Veterans Administration

Journal of Rehabilitation Research and Development Vol. 24 No. 4
Pages 103–110

Multimicrophone adaptive beamforming for interference reduction in hearing aids*

PATRICK M. PETERSON, NATHANIEL I. DURLACH, WILLIAM M. RABINOWITZ, and PATRICK M. ZUREK
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Abstract—To reduce interference in monaural hearing aids from sound sources that are spatially separated from a target source, we are investigating methods for combining information from multiple microphones. In this paper, we describe an adaptive beamforming method that functions to preserve target signals arriving from straight-ahead of a microphone array while minimizing output power from off-axis interference sources. In a preliminary evaluation of a two-microphone system, sentence intelligibility tests were administered to normal-hearing subjects using processed and unprocessed materials from simulated environments in which the target was on-axis, the interference (speech babble) was 45 degrees off-axis, and the reverberation mimicked that of a living room, a conference room, and anechoic space. Compared to listening through a single microphone, the two-microphone beamformer reduced the target-to-interference ratio required to achieve 50 percent keyword intelligibility by 30, 14, and 0 dB in the anechoic, living-room, and conference-room conditions, respectively. The corresponding improvements over binaural listening (one microphone to each ear) were 24, 9, and 0 dB. Further tests in the living-room environment using the same beamforming system but with filter impulse responses shortened by a factor of four (which would decrease the adaptation time by a factor of four) decreased the improvement by 5 dB. These results are sufficiently encouraging to warrant further tests involving more realistic reverberant conditions, multiple sources of interference, and time-varying acoustic environments.

* This research was supported by a grant from the National Institute of Neurological and Communicative Disorders and Stroke, Grant No. RO1 NS21322.

INTRODUCTION

It is well known that impaired listeners have difficulty hearing in acoustic environments that contain background noise and reverberation. For many of these listeners, the underlying causes of this difficulty are not yet fully understood. However, for listeners who perceive the acoustic world monaurally, at least one major cause is obvious. Because such listeners are sampling the acoustic environment at only one point in space, their ability to localize sound sources, and to selectively attend to one of them, is seriously degraded. The importance of sampling the acoustic environment at more than one point in space is indicated not only by straightforward physical considerations, but also by the wealth of data on the advantages of binaural hearing in humans (2, 9, 16, 18, 20, 21, 24) and by the ubiquity of spatially diversified sensing systems in all sense modalities and all animals.

Most previous attempts to reduce the effects of environmental noise using single-microphone systems have failed to improve speech intelligibility (19), even when reverberation is excluded. Given that interfering sound sources are almost always spatially separate from a target source, processing that uses a spatially-extended sensing system to exploit directional differences can help separate target and interference signals for presentation through a hearing aid.

In this paper, we report briefly the results of

preliminary work on the development of an adaptive beamforming microphone array that potentially could be incorporated into a portable, headworn hearing aid for monaural listeners. The inclusion of a spatially-extended microphone array is motivated by the need to resolve the environment spatially in order to attenuate energy from unwanted directions. An adaptive capability is required to compensate for changes in the acoustic environment, movements of the listener's head, and adjustments of the microphone array relative to the head. Directional microphones currently available in hearing aids are limited both by the short distances over which they sample the environment and by their inability to adapt to that environment.

The techniques used here are drawn from the fields of antenna design and signal processing (1, 10, 13, 22). The main new element is the application of these techniques to the hearing aid problem, i.e., to the case in which 1) the dimensions of the antenna are constrained to be roughly comparable to the wavelengths of the received signals, 2) the received signals originate from a few distinct sources but are frequently degraded by reverberation, and 3) the ultimate goal is improved speech intelligibility.

An antenna system with appropriate directional characteristics would be of value not only to aids that stimulate the listener acoustically, but also to aids that stimulate the subject via cochlear implants or tactile input. Independent of the mode of stimulation, interference reduction is beneficial. Note also that achievement of a directional antenna system would complete only the first step in the creation of a spatially satisfactory aid. The listener needs to be able not only to focus on signals arriving from a desired direction, but to change that direction at will and to monitor all directions essentially simultaneously (e.g., for warning signals). In other words, the achievement of a directional antenna system is necessary, but not sufficient, to create a spatially satisfactory aid. Further discussion of these issues is available in Colburn, Zurek, and Durlach (8). Note finally, that our present work is concerned with single-output systems. The application of such systems to hearing-impaired persons with bilaterally-aidable hearing, and the extension of multiple-microphone systems to provide binaural output, involve issues and problems that extend beyond the scope of the present paper.

PROCESSING

An array beamformer generates a single-channel signal from multiple microphone signals by filtering the individual signals and adding them together, i.e., combining the signals linearly with frequency-dependent weights in amplitude and phase. For a given weighting function and for sources in the far-field of the array (i.e., at distances from the array much greater than the extent of the array), the spatial response of such a beamformer can be described by an antenna pattern showing the dependence of array gain on direction. The ideal antenna pattern for our application would have a sharply focused beam in the target direction with very low sensitivity in all other directions. However, since the dimensions of our array are comparable to the signal wavelengths, it is impossible to achieve such a highly directional pattern. Instead, the reduction of interference must be achieved by adapting the weights to the particular interference (e.g., by forming beam pattern "nulls" in the directions of the interference). If the interference is non-stationary, then the weights must continually adapt in time.

Adaptive array methods can be classified according to the criteria used for selecting weights to combat a given interference structure and the algorithms used for changing the weights in time. Among the more popular weight-selection criteria are noise-cancellation (6,13,27), eigenvector decomposition (3,4,25), and constrained beamforming (11,12,26), which is closely related to the known-direction minumum-variance and maximum-likelihood criteria (22). Noise cancellation uses an interference-only signal to remove correlated interference from a target-plus-interference signal (6). In their pure form, such methods are inappropriate for our application because all of our microphones will contain some target signal. Eigenvector decomposition can be used to estimate the directions of signal arrivals and cancel all but one of the signals as long as the number of microphones exceeds the number of arrival directions. Unfortunately, since even a small amount of reverberation will generate many directional echoes, eigenvector decomposition is not robust in complex environments.

Constrained adaptive beamformers, to which we will restrict our attention in the rest of this paper, operate by minimizing array output power under

the constraint that signals from the target direction be preserved (11,22). 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). As long as these assumptions hold, constrained minimization of total output power will maximize target-to-interference ratio because the target and interference output components are independent and the target component cannot change. 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.

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. The major problem we have encountered with the constrained beamforming approach 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. Clearly, it is important to 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. To date, we have concentrated on simple, slowly-adapting algorithms because more work has been done with them and because they will be the first feasible methods to incorporate in hearing aids. Furthermore, the performance of these algorithms after adaptation in stationary environments indicates the ultimate performance to be expected from any constrained adaptive beamformer.

Frost (11) described one of the first practical constrained adaptive beamformers, 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 (least mean squares) principle (27), 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.

An even simpler constrained beamformer, which can be made equivalent to Frost's system, has been proposed by Griffiths and Jim (12). 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 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.

We have implemented the two-microphone Griffiths-Jim beamformer outlined schematically in **Figure 1**. (The dashed lines in **Figure 1a** 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. The beamforming problem is thus transformed into a noise-cancellation problem and the sum and dif-

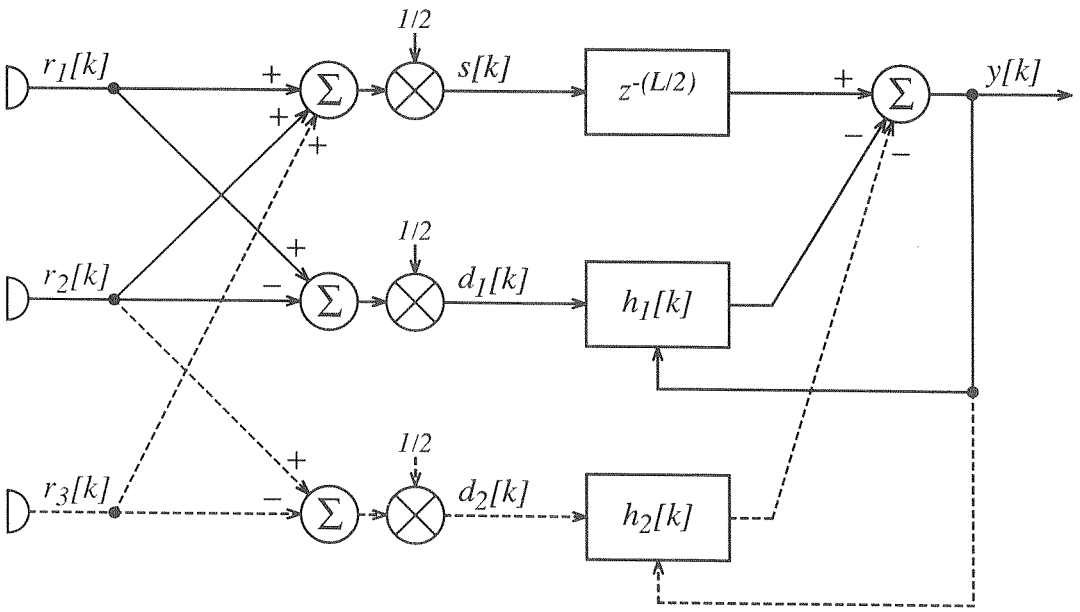

Figure 1a.
The two-microphone Griffiths-Jim beamformer. The microphone signals at sample index k, denoted $r_1[k]$ and $r_2[k]$, are added and subtracted and then scaled to form the sum signal $s[k]$ and the difference signal $d_1[k]$. The sum signal (which contains the target signal plus interference) 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 (interference-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.

ference 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.

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 longer filters, 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 provide a guide in choosing filter lengths, 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 points performed better than 100 points to an extent dependent on the amount of reverberation. In the present study, we employed both 100- and 400-point filters.

In setting μ, we reasoned that the nonstationarity of speech and of the environment would limit eventual performance in any case, so there should be no penalty in choosing a large μ for fast adaptation.

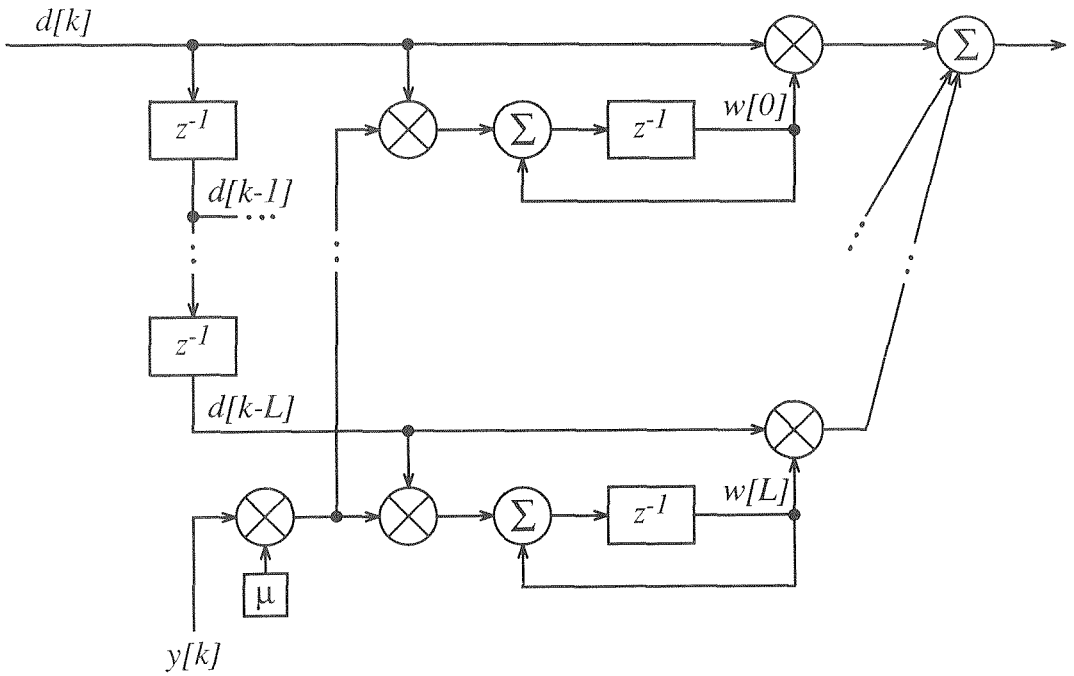

Figure 1b.
The adaptive FIR filter structure. The adaptive FIR filter operates on the $L + 1$ most recent difference-signal samples, $d[k], d[k-1], \ldots, d[k-L]$, which are held in the chain of L unit delays labelled $z^{-(L/2)}$. Each sample, $d[k-l]$, is multiplied by a weight, $w[l]$, and all the 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 $d[k-l]$.

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 had to initialize 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.

INITIAL EVALUATION

We have conducted a preliminary evaluation of the two-microphone adaptive beamformer with intelligibility tests administered to normal-hearing subjects. We compared target intelligibility for three cases: monaural unprocessed, binaural unprocessed, and monaural processed. In the binaural-unprocessed case, the signal from one microphone is fed to one ear and the signal from the other microphone to the other ear. In the monaural-unprocessed case, only one microphone signal is presented. In the monaural-processed case, the signals from the two microphones are combined as described in the previous section before presentation to the listener. In order to eliminate the effects of interaural asymmetries on the comparisons between

monaural and binaural cases, the presentations in the two monaural cases were actually diotic rather than monaural, i.e., the single-channel signal was presented identically to both ears. For listeners whose hearing is perfectly symmetric, diotic and monaural presentations lead to essentially identical results.

The source materials used in the tests were digitized, single-channel, anechoic recordings of IEEE Harvard sentences (14) for the target and SPIN babble (15) 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.

The two simulated microphone signals were generated by passing the sentence and babble source materials through simulated room transfer functions (23). 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. 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 degrees off the normal to the array axis, also at a slightly different height. **Table 1** summarizes the parameters of the simulated environments.

The results of the intelligibility tests are shown in **Figure 2**. In the anechoic 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 (5, 24). 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

Table 1.
Simulated reverberant environments.

	ANECHOIC SPACE	LIVING ROOM	CONFERENCE ROOM
Size (meters)	---	4.6 × 3.1 × 2.4	6.1 × 5.2 × 2.7
Microphone Coordinate[a] (meters)	(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 Coordinates	(0.17, 0.86, 0.17)	(2.93, 2.24, 1.73)	(4.31,4.14, 1.55)
Interference Coordinates	(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[b]	---	0.6	0.3
Reverberation Time[c]	---	120 ms	480 ms
Critical Distance[d]	∞	1.8 m	1.2 m
Direct-to-Reverberant Energy Ratio[e]	∞	5.9 dB	-6.3dB

Notes:

[a] 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.

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

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

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

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

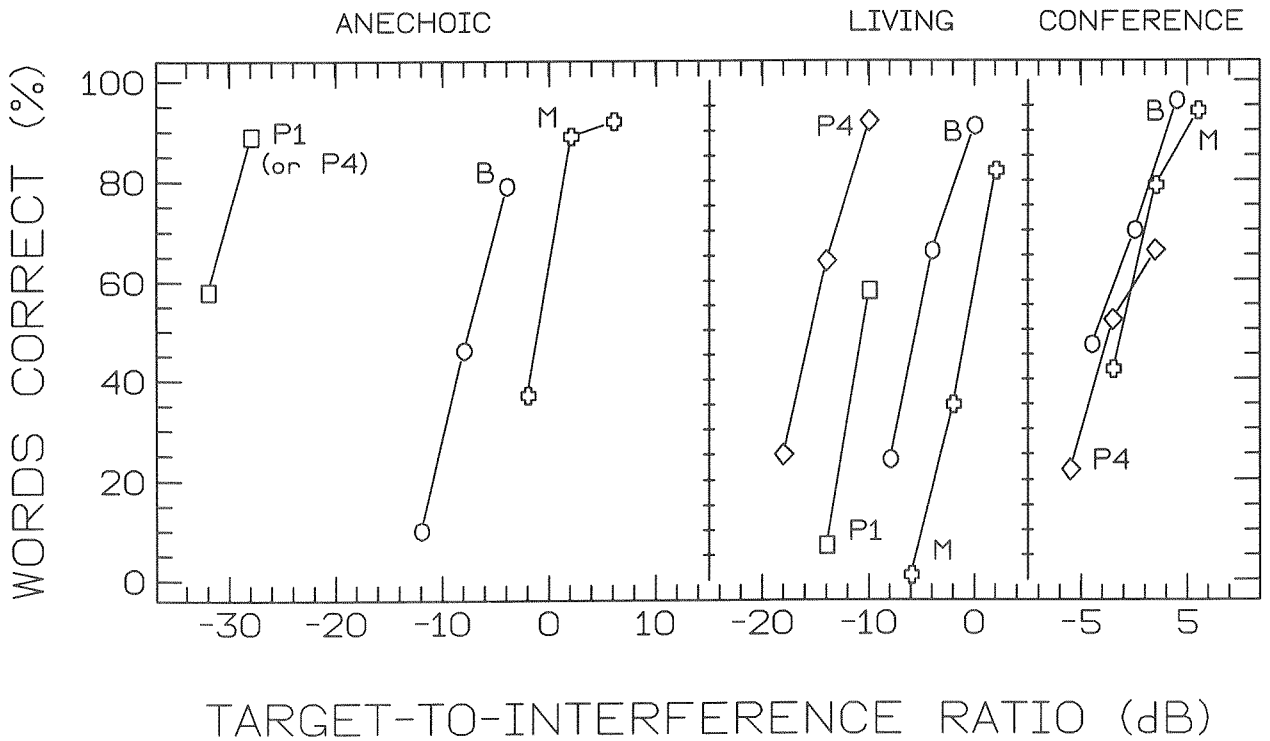

Figure 2.
Percentage keywords correct as a function of Target-to-Interference power ratio in three different environments. Each curve represents data for one of four processing conditions: monaural-unprocessed (M), binaural-unprocessed (B), 100-point adaptive processing (P1), and 400-point adaptive processing (P4). Each data point is the average score of 5 normal-hearing subjects listening to 10 sentences with 5 keywords per sentence.

nonstationarity 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 (21, 24). 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.

COMMENTS

The above results demonstrate the potential of adaptive array beamforming for hearing aids. Under the given conditions, and with zero-to-moderate reverberation, the interference reduction achieved by the array exceeds that achieved by the binaural auditory system. Furthermore, when the reverberation is severe, the array performs no worse than the binaural system.

To determine the generalizability of the above 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 geometries (inteference source at different 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 independ-

ent 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.

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. 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 2** and recall that shorter filters adapt faster.) If a Griffiths-Jim beamformer cannot achieve adequate performance and sufficiently fast adaptation simultaneously, then alternative, fast-adapting methods (7,17) should be explored, although these methods may be more difficult than the LMS method to realize in a practical hearing aid.

REFERENCES

1. **Baggeroer AB**: Space/Time Random Processes and Optimum Array Processing. *NUC TP 506*. San Diego, CA: Naval Undersea Center, 1976.
2. **Blauert J:** *Spatial Hearing*, Cambridge, MA: MIT Press, 1983.
3. **Bordelon DJ**: Complementarity of the Reddi method of source direction estimation with those of Pisarenko and Cantoni and Godara. *J Acoust Soc Am* 69:1355–1359, 1981.
4. **Cantoni A and Godara LC:** Resolving the directions of sources in a correlated field incident on an array. *J Acoust Soc Am* 67:1247–1255, 1980.
5. **Carhart R, Tillman TW, and Greetis ES:** Release from multiple maskers: Effects of interaural time disparities. *J Acoust Soc Am* 45:411–418, 1969.
6. **Christiansen RW, Chabries DM, and Lynn D:** Noise reduction in speech using adaptive filtering I: Signal processing algorithms. *J Acoust Soc Am Suppl 1* 71:S7(A), 1982.
7. **Cioffi JM and Kailath T:** Fast, recursive least squares, transversal filters for adaptive filtering. *IEEE Trans. Acoust. Speech Sig. Proc.* ASSP-32:304–337, 1984.
8. **Colburn HS, Zurek PM, and Durlach NI:** Binaural Directional Hearing—Impairments and Aids. In *Directional Hearing*, Yost and Gourevitch, eds. New York: Springer Verlag, in press.
9. **Durlach NI and Colburn HS:** Binaural Phenomena. In *Handbook of Perception*, Vol. 4, Carterette and Friedman, eds. New York: Academic Press, 1978.
10. **Elliott RS:** *Antenna Theory and Design*. Engelwood Cliffs, NJ: Prentice-Hall, 1981.
11. **Frost OL:** An algorithm for linearly constrained adaptive array processing. *Proc. IEEE* 60:926–935, 1972.
12. **Griffiths LJ and Jim CW:** An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Ant. Prop.* AP-30:27–34, 1982.
13. **Honig ML and Messerschmitt DG:** *Adaptive Filters: Structures, Algorithms, and Applications*. Boston: Kluwer, 1984.
14. **IEEE:** *IEEE Recommended Practice for Speech Quality Measurements*, IEEE No. 297. New York: IEEE, 1969.
15. **Kalikow DN, Stevens KN, and Elliott LL:** Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am* 61:1337–1351, 1977.
16. **Koenig W:** Subjective effects in binaural hearing. *J Acoust Soc Am* 22:61–62(L), 1950.
17. **Lee DTL, Morf M, and Friedlander B:** Recursive square root ladder estimation algorithms. *IEEE Trans. Acoust. Speech Sig. Proc.* ASSP-29:627–641, 1981.
18. **Levitt H and Rabiner LR:** Binaural release from masking for speech and gain in intelligibility. *J Acoust Soc Am* 42:601–608, 1967.
19. **Lim JS:** *Speech Enhancement*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
20. **MacKeith NW and Coles RRA:** Binaural advantages in hearing of speech. *J Laryngol Otol* 85:213–232, 1971.
21. **Moncur JP and Dirks D:** Binaural and monaural speech intelligibility in reverberation. *J Speech Hear Res* 10:186–195, 1967.
22. **Monzingo RA and Miller TW:** *Introduction to Adaptive Arrays*. New York: Wiley, 1980.
23. **Peterson PM:** Simulating the reponse of multiple microphones to a single acoustic source in a reverberant room. *J Acoust Soc Am* 80:1527–1529, 1986.
24. **Plomp R:** Discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise). *Acustica* 34:200–211, 1976.
25. **Reddi, SS:** An adaptive nulling technique for signal reception. *J Acoust Soc Am* 72:159–161, 1982.
26. **Strube HW:** Separation of several speakers recorded by two microphones (cocktail party processing). *Signal Processing* 3:355–364, 1981.
27. **Widrow B et al.:** Adaptive noise cancelling: Principles and applications. *Proc. IEEE* 63:1692–1716, 1975.