Evaluation of an adaptive beamforming method for hearing aids

Julie E. Greenberg and Patrick M. Zurek
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 15 July 1991; revised 27 September 1991; accept 1 November 1991)

In this paper evaluations of a two-microphone adaptive beamforming system for hearing aids are presented. The system, based on the constrained adaptive beamformer described by Griffiths and Jim [IEEE Trans. Antennas Propag. **AP-30**, 27–34 (1982)], adapts to preserve target signals from straight ahead and to minimize jammer signals arriving from other directions. Modifications of the basic Griffiths–Jim algorithm are proposed to alleviate problems of target cancellation and misadjustment that arise in the presence of strong target signals. The evaluations employ both computer simulations and a real-time hardware implementation and are restricted to the case of a single jammer. Performance is measured by the spectrally weighted gain in the target-to-jammer ratio in the steady state. Results show that in environments with relatively little reverberation: (1) the modifications allow good performance even with misaligned arrays and high input target-to-jammer ratios; and (2) performance is better with a broadside array with 7-cm spacing between microphones than with a 26-cm broadside or a 7-cm endfire configuration. Performance degrades in reverberant environments; at the critical distance of a room, improvement with a practical system is limited to a few dB.

PACS numbers: 43.66.Ts, 43.72.Ew, 43.60.Lq

INTRODUCTION

Interference from background noise is one of the primary problems of hearing-aid users and the hearing impaired (Plomp, 1978; Smedley and Schow, 1990). Traditional hearing aids amplify all sounds without discriminating between the desired target and the background noise. Although a variety of single-microphone speech-enhancement techniques have been investigated (Lim and Oppenheim, 1979), the only one shown to produce clear improvements in speech intelligibility involves simple modifications of the frequency response to reduce spectrally localized interference (Van Dijkhuizen *et al.*, 1989; Rankovic *et al.*, 1992). The meager success of single-microphone systems, together with the known advantages of multiple-element sensing systems, has led to interest in multiple-microphone hearing aids. Multiple-microphone hearing aids can be classified as either fixed or adaptive systems. Fixed systems are typically designed to maximize directionality, while adaptive systems can provide additional benefits in time-varying acoustic conditions. This work is restricted to the study of adaptive systems, although it should be noted that a successful adaptive system must perform at least as well as a comparable fixed system. In this paper the performance of an adaptive two-microphone noise-reduction system is evaluated.

I. BACKGROUND

Beamformers process the signals from a spatially distributed array of sensors to improve reception of the *target*, a sound emanating from a specified direction, in the presence of *jammers*, sounds arriving from other directions. An adaptive beamformer is one whose processing depends on, and changes with, characteristics of the input signals. The reader is referred to Van Veen and Buckley (1988) and Widrow and Stearns (1985) for excellent introductions to adaptive beamforming and its applications.

A. The Griffiths–Jim beamformer

The system studied here is based on an adaptive beamforming algorithm described by Griffiths and Jim (1982). A two-microphone version[1] of the system for straight-ahead (0°) targets is shown in Fig. 1.

The operation of this beamformer is most easily described as that of an adaptive noise canceller (Widrow *et al.*, 1975) with a preprocessor that forms the sum and difference of the microphone signals. If the target is straight ahead and there is no reverberation, then subtraction cancels the target and produces a signal that depends on the jammer alone. The sum and difference are then used as inputs to the adaptive noise canceller, which requires a primary input that contains target plus jammer (sum signal), and a reference input that ideally contains a filtered version of the jammer only (difference signal). The reference signal passes through an adaptive finite-impulse-response filter whose weights are adjusted to minimize the power in the output signal. This minimization is achieved by filtering the reference input to approximate the correlated signal in the primary path, and subtracting. The delay in the primary path allows the adaptive filter's response to be noncausal. If the target and jammer are uncorrelated and if the reference input contains no target signal, then minimizing the output power results in an output signal with minimum jammer power and no target distortion.

One of the simplest adaptive algorithms, and the one employed exclusively in this study, is the LMS algorithm (Widrow and Stearns, 1985), which attempts to minimize

 0001-4966/92/031662-15$00.80

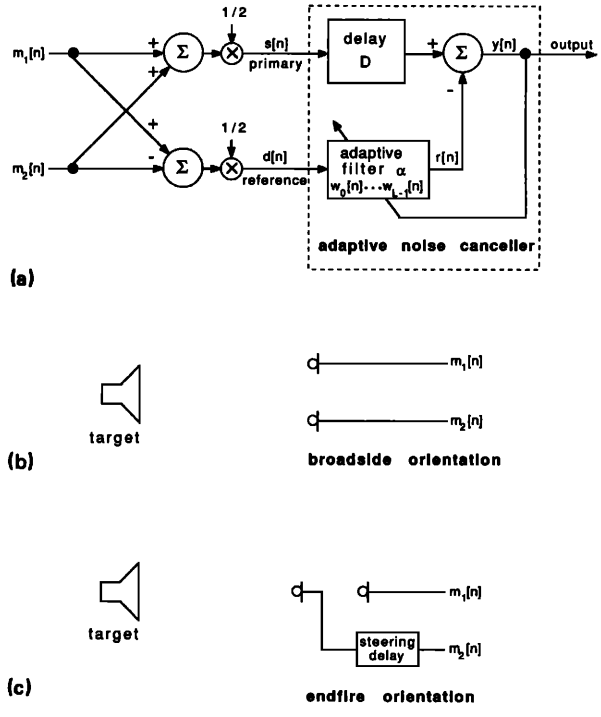

FIG. 1. (a) Two-microphone Griffiths–Jim beamformer. The microphone signals, $m_1[n]$ and $m_2[n]$ are combined to form sum and difference signals, $s[n]$ and $d[n]$. These signals are used as the primary and reference inputs to an LMS adaptive noise canceller. The adaptive noise canceller incorporates a D-sample delay and an L-point adaptive filter with weights $w_0[n],...,w_{L-1}[n]$ and adaptation constant α. The reference input is filtered and subtracted from the delayed primary input. The LMS algorithm adjusts the filter weights to minimize total output power, which, under ideal conditions, also minimizes jammer output power. (b) Broadside array orientation. For straight-ahead targets, the broadside orientation requires no delay to equalize microphone signals. (c) Endfire array orientation. For straight-ahead targets, the array is rotated so the microphones are on a line with the target source. In this case, the front microphone signal is delayed to phase align the target signal.

output noise in the least-mean-square sense. The form of the algorithm implemented here updates the adaptive filter according to the formula

$$w_l[n+1] = w_l[n] + 2\alpha y[n]d[n-l]/LP_d[n], \quad (1)$$

where n is the discrete time index, $w_l[n]$ is the l th coefficient of the L-point adaptive filter ($l = 0,...,L-1$), $d[n]$ is the difference (reference) signal, $y[n]$ is the beamformer output signal, $P_d[n]$ is a running estimate of power in the reference signal, and α is a constant that controls the rate of adaptation. The convolution performed by the adaptive filter is described by

$$r[n] = \sum_{l=0}^{L-1} d[n-l]w_l[n], \quad (2)$$

and the system output $y[n]$ is given by

$$y[n] = s[n-D] - r[n], \quad (3)$$

where $s[n]$ is the sum (primary) signal and D is the delay in the primary channel. In this work, $D = L/2$.

Under ideal conditions, sources that are exactly straight ahead of the array pass through the system with unit gain, since signal components arriving at the two microphones with equal magnitude and phase do not appear at the reference input and are not affected by the adaptive cancellation process. Off-axis sources contribute components to both the reference and primary inputs and therefore are subject to cancellation. The basic characteristic of this system that makes it appealing for application to hearing aids is that the reference signal is obtained from the array itself—which may be mounted on the head—obviating the need for a remote microphone (e.g., Brey *et al.*, 1987; Harrison *et al.*, 1986).

B. Important issues and previous studies

Under realistic conditions, the hearing-aid application challenges the robustness of adaptive beamformers because the usual operating assumptions are all violated to some degree and computational resources are at a minimum. System performance is obviously influenced by violations of the two fundamental assumptions: uncorrelated target and jammer and exact knowledge of target direction. To some extent, performance of the noise canceller is limited by the ability of the preprocessor in Fig. 1 to provide an uncorrelated, target-free reference signal. (See the sections below on array misalignment and reverberation.) On the other hand, performance is also limited by the acoustic environment (degree of reverberation and number of jammer sources), the system configuration (array geometry and placement), and implementation issues (adaptation rate, misadjustment, and filter length). A meaningful evaluation, therefore, must consider not only the performance of the idealized system, but also the impact of imperfections and limited ranges of parameters. Because the effects of errors, of parameter choices, and of resource limitations are all complex and often interactive, there is a web of issues to be considered. The following is a summary of the more important issues, presented with the results of recent studies that have begun to document adaptive-beamformer performance under conditions relevant to hearing aids.

1. Array misalignment

With any deviation from perfectly symmetric alignment of the array to the target source, the target will not be entirely eliminated by subtraction of the microphone signals, and residual target components will "leak" into the reference channel. The importance of target leakage was illustrated by Widrow *et al.* (1975), who showed that, for the case of an unconstrained adaptive filter (one whose impulse response extends infinitely in both time directions), the target-to-jammer ratio at the output of the noise canceller is equal to the jammer-to-target ratio at the reference input. When the leakage path has any nonzero transfer function, the problem clearly worsens as the input target-to-jammer ratio (TJR) increases, leading to poorer system performance at better TJRs.

The degradation in system performance caused by target leakage depends on the degree of leakage, but it is generally seen with TJRs as low as 0 dB and is clearly detrimental at 10–20 dB (Peterson *et al.*, 1990). In many noise-cancelling applications, the TJR is always less than 0 dB, and conse-

quently moderate target leakage is tolerable. However, in the hearing-aid application, TJRs greater than 0 dB are encountered frequently. Since at a minimum hearing aids must do no harm, the tendency to degrade clean targets must be eliminated if the system is to have any chance of success. A solution to this problem is proposed in Sec. II A, and its effectiveness is demonstrated in Sec. IV A.

2. Multiple jammers

Because the two-microphone beamformer has one adaptive filter, only one transformation can be implemented to estimate jammer in the primary channel based on jammer in the reference channel. For a single jammer this transformation is theoretically sufficient for optimal cancellation of that jammer. For multiple jammers from different locations, a single transformation cannot provide optimal cancellation of all jammers. The worst case is expected when multiple jammers have the same spectral shape and power (Peterson, 1989). Then the total reduction in jammer power achieved by the system may be no more than a few dB, whereas the reduction of any one jammer presented alone may be tens of decibels. Experimental confirmation of this effect can be found in Weiss (1987) and Peterson *et al.* (1990). Peterson (1989) gives a theoretical analysis of the effects of numbers of jammers and microphones on optimal performance for head-sized arrays.

Although the effect of an additional jammer can be large, the situation is not lost. While there usually is more than one noise source active in an environment, it is rare that the sources have very similar spectra and levels. The prototypical "cocktail party" is probably the most common approximation to the worst case. Further, the solution to multiple jammers is simply additional microphones (so that the number of adaptive filters equals or exceeds the number of jammers). If warranted, this solution is available at the cost of computational complexity that increases linearly with the number of microphones. In this paper discussion is limited to two-microphone arrays and single jammer sources.

3. Reverberation

It is expected that the performance of a Griffiths–Jim beamformer will be degraded by reverberation. A combination of effects associated with reverberation of the target and of the jammer contribute to this degradation. Off-axis target reflections violate the assumption of target equality at the microphones and lead to target leakage into the reference and subsequent target cancellation, provided they fall within the time window of the adaptive filter. Reflections of either target or jammer that arrive outside the filter's time span are uncorrelated with the primary signal and appear as additional jammers, which may be cancelled less effectively. In contrast, jammer reflections that arrive within the adaptive filter's window are susceptible to cancellation, which leads to a tradeoff in selecting the filter length. Longer filters will cancel jammer reflections more effectively at low TJRs, but will be more prone to target cancellation (from target reflection) at high TJRs.

Recent studies have presented examples of the influence of reverberation on the performance of both beamformers and traditional adaptive noise-cancelling systems. An adaptive noise canceller relies on obtaining a "target-free" reference signal from a directional microphone pointed away from the target or from a remote microphone placed close to the noise source. Comparisons of experimental results are difficult because performance obviously depends heavily on the configuration of noise and target sources, the details of microphone placement, and the characteristics of room reverberation. System gains of 3–6 dB have been reported for a "moderate level of reverberation" (Weiss, 1987), and 7 dB for a "moderately reverberant room" (Schwander and Levitt, 1987). Using a Griffiths–Jim beamformer in a simulated "living room" resulted in an effective 8–16 dB gain in intelligibility (Peterson, 1987; Peterson *et al.*, 1990). Section IV C contains results illustrating beamformer performance for various microphone geometries, filter lengths, and degrees of reverberation.

4. Array geometry and placement

It is well known that the performance of an array is limited by the number, spacing, and orientation of the sensors, in conjunction with their internal noise. Peterson (1989) examined the optimal performance of head-sized, free-space arrays for the two extremes of completely diffuse and purely directional jammer fields. Two of the important findings of that work are that endfire arrays are often superior to broadside arrays, and that a small number of microphones (typically $\leqslant 6$) is sufficient to achieve nearly maximum gain. This work provides valuable insight, but it does not afford a complete account of the factors limiting performance under realistic conditions. Thus far, no direct experimental work has been done to determine how array size and placement on the head and body interact with other parameters in imperfect adaptive arrays. The present work addresses these issues with the measurements reported in Sec. IV B for arrays mounted in free space and on a KEMAR manikin.

5. Adaptation rate and misadjustment

An intrinsic property of the LMS adaptive algorithm is the tradeoff between speed of adaptation and *misadjustment*, which is a residual output noise resulting from fluctuations in the adaptive filter weights. Fast adaptation requires large steps in adjusting the filter weights (obtained by selecting a large value for the adaptation constant α), but the large steps also cause large residual error when the weights reach steady state.

Misadjustment is usually described in the literature as a percentage of the minimum mean-square error of the signal to be cancelled (Widrow and Stearns, 1985). If precision of jammer cancellation were the only factor to consider, then a misadjustment of 10%–20% (corresponding to cancellation within 1 dB of the minimum residual jammer power) might easily be acceptable in the hearing-aid application. However, the presence of a target signal in the output contributes to the misadjustment so that the value of α required for satisfactory jammer cancellation in the absence of target leads to unacceptable performance when the target is present. The source

of this increased misadjustment can be understood from Eq. (1), which shows that the weight adjustments are proportional to both the adaptation constant α and the beamformer output $y[n]$, which includes the target signal. Methods to overcome this problem are proposed in Sec. II and evaluated in Sec. IV A.

Misadjustment and the inversely related adaptation rate have received very little study in the application to hearing aids. The present work attempts to reduce misadjustment, but does not provide a quantitative analysis of adaptation rate. The approach has been to address other problems first; if adaptation speed is found to be a problem after the others have been resolved, then alternative algorithms with faster convergence will be investigated.

6. Filter length and primary delay

Increasing filter length is expected to be beneficial when performance is limited by the inability of the adaptive filter to transform the reference jammer to match the primary jammer. Further, performance can be improved by allowing both past and future reference samples to contribute to the current output through use of a delay in the primary path. Zurek *et al.* (1990) computed the performance of optimal filters of various lengths using real-room transfer functions and found that performance is limited by filter length and primary delay only under near-ideal conditions (i.e., minimal reverberation and low TJR). With realistic amounts of target signal (TJR$\geqslant$0 dB), under either anechoic or reverberant conditions there is little dependence of performance on these parameters.

C. Goals of present study

The present study addresses several of the issues just discussed. First, algorithms for controlling adaptation based on TJR will be described and evaluated. These techniques address the problems of misalignment and misadjustment, both of which are manifested at high target-to-jammer ratios. Second, evaluations will be presented of a sample of two-microphone arrays that differ in intermicrophone spacing and placement on the body. Third, an investigation and more thorough characterization of the effects of reverberation on beamformer performance will be described.

The systems studied include both computer simulations and a microprocessor-based real-time implementation. For static acoustic conditions the two approaches are functionally identical. A real-time system, however, provides immediate feedback (to the experimenter) and necessarily involves a more realistic test than does a simulation. The real-time system also provides the ability to listen to the output in dynamically varying conditions, which will be the subject of a later study.

The evaluations presented here are in terms of physical measurements of signals, with no direct intelligibility tests and no involvement of hearing impairments. To the extent that speech reception by hearing-impaired (or normal-hearing) listeners in noise is limited by the background noise, and not by other factors, reduction of background noise will have a direct benefit. This reduction is assessed through signal measurements having a meaningful relation to speech intelligibility. Possible interactions of system performance with the magnitude or type of hearing loss will be the subject of future studies.

II. MODIFICATIONS FOR CONTROLLING ADAPTATION

Two methods, shown in Fig. 2(a) as "correlation-based inhibition" and "power normalization," have been developed for dealing with the problems of misadjustment and misalignment, both of which are manifested at high target-to-jammer ratios. An essential feature of these methods is that they take advantage of the fact that the target signal in this application—speech—exhibits a high degree of fluctuation, and, in fact, has pause periods during which the target is absent. Both of the modifications can be thought of as attempts to sense the TJR in order to allow adaptation only during intervals when TJR is small. The present description of the adaptation-control algorithms will be qualitative and intuitive; a more complete analysis is in preparation.

A. Correlation-based inhibition of adaptation

As explained above, a basic assumption of beamforming is that the target signals arrive from a known direction, or more specifically, that the target signals can be equalized at the microphone outputs. If this assumption is approximately met, then target components in the two microphone signals will be nearly perfectly correlated. On the other hand, microphone signals from off-axis jammers have a cross-correlation

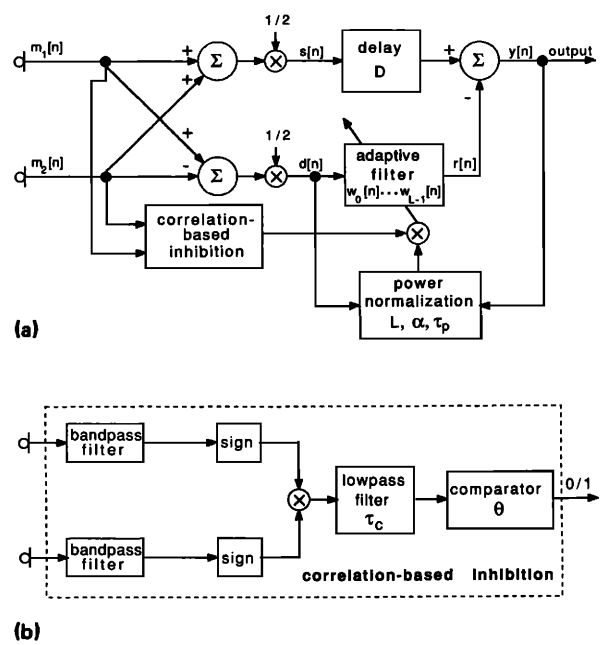

FIG. 2. (a) Beamformer modified by the addition of correlation-based inhibition and power normalization. The calculation of power normalization is performed according to Eq. (4) in the text, and depends on the filter length, L, the adaptation constant, α, and the power normalization time constant, τ_p. (b) Detail of the operations performed in correlation-based inhibition. The microphone signals are filtered by a bandpass filter with cutoff frequencies depending on the particular array geometry. The filtered signals then pass through a one-bit polarity coincidence correlator, which multiplies the sign bits of the two signals. The resulting instantaneous correlation signal is processed by a lowpass filter with time constant τ_c and compared to the inhibition threshold, θ, to produce a binary output.

that, depending on frequency and direction, is less than unity. Thus, the intermicrophone cross-correlation varies with TJR and can be used as an estimate of it.

The operation of this adaptation-control algorithm is shown in Fig. 2(b). In order to maximize the decorrelation induced by the jammer, the microphone signals are first processed by a filter that passes a particular frequency range. That frequency range is chosen to minimize intermicrophone correlation averaged over all jammer directions, and thus depends on intermicrophone spacing (Greenberg, 1989). Next, a one-bit polarity-coincidence correlator and single-pole low-pass filter with time constant τ_c provide a running estimate of intermicrophone cross-correlation. The estimate obtained from the signal polarities is proportional to the arcsine of the true cross-correlation (Papoulis, 1984). This running correlation estimate, a value between $+1$ and -1, is compared to an inhibition threshold, θ, producing an output of "1" or "0" when the correlation is below or above the threshold, respectively. By multiplying the error feedback by this binary signal, adaptation of the weights is inhibited when the TJR, as estimated by the cross-correlation, exceeds the threshold value. This method should have little or no effect on beamformer performance for low input TJR, and should reduce target cancellation and misadjustment for high input TJR. The expected cost for this improved steady-state performance is longer average adaptation time.

B. Adaptation control through power normalization

The second method of controlling the adaptive weights is intended to minimize misadjustment when input TJR is high. Control is achieved by noting that when the input TJR is high, and there is negligible leakage of target signal into the reference, then the output power will be greater than the reference input power. By normalizing the weight update equation with respect to a combination of output power and reference input power, the adaptive steps can be reduced when input TJR is high. In the present implementations, output power is incorporated into the update equation by normalizing the weight increment with the sum of reference and output powers:

$$w_l[n+1] = w_l[n] + 2\alpha y[n] d[n-l]/L(P_d[n] + P_y[n]), \tag{4}$$

where P_y is running-averaged output power, and both P_y and P_d are averaged using the same single-pole low pass with time constant equal to the length of the adaptive filter, $\tau_p = L/(10\,\text{kHz})$. Without the inclusion of output power in the normalization, the adaptive steps grow large along with the output target when the output target-to-jammer ratio is large. Since the target is usually uncorrelated with the jammer, large steps in the weights induced by the target lead to large deviations of the weights away from their optimal values for jammer cancellation and can even increase total output power by increasing misadjustment noise. Normalizing with respect to output power reduces the size of these target-induced weight deviations. Again, the expected cost for this improved steady-state performance is longer average adaptation time.

The two adaptation-control algorithms employed here differ in their operation in that the correlation-based method gives all-or-none control while the power-normalization method uses continuously variable control. Obviously, either method could be designed to operate in either manner. The particular forms of the algorithms used here were based on convenience, simplicity of implementation, and pilot investigations.

The two modifications are similar to algorithms previously suggested by others. Harrison *et al.* (1986), Kaneda and Ohga (1986), and Van Compernolle (1990) discussed inhibiting adaptation during intervals of high TJR in the noise-cancellation application, while Duttweiler (1978) and Sondhi and Berkeley (1980) described its use in adaptive echo cancelling on the telephone network. Power normalization methods similar to the one described here have been proposed by Duttweiler (1982), Shan and Kailath (1988), and Jeyendran and Reddy (1990). A complete analysis of the two modifications and comparison to these other algorithms will be the subject of future work.

III. METHODS

A. Performance metric

Because the output of the beamformer can be described by linear transformations of the input target and jammer signals, present understanding of the effects of linear filtering and additive noise on speech intelligibility, as embodied in the articulation index (Kryter, 1962; ANSI, 1969), is used to form a measure of system performance. Accordingly, the basic signal measurement employed here, as previously (Peterson, 1989; Peterson *et al.*, 1990), incorporates power-spectrum analysis, decibel transformations, and spectral weighting. For any signal s, the intelligibility-weighted level $\Gamma(s)$ is

$$\Gamma(s) = \sum_j a_j B_j(s), \tag{5}$$

where $B_j(s)$ is the decibel level in the jth $\frac{1}{3}$-oct band of signal s, and a_j is the weight assigned to the jth band by articulation theory (ANSI, 1969).

Values of Γ are used only as components in comparisons between signals, as the absolute values depend on the choice of units and have no meaning. Two types of comparisons are defined on the four signals of interest: target input, T_i; target output, T_o; jammer input, J_i; jammer output, J_o. The first comparison shows the *improvement* from input to output in either the target,

$$\Delta\Gamma(T) = \Gamma(T_o) - \Gamma(T_i), \tag{6}$$

or the jammer,

$$\Delta\Gamma(J) = \Gamma(J_i) - \Gamma(J_o). \tag{7}$$

The second comparison is the overall intelligibility-weighted gain, G_I, in the target-to-jammer ratio from input to output:

$$G_I = \Delta\Gamma(T) + \Delta\Gamma(J) = \Gamma(T_o) - \Gamma(J_o) - \Gamma(T_i) + \Gamma(J_i), \tag{8}$$

except that when there is no target present, G_I is defined to equal $\Delta\Gamma(J)$. Throughout this study, the input levels $\Gamma(T_i)$

and $\Gamma(J_i)$ are obtained from the microphone signal in which $\Gamma(T_i) - \Gamma(J_i)$ is higher [or from the one with smaller $\Gamma(J_i)$ when the target is absent]. Note that because of the different spectral shapes of received target and jammer, $\Gamma(T_i) - \Gamma(J_i)$ does not necessarily equal TJR, which depends on broadband power levels of the target and jammer source materials and is a primary experimental variable.

Because of the averaging inherent in estimating power spectra for calculating values of Γ, G_I is only meaningful for measuring performance of a converged system in steady state, and cannot be used to assess intelligibility of transient conditions.

B. Simulations

Computer programs were developed to simulate the modified beamformer systems described in Sec. II. Input signals for the simulation were generated by convolving target and jammer source materials with source-to-microphone impulse responses. The target source material consisted of concatenated sentences (IEEE, 1969) spoken by a male talker, and the jammer source material was 12-talker SPIN babble (Kalikow *et al.*, 1977). Both of these sources were anechoic recordings that were digitized, low-pass filtered with a 4.5-kHz antialiasing filter, sampled at 10 kHz, and approximately whitened with 6-dB/oct high-frequency emphasis. The source-to-microphone impulse responses were generated using a simulation of free-space microphones in a rectangular room with specifiable uniform surface absorption and dimensions (Peterson, 1986).

In order to measure output target and jammer components separately, the simulation incorporated a controlling processor and two yoked processors. The controlling processor operated on target and jammer inputs summed together, while the yoked systems, which have the same structure as the controller, processed the target and jammer separately using adaptive filter weights copied exactly from the controlling processor. Because the filtering operation performed in the controlling processor is linear, superposition holds and the controlling processor output $y[n]$ is equal to the sum of the two yoked processor outputs, $T_o[n]$ and $J_o[n]$.

In arriving at the data points that are shown in later plots, the adaptive filters were initialized with zero weights and the simulation was run for 13.68 s, a period sufficient to reach an approximate steady state for all conditions. The final 1.28 s provided the analysis segment from which target and jammer signals were taken for spectral analysis and calculation of Γ's. In order to avoid extraneous variability from source signal fluctuation, fixed tokens of target and jammer input signals were selected and used for every trial. These tokens were selected so that $\Gamma(T_i) - \Gamma(J_i)$ for the analysis segment was within 0.5 dB of $\Gamma(T_i) - \Gamma(J_i)$ for the entire token.

All calculations were performed in single-precision floating point. The simulations used a value of $\alpha = 0.05$ for the adaptation constant. For a rough interpretation of this value of α, the rule of thumb for a white reference signal is that the time constant of adaptation is five times L, the adaptive filter length, yielding a time constant of adaptation of 50 ms for $L = 100$ at a 10-kHz sampling rate (Widrow and Stearns, 1985).

The bandpass filter used in both channels of the correlation-based inhibition was designed from fifth-order high-pass and low-pass Butterworth filters. The cutoff frequencies of the bandpass filter were selected as a function of the microphone geometry and spacing, as derived in Greenberg (1989). The correlation time constant was determined from a pilot investigation indicating that a value of $\tau_c = 10$ ms allowed the correlation estimate to accurately track the intensity fluctuations of speech. It is desirable to select the lowest value of the inhibition threshold θ that permits adaptation in the absence of target, since performance improves monotonically with decreasing threshold (Greenberg, 1989) until the point where adaptation is inhibited entirely. Issues concerning the selection of θ will be discussed in Sec. III C 5.

C. Real-time measurements

1. Hardware

A real-time implementation of the modified beamformer in Fig. 2 was achieved with a Motorola DSP56001 digital signal processor operating in a DSP56000ADS application development system (ADS) and interfaced to a personal computer. An ADC56000 board (Ariel Corp.), with two channels of 16-bit input and output, was the analog interface to the development system.

Two Motorola ADS boards, each with an Ariel ADC56000, were used to implement the Griffiths–Jim beamformer with correlation-based inhibition. One ADS implemented the Griffiths–Jim algorithm while the other estimated intermicrophone correlation and provided a binary signal to the first ADS, indicating when adaptation should be inhibited. Note that using the second pair of boards to perform the correlation-based inhibition was convenient for this experimental system, but the technique can be (and later was) incorporated onto a single processor along with the beamformer.

In implementing the LMS algorithm in fixed-point arithmetic, the effects of quantization errors due to rounding of the adaptive weights must be considered. One solution to the potential problems created by roundoff errors is to use more bits for the adaptive weights than for the signals (Haykin, 1986). Given the 16-bit input signals, this is trivial to accomplish with the DSP56001, which has 24-bit wide data paths, 24×24-bit multiplication, and 56-bit accumulation.

The real-time system employed two omnidirectional microphones (Knowles Electronics BT-1759), custom preamplifiers, and a sample rate of 10 kHz, for a total system bandwidth of 4.5 kHz. With this sample rate, the adaptive filter length L was limited to a maximum of 169 points.

2. Software

To facilitate measurement of system performance, the microprocessor software was designed to allow the experimenter to select between operating modes via two toggle switches. In addition to a regular "processing" mode, there was also a "no processing" mode in which the system sam-

pled one microphone signal and copied the sample directly to the D/A output. This was used for determining performance baseline. In the "freeze" mode, adaptation was inhibited but the frozen filter continued to process the inputs. This mode was used to measure the component of the system output due either to target or jammer, as described in Sec. III C 8.

3. Source material

The target source was a tape recording of a single male speaker reading continuous discourse and the jammer source was the same SPIN babble used in the simulations. The recordings were approximately whitened with a gain of 6 dB/oct and low-pass filtered with cutoff of 4.5 kHz. After appropriate amplification, the signals were delivered to Radio Shack Minimus-7 loudspeakers.

4. Microphone arrays

The system was evaluated for various two-microphone geometries in free space and mounted on the head and torso of a KEMAR manikin (Burkhard and Sachs, 1975). The arrays, and the signal processing specific to each, follow.

The free-space arrays were the following:

b26: A broadside orientation with 26-cm spacing between microphones. The filters for correlation estimation had a passband of 0.6–1.2 kHz;

b7: Broadside with 7-cm spacing. The correlation filters had a passband of 2–4 kHz;

e7: Endfire orientation with 7-cm spacing. The signal from the front microphone was delayed by two samples (200 μs) to align targets. The correlation filter was high pass with cutoff at 2 kHz.

The four arrays mounted on the KEMAR manikin were the following:

ears: One microphone at each ear, near the microphone locations for behind-the-ear hearing aids. The correlation filters had a passband of 0.6–1.2 kHz.

glasses: On the top front of eyeglass frames with 7-cm spacing. The correlation filters had a passband of 2–4 kHz.

pocket: Microphones spaced by 7 cm, placed on the top of a plastic box worn in the left shirtpocket. For this configuration, the manikin's torso was rotated roughly 30° clockwise in order to align the microphone array to the target. The correlation filter passband was 2–4 kHz.

temple: Microphones mounted on the left temple of eyeglass frames with an 8-cm spacing. This spacing was found to result in intermicrophone delays over all frequencies roughly equal to the 200-μs delay between microphones for the 7-cm endfire array in free space. The correlation filter was high pass with cutoff at 2 kHz.

5. Parameter selection

For all conditions, the adaptive filter length was its maximum value of $L = 169$ with the corresponding power-normalization time constant of $\tau_p = 17$ ms. As in the simulations, the real-time system used values of adaptation constant $\alpha = 0.05$ and correlation time constant $\tau_c = 10$ ms, the bandpass filters for correlation-based inhibition were designed from fifth-order high-pass and low-pass Butterworth filters, and the cutoff frequencies were selected as a function of the microphone geometry and spacing.

In these experiments, the inhibition threshold θ was selected as a function of array geometry and level of reverberation. In anechoic environments, θ was set to slightly above the nominal correlations observed for a jammer at 45°, which resulted in $\theta = -0.4$ for all broadside arrays, and $\theta = +0.5$ and $\theta = +0.6$ for the **e7** and **temple** arrays, respectively. In reverberant environments, nominal correlation values for the jammer alone were observed to be elevated from their anechoic levels. For reverberant conditions, where it was observed that the threshold value determined in the anechoic environment permitted virtually no adaptation, the threshold was increased to a value that permitted adaptation roughly half of the time.

In a practical system, it is reasonable to select θ based on the array geometry, but it is obviously not possible to select θ based on direct knowledge of the degree of reverberation. Methods for optimal or automatic selection of θ will be the subject of future investigations. One possible approach would be to select a baseline value of θ based on theoretical considerations such as the array geometry, and then to dynamically adjust the threshold. This threshold adjustment could be in response to an on-line estimate of reverberation, or it could be performed to ensure, over some time window, a minimum percentage of adapting cycles.

6. Acoustic test conditions

Two acoustic environments were used for testing: an anechoic chamber and a moderately reverberant room. This latter room was $4 \times 5.5 \times 2.7$ m with a tile floor, plasterboard walls, and dropped-ceiling panels. The critical distance of the reverberant room was estimated from recorded impulse responses to be approximately 0.8 m. In the anechoic chamber the speakers were placed 1 m from the microphone array while in the reverberant room both speakers were at distances of either 0.8 or 2.6 m from the array. The speakers were placed in the horizontal plane of either the free-space array or the manikin's head.

System performance was always measured for the target loudspeaker at 0° (straight ahead) with a single jammer located at 45°. (Positive and negative angles refer to azimuths clockwise and counterclockwise from 0°, respectively.) Results for the 45° jammer are representative; additional conditions studied are reported elsewhere (Greenberg, 1989). With each microphone array, two separate conditions were tested: one with the microphone array aligned to the target (by rotating the array to minimize intermicrophone phase differences), and one with the array misaligned by rotating it counterclockwise roughly 10°. Note that with the array misaligned the jammer was located 55° relative to the array.

7. Signal measurement

Output signals from the real-time system were measured with a spectrum analyzer (HP35660A) programmed to calculate Γ according to Eq. (5) based on the average of at least 25 successive spectra.

8. Protocol

For each microphone array, source signals were adjusted in level to provide target and jammer output signals of roughly equal power with the system in the "no processing" mode. Playing the target and jammer individually allowed measurement of the unprocessed target and jammer spectra and calculation of $\Gamma(T_i)$ and $\Gamma(J_i)$. Source levels were then adjusted to provide each of four TJRs: $-\infty$ (jammer alone), -20, 0, and $+20$ dB. For each TJR, overall levels were adjusted to provide analog signals that would occupy the full input range of the analog-to-digital converter (a function that would be approximated using automatic gain control in a practical hearing aid). To determine the response of the system to target or jammer alone, the system was first allowed to reach steady state with both sources playing and then put into "freeze" mode. With the weights frozen, the target and jammer were played separately so that $\Gamma(T_o)$ and $\Gamma(J_o)$ could be computed from the processed target and jammer. Because freezing the weights selects the system response at a single instant in the adaptive process, in order to obtain an accurate representation of the effect of the system it is necessary to freeze the weights at a variety of times and average the resulting values of Γ. Preliminary assessment of variability indicated that taking two to four such measurements was sufficient to reduce the uncertainty of the average to 1–2 dB.

IV. RESULTS

A. Adaptation-control methods

The effect of employing one or the other, or both, of the proposed adaptation-control methods is demonstrated through example simulations. The results in Fig. 3 show the intelligibility-weighted gain, G_I, as a function of TJR for an aligned array and for an array misaligned by 10° counterclockwise rotation.

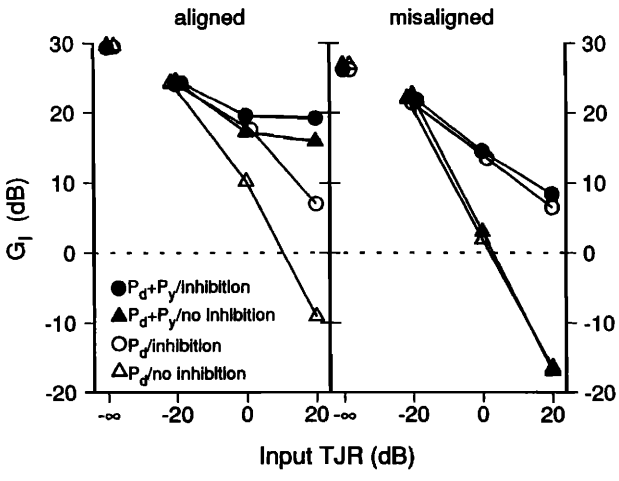

FIG. 3. Results of computer simulations showing the effects of adaptation-control methods with the array perfectly aligned and misaligned by 10°. The plots show intelligibility-weighted gain, G_I, as a function of input TJR. The environment was anechoic, and the target and jammer were separated by 45° in azimuth. The distance to the array center was 0.9 m for both target and jammer sources. The array was a free-space, 26-cm broadside array. System parameters were adaptive filter length $L = 100$ points (10 ms), adaptation constant $\alpha = 0.05$, power-normalization time constant $\tau_p = 10$ ms, correlation time constant $\tau_c = 10$ ms, inhibition threshold $\theta = -0.4$, and correlation filter passband 0.6–1.2 kHz.

Consider first the results with the aligned array. With no target present (TJR $= -\infty$ dB), the system reduces the jammer by about 30 dB, independent of the use of the adaptation-control methods. With the target present, there is a strong dependence of G_I on TJR for the unmodified system (i.e., with no correlation-based inhibition and with power normalized using P_d only). Whereas the system improves the target-to-jammer ratio by 20–25 dB with TJR $= -20$ dB, performance declines rapidly with increasing TJR. With TJR $= +20$ dB, the system degrades the signal by almost 10 dB.

The primary effect of introducing either of the adaptation-control methods is to improve performance at high TJRs, with more improvement for higher TJR. The change is clearly seen, however, with TJRs as low as 0 dB. It is also evident that, with the array aligned, introducing either modification alone leads to considerable improvement, with power normalization providing a larger G_I at TJR $= +20$ dB. Introducing both modifications, however, improves performance beyond that obtained with either alone.

In the simulation of the aligned array there is no target leakage into the reference, and therefore no possibility of target cancellation. The decline in performance with increasing TJR in this case is due entirely to misadjustment induced by the target in the output. The target passes through the system unchanged [i.e., $\Delta\Gamma(T) = 0$ and therefore $G_I = \Delta\Gamma(J)$]. In this case, observation of a negative G_I indicates that target-induced misadjustment of the adaptive-filter weights has caused output noise power greater than input jammer power.

Now consider the results with a misaligned array. With no target present, the jammer was reduced by about 27 dB, again with negligible differences between adaptation-control methods. The 3-dB difference between this value and the corresponding measurement with the aligned array represents the difference in system performance for a jammer at 55° versus a jammer at 45°. Such fluctuations in performance are not unexpected given similar variations of optimum performance with a single jammer (Peterson, 1989).

The degrading effect of TJR on the performance of the unmodified system is slightly greater for the misaligned than for the aligned array. In this case there is only a 2-dB net improvement at TJR = 0 dB, and a degradation of about 16 dB at TJR $= +20$ dB. Because of target leakage in the misaligned case, G_I can reflect changes from input to output in target, or jammer, or both. Analysis of these components is discussed below.

The effects of the adaptation-control methods on performance with the misaligned array show a different pattern from that of the aligned array. With the misaligned array, the effect of power normalization is insignificant compared to that of correlation-based inhibition. The reason for this difference is that normalization by $P_d + P_y$ does not prevent target cancellation, it only makes the size of the steps taken toward cancellation more uniform. On the other hand, correlation-based inhibition is still effective at maintaining G_I

positive at TJR = + 20 dB. In the present example with the misaligned array and TJR = + 20 dB, adding P_y to the normalization increased jammer cancellation by about 2 dB when correlation-based inhibition was in effect.

The performance of the unmodified system with aligned and misaligned arrays can be analyzed in terms of $\Delta\Gamma(T)$ and $\Delta\Gamma(J)$ to illustrate the three effects of high TJR. The first effect, alluded to above, is *misadjustment* induced by the target. Because target and jammer are uncorrelated, the presence of a strong target in the output causes large excursions in the filter weights away from the set that optimally cancels the jammer. For the aligned array, misadjustment is the sole cause of reduced G_I.

Despite the similar trends in G_I for the unmodified system with aligned and misaligned arrays, the degradation in the latter case is due to the two other effects and has (almost) nothing to do with misadjustment. Table I gives the values of $\Delta\Gamma(T)$ and $\Delta\Gamma(J)$ for the misaligned-array condition. Since G_I is obtained by adding $\Delta\Gamma(T)$ and $\Delta\Gamma(J)$, it can be seen that the G_I of about 2 dB at TJR = 0 dB results from the sum of about 8 dB of jammer reduction and 6 dB of target cancellation, and that the G_I of about − 16 dB at TJR = + 20 dB was almost entirely due to target cancellation. Thus the second effect, *target cancellation*, accounts for some of the reduction in G_I with increasing TJR. The remainder of the change in G_I, that is, change in $\Delta\Gamma(J)$, is due to the third effect, *jammer decancellation*, which is the counterpart to target cancellation. As a result of the filter's weights shifting to cancel the target, they shift away from the optimal set of weights for jammer cancellation, so that less jammer is cancelled. Because the algorithm seeks a minimum in total output power, small dB changes in the target, the dominant component, can be accompanied by large dB increases in output jammer.

The attribution of $\Delta\Gamma(J)$ to jammer decancellation, and not to misadjustment, was made by estimating[2] the optimal filter weights and using those fixed weights in the system. The $\Delta\Gamma(T)$ and $\Delta\Gamma(J)$ performance components of this "fixed optimal" system are seen in Table I to match fairly well those of the unmodified system at TJRs up to 0 dB. Because there is no misadjustment in the fixed optimal system, the similarity suggests that misadjustment is not the controlling effect in the performance of the unmodified system for TJR$\leqslant$0 dB. It is interesting to note that, at TJR = + 20 dB, use of the fixed optimal weights leads to much more target cancellation than the simulated system. It can be surmised that in this case misadjustment has the beneficial effect of limiting the amount of target cancellation.

The price to pay for the improvements in asymptotic performance comes in the speed of adaptation. Because both adaptation control methods act to reduce the weight-update step size at high TJRs, the time required to reach steady state will be longer, on average, than if these controls are not used. The effect on adaptation rate can be readily seen for correlation-based inhibition by examining the percentage of time steps on which the adaptive filter was allowed to adapt. For the cases shown in Fig. 3 these percentages are roughly 88% at TJR = − ∞ dB, 81% at − 20 dB, 56% at 0 dB, and 24% at + 20 dB (percentages are similar for aligned and misaligned arrays in these conditions). What is not captured by these percentages is the pattern of inhibition. The process is characterized not by a continuous adaptation with a longer time constant, but rather by an adaptation that proceeds at the same rate as for low TJRs but is interrupted for varying lengths of time. With high TJRs, adaptation is allowed during sporadic intervals corresponding to pauses in the target speech. Depending on the intensity pattern of the target following a transition, adaptation may proceed very quickly or may be delayed.

Although a single condition was used as an example here, similar effects of the adaptation-control methods have been seen with other source conditions, arrays, and environments. For the real-time evaluations presented in the following section, all conditions were tested with and without correlation-based inhibition (but always with normalization by $P_d + P_y$). Because the effect is similar (as a percentage of the difference in G_I at low and high TJRs) for all cases, only the results with inhibition will be presented. The full complement of results can be found elsewhere (Greenberg, 1989).

Finally, the perceptual consequences of these adaptation-control methods were judged through informal listening using the real-time system. The two noticeable effects of the adaptation-control methods were (1) a marked improvement in intelligibility for conditions of high TJR, par-

TABLE I. Entries are intelligibility-weighted gain, G_I, and its components, $\Delta\Gamma(T)$ and $\Delta\Gamma(J)$, in dB, for the misaligned condition. Values of G_I with and without inhibition correspond to the right side of Fig. 3. The fixed optimal condition is described in the text.

		Without inhibition			With inhibition			Fixed optimal		
TJR (dB)		$\Delta\Gamma(T)$	$\Delta\Gamma(J)$	G_I	$\Delta\Gamma(T)$	$\Delta\Gamma(J)$	G_I	$\Delta\Gamma(T)$	$\Delta\Gamma(J)$	G_I
	− ∞	0	26.9	26.9	0	26.2	26.2	0	26.2	26.2
P_d	− 20	− 0.2	22.3	22.1	− 0.1	21.5	21.4	0.3	26.1	26.4
	0	− 6.3	8.2	1.9	− 1.9	15.4	13.5	− 6.5	9.1	2.6
	+ 20	− 16.3	− 0.1	− 16.4	− 3.0	9.5	6.5	− 32.5	− 2.4	− 34.9
	− ∞	0	26.9	26.9	0	26.2	26.2			
$P_d + P_y$	− 20	− 0.1	22.6	22.5	0.0	21.8	21.8			
	0	− 5.8	8.8	3.0	− 1.8	16.3	14.5			
	+ 20	− 16.8	− 0.1	− 16.9	− 3.2	11.6	8.4			

ticularly dramatic when the array is misaligned, since in this case the unmodified algorithm causes audible cancellation of the target while the modified algorithm eliminates that cancellation; and (2) a longer time to reach steady state, as expected, with high TJR. There were no audible effects or distortion relating to the on/off pattern of correlation-based inhibition. Although adaptation of the filter weights switches on and off, when the correlation estimate is above threshold the input signals are processed with the most recent set of adaptive weights, determined during the last cycle with the correlation estimate below threshold. Furthermore, the effect of failure of the inhibition mechanism would be less effective noise cancellation, but not distortion of the target signal.

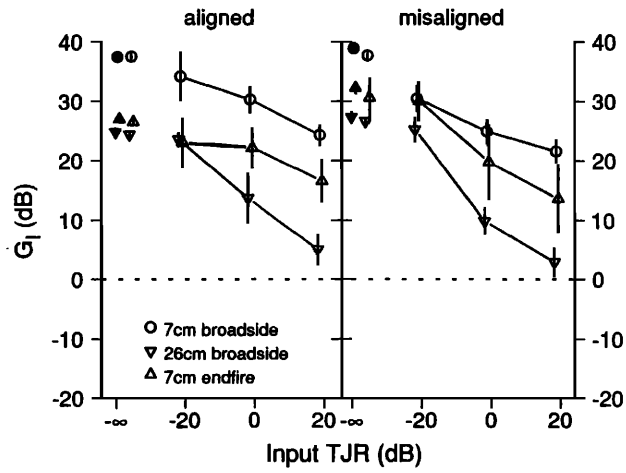

FIG. 4. Measurements with real-time system comparing three free-space arrays in an anechoic chamber. The plots show intelligibility-weighted gain, G_I, as a function of input TJR. The arrays were aligned and misaligned by 10°. The target and jammer were separated by 45° in azimuth.

B. Arrays in an anechoic environment

1. Free-space arrays

Figure 4 shows G_I as a function of input TJR for the real-time system in an anechoic chamber using both adaptation-control methods with target and jammer sources separated by 45°. The left and right panels present measurements made with aligned and misaligned microphone arrays, respectively. The vertical bars on each point indicate plus or minus two standard deviations of G_I, based on the standard error of the mean for the measured Γ's. The uncertainty of the measurements can be attributed to two sources of variability: the different tokens of source material and the different sets of frozen filter weights used for analysis. Neither of these sources of variability existed in the simulations, but if the simulations had not employed fixed tokens, the resulting variability would have been less than that exhibited by the real-time system, since the filter weights were not frozen in the simulations.

The filled symbols at the left of each panel in Fig. 4 represent measurements made while the weights were adapting (which could only be done with no target), and the unfilled symbols represent measurements made while the weights were frozen. Note that for all conditions the filled and unfilled points are within 1.5 dB so, as expected, there is no difference between measurements made with adapting and frozen weights for the no-target case. The differences in no-target performance between the aligned and misaligned arrays is due, as before, to the change in the effective jammer angle from 45° to 55°.

The 7-cm broadside array, **b7**, consistently performs at least 5 dB better than the 26-cm array, **b26**, and for some conditions the difference is as much as 20 dB. The primary reason is that, for any geometry, some frequency components of a jammer at a given angle arrive in phase (and nearly equal in amplitude) at the two microphones, and therefore cannot be cancelled. For the **b7** array, with a jammer at 45° (producing an intermicrophone delay of 143 μs), the lowest in-phase frequency (other than those near dc) is about 7 kHz. Since this frequency is outside the 0.18–4.5 kHz passband of the system, in-phase frequencies did not substantially affect system performance. In contrast, for the **b26** array with a jammer at 45°, frequency components at 1.87 and 3.7 kHz arrive in phase at the microphones and cannot be rejected by the system. From examination of the jammer output spectra it is apparent that the peaks at in-phase frequencies account for the difference in performance between the **b7** and **b26** arrays.

The **e7** array performs better than the **b26** array, but not as well as the **b7** array. This is also explained by considering intermicrophone delay. After the 200-μs delay used to align straight-ahead sources, the intermicrophone delay for a jammer at 45° is 57 μs. Like the **b7** array, the lowest in-phase frequency is outside the system passband, accounting for performance superior to that of the **b26** array. However, with this smaller delay between microphone signals, the system is less effective than the **b7** array at canceling the jammer, particularly at low frequencies. Although the **b7** array performs substantially better than the **e7** array for a jammer located at 45°, endfire arrays perform well against 180° jammers, while free-space broadside arrays cannot cancel a jammer arriving from 180°, due to front-to-back symmetry.

Comparing Figs. 3 and 4 for the **b26** array with both adaptation-control methods reveals some discrepancies between results from the simulations and the real-time system. For most conditions, the difference is less than 5 dB and can be attributed to differences in a variety of factors, including source materials, filter length, data precision, sensor noise, and measurement variability. However, the larger differences seen for the aligned array at TJR$\geqslant$0 dB is due to imperfect alignment of the real-time system. In the simulations, the array is perfectly aligned and no target leakage occurs. With the real-time system, the array cannot be perfectly aligned, resulting in target leakage, which causes target cancellation [evidenced by nonzero values of $\Delta\Gamma(T)$] and jammer decancellation. [For the aligned **b26** array at TJR = 0 dB, $\Delta\Gamma(T) = -2.5$ and $\Delta\Gamma(J) = 16.2$; at TJR = +20 dB, $\Delta\Gamma(T) = -6.5$ and $\Delta\Gamma(J) = 11.5$.] As a result, real-time system performance is worse than the simulation results for aligned arrays at high TJR.

2. Arrays on KEMAR

G_I's for the same source conditions measured with KEMAR-mounted arrays are shown in Fig. 5. The general trends of the results shown in Fig. 5 for the broadside arrays mounted on the manikin (**ears, glasses,** and **pocket**) do not differ greatly from those shown in Fig. 4 for their free-space counterparts. The two configurations corresponding to the **b7** array (**glasses** and **pocket**) consistently perform better than the **ears** configuration, which corresponds to the **b26** array.

Performance of the **temple** configuration does differ from that of the **e7** array. Although the general trend of decreasing performance with increasing TJR is preserved, the values of G_I for the **temple** array are 10–20 dB below the corresponding values for the **e7** array. This difference can be attributed to the presence of the head. Since the **temple** array is mounted on the manikin's left, the sound from a 45° (to the right) jammer source travels around the head before impinging on the array. Examining the intermicrophone transfer function for the **temple** array and a 45° source showed that it is closely approximated by a 200-μs delay. After the two-sample delay implemented to align straight-ahead sources, the jammer signal is nearly in phase and is indistinguishable from a straight-ahead target. As a result, the system provides moderate cancellation of both target and jammer, as illustrated at TJR = 0 dB, where $\Delta\Gamma(T) = -4.4$ and $\Delta\Gamma(J) = 8.8$.

The effect of correlation-based inhibition is also worth mentioning briefly. As stated earlier, the conditions of Figs. 4 and 5 were also tested without correlation-based inhibition (but always with normalization by $P_d + P_y$). The improvement in G_I due to correlation-based inhibition averaged 16.9 dB for the 14 measurements (seven arrays × two alignments) taken at TJR = + 20 dB.

The various microphone configurations on KEMAR were compared over a wide range of jammer-incidence angles through measurements of the no-target (TJR = − ∞ dB) system performance. The two configurations that are mounted symmetrically on the manikin, **glasses** and **ears**, were measured for jammer angles from − 15° to 180°. The two configurations not mounted symmetrically, **temple** and **pocket**, were measured for a complete rotation of jammer angles. The measurements were made by rotating the manikin in increments of 15°.

The resulting values of G_I are shown in Fig. 6. In order to include the effects of head and body shadow in this comparison, the input signal level, $\Gamma(J_i)$, was determined from a free-space microphone placed at the location of the center of KEMAR's head. This change in procedure accounts for the discrepancies between Figs. 5 and 6 for TJR = − ∞ dB and jammer angles of ± 45°.

Figure 6 shows that the **ears** array gives jammer reduction of roughly 30 dB for all angles except 180° and performs much better than the other three configurations at small jammer angles (15° and 30°). This is because for the same jammer angle, the wider spacing provides larger intermicrophone delays, and as a result the system can better cancel jammers at small angles. On the other hand, the relatively large intermicrophone delays for small jammer angles also make this configuration more sensitive to misalignment, as seen in Fig. 5.

As expected, the broadside arrays placed symmetrically on the body provide little cancellation of a jammer at 180°. The **pocket** array, however, provides G_I's of roughly 20 dB for a jammer at 180°. While this performance is 10–15 dB below the performance for jammers at 180 ± 15°, it is much better than the 7–8 dB provided by the other broadside configurations. This effect may be due to the asymmetry of acoustic paths for a jammer at 180°, since the array is mounted in the left pocket, not centered on the torso.

The **temple** array performs consistently well for a wide range of jammer angles, providing 20–30 dB of improvement between − 45° and + 45°. Unlike the broadside configurations, there is no reduction in performance for a jammer at

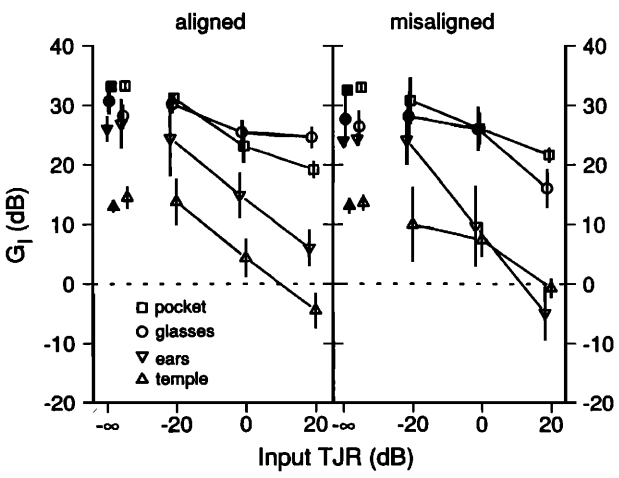

FIG. 5. Measurements with real-time system comparing four KEMAR-mounted arrays in an anechoic chamber. The plots show intelligibility-weighted gain, G_I, as a function of input TJR. The arrays were aligned and misaligned by 10°. The target and jammer were separated by 45° in azimuth.

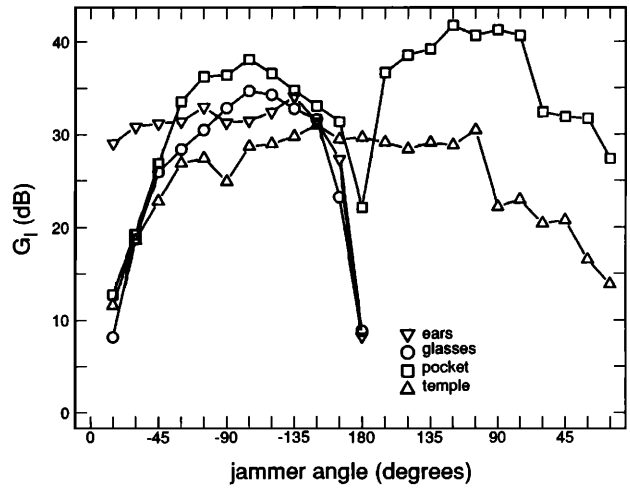

FIG. 6. Measurements with real-time system comparing four KEMAR-mounted arrays in an anechoic chamber with no target signal. The plots show intelligibility-weighted gain, G_I, as a function of jammer angle.

180°. There is an asymmetry in performance at angles beyond ± 105°. For positive angles less than 105° (head between the loudspeaker and the microphone array) scattering caused by the head impairs the system's ability to cancel the jammer, as discussed above. For negative angles (microphone array turned toward the loudspeaker), the presence of the head is not a problem, and performance is comparable to that of the **glasses** and **pocket** arrays.

Overall, the results of Figs. 5 and 6 indicate that the 7-cm broadside arrays, **glasses** and **pocket**, are roughly equivalent in performance and generally superior to the **ears** and **temple** arrays. The **pocket** configuration enjoys an added benefit from body shadowing of rearward jammers. The **ears** array is inferior, primarily because of susceptibility to target cancellation, whereas the **temple** array suffers from reduced jammer cancellation due to scattering caused by the head.

C. Reverberation

As discussed in Sec. I B 3, reverberation must limit the performance of adaptive beamformers. In this section the effect of reverberation is first illustrated with arrays on KEMAR as described above. Second, a simulation study is described in which the interactive effects of reverberation, TJR, and filter length are illustrated.

1. Arrays on KEMAR in moderate reverberation

System performance in the moderately reverberant room was measured for three array configurations on KEMAR (**ears, glasses**, and **pocket**), with target and jammer separated by 45° in azimuth. System parameters were the same as those used in the anechoic environment. The results for the 0.8-m loudspeaker distance are shown in Fig. 7. At this distance, G_I for the three arrays ranged from 7–10 dB with no target. Performance of the **glasses** and **pocket** configurations remained nearly constant with increasing TJR (8–9 dB for aligned arrays and 5–7 dB for misaligned ones), while performance for the **ears** configuration decreased but remained non-negative with increasing TJR.

The results for the 2.6-m speaker distance are shown in Fig. 8. The same system parameters were used except that the inhibition threshold, θ, was changed from -0.4 to -0.2 to allow adaptation when no target was present. For this speaker distance, G_I for no-target conditions was 3–6 dB. As before, performance varied little with increasing TJR.

The noteworthy results in Fig. 8 are (1) performance with the **glasses** configuration was better than with the **pocket** configuration; and (2) performance with the **glasses** configuration when target was present was often better (by up to 4 dB) than when it was absent. Examining the values of Γ (not shown) used to calculate G_I reveals that when G_I for TJR = 0 dB was larger than G_I for TJR = $-\infty$ dB, this improvement was primarily due to unexpected gain in the target from input to output, and not from increased jammer cancellation. This effect and the differences in performance between the **glasses** and **pocket** configurations illustrate the complex interactions that can result from strong reverberation.

2. Simulation study of reverberation

In order to simplify and gain better control over the environment, the effects of reverberation were studied with computer simulations. A sampling of rooms was simulated to study the joint effects of TJR, degree of reverberation, and, to a limited extent, filter length.

Descriptions of the simulated reverberant rooms are given in Table II. The source/array geometries correspond to the **b7** array with a jammer at 45°. The first series of simulations employed condition A, which also includes a target aligned to 0°. The correlation-based inhibition and power normalization modifications were both used.

Figure 9 shows the results with absorption coefficients equal to 1.0 (anechoic), 0.6 and 0.2, and with 100- and 1000-

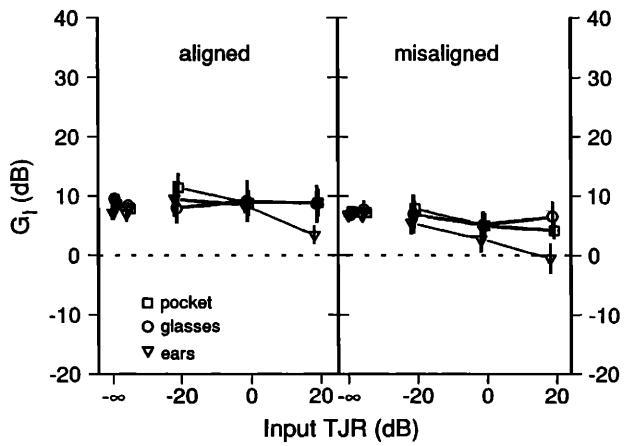

FIG. 7. Measurements with real-time system comparing three KEMAR-mounted arrays in a moderately reverberant room, with source-to-array distance of 0.8 m. The plots show intelligibility-weighted gain, G_I, as a function of input TJR. The arrays were aligned and misaligned by 10°. The target and jammer were separated by 45° in azimuth.

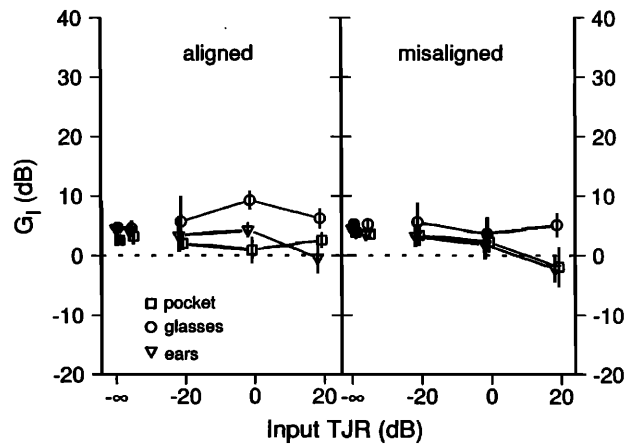

FIG. 8. Measurements with real-time system comparing three KEMAR-mounted arrays in a moderately reverberant room, with a source-to-array distance of 2.6 m. The plots show intelligibility-weighted gain, G_I, as a function of input TJR. The arrays were aligned and misaligned by 10°. The target and jammer were separated by 45° in azimuth.

TABLE II. Coordinates, in meters, describing the dimensions of the simulated rooms and the locations of sources and microphones. In this coordinate system one corner of the rectangular room corresponds to the origin and the room is oriented squarely with the three orthogonal planes. The simulated rooms all have uniform surface absorption. The distance from the sources to the center of the array is 0.9 m, the microphone spacing is 0.07 m, and the target and jammer angles of incidence are 0° and 45°, respectively.

	Condition A	Condition B	Condition C
Dimensions	5.2, 3.4, 2.8	5.2, 3.4, 2.8	6.2, 5.2, 3.1
Target	2.9, 2.3, 1.7	...	...
Jammer	3.5, 1.9, 1.7	3.6, 0.7, 1.7	3.5, 1.9, 1.7
Right microphone	2.79, 1.37, 1.6	3.43, 1.52, 1.6	2.79, 1.37, 1.6
Left microphone	2.72, 1.39, 1.6	3.47, 1.58, 1.6	2.72, 1.39, 1.6

point filters.[3] The latter two values of absorption result in direct-to-reverberant energy ratios at the array of 5.7 and -2.4 dB, respectively. The anechoic results for both filter lengths show a dip at TJR = 0 dB that is unlike previous results in anechoic conditions (Figs. 3–5), which have declined monotonically with TJR. This dip was traced to the use of a less conservative inhibition threshold, resulting in greater target-induced misadjustment at TJR = 0 dB than at TJR = + 20 dB, where the target signal has a greater impact on intermicrophone correlation.

The effect of adding reverberation is striking at all values of TJR, and interacts strongly with filter length. An absorption coefficient of 0.6, which produces perceptual coloration that is barely noticeable, has a dramatic effect on system performance—a drop in G_I from about 43 to 10 dB in the no-target, $L = 100$, condition. The drop is not as large with the longer filter, as expected.

The observation that G_I is always positive for the shorter filter is attributable to the fact that, for this particular case, there was no target leakage into the adaptive filter and hence no target cancellation. There was one reflection in the target impulse response within 50 points of the direct wave, but it so happened that it was also in phase at the microphones. Although this is a special case, it is indicative of the expected trend for shorter filters to be less susceptible to target leakage. With the longer filter, better jammer cancellation is achieved at low TJRs at the expense of substantial target cancellation at high TJRs.

To gain an overview of the effects of reverberation, a variety of room conditions were simulated, all with TJR = $-\infty$ dB. Condition B used the same room as condition A, but with the array and jammer source shifted to another location (maintaining their relative orientation). Condition C used a room with twice the volume of the previous room. In all three conditions the uniform absorption coefficient was varied.

The results are shown in Fig. 10 with G_I plotted as a function of the direct-to-reverberant ratio at the array. A direct-to-reverberant ratio of 0 dB corresponds to the critical distance of the room, where the direct and reverberant energies are equal; direct-to-reverberant ratios greater than 0 dB correspond to source-array distances less than the critical distance.

The first point to note from Fig. 10 is that the differences due to different room conditions are small compared to the effects of the other variables, so, as a first approximation, performance can be summarized as being dependent only on the direct-to-reverberant ratio, and not on the detailed structure of the reverberation, for a given system and array/source geometry. As before, with TJR = $-\infty$ dB, better

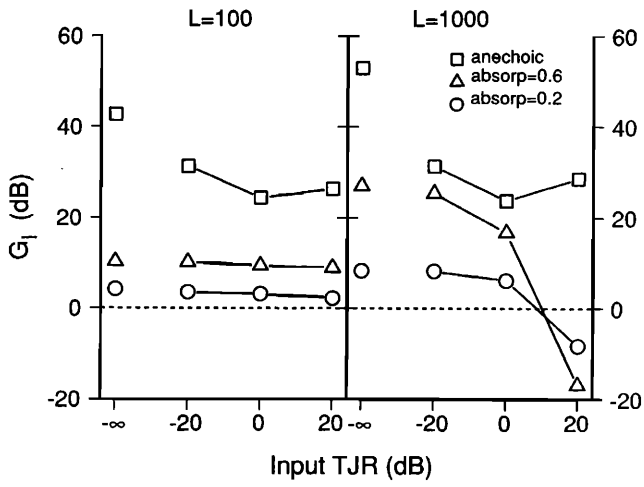

FIG. 9. Results of computer simulations showing the effects of acoustic absorption and adaptive filter length. The plots show intelligibility-weighted gain, G_I, as a function of input TJR. The environment was simulated room A, which contains a free-space, 7-cm broadside array aligned to the target, and a jammer located at 45°. System parameters were adaptation constant $\alpha = 0.05$, power-normalization time constant $\tau_p = 10$ ms for $L = 100$, $\tau_p = 100$ ms for $L = 1000$; correlation time constant $\tau_c = 10$ ms, inhibition threshold $\theta = 0.0$, and correlation filter passband 2–4 kHz.

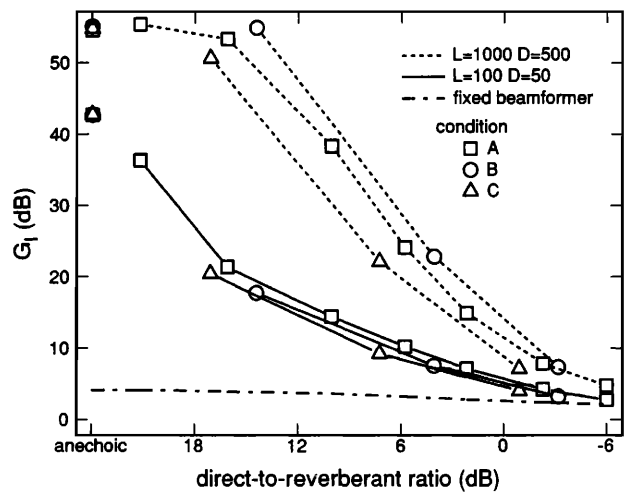

FIG. 10. Results of computer simulations showing the effects of reverberation and adaptive filter length with no target signal. Three simulated room/array geometries were used in conjunction with varying absorption coefficients to generate a variety of reverberant conditions. The plots show intelligibility-weighted gain, G_I, as a function of the direct-to-reverberant ratio. There was no target present, the jammer was located at 45°, and the array was a free-space, 7-cm broadside array. System parameters were adaptation constant $\alpha = 0.05$, power-normalization time constant $\tau_p = 10$ ms for $L = 100$, $\tau_p = 100$ ms for $L = 1000$; correlation time constant $\tau_c = 10$ ms, inhibition threshold $\theta = 0.0$, and correlation filter passband 2–4 kHz.

performance is seen with the longer filter, but for both filter lengths performance degrades as the direct-to-reverberant ratio is reduced. Although much longer filters may provide substantial improvement at low direct-to-reverberant ratios, filter lengths attainable with a practical system only provide significant improvements at source-to-array distances less than the critical distance of the room.

Also shown in Fig. 10 is the performance that is achieved with the adaptive filter disconnected—that is, where the system output is the primary signal, the sum of the two microphones. This corresponds to the performance of a fixed delay-and-sum beamformer. In the cases shown here, as well as in all other cases that have been examined, performance with no target approaches that of the delay-and-sum beamformer as reverberation increases.

V. DISCUSSION

This study has illustrated the primary factors determining the performance of an adaptive beamformer for use in hearing aids. The roles of these factors, initially considered likely to limit performance, are now becoming clearer.

Previous studies found misalignment and misadjustment to be serious problems at moderate and higher target-to-jammer ratios. The methods proposed in this paper for controlling adaptation allow robust performance in the presence of misalignment errors and strong targets, with what seems so far to be minimal costs. Much of the potential benefit of adaptation control has been realized with the present *ad hoc* methods, but further work to develop optimal methods is warranted. Also, adaptation speed and its dependence on adaptation control has not yet been examined thoroughly. Such effects are best judged perceptually, and are to be the subject of a later study.

The results of the comparison among arrays gave indications of the role of microphone spacing and placement. The wider-spaced arrays, **b26** and **ears**, showed greater susceptibility to target leakage than the 7-cm arrays and so, if only two microphones are used, the smaller spacing is preferred. It was also seen that, despite the theoretical superiority of an endfire configuration in free space (Peterson, 1989), placing that array on the temple of eyeglasses leads to inferior performance for certain jammer angles. Broadside arrays benefit from the approximate left–right symmetry of the head and torso.

It would be desirable, of course, to reduce the size of the array as much as possible. From Peterson's (1989) results, one would predict that a two-microphone spacing as small as 2 cm would perform as well as a 7-cm spacing. This possibility certainly deserves further study with a realistic evaluation of the adaptive system.

Taken as a whole, the results of this study suggest that the predominant limitation to the performance of adaptive beamformers in the hearing-aid application comes from the structure of the acoustic environment. In particular, reverberation has a systematic degrading effect on system performance, and although conditions involving multiple jammers were not tested here, they are also certain to show degrading effects. It seems that all the other problems—aside from reverberation—either have been solved or can be neglected because they rarely limit performance.

The findings on reverberant effects can be summarized as indicating that, with practical systems, significant improvements are attained only at source-to-array distances less than the critical distance of the room. Further, these improvements will be obtained only for the best case of no (or very low) target signal. With a moderate or strong target, the system operating in even modest reverberation can severely degrade the target. For example, note that the no-target G_I of 28 dB achieved with the 1000-point filter at a direct-to-reverberant ratio of 6 dB (or absorption = 0.6) goes to -17 dB with an input TJR of 20 dB.

It was known from the start that reverberation would be a limiting factor, but not how limiting compared to the other potential problems, nor how much of an impact a given degree of reverberation would have on performance. Those relations are now at least qualitatively described and clearly show that performance in reverberation must be improved before adaptive beamforming can provide an acceptable solution to the hearing-aid problem.

Fixed beamformers provide one alternative to the adaptive systems considered in this paper. In fixed beamforming, the weights are typically selected to maximize the directivity index, which corresponds to optimum performance in the presence of isotropic noise (Peterson, 1989). This provides some indication of performance in extreme reverberation, since the diffuse sound field in a reverberant room can be accurately modeled by isotropic noise (Cremer and Müller, 1982). Performance of fixed beamformers also provides a minimum standard that adaptive systems must exceed to justify the increase in computational complexity. Soede (1990) evaluated fixed arrays consisting of five cardioid microphones and measured improvements of roughly 7 dB in a diffuse noise field.

It is important to recognize that when reverberation is strong, the adaptive system tends to reduce its filter weights to zero, collapsing to a delay-and-sum beamformer with little expected target cancellation when short filters are used. This outcome is encouraging and can be exploited in future designs to produce a hybrid beamformer combining the advantages of fixed and adaptive beamforming. Such a system would incorporate an adaptive beamformer for optimal performance against directional jammers when reverberation is low, and a fixed design to optimize performance against a diffuse field when reverberation is strong.

ACKNOWLEDGMENTS

This work was supported by Grant No. DC00270 from the National Institute of Deafness and Other Communicative Disorders, by a Fellowship to the first author from AFOSR, and by equipment gifts from Motorola, Inc. The authors wish to thank N. I. Durlach, X. D. Pang, P. M. Peterson, and W. M. Rabinowitz for their comments on a draft of this paper, and P. M. Peterson in particular for many fruitful discussions.

[1] The general form of the Griffiths–Jim beamformer has M microphones, $M-1$ adaptive filters, and a common primary signal from which the sum of adaptive filter outputs is subtracted. In addition, the general system in-

corporates steering transformations of the microphone signals to align targets from a specified "look" direction. See Griffiths and Jim (1982) and Widrow and Stearns (1985) for further details.

[2] Optimal weight estimation consisted of (1) estimating the primary-reference cross-correlation function and the reference input autocorrelation function of the analysis signals (Oppenheim and Schafer, 1975); (2) defining the cross-correlation vector $\mathbf{p}$ and the reference autocorrelation matrix $\mathbf{R}$ based on these functions; and (3) taking the optimal weight vector as $\mathbf{R}^{-1}\mathbf{p}$ (Widrow and Stearns, 1985). The fact that the jammer reduction by the simulated system was sometimes slightly better (at TJR $= -\infty$ dB) than the fixed optimal filter in Table I can be attributed to the ability of the adaptive system to track spectral changes in the nonstationary signals.

[3] The simulation method had to be modified for conditions in Fig. 9 with $L = 1000$ because of the very long adaptation time. For that case, the optimal weights were calculated (see footnote) and used to initialize the adaptive filter.

ANSI (1969). *American National Standard Methods for the Calculation of the Articulation Index*, ANSI S3.5-1969 (American National Standards Institute, New York).

Brey, R. H, Robinette M. S, Chabries, D. M., and Christiansen, R. W. (1987). "Improvement in speech intelligibility in noise employing an adaptive filter with normal and hearing-impaired subjects," J. Rehab. Res. Dev. **24**, 75–86.

Burkhard, M. D., and Sachs, R. M. (1975). "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. **58**, 214–222.

Cremer, L., and Müller, H. A. (1982). *Principles and Applications of Room Acoustics*, translated by T. J. Schultz, (Applied Science, London).

Duttweiler, D. L. (1978). "A twelve-channel digital echo canceler," IEEE Trans. Commun. **COM-26**, 647–653.

Duttweiler, D. L. (1982). "Adaptive filter performance with nonlinearities in the correlation multiplier," IEEE Trans. Acoust. Speech Signal Process. **ASSP-30**, 578–586.

Greenberg, J. E. (1989). "A real-time adaptive-beamforming hearing aid," M.S. thesis, Department of Electrical Engineering and Computer Science, M.I.T.

Griffiths, L. J., and Jim, C. W. (1982). "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. Antennas Propag. **AP-30**, 27–34.

Harrison, W. A., Lim, J. S., and Singer, E. (1986). "A new application of adaptive noise cancellation," IEEE Trans. Acoust. Speech Signal Process. **ASSP-34**, 21–27.

Haykin, S. S. (1986). *Adaptive Filter Theory* (Prentice–Hall, Englewood Cliffs, NJ).

IEEE (1969). "IEEE recommended practice for speech quality measurements," IEEE, New York.

Jeyendran, B., and Reddy, V. U. (1990). "Recursive system identification in the presence of burst disturbance," Signal Proc. **20**, 227–245.

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," J. Acoust. Soc. Am. **61**, 1337–1351.

Kaneda, Y., and Ohga, J. (1986). "Adaptive microphone-array system for noise reduction," IEEE Trans. Acoust. Speech Signal Process. **ASSP-34**, 1391–1400.

Kryter, K. D. (1962). "Methods for calculation and use of the articulation index," J. Acoust. Soc. Am. **34**. 1689–1697.

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

Oppenheim, A. V., and Schafer, R. W. (1975). *Digital Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ).

Papoulis, A. (1984). *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York).

Peterson, P. M. (1986). "Simulation of the impulse response between a single source and multiple, closely-spaced receivers in a reverberant room," J. Acoust. Soc. Am. **80**, 1527–1529(L).

Peterson, P. M. (1987). "Using linearly constrained adaptive beamforming to reduce interference in hearing aids from competing talkers in reverberant rooms," Proc. IEEE Int. Conf. Acoust. Speech Signal Process. **ICASSP-87**, 2364–2367.

Peterson, P. M. (1989). "Adaptive array processing for multiple microphone hearing aids," Ph.D. thesis, Department of Electrical Engineering and Computer Science, M.I.T.; available as Technical Report No. 541 of the Research Laboratory of Electronics, M.I.T.

Peterson, P. M., Wei, S.-M., Rabinowitz, W. M., and Zurek, P. M. (1990). "Robustness of an adaptive beamforming method for hearing aids," Acta Otolaryngol. Suppl. **469**, 85–90.

Plomp, R. (1978). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.

Rankovic, C. M., Freyman, R. L., and Zurek, P. M. (1992). "Potential benefits of adaptive frequency-gain characteristics for speech reception in noise," J. Acoust. Soc. Am. **91**, 354–362.

Schwander, T. J., and Levitt, H. (1987). "Effect of two-microphone noise reduction on speech recognition by normal-hearing listeners," J. Rehab. Res. Dev. **24**, 87–92.

Shan, T. J., and Kailath, T. (1988). "Adaptive algorithms with an automatic gain feature," IEEE Trans. Circuits Syst. **CAS-35**, 122–127.

Smedley, T. C., and Schow, R. L. (1990). "Frustrations with hearing aid use: Candid reports from the elderly," Hear. J. **43**, 21–27.

Soede, W. (1990). "Improvement of speech intelligibility in noise," Ph.D. thesis, Delft University of Technology.

Sondhi, M. M., and Berkley, D. A. (1980). "Silencing echoes on the telephone network," Proc. IEEE **68**, 948–963.

Van Compernolle, D. (1990). "Hearing aids using binaural processing principles," Acta Otolaryngol. Suppl. **469**, 76–84.

Van Dijkhuizen, J., Festen, J. M., and Plomp, R. (1989). "The effect of varying the amplitude-frequency response on the masked speech-reception threshold for sentences for hearing-impaired listeners," J. Acoust. Soc. Am. **86**, 621–628.

Van Veen, B. D., and Buckley, K. M. (1988). "Beamforming: A versatile approach to spatial filtering," IEEE ASSP Mag. **5**(2), 4–24.

Weiss, M. (1987). "Use of an adaptive noise canceler as an input preprocessor for a hearing aid," J. Rehab. Res. Dev. **24**, 93–102.

Widrow, B., and Stearns, S. D. (1985). *Adaptive Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ).

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

Zurek, P. M, Greenberg, J. E., and Peterson, P. M. (1990). "Sensitivity to design parameters in an adaptive-beamforming hearing aid," Proc. IEEE Int. Conf. Acoust. Speech Signal Process. **ICASSP-90**, 1129–1132.