A comparison of hearing-aid array-processing techniques

James M. Kates[a)] and Mark R. Weiss
Center for Research in Speech and Hearing Sciences, City University of New York, Graduate Center, Room 901, 33 West 42nd Street, New York, New York 10036

(Received 8 May 1995; revised 5 September 1995; accepted 9 January 1995)

Microphone arrays have proven effective in improving speech intelligibility in noise for hearing-impaired listeners, and several array processing techniques have been proposed for hearing aids. Among the signal-processing approaches are classical delay-and-sum beamforming, superdirective arrays, and adaptive arrays. To directly compare the effectiveness of these different processing strategies, a 10-cm-long linear array was built using five uniformly spaced omnidirectional microphones. This array was used in the end-fire orientation to acquire speech and noise signals for a variety of array placements in two representative rooms. Both digital and simulated analog processing techniques were considered, with the array processing implemented in the frequency domain. The performance metric was the steady-state array gain weighted to represent the relative importance of the different frequency regions in understanding speech. The processing comparison indicates that digital systems are more effective than the simulated analog processing, and that both superdirective and adaptive digital array processing can provide more than 9 dB of weighted array gain.

PACS numbers: 43.66.Ts, 43.60.Gk

INTRODUCTION

In this paper, several fixed-coefficient and adaptive processing algorithms are compared for a short microphone array suitable for hearing-aid applications. The processing effectiveness is evaluated using acoustic data acquired in two representative rooms, with the processing performed off-line. End-fire array placements on the side of the head or near reflecting surfaces were used to give conditions similar to those that could be experienced in everyday use. The data from a real room avoids the limitations of computer simulations that are typically used in evaluating array-processing algorithms, and permits an accurate comparison of the different processing strategies that have been proposed for hearing aids.

A short microphone array is attractive for hearing-aid applications since it is one of the few approaches, among the many that have been proposed, that has actually improved speech intelligibility in noise for the hearing impaired. The improvement in signal-to-noise ratio (SNR) for a 10-cm long array using five uniformly spaced cardioid microphones with delay-and-sum beamforming is 5–12 dB (Soede *et al.*, 1993a, 1993b), with the greatest improvement occurring at the highest frequencies. Such an array can be hand-held or can be built into an eyeglass frame, and the performance of the array does not appear to be affected by the head to any great extent. The directional arrays used by Soede *et al.* improved the speech reception threshold (SRT) by 7 dB in a diffuse noise field, so the improvement in SNR is directly related to a comparable improvement in speech intelligibility in noise.

The performance offered by delay-and-sum beamforming can be bettered by using superdirective array processing (Cheng, 1971; Cox *et al.*, 1986), in which the array performance is optimized for noise coming uniformly from all directions. A sensitivity constraint (Newman *et al.*, 1978; Cox *et al.*, 1986) can be used in designing the superdirective array weights to reduce the effects of microphone position errors, wavefront perturbations, and the sensor internal noise. The constraint, however, causes a small reduction in the array gain. Simulation studies (Kates, 1993; Stadler and Rabinowitz, 1993) have shown that a constrained superdirective array can offer substantially more array gain than classical delay-and-sum beamforming, but the performance in a real room has not been ascertained. A further processing option is an oversteered array, similar to delay-and-sum beamforming except that the time delays used in combining the microphone output signals are greater than the acoustic propagation times between the microphones. An oversteered array can offer performance very close to that of the optimal superdirective array (Cox *et al.*, 1986), and can be realized with a relatively simple analog system.

Adaptive algorithms have also been proposed for hearing-aid arrays (Peterson *et al.*, 1987; Greenberg and Zurek, 1992; Link and Buckley, 1993; McKinney and DeBrunner, 1993; Hoffman *et al.*, 1994). Adaptive array processing offers the possibility of improved performance over arrays using fixed coefficients, but a perturbed wavefront, as can be caused by sensor misalignment or by a specular reflection, can result in signal cancellation (Cox, 1973). The scaled projection algorithm (Cox *et al.*, 1987) can be used to prevent signal cancellation, and its application to adaptive hearing-aid arrays (Link and Buckley, 1993; Hoffman *et al.*, 1994) has resulted in improved performance. However, the improvement in speech SNR due to the array processing can be substantially reduced at low ratios of direct to reverberent sound even when the scaled projection constraint is used

[a)]Corresponding author. Tel: 212-642-2179; Fax: 212-642-2379; E-mail:

The desire for immunity to correlated interference such as specular reflections has lead to modifications of the basic adaptive array-processing algorithms. One technique is to force the correlation matrix used in the array processing to have a Toeplitz structure (Godara and Gray, 1989; Godara, 1991) in which the entries of the correlation matrix are replaced by the values averaged along the diagonals. Simulation studies have shown that the resulting structured correlation matrix offers improved performance in the presence of correlated interference for an array several wavelengths long (Godara, 1991). A further modification is to form a composite correlation matrix, using the structured correlation matrix in the scaled projection algorithm at a low estimated input SNR value and gradually changing to the correlation matrix corresponding to an ideal isotropic noise field at high input SNR values. This approach is designed to give the benefits of an adaptive system at low input SNR values, but to smoothly shift to a superdirective array at high input SNR values where adaptive systems have exhibited reduced performance.

In this paper, five frequency-domain processing algorithms are compared for the same set of microphone data. The algorithms are classical delay-and-sum beamforming, an oversteered superdirective array, an optimal superdirective array, an adaptive system using the scaled projection algorithm, and an adaptive system using the scaled projection algorithm combined with the composite structured correlation matrix. In order to directly compare the effectiveness of these different processing strategies in a real room, a 10-cm-long linear array was built using five uniformly spaced omnidirectional microphones. This array was used in the end-fire orientation to acquire speech and noise signals for three array placements in two representative rooms. The methods used for the data acquisition, signal processing, and performance evaluation are described in the remainder of the paper, along with the performance results.

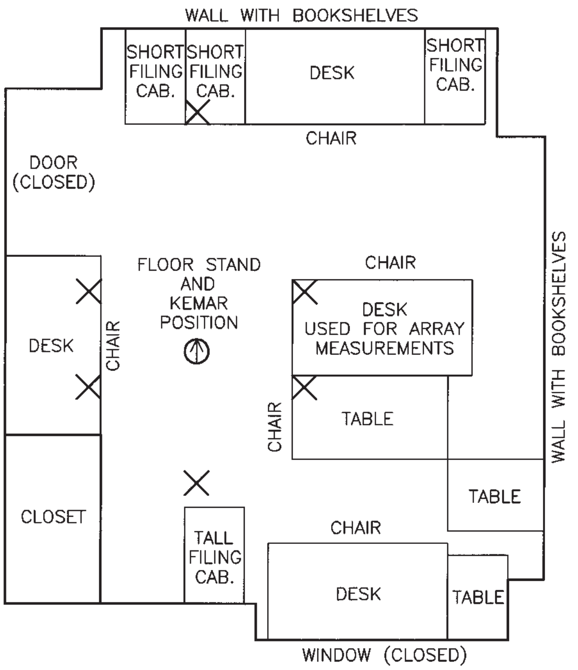

FIG. 1. Floor plan of the office used for the array measurements. The array position and orientation for the floor-standing and KEMAR measurements are indicated by the arrow within the circle, and the loudspeaker positions for the speech and noise are indicated by the crosses. Angles are measured counterclockwise from the array orientation, with the speech loudspeaker position at 0 deg. The desk used for the desk-top measurements is also identified. For the desk measurements, the array was positioned at the ''U'' in ''USED'' with the speech loudspeaker positioned at the ''Y'' in ''ARRAY.''

I. METHOD

A. Data acquisition

The array used for experiments was 10 cm long and consisted of five uniformly spaced Knowles EK-3033 omnidirectional microphones. The array was used in the end-fire orientation. The outputs of the microphones were found to be matched to within ± 1 dB, and no amplitude or phase equalization was provided. The microphone outputs were sampled at 10 kHz using an A/D converter with simultaneous sample-and-hold circuits having a ± 25 ns aperture uncertainty.

Stimuli were presented one at a time over a loudspeaker with the microphone responses sampled and stored on the computer for later processing. Speech stimuli were presented at an azimuth of 0 deg, and the noise stimuli were presented at azimuths of 60, 105, 180, 255, and 300 deg counterclockwise around the array. The speech stimulus consisted of the sentence ''The candy shop was empty.'' spoken by a male talker. The uncorrelated noise stimuli at the other azimuths consisted of multitalker speech babble. A combined noise source was formed by summing the babble signals from the five noise azimuths at equal intensities; this combination pro-

a restaurant or similar environment where several people are talking simultaneously. The test stimuli were bandlimited to 5 kHz.

Two rooms, an office and a conference room, were used for the measurements. A floor plan of the office showing the location of the furniture, the microphone array, and the loudspeaker positions, is presented in Fig. 1. The office walls are painted plasterboard, the floor is carpeting over a concrete slab, and the ceiling is acoustical tile beneath a plenum. Two of the walls are covered with bookshelves, and the office contains several desks, tables, and chairs, thus providing a complex acoustical environment. A floor plan of the conference room showing the location of the furniture, the microphone array, and the loudspeaker positions, is presented in Fig. 2. The construction of the conference room is the same as for the office with the exception that the floor is covered with cork tiles instead of carpeting.

Three array positions were used for the data acquisition in each room. A quasi-''free-field'' position was obtained by placing the array at a height of 1.4 m on a floor stand near the middle of the room and as far as possible from any reflecting surface. A desktop position was obtained by placing the array on a microphone stand at a height of 15 cm above the surface of a desk (office) or group of tables (conference room), with the array at one end and the speech loudspeaker at the opposite end of the desk or tables. Measurements were also made using the KEMAR anthropometric manikin

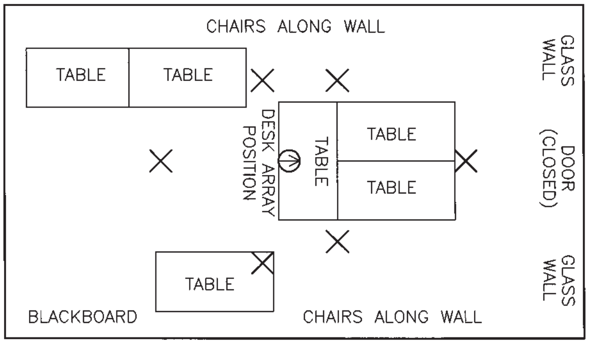

FIG. 2. Floor plan of the conference room used for the array measurements. The array position and orientation for the desk-top measurements is indicated by the arrow within the circle, and the loudspeaker positions for the speech and noise are indicated by the crosses. For the floor-standing and KEMAR measurements, the tables were moved to the periphery of the room and approximately the same loudspeaker and array positions within the room were used.

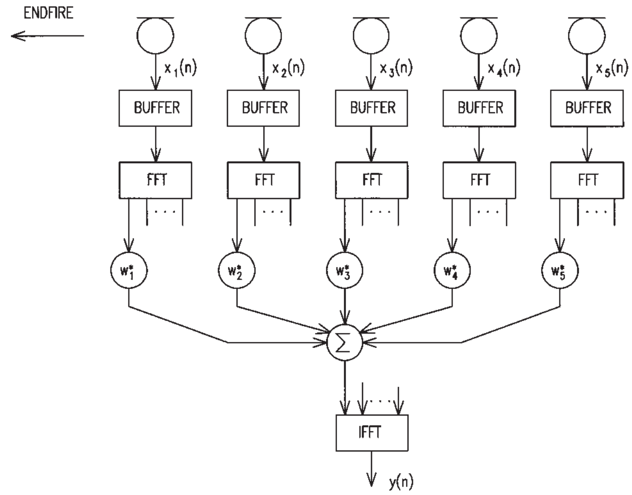

FIG. 3. Block diagram for the frequency-domain array processing.

room with the array positioned just above the left ear at a height of 1.2 m above the floor. The power for each speech or noise test signal at each microphone array position was normalized by forming the rms average across the five microphones in the array and setting this average to 1 V.

The physical and acoustic properties of the rooms are summarized in Table I. The reverberation time was estimated by observing the decay of a speech-shaped noise signal that was allowed to reach steady state in the room and was then switched off. The test signal was output by the speech loudspeaker in the room and the response was measured at the floor microphone array position. Since the ambient noise levels in the rooms did not permit an accurate measurement of the entire 60-dB decay of the test signal, the time to reach a level 20 dB below the steady-state level was measured and then tripled to give the indicated 60-dB reverberation time. The calculated quantities were computed from the physical measurements and reverberation time using the room acoustics formulae given by Beranek (1954).

B. Array processing

All of the array processing was implemented using a block frequency-domain approach as shown in Fig. 3. A frequency-domain implementation of the adaptive processing

TABLE I. Acoustic properties of the rooms used for the processing evaluation.

Property	Office	Conference room
Measured:		
Length, m	5.1	10.7
Width, m	4.5	6.2
Height, m	2.8	2.8
Volume, m^3	60	185
Reverb. time T_{60}, ms	250	600
Calculated:		
Ave. absorption coef.	0.332	0.197
Mean free path, m	2.50	3.26
Critical distance, m	0.97	1.05
Speech direct/reverb., dB	−6.3	−10.5
Noise direct/reverb., dB	−0.2	−5.0

generally offers faster convergence than a time-domain version due to the reduced eigenvalue spread in the correlation matrices (Narayan *et al.*, 1983). A block frequency-domain implementation was chosen to reduce the computational burden (Mansour and Gray, 1982), and a time-domain constraint was added to the weight computation to ensure a causal adaptive filter (Clark *et al.*, 1983). To implement the equivalent of an L-tap time-domain filter, a $2L$-sample block of data is acquired from each microphone. A fast Fourier transform (FFT) of size $2L$ is performed on each $2L$-sample data buffer, after which the weights are computed independently for each positive FFT frequency bin. The frequency-domain signal is multiplied by the weights, summed across microphones at each frequency, and a $2L$-point inverse FFT returns the weighted signal to the time domain. An overlap-save implementation (Clarke *et al.*, 1983) was used, with the buffer contents and weights updated every L input samples. Relatively short adaptive filters, varying in length from $L=8$ to $L=32$ samples, were used in the experiments since work on adaptive microphone arrays (Sondhi and Elko, 1986) has indicated that a short filter offers better immunity to deleterious reflection effects than does a long filter.

The weight vectors for the different processing approaches can all be expressed using the same basic equation (Cox, 1973; Monzingo and Miller, 1980). The set of microphone weights in each FFT frequency bin is chosen to optimize the array output SNR subject to a constraint that a signal from the end-fire direction be passed with unit gain. The processing strategies differ primarily in the description of the noise field. The equation for the steady-state weights for all of the processing approaches is given by

$$\mathbf{w}(k)=\frac{\mathbf{R}^{-1}(k)\mathbf{d}(k)}{\mathbf{d}^{*}(k)\mathbf{R}^{-1}(k)\mathbf{d}(k)}, \tag{1}$$

where $\mathbf{d}(k)$ is the steering vector (vector giving the phase shift from one microphone to the next as a wave arriving from 0 deg propagates across the array) for FFT frequency index k, $\mathbf{R}(k)$ is the noise correlation matrix, and the asterisk

Classical delay-and-sum beamforming is based on the assumption that the dominant source of noise is the self-noise of the microphones and not the ambient noise field. This assumption leads to the system noise correlation matrix being the identity matrix, that is, $\mathbf{R}(k)=\mathbf{I}(k)$, since the assumed Gaussian noise has equal intensity and is completely independent at each microphone (Cox, 1973). The solution of Eq. (1) for this form of assumed interference reduces to $\mathbf{w}(k)=\mathbf{d}(k)/M$, where M is the number of microphones in the array. The weight vector is independent in each FFT frequency bin. The oversteered weight vector is similar to that for delay-and-sum beamforming, but uses a modified steering vector having time delays multiplied by a scale factor greater than one. The oversteered delay factor was implemented by approximating the phase response of a cascade of analog one-zero/one-pole all-pass networks, with the group delay chosen to double the normal propagation time between the microphones at low frequencies and to reduce to the normal propagation time at 5 kHz. This degree of oversteering gave a minimum white noise gain of 0 dB.

The weights for the superdirective and adaptive algorithms are also similar in form, but use correlation matrices that optimize the array performance for the ambient noise field rather than for the sensor self-noise. The superdirective processing is based on the correlation matrix $\mathbf{R}(k)$ calculated *a priori* for an assumed ideal spherically isotropic noise field (Cheng, 1971), while the adaptive system uses for $\mathbf{R}(k)$ the signal-plus-noise correlation matrix estimated directly from the incoming microphone signals (Cox *et al.*, 1987). The superdirective processing therefore determines the array weights based on assumed noise-field characteristics, while the adaptive processing determines the array weights in response to the actual noise field found in the room.

The adaptive processing was implemented using the scaled projection algorithm of Cox *et al.* (1987). This algorithm imposes a constraint on the magnitude of the weights so that

$$\mathbf{w}^*(k)\mathbf{w}(k) \leq 1/\delta^2(k), \tag{2}$$

which has been shown to minimize the amount of signal cancellation that will occur under perturbed wavefront conditions. The constraint is equivalent to adding a constant to the elements of the main diagonal of the system correlation matrix $\mathbf{R}(k)$, with the result that the array response approaches that of delay-and-sum processing when tightly constrained. Because of the frequency-domain implementation, the weight constraint can easily be made frequency-dependent. At low frequencies, where the array is shortest with respect to the acoustic wavelength and thus has the poorest directivity, the constraint can be adjusted to allow a higher degree of directionality in the array response. Conversely, at high frequencies, where delay-and-sum beamforming can give adequate amounts of array gain, the constraint can be tightened to guarantee that no signal cancellation will occur. The weight constraint was thus set to

$$\delta^2(k) = \begin{cases} 0 \ \text{dB} \ re{:}1, & f<1 \ \text{kHz}, \\ \ldots \end{cases} \tag{3}$$

The correlation matrix was computed separately at each FFT analysis frequency, and each matrix was smoothed by a low-pass filter having a time constant of 500 ms. The weight adaptation was independent at each of the FFT bin frequencies, with convergence for the equivalent of a 16-tap filter taking about 200 ms. The algorithm was allowed to adapt for 2 s to ensure full convergence prior to computing the performance metrics.

The superdirective processing used a variant on the scaled projection algorithm to produce the array weights. The scaled projection algorithm uses the signal-plus-noise correlation matrix in computing the set of array weights that minimizes the array output power; the processing is adaptive because the weights are computed iteratively using a correlation matrix that can change over time. To produce the superdirective weights, the correlation matrix for the ideal spherically isotropic noise field, computed *a priori* from the array geometry (Cheng, 1971) and unvarying in time, was substituted for the matrix measured from the input signal. The weight calculation was then iterated until convergence was reached, and the converged weights were used for the superdirective array performance measurements. The superdirective weights used the same scaled projection constraint on the magnitude of the weight vector as was used for the adaptive processing in order to prevent any potential signal cancellation caused by system misalignment (Cox *et al.*, 1986).

For the composite structured correlation matrix, the scaled projection algorithm framework was again used, but with a modified correlation matrix. The values of the measured signal correlation matrix were first averaged along each diagonal to give a Toeplitz structure (Godara, 1991); this matrix was then combined with the correlation matrix calculated for the spherically diffuse noise field, with the proportion of the diffuse noise-field matrix increasing with increasing input SNR.

C. Performance metric

The performance metric used in this paper is the articulation index (AI) weighted array gain, which is similar to intelligibility-weighted gain (Greenberg *et al.*, 1993). Experimental results have shown a strong correlation between the array gain and the improvement in speech intelligibility (Soede *et al.*, 1993b), and an even better correlation between the weighted array gain and intelligibility (Hoffman *et al.*, 1994). The AI-weighted array gain is calculated from the array gain computed at each frequency of the transformed data, and the array gains are combined using weights for each frequency band derived from the articulation index importance function given by Kryter (1962a).

The array gain (Cox *et al.*, 1987) for the kth FFT bin is given by

$$G(k) = \frac{|\mathbf{w}^*(k)\mathbf{d}(k)|^2}{\mathbf{w}^*(k)\mathbf{Q}(k)\mathbf{w}(k)}, \tag{4}$$

where $\mathbf{Q}(k)$ is the noise-alone correlation matrix normalized so that $\mathrm{Tr}[\mathbf{Q}(k)]=M$, the number of microphones in the

TABLE II. AI-weighted array gain in dB for the single noise source in the office. The data are presented as a function of the microphone array position and the filter length L, and are averaged over source location.

Position and length	Delay and sum	Oversteered	Optimal superdirective	Scaled projection Input SNR, dB			Composite struct. Input SNR, dB		
				−10	0	+10	−10	0	+10
Floor									
$L=8$	5.5	7.6	8.9	10.6	10.3	9.7	9.0	9.5	9.9
16	5.3	7.5	9.5	11.4	10.9	9.9	9.7	9.8	10.0
32	5.1	7.3	9.3	11.1	10.6	9.4	9.1	9.2	9.4
Desk									
$L=8$	6.0	8.2	10.1	11.8	11.5	11.1	10.5	10.5	10.2
16	5.6	7.9	10.6	12.4	11.8	10.7	10.7	10.9	10.9
32	5.3	7.6	10.1	11.8	11.1	9.8	9.9	10.0	10.0
KEMAR									
$L=8$	5.3	7.3	8.1	10.0	9.5	8.4	8.2	8.7	8.9
16	5.0	7.0	8.6	10.7	10.2	8.5	8.5	8.8	9.0
32	4.8	6.8	8.4	10.6	9.9	7.8	8.1	8.3	8.4

the spatial distribution of the noise, but is independent of the actual signal and noise powers. An array consisting of a single omnidirectional microphone has an array gain of 1. The estimated noise-alone correlation matrix used in the array-gain calculation was smoothed using a low-pass filter having a time constant of 500 ms. Since both the speech and noise were measured in reverberent rooms, this metric gives the ratio of the power in the direct portion of the speech signal to the total direct-plus-reverberent noise power at the array output, normalized by the SNR at the array input. This measure thus represents the directional gain of the array in the noise field. It is also possible, under conditions of a perturbed signal wavefront, for the array gain to appear to be favorable even though signal cancellation is occurring. The output signal power in each FFT frequency bin was monitored as a check for this condition, and no measurable signal cancellation was observed.

The AI-weighted array gain is then given by

$$G_{\mathrm{AI}}=\sum_{k=0}^{K} a(k)[10 \log_{10} G(k)] \ \ \mathrm{dB}, \tag{5}$$

where the set of weights $\{a(k)\}$ is the AI importance function weights given by Kryter (1962a) reinterpolated for the FFT band edges. Spread of masking effects are ignored in this metric. The AI-weighted array gain G_{AI} is expressed in dB *re*: the array gain for a single omnidirectional microphone.

The array gain given in Eq. (4) differs from the ratio of array output SNR to input SNR used by other authors (Greenberg and Zurek, 1992; Hoffman *et al.*, 1994) as the basis of the performance metric. The array output SNR is the ratio of the total speech power to the total noise power at the output of the array. The speech and noise powers both include the reverberated as well as the direct components. The array output SNR is given by

$$\mathrm{SNR}(k)=\frac{\mathbf{w}^*(k)\mathbf{S}(k)\mathbf{w}(k)}{\mathbf{w}^*(k)\mathbf{N}(k)\mathbf{w}(k)}, \tag{6}$$

where $\mathbf{S}(k)$ is the speech-alone correlation matrix and $\mathbf{N}(k)$

using this metric would then be calculated as the ratio of the array output SNR to the array input SNR, converted to dB and summed over frequency using the AI weights.

The preference for array gain *versus* the ratio of output to input SNR as the basis of the performance metric depends on the assumptions made about the effects of reverberation on speech intelligibility. The SNR-based metric assumes that all speech power, reverberated as well as direct, contributes equally to speech intelligibility. Experiments in speech intelligibility in reverberation, however, indicate that reverberation times typical of the rooms used in this paper lead to reduced speech intelligibility, and the longer the reverberation time the greater the reduction in intelligibility (Moncur and Dirks, 1967; Houtgast and Steeneken, 1972). This reduction of speech intelligibility with increasing reverberation time applies to hearing-impaired as well as to normal-hearing subjects (Duquesnoy and Plomp, 1980; Nábělek, 1982; Nábělek, 1988). The effects of reverberation on speech intelligibility have been accurately modeled by the speech transmission index (STI), based on the modulation transfer function within the room for speech envelope modulation frequencies (Houtgast and Steeneken, 1973; Steeneken and Houtgast, 1980). Even small amounts of reverberation within the room will reduce the envelope modulation depth and will therefore reduce the speech intelligibility predicted by the STI. These results indicate that the effects of reverberation are similar to those of noise in reducing speech intelligibility in rooms. Thus the array gain, by excluding the reverberated components in the estimated speech power, may lead to a more valid estimate of the array benefit for speech intelligibility in rooms than an estimate that assumes that all of the reverberated speech is beneficial.

II. RESULTS

The data from the experiment are presented in Tables II–V. Five array processing approaches were considered in the experiment. The three fixed-coefficient approaches of delay-and-sum beamforming, oversteered delay-and-sum beamforming, and optimal superdirective processing were

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.