composed not of one but of multiple equal-power jammers whose angles are independent, uniformly-distributed random variables. Figure 4.12 shows $\langle G_z \rangle_I$, the mean intelligibility-averaged gain against total noise, for 1000 multiple-jammer configurations as a function of array length for 2, 4, 6, and 8-microphone endfire and broadside arrays with 2, 4, and 6 jammers

[8] The distributions shown here were compared to distributions derived from uniformly-spaced jammer angles and the match was consistent with the 0.5 dB confidence limit.

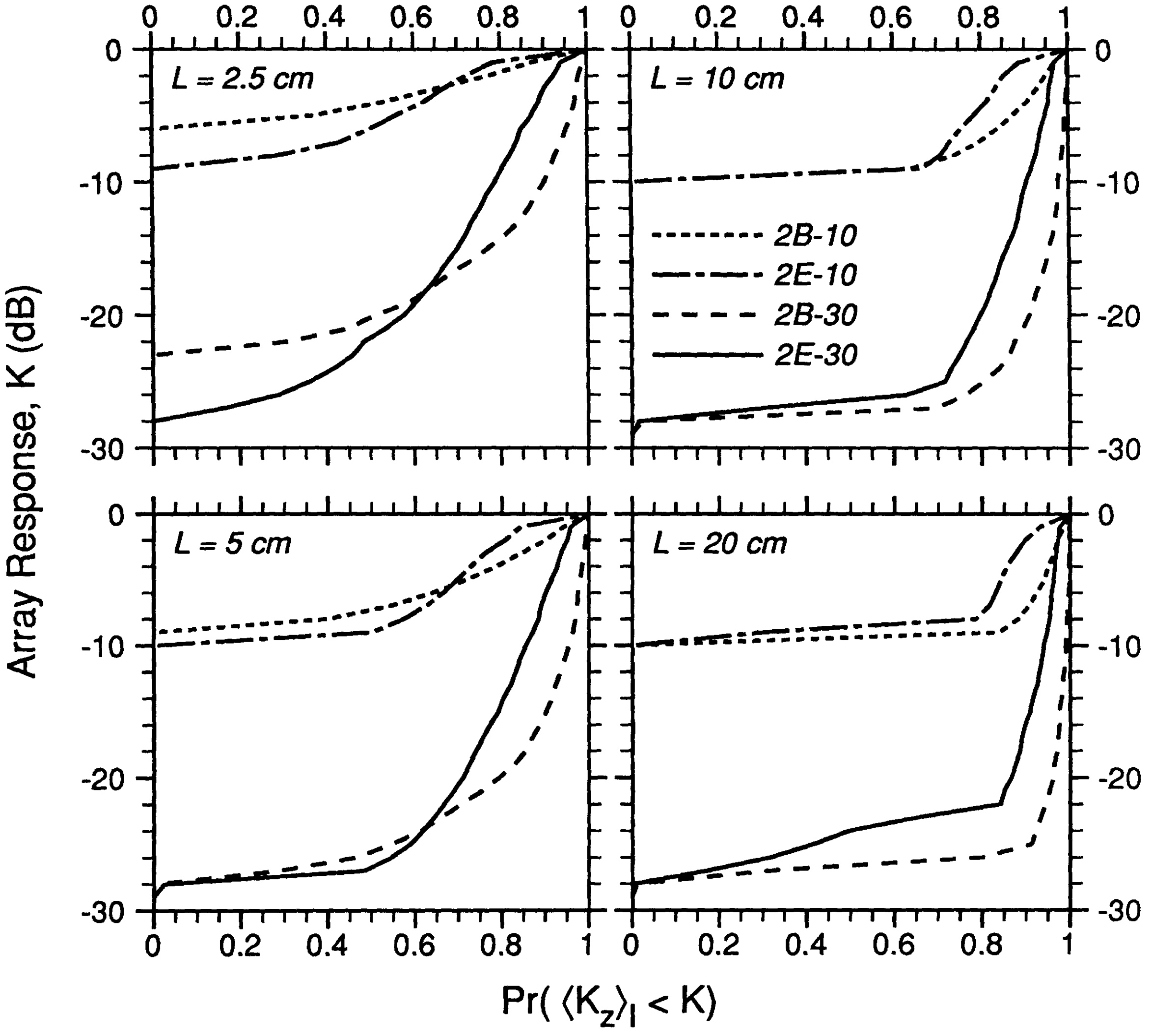


Figure 4.11: Distribution of array total-noise response for 1000 randomly chosen single-jammer angles, 2-microphone endfire and broadside arrays, various array lengths, and 2 levels of sensor noise. The annotation 2B-10 refers to a 2-microphone broadside array with relative sensor noise of β = -10 dB.

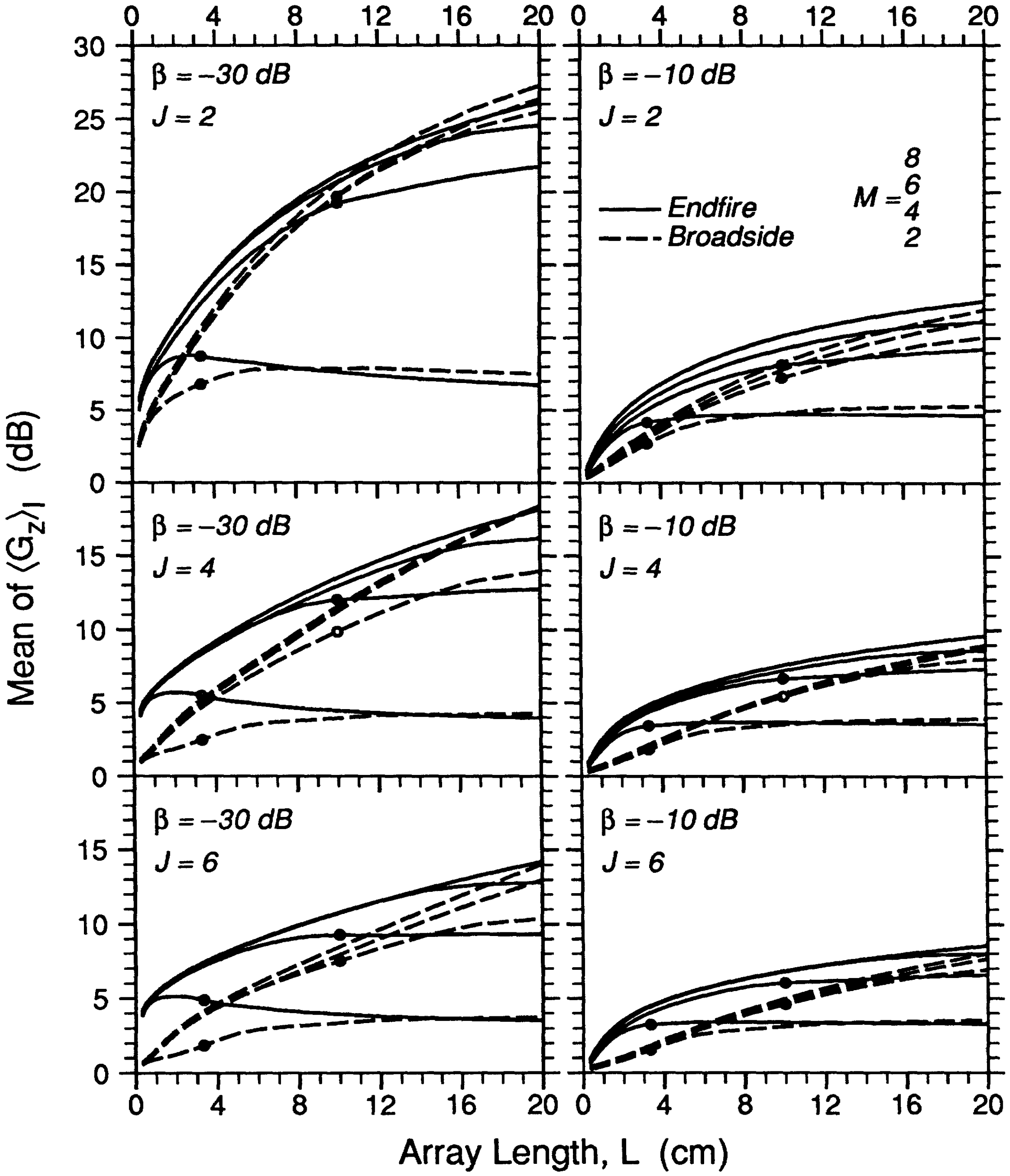


Figure 4.12: Mean intelligibility-averaged broadband array gain against total noise, $\langle G_z \rangle_I$, as a function of array orientation, array length L, number of microphones M, number of jammers J, and relative sensor noise level β. The sample mean was determined using 1000 randomly-sampled jammer configurations.

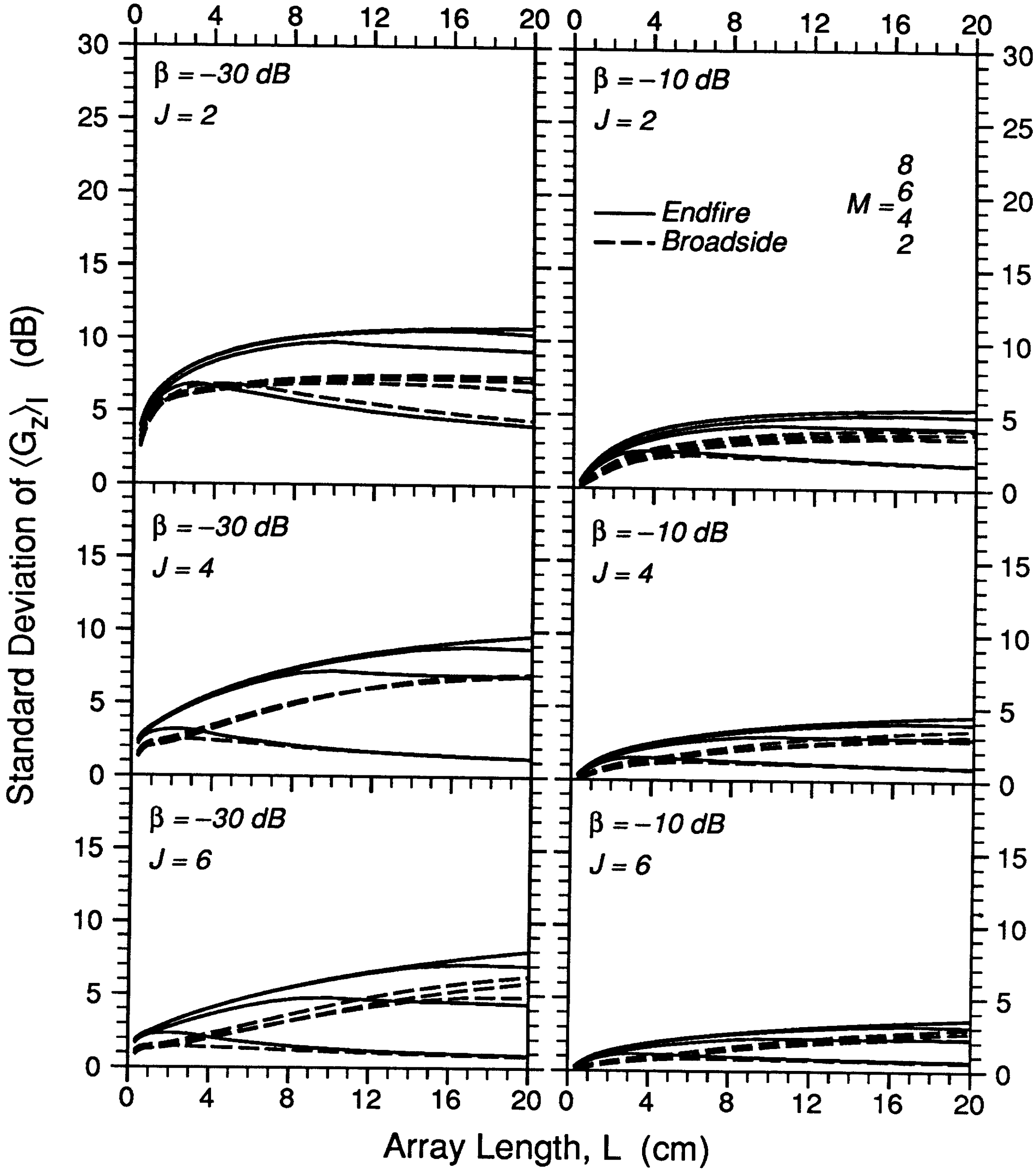


Figure 4.13: Standard deviation of the broadband array gain shown in Figure 4.12. As discussed in the text, larger standard deviations may correlate with "flatter" response near 0°.

and $\beta = -10$ and -30 dB. Figure 4.13 shows the corresponding standard deviations of gain. Directional-noise performance in Figure 4.12 can be compared directly with isotropic noise performance in Figure 4.3, although some care is necessary because the sets of microphone numbers and sensor-noise values are not identical. Based on these figures we make the following observations.

- Even when arrays are short and sensor noise is only 10 dB less than directional noise and the number of jammers is large, two microphones can give 3 dB of gain and four microphones can give 6 dB of gain. For less sensor noise or fewer jammers, the gains can be much greater.

- As in the isotropic noise case, performance saturates when the number of microphones exceeds 4 for short arrays and 6 for long arrays.

- Performance does not degrade drastically when the number of jammers exceeds the number of microphones, probably because the presence of sensor noise limits the best possible performance for small numbers of jammers.

- Endfire arrays generally outperform broadside arrays, especially if the arrays are short, but the difference between isotropic and broadside performance in many-jammer directional noise fields is less than the difference in isotropic fields.

- Endfire performance in many-jammer fields is roughly equivalent to endfire performance in an isotropic field, but broadside arrays perform significantly better against many jammers (at least up to 6 jammers) than against isotropic noise.

- The mean gains of long broadside and endfire configurations are nearly equal but their response pattern shapes may be significantly different, as indicated by differing standard deviations of gain.

In general, it is clear that optimum array processing can provide significant benefit in a wide range of directional-noise fields. Once again, the endfire configuration seems superior to the broadside configuration, although performance differences are less compelling in directional noise fields than in isotropic noise fields.

4.4 Directional Noise in Reverberation

In reverberant environments we can no longer use the directional plane-wave assumption and write the simple source-microphone transfer function of equation (4.16). This makes it harder to analyze performance; however, it does not necessarily make it harder to determine the optimum processor. In particular, note that the only properties of the interference that influence the choice of optimum weights are those that are revealed in the total-noise cross-spectral matrix; and that this matrix can be estimated without knowing the individual jammer transfer functions. If the target transfer function is known, it should make no fundamental difference whether the jammer-microphone transfer functions are reverberant or not[9]. Thus, for example, if sensor noise were absent, it should still be possible to cancel the interference from J reverberant jammers with an array of $J+1$ microphones.

The problems created by reverberation are due mainly to *target* reverberation. If one regards reverberated target as "interference", then the assumption that target and interference are uncorrelated does not hold. On the other hand, if one regards reverberated target as "desired" signal, then the assumption of known target transfer function is likely to be violated.

Rather than analyze optimum performance in reverberant environments, we chose to pursue an empirical approach. The adaptive beamformer evaluation in Chapter 6 was carried out in different reverberant environments to indicate the magnitude of target reverberation effects.

[9] Of course, we are ignoring practicality here. A practical system based on a finite observation time, T, must fail in rooms with reverberation times greater that T.

4.5 Head Shadow and Other Array Configurations

The major results of this chapter were derived for a few specific array configurations in extremely simple environments. We now consider the effects of head-shadow and other array configurations on optimum performance.

We can establish upper bounds on the effects of head-shadow by considering measurements of interaural amplitude and phase differences (Durlach and Colburn, 1978; Shaw, 1974). We will presume that no two microphones mounted anywhere about the head could experience more of a "head effect" than two ears located on opposite sides of the head. For any incident angle, interaural arrival-time differences are only slightly frequency-dependent and always fall in the range of 1.0 to 1.5 times the free-field arrival-time differences. Interaural amplitude differences are strongly dependent on both frequency and incident angle, but never amount to more than 5 dB at 500 Hz, 10 dB at 2.5 KHz, or 17 dB at 5 KHz. For many incident angles, amplitude differences are considerably smaller.

Head shadow can affect both the target transfer function, $\underline{\mathcal{H}}(f)$, and the noise cross-spectral matrix, $\mathcal{S}_{zz}(f)$. We will assume that $\underline{\mathcal{H}}(f)$ can be measured *a priori* to calibrate for the effects of head shadow[10]. Admittedly, for arrays mounted close to the head, $\underline{\mathcal{H}}(f)$ may be sensitive to differences in mounting position and, therefore, difficult to calibrate in practice. If $\underline{\mathcal{H}}(f)$ is known, the essence of adaptive beamforming is that $\mathcal{S}_{xx}(f)$ and, by implication, $\mathcal{S}_{zz}(f)$, can be estimated from the received microphone signals. In other words, beamformers do not need *a priori* information about $\mathcal{S}_{zz}$, whether head-shadow is present or not. However, we still need to know the effects of head-shadow on $\mathcal{S}_{zz}$ to predict the asymptotic beamformer performance.

If we consider intermicrophone phase only, adding a head to an array is equivalent to stretching the array, perhaps non-uniformly. From Figures 4.3 and 4.12 it is clear that changing the size of an array *uniformly* by 50% has only minor effects

[10] For this strategy to be feasible, we must restrict $\underline{\mathcal{H}}(f)$ to represent only direct arrivals, which can be measured in an anechoic environment.

on performance. We presume the same will be true for non-uniform changes if the non-uniformity is not too great.

Intermicrophone amplitude differences have the potential to cause much greater alterations in performance. To see this, consider a directional jammer. As long as different microphone signals are completely correlated, regardless of their amplitudes or phases, equivalent information is available at all microphones and performance is not compromised. For correlation to be preserved, however, the level of the directional signal at all microphones must be significantly above the level of sensor noise. When the directional signal at a shadowed microphone falls below sensor noise, it cannot contribute to nulling of the directional jammer. For the simple case of a 2-microphone system, the directional signal at the non-shadowed microphone must then be treated as noise.

Now, to be more specific with the 2-microphone example, suppose that the sensor-to-directional noise ratio, β, is -10 dB in the free field. Let us compare the performance of a free-field array with that of a head-mounted array for which the microphone amplitudes are +5 dB and -12 dB relative to free-field amplitudes [an extreme case in the spirit of available data (Shaw, 1974)]. The performance of the free-field array will be sensor-noise limited at most frequencies, the jammer will be essentially cancelled, and broadband total-noise gain, $\langle G_z \rangle_I$, will be about 10 dB (from Figures 4.7, 4.8 and 4.4). The head-mounted array must treat the two microphone signals as uncorrelated and unequal noises, a generalization of the equal noise example solved in Section 4.1. The solution for unequal noises is for the optimum processor to align the target in both channels, just as before, and then add the channels in inverse proportion to the noise power in each channel. In our example the processing amounts to weights of 0 dB and -15 dB on the weak and strong microphone signals, respectively. Since the output noise power from each channel is proportional to the weight *squared*, the weak-channel noise, which we assume is still -10 dB relative to the free-field jammer, will dominate. The net result is broadband total-noise gain, $\langle G_z \rangle_I$, relative to the free-field jammer, of 10 dB, almost identical to that of the free-field processor.

Intuitively, we can summarize this example by saying that, when the combination of head-shadow and sensor noise is not enough to destroy intermicrophone correlation, the directional jammer can be cancelled and performance is limited by sensor noise. When head-shadow and sensor noise make the directional jammer unobservable in one of the microphones, then that microphone signal is *already* sensor-noise limited and the strong-jammer signal can be almost ignored. To a first approximation, then, it seems that amplitude and phase effects due to head-shadow may not be very detrimental to array performance. However, this conclusion is based on a simple, one-jammer example and ought to be tested with multiple jammers and isotropic noise. It can also be argued that, for systems with more that two microphones, the loss of jammer information in a shadowed microphone may be more than offset by better jammer information in other microphones whose jammer-to-sensor-noise ratio is increased by the presence of the head.

Even if head shadow does influence performance against more complicated interference, it is unlikely that it would change our picture of the relative benefits of endfire and broadside arrays. Head shadow effects over the extent of a linear array (i.e., on one side of a head) are not large enough to greatly alter the fundamental properties of broadside and endfire arrays, namely, that target signals arrive simultaneously in one case and with maximum possible intermicrophone delay in the other.

We only considered equispaced, linear, endfire and broadside arrays because, as shown in Table 4.1, these two configurations represent extremes of performance for small, sensor-noise free, M-element arrays operating in free-field isotropic noise. As indicated in the table, performance was limited only by array configuration and number of microphones. When sensor noise is present, however, our own results indicate that performance is limited by something other than number of microphones, since performance saturates beyond 4 to 6 microphones. In fact, the data in Figures 4.3 and 4.12 suggest that, for a given array length, the maximum number of useful microphones corresponds to spatial oversampling by a factor of slightly more than 1 when sensor noise is high to a factor of perhaps 2 when sensor noise is low. If spatial sampling, rather than number of microphones, is the limiting factor, then the best arrangement of M microphones is probably not a linear array but some other configuration that makes better use of the spatial extent of the head. Based on spatial-sampling considerations, an array around the circumference of the head might be able to use information from as many as 12–15 microphones, while a spherical surface array might saturate at 50–60 microphones.

Chapter 5

Adaptive Beamforming

The optimum processors of the preceding chapters were all based on *a priori* knowledge of, at least, the target-to-array transfer function and the total noise correlation matrix (or cross-spectral matrix). This chapter describes a realizable processing system that continuously adjusts its parameters based on the received microphone signals and approaches optimum performance for stationary interference configurations.

Constrained adaptive beamformers[1] operate by minimizing array output power under the constraint that signals from the target direction be preserved (Monzingo and Miller, 1980; Frost, 1972). The method assumes that the target and interference are uncorrelated and that the target direction is known (i.e., the relative amplitudes and phases of the target signal at the microphones are known)[2]. As long as these assumptions hold, constrained minimization of total output power necessarily minimizes interference output power. Since constrained beamformers make no assumptions about the structure of the interference environment (e.g., number and directionality of sources), they should not be overly sensitive to interference complexity.

A major problem with application of constrained beamforming to hearing aids concerns the presence of reverberated target energy. If one regards reverberated target as "interference", then the assumption that target and interference are uncorrelated does not hold. On the other hand, if one regards reverberated target as "desired" signal, then the assumption of known target direction is violated. In

[1] Although these methods are called beamformers, due to limited array size they cannot form sharp beams in our application. We are capitalizing on their ability to adaptively steer nulls.

[2] In our application, the target direction is straight ahead of the listener and the target signals at the microphones are assumed to have the relative amplitudes and phases of a target straight-ahead in anechoic space.

Chapter 6 we evaluate adaptive beamformers in various reverberant environments to determine the effects of target reverberation.

Among the algorithms using the constrained adaptive beamforming criterion, there is considerable variety in the strategies for adapting the microphone weights in time. At one extreme, it is possible to calculate the optimum weights directly after each signal sample, but the amount of calculation per sample can be prohibitive. At the other extreme, some algorithms make very simple calculations with each sample, eventually converging on the optimum weights, but only after many samples. Thus, the fundamental tradeoff is between speed of calculation and speed of convergence. Our eventual decision to implement the constrained adaptive beamforming method of Griffiths and Jim (Griffiths and Jim, 1982) was based on two considerations. First, it required the minimum amount of computation for a given filter length and would be among the first candidates for inclusion in a wearable aid. Second, it has the same ultimate performance as any beamforming method and, since our initial evaluation involved stationary environments, its performance after adaptation would indicate the ultimate performance to be expected of adaptive beamformers in general.

5.1 Frost Beamformer

Frost described one of the first practical constrained adaptive beamformers (Frost, 1972), a sampled-data system suitable for digital implementation. It is a time-domain beamformer composed of tapped delay lines following each microphone, adaptive amplitude weights at each tap, and a summer that forms the output from the weighted delayed samples. The weight-adaptation procedure is based on Widrow's LMS principle (Widrow, Glover, McCool, et al., 1975), but modified to incorporate the target preservation constraint. The LMS adaptation procedure is a stochastic gradient method that depends on the fact that average output power is quadratically related to the array weights. Therefore, the weights can be adjusted directly to give minimum output power by following the gradient of the quadratic power function. This adaptation is slow but simple and eventually converges to within a "misadjustment" factor of the optimum weights. This misadjustment arises because the gradient of the quadratic power function can only be estimated from stochastic data. Misadjustment can be reduced by increased averaging in the stochastic gradient estimate, but at the cost of longer adaptation times.

5.2 Griffiths-Jim Beamformer

An even simpler constrained beamformer, which can be made equivalent to Frost's system, has been proposed by Griffiths and Jim (Griffiths and Jim, 1982). Instead of adjusting the array weights directly with a constrained LMS algorithm, they propose a two-stage system in which an initial linear transformation of array signals constrains the target gain and a subsequent *unconstrained* LMS filtering removes interference. Since the system is composed of separate, standard, single-channel LMS noise-cancelling filters, extension to an arbitrary number of microphones is trivial and implementation in both hardware and software should be straightforward.

A broadside two-microphone Griffiths-Jim beamformer is outlined schematically in Figures 5.1 and 5.2. (The dashed lines in Figure 5.1 show an extension of the system to three microphones.) The two microphone signals are transformed into a sum signal, $s[k]$, which contains target plus interference, and a difference signal, $d_1[k]$, which contains *no target* for straight-ahead targets in an anechoic field[3]. The beamforming problem is thus transformed into a noise-cancellation problem and the sum and difference signals can be fed to a standard LMS noise-canceller composed of a sum-signal delay, $z^{-(L/2)}$, an adaptive filter, $h_1[k]$, and an output summer. The method can be simply extended to more microphones by summing all microphone signals into one sum signal and forming pairwise microphone difference signals that feed separate LMS noise-cancelling filters, each of which operates to cancel noise in the sum signal.

[3] For orientations other than broadside, the microphone signals can be "steered" (i.e. multiplied by amplitude and phase factors to equalize and time-align the target signals in each channel), effectively transforming them into broadside signals.

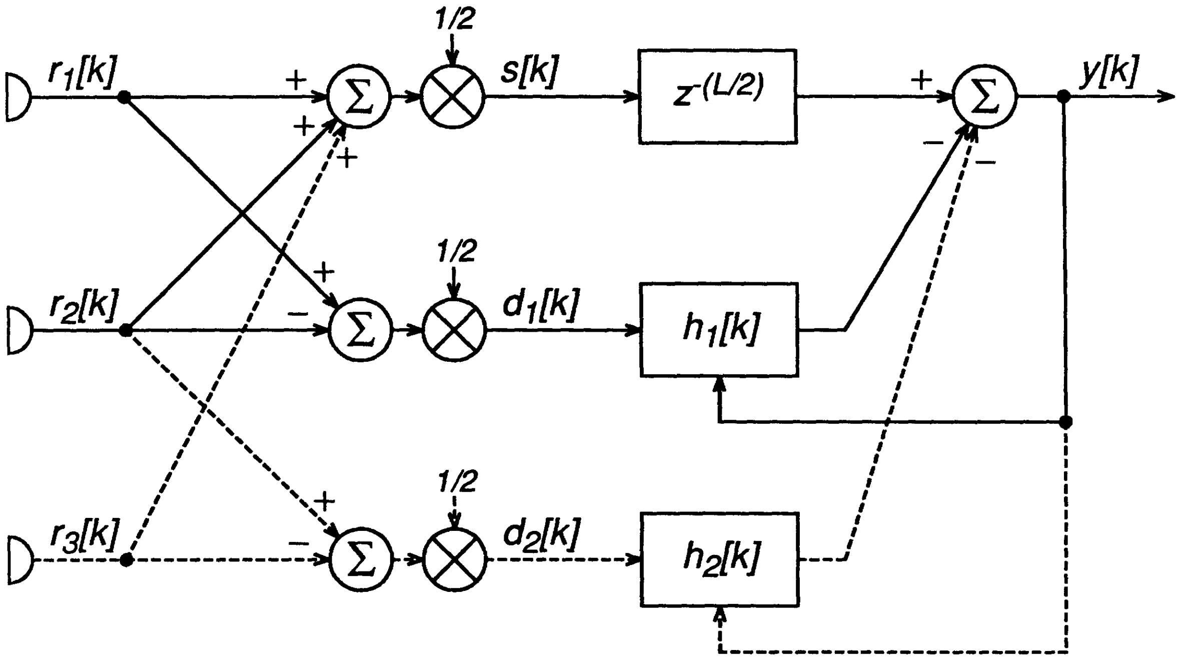


Figure 5.1: A Griffiths-Jim beamformer for a two-microphone broadside array. The microphone signals at sample index k, denoted $r_1[k]$ and $r_2[k]$, are added and subtracted and then scaled to form the sum signal $s[k]$ and the difference signal $d_1[k]$. The sum signal (which contains the target signal plus interference) is delayed by $L/2$ samples in the delay element labelled $z^{-(L/2)}$. The difference signal (which should contain only interference) is passed through the $L+1$-point FIR adaptive filter $h_1[k]$ to form an interference cancellation signal which is subtracted from the delayed sum signal to form the output $y[k]$. The output is then used in adjusting the adaptive filter coefficients to further reduce output interference. This is accomplished by, in effect, correlating the output with the past $L+1$ samples of the (inteference-only) difference signal, and then adjusting the FIR filter weights to drive that correlation to zero. At zero correlation, none of the output interference can be predicted from the past difference-signal samples and the adaptive filter has transformed the difference-signal interference to most closely resemble the interference in the sum signal. To incorporate a third microphone signal, $r_3[k]$, the sum signal summation is extended and a new difference signal, $d_2[k]$, is formed. The difference signal is passed through an identically-constructed adaptive FIR filter $h_2[k]$ before being subtracted from the delayed sum signal.

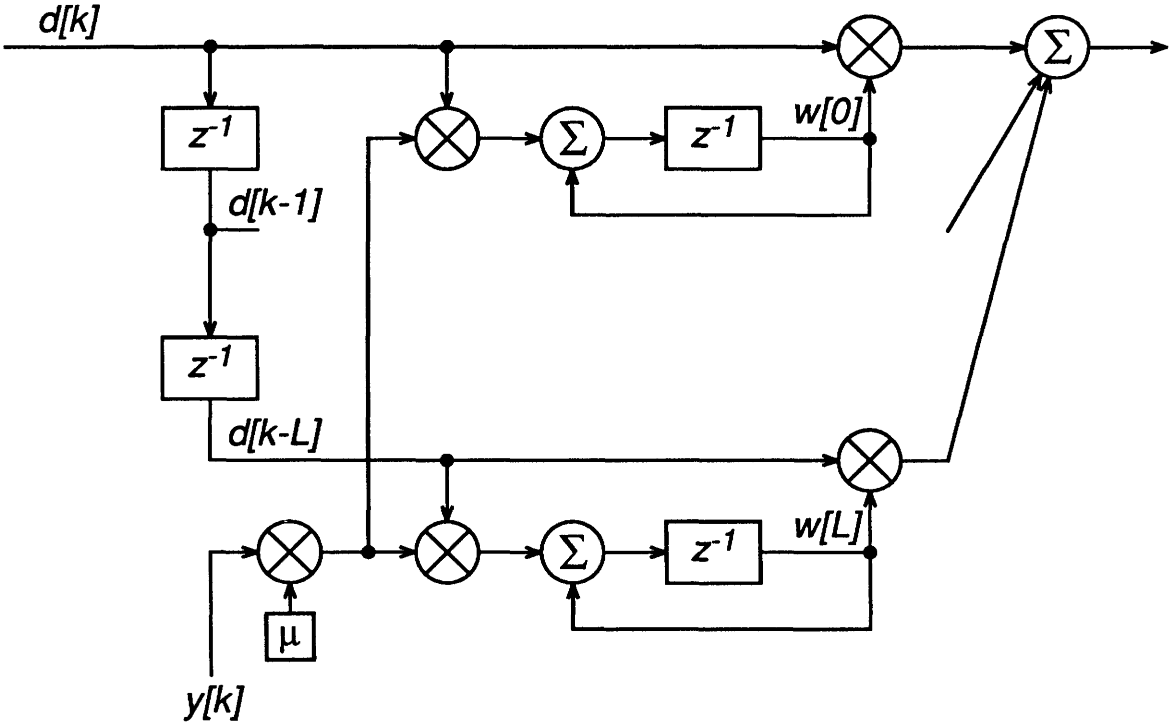


Figure 5.2: The adaptive FIR filter structure. The adaptive FIR filter operates on the $L+1$ most recent difference-signal samples, $d[k], d[k-1], ..., d[k-L]$, which are held in the chain of L unit delays labelled z^{-1}. Each sample, $d[k-l]$, is multiplied by a weight, $w[l]$, and all the weighted samples are added together to form the filter output. The adaptation of weight $w[l]$ is driven by the product $\mu\, y[k]\, d[k-l]$, which depends on the fixed parameter μ and the beamformer output $y[k]$ and is added to $w[l]$ to form the weight for the next sample index. The accumulation of product terms in $w[l]$ can be viewed as a stochastic estimate of the correlation between $y[k]$ and $d[k-l]$.

5.3 Other Methods

The Griffiths-Jim processor that we eventually implemented is a constrained adaptive beamformer with an especially simple time-domain realization. Frost's time-domain method is equivalent to a particular Griffiths-Jim implementation but uses a more complex adaptation algorithm. Strube's method (Strube, 1981) is identical in principle but uses block processing in the frequency domain and tends to generate distracting artifacts. Computationally-efficient fast-adapting methods (Cioffi and Kailath, 1984; Lee, Morf and Friedlander, 1981) may eventually be needed to follow changes in interference environments, but such methods are quite complex. In addition, faster adaptation can create problems by adapting not only to the changing environment, but also to momentary changes in the target and interference signals themselves (Honig and Messerschmitt, 1984). The fast-adapting methods will only be attractive if simpler methods cannot cope with the variability of hearing-aid environments

Chapter 6

Experimental Evaluation

In preceding chapters we have shown that, in theory, adaptive beamformers ought to reduce interference in hearing aids. In this chapter we test this hypothesis by, first, simulating an adaptive beamforming hearing aid in some representative environments and, second, evaluating the intelligibility of that simulated aid. By simulating the multimicrophone aid, we eliminate many possible confounding effects, such as head-shadow or microphone mismatch, which do not represent fundamental problems and can be studied in more detail later. By simulating the environments, we control the amount and type of reverberation, which allows us to study empirically a potential major problem that we were not able to treat theoretically. The choice of a two-microphone, head-width, broadside array and the inclusion of a no-processing, binaural-hearing condition in the intelligibility tests allowed us to compare adaptive-beamforming performance with normal human binaural performance.

6.1 Methods

6.1.1 Microphone Array Processing

We implemented a two-microphone Griffiths-Jim beamformer as described in the previous chapter. The system is characterized by three parameters: the sampling rate, which was fixed at 10 kHz; L, the length of the adaptive noise-cancelling filter (i.e., the number of samples in its impulse response); and μ, which controls the adaptive step size. With larger L, the system can potentially remove more interference, but at the cost of more computation and longer adaptation time. With larger μ, adaptation time shortens but misadjustment increases and the filter approaches instability.

To guide the choice of L, we fed interference alone to the system and measured total output power with 20-, 100-, and 400-point filters. In an anechoic environment, the 20-point filter was clearly inferior while the 100- and 400-point filters gave identical performance. In reverberant environments, the 20-point filter was still clearly inferior while 400-point filters performed better than 100-point filters to an extent dependent on the amount of reverberation. In the present study, we used both 100 and 400 for L.

In setting μ, we reasoned that the time-variability of speech (and of the environment) would limit eventual performance, so there should be no penalty in choosing a large μ for fast adaptation. Preliminary experiments with a range of μ values confirmed this behavior and we finally chose a value 10 times smaller than that which would cause instability. The value of μ meeting this criterion depends on overall input power and, in a practical algorithm, would be calculated as $\mu = \alpha/P(t)$, where $P(t)$ is a running measure of input power and α is a normalized adaptation parameter. Our choice of μ, which was made in fixed power experiments, corresponds to $\alpha = 0.0004$.

These choices for L and μ gave empirical adaptation times of a few seconds. Since our intelligibility test stimuli last only a few seconds, and since we sought to evaluate the asymptotic (adapted) performance of the system, we initialized the weights to values near their adapted values. For the anechoic environment, we were able to calculate the optimum weights *a priori* and initialize with these values. For reverberant environments, we used "tuned" initial weights obtained by initializing with optimum anechoic weights, running the system for 3 or 4 seconds, and then measuring the adapted weights.

6.1.2 Simulated Reverberant Environments

The two simulated microphone signals were generated by passing anechoic source materials through simulated room transfer functions (Peterson, 1986). To obtain a range of reverberant environments, we simulated the transfer functions from target and interference locations to two microphone locations in three spaces: an anechoic space, a living-room space, and a conference-room space. Figure 6.1 illustrates a

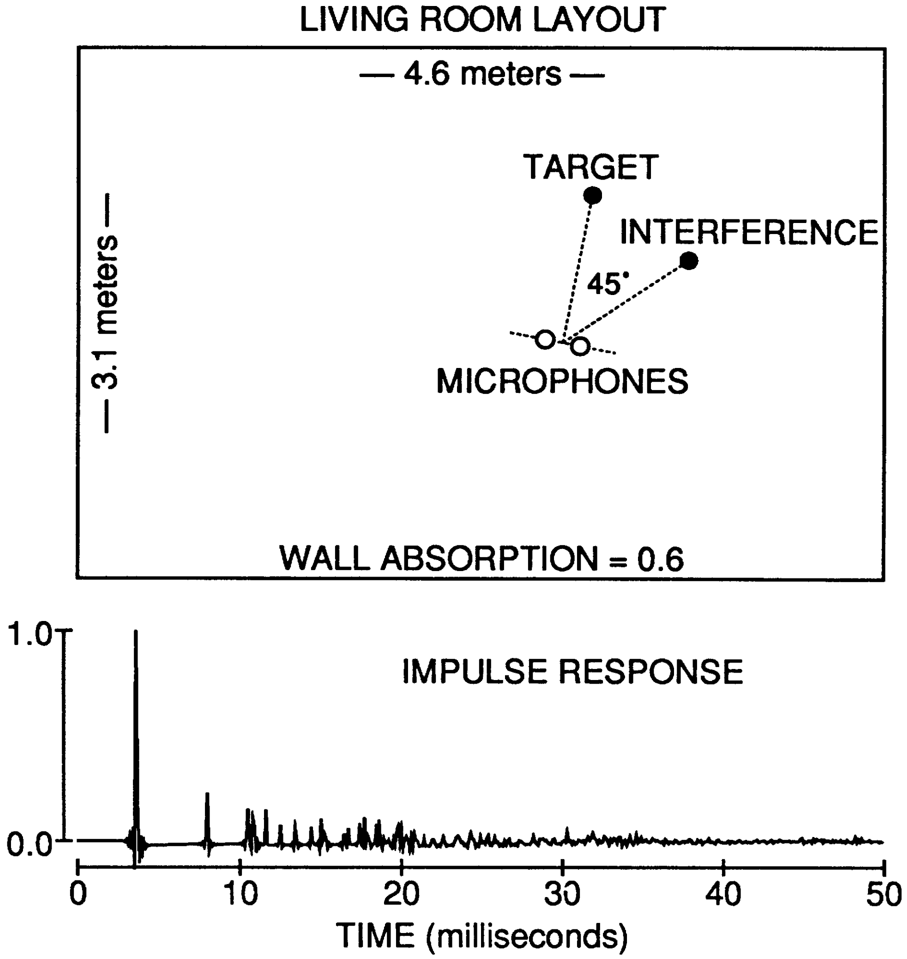


Figure 6.1: The layout used in simulating a living room environment and the first 50 milliseconds of one of the source-to-microphone impulse responses. The room height was 2.4 meters.

typical room and transfer function. Both the sound sources and the microphones were assumed to be omnidirectional. Thus, our simulation included neither transducer directivity nor head-shadow effects. The microphones were spaced 20 cm apart and, to make the simulation less sensitive to room modes, their connecting axis was not parallel to any wall. The target source was always located on a normal bisecting the axis connecting the microphones but at a slightly different height. The interference source was located at 45° off the normal to the array axis, also at a slightly different height. Table 6.1 summarizes the parameters of the simulated environments.

6.1.3 Intelligibility Tests

We administered intelligibility tests to normal-hearing subjects to compare target intelligibility for three cases: monaural unprocessed, binaural unprocessed, and monaural processed. In the binaural-unprocessed case, the signals from the two microphones were fed separately to the two ears. In the monaural-unprocessed case, only one microphone signal was presented. In the monaural-processed case, the signals from the two microphones were processed by the Griffiths-Jim beamformer and the beamformer output was presented to the listener[6].

The source materials used in the tests were digitized, single-channel, anechoic recordings of IEEE Harvard sentences (IEEE, 1969) for the target and SPIN babble (Kalikow, Stevens and Elliott, 1977) for the interference. Both sets of materials were low-pass filtered with a 4.5-kHz anti-aliasing filter and approximately whitened with 6-dB-per-octave high-frequency emphasis to increase intelligibility in the unprocessed conditions.

[6] In both "monaural" cases the presentations were actually diotic (identical in both ears) rather than monaural. For listeners whose hearing is perfectly symmetric, diotic and monaural presentations lead to essentially identical results.

Room	ANECHOIC	LIVING	CONFERENCE
Size (meters)	—	4.6 × 3.1 × 2.4	6.1 × 5.2 × 2.7
Microphone Locations (x,y,z in meters)[1]	(0, 0, 0) ±(0.10, -0.02, 0)	(2.76, 1.38, 1.55) ±(0.10, -0.02, 0)	(3.80, 1.73, 1.38) ±(0.10, -0.02, 0)
Target Location	(0.17, 0.86, 0.17)	(2.93, 2.24, 1.73)	(4.31, 4.14, 1.55)
Jammer Location	(0.72, 0.48, -0.17)	(3.48, 1.86, 1.38)	(5.87, 3.07, 1.21)
Target-to-Microphone Distance	0.9 m	0.9 m	2.5 m
Wall Absorption[2]	—	0.6	0.3
Reverberation Time[3]	—	120 ms	480 ms
Critical Distance[4]	∞	1.8 m	1.2 m
Direct-to-Reverberant Energy Ratio[5]	∞	5.9 dB	-6.3 dB

Notes:

[1] Specified as the coordinates of the midpoint between microphones plus or minus an offset to each microphone. Distances were originally specified in sample times and converted to meters based on a 10-kHz sampling rate and a sound speed of 345 m/sec.

[2] The ratio of energy absorbed to energy incident for each wall reflection; assumed uniform over all walls.

[3] The source-to-receiver distance at which the energy received directly from a source equals the energy received from all reflected paths.

[4] Time required for reverberant energy to decay by 60 dB.

[5] At the point midway between the two microphones.

Table 6.1: Characteristics of the three simulated reverberant environments.

6.2 Results

6.2.1 Intelligibility Measurements

The results of the intelligibility tests are shown in Figure 6.2. In the anechoic

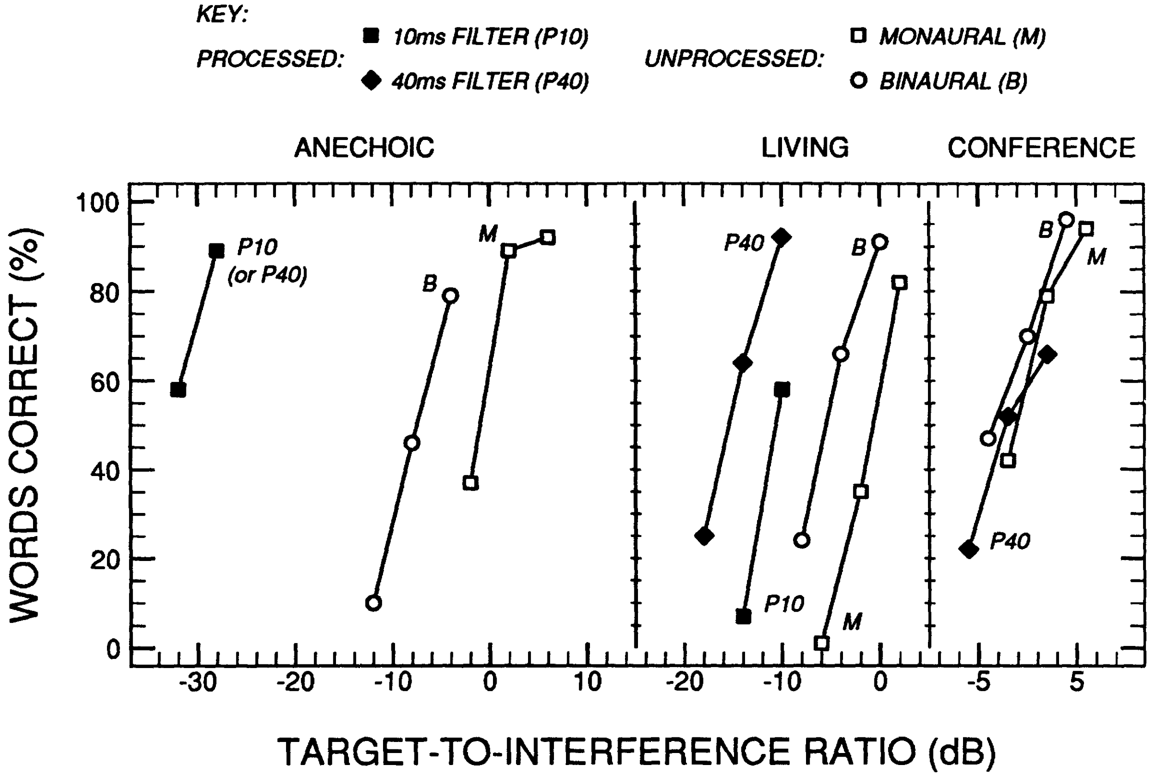


Figure 6.2: Percentage keywords correct as a function of Target-to-Interference power ratio in three different environments. Each curve represents data for one of four processing conditions: monaural-unprocessed (M), binaural-unprocessed (B), 100-point adaptive processing (P1), and 400-point adaptive processing (P4). Each data point is the average score of 5 normal-hearing subjects listening to 10 sentences with 5 keywords per sentence.

environment, listeners using unprocessed signals needed 6 dB less target power binaurally than monaurally for equivalent keyword intelligibility, a result roughly consistent with data in the literature (Carhart, Tillman and Greetis, 1969; Plomp, 1976). A 100-point beamforming system, on the other hand, achieved equivalent intelligibility with 30 dB less input target power than that required for the monaural-unprocessed case. Although, in theory, a beamformer could achieve perfect cancellation of one interference source in an anechoic environment, for our system, cancellation is limited by the misadjustment error of the LMS adaptation algorithm and by the time-variability of the input signals. In the living room environment, the binaural advantage for unprocessed signals fell slightly to 5 dB, while 100- and 400-point beamformers showed 9 and 14 dB of improvement, respectively, over the monaural-unprocessed condition. In the simulated conference room, the differences among tested conditions were less than 1 dB. Again, the results for the two unprocessed conditions are roughly consistent with other intelligibility experiments in highly reverberant environments (Moncur and Dirks, 1967; Plomp, 1976). These comparisons cannot be exact because the related studies were done with listeners in the acoustic field (thereby including head-shadow, pinna, and head-movement cues) and using different reverberant conditions.

6.2.2 Response Measurements

To illuminate the reasons for the measured intelligibility results, we made some objective measurements of system performance. The directional response was calculated from "snapshots" of the time-varying filter weights, and the magnitude of the frequency response in the target and interference directions was determined from input and output spectra.

Figure 6.3 shows the "power-averaged" broadband directional response[7] of the system after adaptation to interference alone in the three environments. The gain at 0° is 0 dB, as it should be, and in the direction of the interference there is a response null, whose depth depends on the environment. (In reverberant environments, interference echoes arrive from many directions.) The pattern is symmetrical about the axis connecting the two microphones. Although we observed 30 dB anechoic

[7] The data and analysis in this chapter predated the development of our intelligibility-averaging technique. By "power-averaging" we mean the averaging of power across frequency with equal weight given to all frequencies.

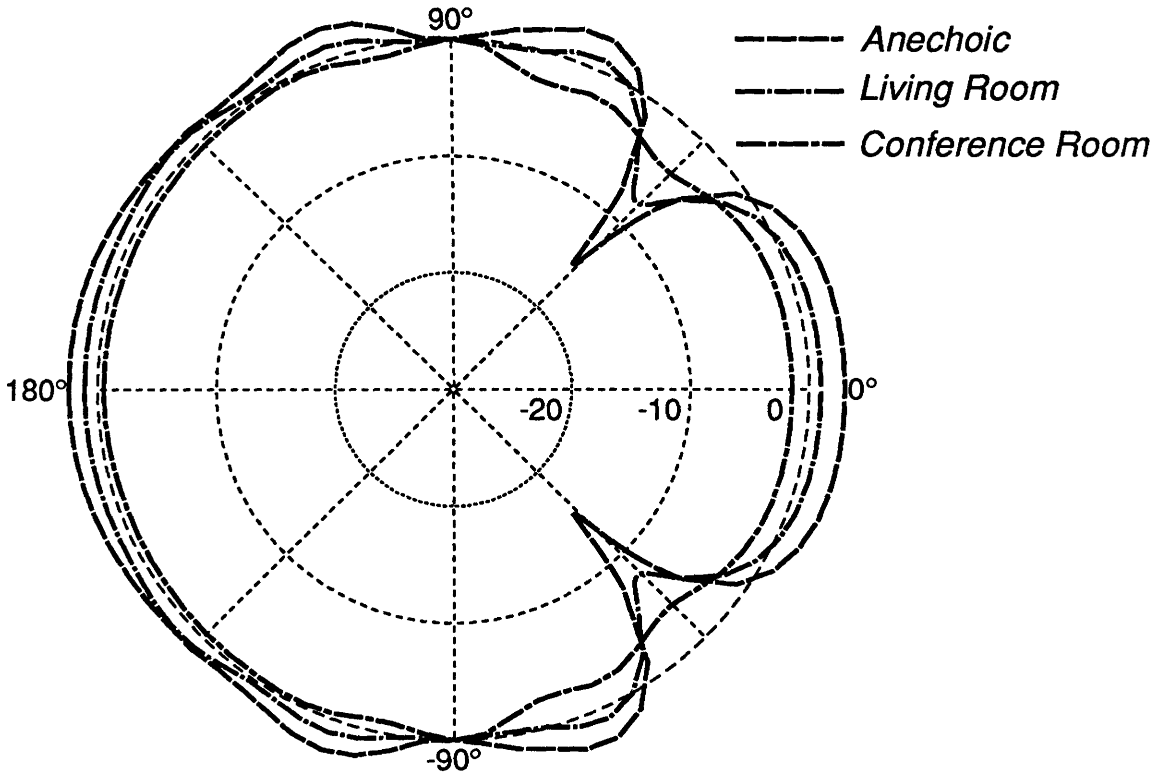


Figure 6.3: Power-averaged broadband beampatterns based on weight "snapshots" of the 2-microphone Griffiths-Jim beamformer in the three environments.

interference rejection in the intelligibility experiments, the power-averaged broadband anechoic null is less than 30 dB. A small part of this discrepancy is due to a poor method of capturing the time-varying filter weights. The major part of the discrepancy is due to the use of power-averaging to characterize broadband response. When intelligibility-averaging is applied to the frequency responses described next, the anechoic null is seen to be, effectively, 31.6 dB deep.

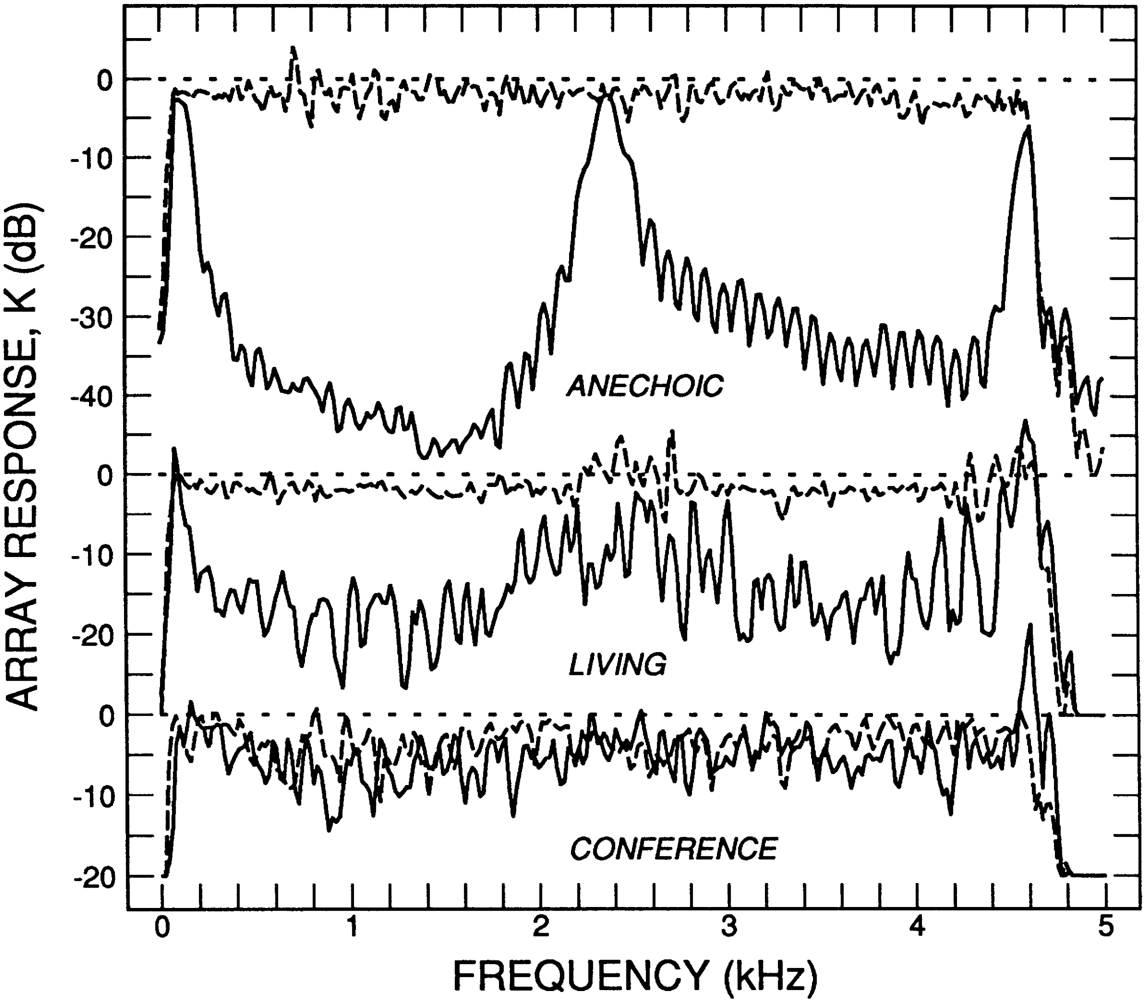


Figure 6.4: Time-averaged frequency-dependent beamformer responses, K_t and K_j for target and interference sources in the three environments. K_t is indicated with a dashed line; K_j with a solid line.

Figure 6.4 shows the magnitude of the beamformer frequency response for target

Environment	Adaptive filter length (ms)	ΔSRT (dB)	$\langle G \rangle_I$ (dB)
Anechoic (AN)	10	30	31.6
	40	–	31.2
Living (LV) Room	10	9	9.5
	40	14	16.7

Table 6.2: Comparison of the measured improvement in speech-reception threshold, ΔSRT, due to the adaptive beamformer, with computed estimates of intelligibility-averaged gain, $\langle G \rangle_I$, for the anechoic and living room environments.

and interference signals, calculated by taking the ratio of output to input average magnitude spectra. The output spectra were obtained from beamformers operating on target and interference signals separately but using filter coefficients from an identical beamformer that was adapting to target and interference at target-to-interference ratios corresponding to 50% intelligibility (-32 dB, -14 dB, and -2 dB for anechoic, living, and conference rooms). The beamformer is not able to reject interference at 2.15 and 4.3 kHz because the microphone separation causes signals from 45° to arrive in-phase at these frequencies. Target gain is not exactly 0 dB because position roundoff errors and reverberation cause the target signals to be different at the two microphones.

Using the measured frequency-responses in Figure 6.4 we can calculate the expected intelligibility-averaged gain, $\langle G \rangle_I$, due to beamforming and then compare $\langle G \rangle_I$ with ΔSRT, the change in speech-reception threshold, measured as the difference in target level necessary for 50% intelligibility. The comparison in Table 6.2 indicates that intelligibility-averaging can be used to predict speech-reception thresholds within about 2 dB.

Chapter 7

Summary and Discussion

We have now shown, both theoretically and experimentally, the potential of adaptive array processing for improving the intelligibility of a target talker by reducing interference from other spatially-distinct sources of sound. Such a "source separator" could be a useful component in multiple-microphone sensory aids for impaired listeners who cannot distinguish between multiple sources of sound in a room.

In Chapter 1 we described the problems of hearing-impaired listeners in complex acoustic environments. We contrasted these problems with the ability of non-impaired listeners to separate distinct sound sources into different "directional channels" and then consciously attend to one channel while subconsciously monitoring the other channels. We then described a strategy for designing sensory aids based on, first, using the information from multiple microphones to resolve separate sound sources and, then, somehow coding this information so the user could perform the concentrate/monitor tasks. This thesis addressed the source separation problem with the understanding that, even if the coding problem could not be solved, one or more directional channels, perhaps with some type of user control, might still contribute to improved sensory aids.

In Chapter 2 we reviewed human performance to establish a point of reference and also reviewed the state-of-the-art in multimicrophone hearing aids. We then described the basic signal processing scheme considered in this thesis, namely, linear combination of a finite number of past samples from multiple microphones into one output signal. With this processor in mind, we developed a stochastic received-signal model for use in later derivations. In its simplest form, this model involved the desired source signal, the transfer function from desired source to array samples, and the total noise observations. Total noise consisted of propagating and sensor noise and could be characterized by a cross-correlation or cross-spectral-density matrix. Finally, we introduced response measures to characterize an array's directional response to single plane-wave signals and its response to signal fields, which could be generated by multiple sources and, possibly, room echoes. In formulating broadband response measures we developed the powerful technique of "intelligibility-averaging" to characterize the potential effect of a given frequency response on speech intelligiblity.

We began our theoretical considerations in Chapter 3 with the description of frequency-domain, unlimited-observation optimum processors for many different criteria and showed that output speech intelligibility was relatively insensitive to choice of optimization criterion. Although unrealizable, frequency-domain processors are important because they facilitate the determination of asymptotic performance limits for related time-domain, limited-observation processors. We then described two (also nearly equivalent) time-domain optimum processors, one of which could be used to characterize the asymptotic behavior of the adaptive beamformers that we would eventually implement. Finally, we worked out a simple example to show how closely a time-domain processor for a given number of past microphone samples would approach asymptotic, unlimited-observation (frequency-domain) performance.

In Chapter 4 we evaluated asymptotic performance limits for head-sized line arrays mounted in free-space with both endfire and broadside orientations in the presence of either isotropic or multiple-source directional noise. (We argued later that head-shadow might not alter our results while other array configurations might significantly improve performance.) We evaluated performance as a function of the number of jammers, number of microphones, array orientation, and assumed sensor-noise level. The sensor-noise parameter was used to limit the sensitivity of a given processor to unmodelled noise and random implementation errors. Without this limit, processing would become extremely "superdirective" and infinitely sensitive to noise and errors. Because sensor noise was present, the performance of head-sized arrays did not increase indefinitely with number of microphones but saturated when the number of microphones reached 4 to 6 for small linear arrays. With other array configurations, such as circular or spherical, the number of useful microphones and the performance level might be higher. For both endfire and linear arrays (and presumably for any array), substantial "superdirective" increases in gain (i.e. at least a few dB) were achievable without excessive noise sensitivity. Linear endfire arrays often outperformed broadside arrays by a significant margin. One implication of this statement, that all arrays should have some extent in the target direction, may or may not be justified.

In Chapter 5 we described various adaptive beamformer implementations, all of which approach optimum performance, as described in the previous chapter, as long as the environment is stationary. Adaptive beamformers are especially attractive, however, because they can adapt automatically to changing environments.

In Chapter 6 we tested an adaptive beamformer experimentally by simulating a two-microphone system in rooms with target and interference sources and various amounts of reverberation. Reverberation was included because target reverberation violates the assumptions upon which adaptive beamformers are based and we were not able to analyze theoretically the effects of this violation on performance. We evaluated the simulated systems by conducting intelligibility tests with human listeners. The results of the intelligibility tests demonstrated the potential of adaptive array beamforming for hearing aids. Under the test conditions, and with zero-to-moderate reverberation, the interference reduction achieved by the array exceeded that achieved by the binaural auditory system. Furthermore, when the reverberation was severe, the array performed no worse than the binaural system.

To determine the generalizability of these results and their implications for a practical hearing aid, a variety of further studies must be performed. For example, interference reduction must be measured using different interference source angles and different reverberant conditions. Similarly, the effects of head shadow and transducer directivity must be included. Of even greater importance, performance with multiple independent interference sources must be studied. We would expect interference reduction to decrease dramatically for both the two-microphone beamformer and the binaural system when multiple sources (covering a range of angles) are introduced. However, whether the array maintains superiority over the binaural system under such conditions is unknown. In principle, performance with

N independent sources of interference can be greatly enhanced by using arrays with $N + 1$ microphones. Such arrays should combat multiple noise sources much more effectively than the binaural system. Detailed studies are required, however, to evaluate practical realizations of such systems.

Some of these studies are now being conducted (or have already been completed) in collaboration with other students in our group. We have looked at the possibility of including head-shadow in the reverberant room simulation (Hammerschlage, 1988). We have done an initial study of 2- and 4-microphone systems with multiple jammers that demonstrated a substantial improvement with more microphones but also demonstrated performance degradation in the presence of strong, misaligned targets (Wei, 1988; Peterson, Wei, Rabinowitz and Zurek, 1989). Finally, we have begun to construct a real-time beamformer with algorithms modified to tolerate strong and misaligned targets (Greenberg, Zurek and Peterson, 1989). This system will be used for realistic evaluations in many more situations than we were able to consider in this thesis.

An issue that has not yet been addressed is the question of adaptation time. A practical system will have to adapt to changing environments quickly enough to keep interference low. One obvious danger is that interference may suddenly appear from a direction (distinct from the target direction) in which the array has greater than normal sensitivity. If adaptation is too slow, the benefit of adaptive beamforming will be lost. At the present time, we have little data on the magnitude and time-scale of environmental variability. Consequently, it is unclear how best to evaluate the adaptation characteristics of various proposed adaptive beamforming arrays, although measurements of array response to the sudden appearance of a source or to modulation of source position would certainly be valuable[1]. In the case of the Griffiths-Jim beamformer, parameters can be adjusted to reduce adaptation time at the cost of steady-state performance. (Compare, for example, the results for filter lengths of 400 and 100 points in Figure 6.2 and recall that shorter filters adapt faster). If a Griffiths-Jim beamformer cannot achieve adequate performance and

[1] Informal measurements indicate that the beamformers used in this evaluation accomplish most of their adaptation to a new jammer within 1 second.

sufficiently fast adaptation simultaneously, then alternative, fast-adapting methods (Cioffi and Kailath, 1984; Lee, Morf and Friedlander, 1981) should be explored, although these methods may be more difficult than the LMS method to realize in a practical hearing aid.

Appendix A

Useful Matrix Identities

The following matrix identities are based on the Sherman-Morrison-Woodbury matrix inversion lemma(Golub and Van Loan, 1983). Assuming that P and Q are invertible,

$$\left(P^{-1}+M^{\dagger}Q^{-1}M\right)^{-1}\equiv P-PM^{\dagger}\left(MPM^{\dagger}+Q\right)^{-1}MP \tag{A.1}$$

can be verified by direct multiplication with $\left(P^{-1}+M^{\dagger}Q^{-1}M\right)$.

$$\left(P^{-1}+M^{\dagger}Q^{-1}M\right)^{-1}M^{\dagger}Q^{-1}\equiv PM^{\dagger}\left(MPM^{\dagger}+Q\right)^{-1} \tag{A.2}$$

can be derived from (A.1) by algebraic manipulation and can be verified by direct substitution.

$$\left(MPM^{\dagger}+Q\right)^{-1}\equiv Q^{-1}-Q^{-1}M\left(P^{-1}+M^{\dagger}Q^{-1}M\right)^{-1}M^{\dagger}Q^{-1} \tag{A.3}$$

is a restatement of (A.1) after the transformation $\left\{M,P,Q\mapsto M^{\dagger},Q^{-1},P^{-1}\right\}$.

Appendix B

Equivalence of Two Optimum Weights

Our goal is to show that, when the target is a stationary random process,

$$\left(\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{xx}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{xx}}^{-1} \;=\; \left(\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,.$$

The received signal model and assumptions of stationarity and independence allow us to write

$$\mathcal{S}_{\boldsymbol{xx}}^{-1} \;=\; \left(\underline{\mathcal{H}}\,\mathcal{S}_{ss}\,\underline{\mathcal{H}}^{\dagger} + \mathcal{S}_{\boldsymbol{zz}}\right)^{-1}\,.$$

Application of (A.3) gives

$$\mathcal{S}_{\boldsymbol{xx}}^{-1} \;=\; \mathcal{S}_{\boldsymbol{zz}}^{-1} - \mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\left(\mathcal{S}_{ss}^{-1} + \underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,.$$

Using this equivalence and then (A.3) with $\left\{M, P, Q \mapsto 1, \mathcal{S}_{ss}, \left(\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\right\}$,

$$\begin{aligned}
\left(\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{xx}}^{-1}\,\underline{\mathcal{H}}\right)^{-1} &= \left(\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}} - \underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\left(\mathcal{S}_{ss}^{-1} + \underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1} \\
&= \mathcal{S}_{ss} + \left(\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\,.
\end{aligned}$$

Finally, applying these expansions to the original expression,

$$\begin{aligned}
&\left(\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{xx}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{xx}}^{-1} \\
&\quad= \left[\mathcal{S}_{ss} + \left(\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\right]\left[\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1} - \underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\left(\mathcal{S}_{ss}^{-1} + \underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\right] \\
&\quad= \left(\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1} \\
&\qquad + \left[\mathcal{S}_{ss} - \left(\mathcal{S}_{ss}^{-1} + \underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1} - \mathcal{S}_{ss}\,\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\left(\mathcal{S}_{ss}^{-1} + \underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\right]\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1} \\
&\quad= \left(\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1} \\
&\qquad + \left[\mathcal{S}_{ss} - \mathcal{S}_{ss}\left(\mathcal{S}_{ss}^{-1} + \underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)\left(\mathcal{S}_{ss}^{-1} + \underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\right] \\
&\quad= \left(\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,\underline{\mathcal{H}}\right)^{-1}\underline{\mathcal{H}}^{\dagger}\,\mathcal{S}_{\boldsymbol{zz}}^{-1}\,,
\end{aligned}$$

and the original assertion is proven.

Bibliography

Anderson, B. D. O. and Moore, J. B., (**1979**). *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, New Jersey.

ANSI, (**1969**). *American National Standard Methods for the Calculation of the Articulation Index*, Technical Report ANSI S3.5-1969, American National Standards Institute, Inc., 1430 Broadway, New York, NY 10018.

Baggeroer, A. B., (**1976**). *Space/Time Random Processes and Optimum Array Processing*, Technical Report NUC TP 506, Naval Undersea Center, San Diego, California.

Beranek, L. L., (**1954**). *Acoustics*, McGraw-Hill, New York.

Beranek, L. L., (**1988**). *Acoustical Measurements*, American Institute of Physics, New York.

Blauert, J., (**1983**). *Spatial Hearing*, MIT Press, Cambridge, Massachusetts.

Bloch, A., Medhurst, R., and Pool, S., (**1953**). "A new approach to the design of super-directive aerial arrays", *Proceedings of the IEE (London)*, **100**(3),303–314.

Boll, S. F., (**1979**). "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Transactions on Acoustics, Speech and Signal Processing*, **ASSP-27**(2),113–120.

Brey, R. H. and Robinette, M. S., (**1982**). "Noise reduction in speech using adaptive filtering II: Speech intelligibility improvement with normal and hearing impaired students", *Journal of the Acoustical Society of America*, **71**(Suppl. 1),S8(A).

Bryn, F., (**1962**). "Optimum signal processing of three-dimensional arrays operating on gaussian signals and noise", *Journal of the Acoustical Society of America*, **34**(3),289–297.

Carhart, R., Tillman, T. W., and Greetis, E. S., (**1969**). "Release from multiple maskers: Effects of interaural time disparities", *Journal of the Acoustical Society of America*, **45**(2),411–418.

Christiansen, R. W., Chabries, D. M., and Lynn, D., (**1982**). "Noise reduction in speech using adaptive filtering I: Signal processing algorithms", *Journal of the Acoustical Society of America*, **71**(Suppl. 1),S7(A).

Chu, L. J., (**1948**). "Physical limitations of omni-directional antennas", *Journal of Applied Physics*, **19**,1163.

Cioffi, J. M. and Kailath, T., (**1984**). "Fast, recursive-least-squares transversal filters for adaptive filtering", *IEEE Transactions on Acoustics, Speech and Signal Processing*, **32**(2),304–337.

Cook, R. K., Waterhouse, R. V., Berendt, R. D., Edelman, S., and Thompson, M. C., Jr., (**1955**). "Measurement of correlation coefficients in reverberant sound fields", *Journal of the Acoustical Society of America*, **27**(6),1072–1077.

Corbett, C. R., (**1986**). *Filtering Competing Messages to Enhance Mutual Intelligibility*, Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Cox, H., (**1968**). "Interrelated problems in detection and estimation i and ii", in *NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics*, pages 23–16 to 23–64, Enschede, the Netherlands.

Cox, H., (**1973a**). "Resolving power and sensitivity to mismatch of optimum array processors", *Journal of the Acoustical Society of America*, **54**(3),771–785.

Cox, H., (**1973b**). "Sensitivity considerations in adaptive beamforming", in Griffiths, J. W. R., Stocklin, P. L., and vanSchooneveld, C., editors, *Signal Processing*, pages 619–645, NATO Advanced Study Institute on Signal Processing, Loughborough, U. K., August, 1972, Academic Press, New York.

Cox, H., Zeskind, R. M., and Kooij, T., (**1985**). "Sensitivity constrained optimum endfire array gain", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, ,46.12.

Cox, H., Zeskind, R. M., and Kooij, T., (**1986**). "Practical supergain", *IEEE Transactions on Acoustics, Speech and Signal Processing*, **ASSP-34**(3),393–398.

Cron, B. F. and Sherman, C. H., (**1962**). "Spatial correlation functions for various noise models", *Journal of the Acoustical Society of America*, **34**(11),1732–1736.

Davenport, W. B., Jr. and Root, W. L., (**1958**). *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York.

Dirks, D. and Moncur, J. P., (**1967**). "Interaural intensity and time differences in anechoic and reverberant rooms", *Journal of Speech and Hearing Research*, **10**,177–185.

Dirks, D. D. and Wilson, R. H., (**1969**). "Binaural hearing of speech for aided and unaided conditions", *Journal of Speech and Hearing Research*, **12**,650–664.

Doob, J. L., (**1953**). *Stochastic Processes*, John Wiley and Sons, New York.

Duhamel, R. H., (**1953**). "Optimum patterns for endfire arrays", *Proceedings of the IRE*, **41**,652–659.

Durlach, N. I. and Colburn, H. S., (**1978**). "Binaural phenomena", in Carterette, E. C. and Friedman, M. P., editors, *Handbook of Perception, Volume 4: Hearing*, chapter 10, pages 360–466, Academic Press, New York.

Durlach, N. I., Corbett, C. R., McConnell, M. V. C., Rabinowitz, W. M., Peterson, P. M., and Zurek, P. M., (**1987**). "Multimicrophone monaural hearing aids", in *RESNA 10th Annual Conference*, RESNA, San Jose, California.

Elliott, R. S., (**1981**). *Antenna Theory and Design*, Prentice-Hall, New Jersey.

Foss, K. K., (**1984**). *Hearing Aid Application of Adaptive Noise Cancellation Utilizing a Two Microphone Array*, Master's thesis, Sever Institute of Washington University, Saint Louis, Missouri.

Frazier, R. H., Samsam, S., Braida, L. D., and Oppenheim, A. V., (**1976**). "Enhancement of speech by adaptive filtering", in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Philadelphia, Pennsylvania.

French, N. R. and Steinberg, J. C., (**1947**). "Factors governing the intelligibility of speech sounds", *Journal of the Acoustical Society of America*, **19**(1),90–119.

Frost, O. L., (**1972**). "An algorithm for linearly constrained adaptive array processing", *Proceedings of the IEEE*, **60**,926–935.

Gardner, W. A., (**1986**). *Introduction to Random Processes*, Macmillan, New York.

Gelfand, S. A. and Hochberg, I., (**1976**). "Binaural and monaural speech discrimination under reverberation", *Audiology*, **15**,72–84.

Golub, G. H. and Van Loan, C. F., (**1983**). *Matrix Computations*, Johns Hopkins, Baltimore.

Goodman, N. R., (**1963**). "Statistical analysis based on a certain multivariate complex distribution (an introduction)", *Annals of Mathematical Statistics*, **34**,152–157.

Graupe, D., Grosspietsch, J. K., and Basseas, S. P., (**1987**). "A single-microphone-based self-adaptive filter of noise from speech and its performance evaluation", *Journal of Rehabilitation Research and Development*, **24**(4),119–126.

Gray, R. M., (**1972**). "On the asymptotic eigenvalue distribution of toeplitz matrices", *IEEE Transactions on Information Theory*, **IT-18**(6),725–730.

Greenberg, J. E., Zurek, P. M., and Peterson, P. M., (**1989**). "Reducing the effects of target misalignment in an adaptive beamformer for hearing aids", *Journal of the Acoustical Society of America*, submitted to Supplement for 117th Meeting.

Griffiths, L. J. and Jim, C. W., (**1982**). "An alternative approach to linearly constrained adaptive beamforming", *IEEE Transactions on Antennas and Propagation*, **30**(1),27–34.

Hammerschlage, R., (**1988**). "Using a rigid-sphere model of the head to simulate binaural impulse responses in a reverberant room". MIT Bachelor's Thesis.

Hansen, R. C., (**1981**). "Fundamental limitations in antennas", *Proceedings of the IEEE*, **69**(2),170–182.

Harford, E. and Dodds, E., (**1974**). "Versions of the CROS hearing aid", *Archives of Otolaryngology*, **100**,50–57.

Hirsh, I. J., (**1950**). "The relation between localization and intelligibility", *Journal of the Acoustical Society of America*, **22**(2),196–200.

Honig, M. L. and Messerschmitt, D. G., (**1984**). *Adaptive Filters: Structures, Algorithms, and Applications*, Kluwer, Boston.

IEEE, (**1969**). *IEEE Recommended Practice for Speech Quality Measurements*, Technical Report IEEE 297, Institute of Electrical and Electronics Engineers, New York.

Jerger, J. and Dirks, D., (**1961**). "Binaural hearing aids. An enigma", *Journal of the Acoustical Society of America*, **33**(4),537–538.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L., (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability", *Journal of the Acoustical Society of America*, **61**(5),1337–1351.

Knowles Electronics, (**1980**). *EB Directional Hearing Aid Microphone Application Notes*, Technical Report TB-21, Knowles Electronics Inc.

Kock, W. E., (**1950**). "Binaural localization and masking", *Journal of the Acoustical Society of America*, **22**(6),801–804.

Koenig, W., (**1950**). "Subjective effects in binaural hearing", *Journal of the Acoustical Society of America*, **22**(1),61–62.

Kryter, K. D., (**1962a**). "Methods for the calculation and use of the articulation index", *Journal of the Acoustical Society of America*, **34**(11),1689–1697.

Kryter, K. D., (**1962b**). "Validation of the articulation index", *Journal of the Acoustical Society of America*, **34**(11),1698–1702.

Lee, D. T., Morf, M., and Friedlander, B., (**1981**). "Recursive least squares ladder estimation algorithms", *IEEE Transactions on Acoustics, Speech and Signal Processing*, **29**(3),627–641.

Lim, J. S., (**1983**). *Speech Enhancement*, Prentice-Hall, New Jersey.

Lim, J. S. and Oppenheim, A. V., (**1979**). "Enhancement and bandwidth compression of noisy speech", *Proceedings of the IEEE*, **67**(12),1586–1604.

Lindevald, I. M. and Benade, A. H., (**1986**). "Two-ear correlation in the statistical sound fields of rooms", *Journal of the Acoustical Society of America*, **80**(2),661–664.

Lochner, J. P. A. and Burger, J. F., (**1961**). "The intelligibility of speech under reverberant conditions", *Acustica*, **7**,195–200.

Lotterman, S. H., Kasten, R. N., and Revoile, S. G., (**1968**). "An evaluation of the CROS-type hearing aid", *Bulletin of Prosthetics Research*, ,104–109.

MacKeith, N. W. and Coles, R. R. A., (**1971**). "Binaural advantages in hearing of speech", *Journal of Laryngology and Otology*, **85**,213–232.

Madison, T. K. and Hawkins, D. B., (**1983**). "The signal-to-noise ratio advantage of directional microphones", *Hearing Instruments*, **34**(2),18.

Markides, A., (**1977**). *Binaural Hearing Aids*, Academic Press, New York.

Moncur, J. P. and Dirks, D., (**1967**). "Binaural and monaural speech intelligibility in reverberation", *Journal of Speech and Hearing Research*, **10**,186–195.

Monzingo, R. A. and Miller, T. W., (**1980**). *Introduction to Adaptive Arrays*, John Wiley and Sons, New York.

Morse, P. M., (**1976**). *Vibration and Sound*, American Institute of Physics, New York.

Mueller, H. G., Grimes, A. M., and Erdman, S. A., (**1983**). "Subjective ratings of directional amplification", *Hearing Instruments*, **34**(2),14–16.

Nabelek, A. K., (**1982**). "Temporal distortions and noise considerations", in Studebaker, G. A. and Bess, F. H., editors, *The Vanderbilt Hearing-Aid Report*, Monographs in Contemporary Audiology, Upper Darby, Pennsylvania.

Nabelek, A. K. and Pickett, J. M., (**1974**). "Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners", *Journal of Speech and Hearing Research*, **17**(4),724–739.

Neuman, A. C. and Schwander, T. J., (**1987**). "The effect of filtering on the intelligibility and quality of speech in noise", *Journal of Rehabilitation Research and Development*, **24**(4),127–134.

Newman, E. and Shrote, M., (**1982**). "A wide-band electrically small superdirective array", *IEEE Transactions on Antennas and Propagation*, **AP-30**,1172–1176.

Ono, H., Kanzaki, J., and Mizoi, K., (**1983**). "Clinical results of hearing aid with noise-level-controlled amplification", *Audiology*, **22**,494–515.

Oppenheim, A. V. and Johnson, D. H., (**1972**). "Discrete representation of signals", *Proceedings of the IEEE*, **60**(6),681–691.

Owsley, N. L., (**1985**). "Sonar array processing", in Haykin, S., editor, *Array Signal Processing*, chapter 3, pages 115–193, Prentice-Hall, Englewood Cliffs.

Peterson, P. M., (**1986**). "Simulating the response of multiple microphones to a single acoustic source in a reverberant room", *Journal of the Acoustical Society of America*, **80**(5),1527–1529.

Peterson, P. M., (**1987**). "Using linearly-constrained beamforming to reduce interference in hearing aids from competing talkers in reverberant rooms", in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, Texas.

Peterson, P. M., Durlach, N., Rabinowitz, W. M., and Zurek, P. M., (**1987**). "Multimicrophone adaptive beamforming for interference reduction in hearing aids", *Journal of Rehabilitation Research and Development*, **24**(2),103–110.

Peterson, P. M., Wei, S., Rabinowitz, W. M., and Zurek, P. M., (**1989**). "Robustness of an adaptive beamforming method for hearing aids", *Acta Otolaryngologica*, in press.

Plomp, R., (**1976**). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)", *Acustica*, **34**,200–211.

Plomp, R. and Mimpen, A. M., (**1981**). "Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences", *Acustica*, **48**,325–328.

Pritchard, R. L., (**1954**). "Maximum directivity index of a linear point array", *Journal of the Acoustical Society of America*, **26**(6),1034–1039.

Rabinowitz, W. M., Frost, D. A., and Peterson, P. M., (**1985**). "Hearing-aid microphone systems with increased directionality", *Journal of the Acoustical Society of America*, **78**(Suppl. 1),S41.

Reed, I. S., (**1962**). "On a moment theorem for complex gaussian processes", *IEEE Transactions on Information Theory*, **8**,194–195.

Santon, F., (**1976**). "Numerical prediction of echograms and of the intelligibility of speech in rooms", *Journal of the Acoustical Society of America*, **59**(6),1399–1405.

Schelkunoff, S. A., (**1943**). "A mathematical theory of linear arrays", *Bell System Technical Journal*, **22**,80–107.

Selby, S. M., editor, (**1975**). *CRC Standard Mathematical Tables*, CRC Press, Cleveland, 23 edition.

Shaw, E. A. G., (**1974**). "Transformation of sound pressure from the free field to the eardrum in the horizontal plane", *Journal of the Acoustical Society of America*, **56**,1848–1861.

Siegenthaler, B., (**1979**). "The non-universal binaural hearing advantage", in Yanick, P., Jr., editor, *Rehabilitation Strategies for Sensorineural Hearing Loss*, chapter 5, Grune and Stratton, New York.

Strang, G., (**1976**). *Linear Algebra and its Applications*, Academic Press, New York.

Strube, H. W., (**1981**). "Separation of several speakers recorded by two microphones (cocktail party processing)", *Signal Processing*, **3**,1–10.

Taylor, T. T., (**1948**). "A discussion of the maximum directivity of an antenna", *Proceedings of the IRE*, **36**,1135.

Tonning, F., (**1971**). "Directional audiometry II. The influence of azimuth on the perception of speech", *Acta Otolaryng.*, **72**,352–357.

Uzkov, A. I., (**1946**). "An approach to the problem of optimum directive antennae design", *Comptes Rendus (Doklady) de l'Academie des Sciences de l'URSS*, **53**(1),35–38.

Van Tassell, D. J., Larsen, S. Y., and Fabry, D. A., (**1988**). "Effects of an adaptive filter hearing aid on speech recognition in noise by hearing-impaired subjects", *Ear and Hearing*, **9**(1),15–21.

Van Trees, H. L., (**1968**). *Detection, Estimation, and Modulation Theory*, Volume 1, John Wiley and Sons, New York.

Vanderkulk, W., (**1963**). "Optimum processing for acoustic arrays", *Journal of the British Institution of Radio Engineers*, October, 285–292.

Wei, S., (**1988**). *Sensitivity of Multimicrophone Adaptive Beamforming to Variations in Target Angle and Microphone Gain*, Master's thesis, MIT.

Weston, D. E., (**1986**). "Jacobi sensor arrangement for maximum array directivity", *Journal of the Acoustical Society of America*, **80**(4),1170–1181.

Widrow, B., Glover, J. R., Jr., McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., Zeidler, J. R., Dong, E., Jr., and Goodlin, R. C., (**1975**). "Adaptive noise cancelling: Principles and applications", *Proceedings of the IEEE*, **63**(12),1692–1716.

Wiener, N., (**1930**). "Generalized harmonic analysis", *Acta Mathematica*, **55**,117–258. Reprinted in: Selected Papers of Norbert Wiener, MIT Press, Cambridge, 1964.

Wiener, N., (**1949**). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, John Wiley and Sons, New York.

Zurek, P. M., (**1980**). "The precedence effect and its possible role in the avoidance of interaural ambiguities", *Journal of the Acoustical Society of America*, **67**(3),952–964.

Zurek, P. M., (**1988** (in revision)). "A predictive model for binaural advantages and directional effects in speech intelligibility", *Journal of the Acoustical Society of America*, in revision.