COMPARISON OF MICROPHONE ARRAY DESIGNS FOR HEARING AID

D. Korompis, A. Wang, and K. Yao

E. E. Department, University of California
Los Angeles, CA 90024-1594

ABSTRACT

We perform simulation of various digital signal processing microphone array architectures to compare their suitibality for a hearing aid pre-processor. The architectures include fixed narrowband array, General Sidelobe Canceler array and Maximum Energy array. The arrays are all equi-spaced linear array. In particular, arrays with 6 microphones and various number of taps have been simulated under typical hearing aid environment of reverberance, short speaker distance and competing speakers. The sampling frequency is chosen to be 10 KHz, with the physical length of the array being 17 cm. The improvements of various array designs are demonstrated and discussed.

1. INTRODUCTION

There are estimated 25 millions of people in the U.S. who are hearing impaired. Of these, less than 10 percent use a hearing aid regularly. While people with normal hearing may have difficulty understanding speech with background noises and or reverberations, this problem becomes even more severe for hearing impaired people. A hearing aid with a microphone array pre-processor can act as a spatial filter that let the desired speaker's signal through, while rejecting unwanted disturbance. Thus, the signal to interference-noise ratio can increase markedly with a significant improvement in speech intelligibility. The hearing-aid environment imposes a limit on the overall length of a portable microphone array (about 20 cm), as well as the minimum bandwidth of the array (at least 6 octaves in the speech band).

2. ARRAY DESIGNS

2.1. Narrowband Fixed Arrays

In recent years, the works on fixed acoustic beamformer have mainly concentrated on delay-steered arrays, which make the signal from the look-direction arrive at the same time to all the microphones [1]. The look direction is usually the broadside or the endfire direction. By adding the signal with some weighting, a mainlobe is directed toward the look-direction. However, there is no way to control the locations of the sidelobe peaks. In this paper, we will consider the uniform weighted and the Maximum Directivity (MD) weighted broadside array. Let the response of the array to a plane-wave from target azimuth ψ, target elevation θ and at frequency ω for a equi-linear array with R sensors be expressed as $W(\psi,\theta,\omega) = \mathbf{E}^*(\psi,\theta,\omega)\mathbf{w}$, where $\mathbf{E}(\psi,\theta,\omega)$ is the array response vector. The directivity index at frequency ω is the ratio of the array output power due to sound incident from the target direction ($\phi = 0, \theta = 0$) to the average array output power to sound incident from all azimuth and elevation directions:

$$D(\omega) \quad = \quad \frac{\mathbf{w}^*\mathbf{E}_\omega{}^\dagger\mathbf{E}_\omega{}^T\mathbf{w}}{\mathbf{w}^*\mathbf{B}_\omega\mathbf{w}},$$

where ^{T} denotes a transposition, † denotes a complex--conjugation, and * denotes a complex-conjugate transposition. $\mathbf{E}_\omega$ is an abbreviation for $\mathbf{E}(0,0,\omega)$ and $\mathbf{B}_\omega$ is the cross-spectral density matrix for isotropic noise.

Stadler [1] has used the weights that maximize $D(\omega)$ given by:

$$\hat{\mathbf{w}} = (\mathbf{E}_\omega{}^T(\mathbf{B}_\omega + \sigma\mathbf{I})^{-1}\mathbf{E}_\omega{}^\dagger)^{-1}\mathbf{E}_\omega{}^T(\mathbf{B}_\omega + \sigma\mathbf{I})^{-1},$$

where the term $\sigma\mathbf{I}$ is inserted to control the white noise gain.

2.2. Adaptive Arrays

This class of adaptive beamformer has the property that it adapts to the changing statistics of the noise field. The only knowledge required is the desired signal direction, and in some cases, its frequency band of interest. Consider an equi-spaced linear array of R omni-directional sensors with L taps at each sensor, where the tap delay is denoted by τ as shown in Fig. 1. The input to the array denoted by $\mathbf{x}(n)$ is a stacked snapshot vector of microphones' input of size M with $M = LR$. Similarly, the weight vector is

$$\mathbf{w} = [w_{1,1}, w_{1,2}, \ldots, w_{1,L}, w_{2,1}, \ldots, w_{2,L}, \ldots, w_{R,L}]^T.$$

The Linearly Constraint Minimum Variance (LCMV) formulation minimizes the output power while satisfying the linear constraints, which usually ensure certain response from the look direction. It can be stated as:

$$\bar{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{\mathbf{XX}}\mathbf{w}, \qquad \text{subject to } \mathbf{C}^T\mathbf{w} = \mathbf{f}.$$

where the matrix $\mathbf{R}_{xx} = E\{\mathbf{x}^\dagger\mathbf{x}^T\}$ is the data correlation matrix.

The General Sidelobe Canceler (GSC) incorporates the linear constraints in the structure of the array. The optimum weight is decomposed into

$$\bar{\mathbf{w}} = \mathbf{w_q} - C_n\mathbf{w}_a,$$

This work is partially supported by The House Ear Institute.

 0-7803-2431-5/95 $4.00 © 1995 IEEE

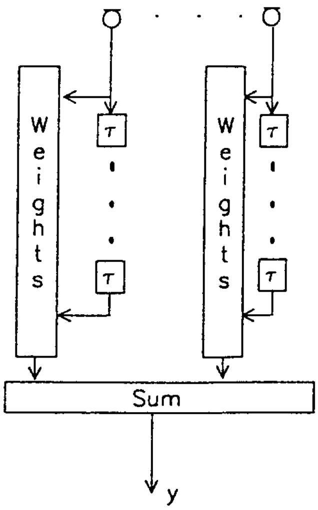


Figure 1: Broadband Arrays

where the branches are referred to as the primary channel and the secondary channel respectively. The primary channel weights are fixed and serve to satisfy the linear constraint, which usually ensure a certain look-direction response. The secondary channel weights are of lower-dimension, and are adapted to cancel the interference that is present in the primary channel. The matrix $\mathbf{C}_n$ is an orthonormal matrix that spans the null space of the constraint matrix $\mathbf{C}$ and is often called the *signal blocking matrix* because its columns are orthogonal to the space in which the signal lies. In the GSC structure, the data stream $\mathbf{x}$ is "convoluted" with the matrix $\mathbf{C}_n$ to form the input to the secondary channel, $\mathbf{x}_a = \mathbf{C}_n\mathbf{x}$. The secondary weights can be updated using unconstrainted LMS algorithm at time k:

$$\mathbf{w}_{a,k+1} = \mathbf{w}_{a,k} - \frac{\mu}{P_e}\, \mathbf{x}_{a,k}\, y_k \,,$$

where P_e is the power estimate on the secondary channel.

The basic assumption of all LCMV formulation is the absence of the desired signal in the secondary channel. The presence of the desired signal in the primary channel will increase the *misadjustment* of the LMS, which will decrease noise cancellation. This problem will be worsened by high SIR at the input. A potentially more severe problem is leakage of desired signal in the secondary channel. With high leakage, the adaptation process will be driven to cancel the desired signal, a phenomenon known as *power inversion.* The leakage problem is apparent in reverberant rooms, where the reflections of the desired signal coming from different directions will not be blocked by $\mathbf{C}_n$. The proposed remedies [8] include:

1. Use filter length that is shorter than the difference in travel time between the direct path and the first reflection of the desired signal.
2. Limit the adaptation only during period of silence of the desired speaker.
3. Evaluate the power estimate to include the power of the output P_y. This power normalization will reduce the rate of adaptation when the desired speaker is present.
4. Limit the norm of the weight, so that the output of the secondary channel would not overwhelm the output of the primary channel.

2.3. Maximum Energy Array

The shortcoming of the adaptive arrays in reverberant environment and the narrowband nature of the conventional fixed array motivate the search for broadband fixed array that is steerable and are able to put nulls in known interference directions. The Maximum Energy (ME) Array [3] maximally collects the energy from a specified spatial region, with the following additional properties:

- a steerable main-beam spatial "look-direction"
- a user selectable spatial attenuation band
- a flat response at the look-direction over a large frequency band

The maximization of the energy in spatial region of $[-\psi_o, \psi_o]$ about the look-angle ψ_c and in the frequency region of $[\omega_1, \omega_2]$ of a broadband array can be measured by

$$\beta = \frac{\int_{\psi_c-\psi_o}^{\psi_c+\psi_o}\int_{\omega_1}^{\omega_2} |\, W(\psi,\omega)\,|^2\, d\omega\, d\psi}{\int_{-\pi/2}^{\pi/2}\int_{\omega_{low}}^{\omega_{high}} |\, W(\psi,\omega)\,|^2\, d\omega\, d\psi} = \frac{\mathbf{w}^* A \mathbf{w}}{\mathbf{w}^* B \mathbf{w}}\,,$$

where the dependence on elevation angle θ has been surpressed. The placement of nulls can be achieved by linear constraints, both on the array value or its derivative. The array weights will then be constrained to lie on some subspace H. The weight is then chosen as :

$$\hat{\mathbf{w}} = \arg\max_{\mathbf{w}\in \mathrm{H}} \frac{\mathbf{w}^*(A-\alpha C)\mathbf{w}}{\mathbf{w}^*(B+\sigma I)\mathbf{w}}\,,$$

where the matrix C quantify the variation of the frequency response at the look-direction, and the term α controls it. The term σI is inserted to control the white noise gain. This generalized eigenvalue problem can then be converted to full-rank, lower-dimensional unconstrained problem, which can then be solved by the Simultaneous Iterative Method.

3. SIMULATION

3.1. General Procedure

A simulation program has been written that generates impulse response functions between sources and array element in a reverberant room. The room is characterize using the image model. Impulses that appear between sampling instances are low-pass interpolated [7]. Also, only images from the front hemisphere are used. We choose to simulate a medium size rectangular room of 5 meter by 4.9 meter by 2.44 meter. As a room becomes more reverberant, the sound field becomes more diffuse and directivity of an array will become less useful. For the adaptive arrays, reverberant room will degrade performance by introducing the desired signal in the secondary channel. For the ME array, putting a null in the direction of the interferer will not cancel it totally, because its reflection will come from other directions. The reverberation condition of a room can be quantified by several measures, including *reverberation time* and *critical* *distance*. Reverberation time is defined as the time it takes for the square pressure of an instantenuous signal to decay to 10^{-6} of its original value. Critical distance is defined as the distance between source and receiver in the room, from which the square pressure from the direct path signal is equal to the square pressure from the reverberant signal. The critical distance is approximately given by the formula

$$r_c = \frac{1}{4}\sqrt{\frac{\hat{\alpha_s} S}{\pi(1-\hat{\alpha_s})}},$$

where S is the total surface of the room and $\hat{\alpha_s}$ is the average energy absorption coefficient α_s of the surfaces. In this experiment, we consider an array of 6 elements with a sampling frequency of 10 KHz. The interelement spacing is 3.45 cm, resulting in overall length of 17.25 cm. The array is placed in the middle of the room, and the speaker and interferer at direction 0 deg and 50 deg respectively, both at the distance 1.5 meter. A pair of HINT speech files is used as speaker and interferer. The long-term average spectrum of these files are centered around 1 KHz. The absoption coefficients are first set to $\alpha_s = [0.20\,0.15\,0.42\,0.20\,0.29\,0.20]$ for the six sides. This corresponds to $r_c = 0.78$ meter. We then reduce α_s linearly to achieve varying r_c. The longest reverberation time of the room corresponding to the above α_s is about 400 ms. However, in all the simulations, the impulse response of the room is truncated at 200 ms. White noise at -30 dB from the combined speaker and interferer signal is also added to the microphone input signal, which is then passed through a DC notch filter with cutoff frequency of 150 Hz. All the simulations are 5 seconds long.

3.2. Performance Measure

Previous works [8],[2] have used the *intelligibility-weighted* SNR measure. However, due to a lack of simple way to compute that measure, this work measures the improvement of the array based on the L2-norm. Based on the speech transmission index theory, the first 50 ms of reverberation is known to be helpful to speech intelligibility. Let $y_{s,50}$ be the output of the array due to the direct path speaker's signal plus its first 50 msec reverberation. Similarly $x_{1,s,50}$ denotes the desired speaker's signal and its 50 msec reverberation at the input of microphone 1. Let

$$SNR_{in} = 20\log_{10}\frac{\| x_{1,s,50} \|_2}{\| x_1 - x_{1,s,50} \|_2} \qquad (dB)$$

and

$$SNR_{out} = 20\log_{10}\frac{\| y_{s,50} \|_2}{\| y - y_{s,50} \|_2} \qquad (dB)$$

The improvement of SNR in dB is defined by :

$$\Delta SNR = SNR_{out} - SNR_{in}.$$

To capture the effect of the adaptive arrays after the transient period, only the last 3 seconds of data are used for SNR calculations.

3.3. Fixed Arrays Procedure

The uniform, the MD and the ME arrays are simulated. In both the MD and ME array, the value of σ is chosen to give white noise gain of approximately 0 dB. The MD array is chosen to maximize directivity at 2 KHz. For the ME array, the tap length is chosen to be 40. Calculations show that longer tap length does not lead to further performance gain. The value of α is chosen to ensure that frequency response variation at the look-direction is within 5 dB. For the linearly constrainted ME, 2 pairs of magnitude and derivative nulls at 160 Hz and 478 Hz are placed at the direction of the interferer. We found that nulls at low frequencies provide best overall rejection.

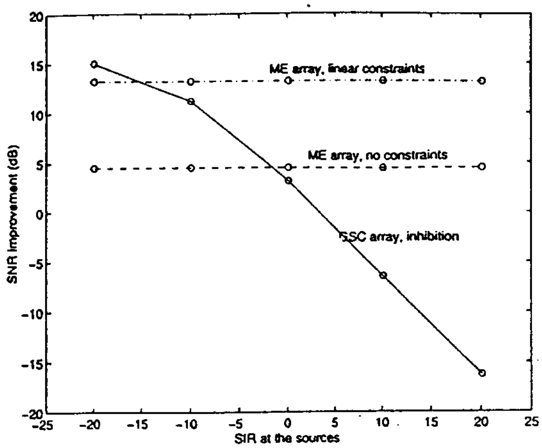
3.4. Adaptive Array Procedure

The GSC array simulations used an all-pass, broadside array with no primary channel delay($D = 0$) and tap length of 100 ($L = 100$) corresponding to 10 ms. The power estimates with power normalization is based on $P_e + P_y$. The power estimate is summed over the length of the array with forgetting factor of 0.99, which corresponds to time constant τ_p of 10 ms. The adaptation constant μ that gives the best Δ SNR for a mildly reverberant room with r_c of 3 meter and 0 dB input SIR is searched. It is found that Δ SNR is not sensitive to variation of μ in the range of 0.2 to 0.5, so μ is then fixed at 0.2 for all simulations. In [8], adaptation is inhibited when the correlation between microphone input is high, which indicate the presence of the speaker. However, the threshold value is dependant on the degree of reverberation. In this work, inhibition is achieved by computing the average exponentially weighted SIR at the microphones, which of course is not possible in real life. By choosing the threshold for input SIR, we can get approximately the same inhibition percentages as [8]. In free-space, the allowed adaptation time is 16% at +20 dB source SIR, 56% at 0 dB and 85% at -20 dB. Previous works have applied prewhitening [4] filter on the microphone input signal. However, we feel that this would affect the improvement of all the array approximately to the same degree, due to the fact that directivity is inherently better at higher frequencies. Thus, no prewhitening filter is implemented.

3.5. Simulation Results

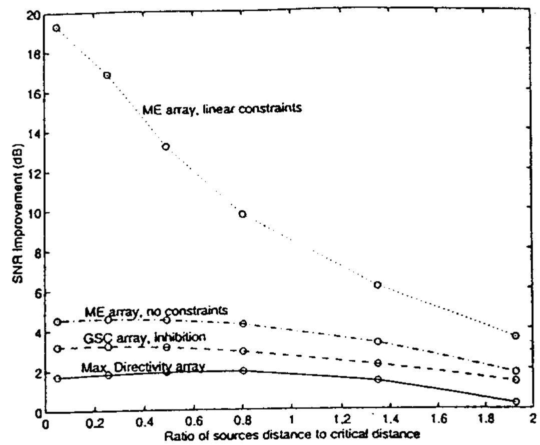
In the first experiment, we investigate the sensitivity of the GSC array to the input SIR. In this experiment, the value of α_s is fixed to give $r_c = 3$ meter. Sources are placed at 1.5 meter. Then the SIR at the sources are varied from -20 dB to 20 dB. It is observed that the GSC array is very sensitive to misadjustment and speaker signal cancellation at even moderate input SIR. This is true even when the adaptation process is control by inhibition. For comparison, the same scenario is simulated with ME arrays, with and without linear constraints. The ME arrays are of course unaffected by the input SIR except for the shifting noise floor. Fig. 2 shows the result of this experiment. We observe that for low SIR, the GSC performs very well. However, for equal strength sources, even the ME array without nulls outperforms the GSC.

Fig. 3 shows the result of the second experiment, where the value of α_s is varied to achieve different room reverberation. The SIR at the sources is fixed at 0 dB. The result for uniform array is very close to the MD array and is not shown. We observed that in all reverberation conditions,



r_c: 3 meter
Speaker distance 1.5 meter
GSC: $L = 100$, inhibition, $\tau_p = 10ms$
ME: $L = 40$, 2 pairs of const. at 160 Hz and 478 Hz

Figure 2: Δ SNR as a function of source SIR



Source SIR: 0 dB
Speaker distance 1.5 meter
GSC: $L = 100$, inhibition, $\tau_p = 10ms$
ME: $L = 40$, 2 pairs of const. at 160 Hz and 478 Hz
MD: $L = 1$, Max. Dir. at 2 KHz

Figure 3: Δ SNR as a function of distances ratio

the ME arrays outperform the GSC, which outperforms the single tap MD array. In free-space, the ME array with nulls are able to provide 18 dB improvement, which decreases to 8 dB when the speakers are at critical distance. The GSC array has improvement of 3.2 dB in free space, which does not change much with reverberation. The MD array stays at about 1.7 dB for all cases.

The surprisingly low improvement figures of the GSC array urges us to speculate that it is due to the nonstationary nature of speech signals. Thus, we conducted several simulations using sinusoid signals at 1.7 KHz and 1.5 KHz as speaker and interferer respectively. Table 1 compares the results of simulation for sinusoid and speech environment. We see that with sinusoid (stationary) sources, the GSC can

Δ SNR (dB)	Free-space	Sour. $= \frac{1}{2} r_c$ Dist.	Sour. $= 2\ r_c$ Dist.
Sinusoid signals	10.0	10.4	6.8
Speech signals	3.2	3.2	1.4

Table 1: Δ SNR for different sources

provide higher improvements even in severely reverberant conditions. Viewed this way, our results are consistent with results in [6], which obtained 7 to 10 dB improvements for source SIR of 0 dB[1] with semi-stationary noise.

4. CONCLUSION

This work compared some simulation results of several array designs, mostly in the competing equal strength speakers under reverberant conditions. It is found that the GSC arrays are very susceptible to the bursty nature of interfering speech. Also, it is susceptible to high source SIR. Under this severe condition, it perform only moderately better than single tap Maximum Directive array. The Maximum Energy array with linear constraints can provide substantially better cancellation. Also, it can be steered to the speaker at non-broadside direction. This is of course assuming that the direction of the speaker and interferer can be located either by user's direction or by some wideband direction of arrival algorithm that can perform well under reverberant conditions.

[1] The SIR figure used in [6] is defined as difference between 5-percentile peak signal level to the mean noise level. From [5], this SIR is about 8 dB higher that our SIR in absolute term.

5. REFERENCES

[1] R.W. Stadler et. al., *On the potential of fixed arrays for hearing aids*, JASA, 1993, pp. 1332-42.

[2] M.W. Hoffman, et. al., *Desired-Speech Signal Cancellation by Microphone Arrays in Reverberation Rooms*, Proc. Asilomar, 1991, pp. 344-348.

[3] D. Korompis and K. Yao, *Broadband Maximum Energy Array with user imposed spatial and frequency constraints* ICASSP, 1994, vol. 4, pp. 529-532.

[4] M.J. Link, et. al., *Prewhitening for intell. gain in hearing aid arrays*, JASA, 1993, pp. 2139-45.

[5] N.R. French, et. al., *Factors Governing the Intelligibility of Speech Sounds*, JASA, 1947, pp. 90-119.

[6] D. Van Compernolle, *Switching Adaptive Filters for Enhancing Noisy and Reverberant Speech from Microphone Array Recordings*, ICASSP 1990, pp. 833-836.

[7] P. Peterson *Simulating the response of multiple microphones to a single acoustic source in a reverberant room*, JASA, 1986, pp. 1527-1529.

[8] J. Greenberg et. al. *Eval. of an adaptive beamforming method for hearing aid*. JASA. 1992, pp. 1662-76.