LARGE VS SMALL APERTURE MICROPHONE ARRAYS: PERFORMANCE OVER A LARGE FOCAL AREA

Joshua M. Sachar, Harvey F. Silverman, and William R. Patterson III

Laboratory for Engineering Man/Machine Systems (LEMS)
Division of Engineering, Brown University
Providence, RI 02912 [*email: jms, hfs or wrp @lems.brown.edu]*

ABSTRACT

Microphone arrays offer the potential of obtaining a high-quality speech signal from a remote talker in a noisy, multi-source environment. In many important applications, sources of interest may be located anywhere within a large focal area and thus it is desirable that an array's performance be uniform over that area. In this paper, simulated and measured results are presented indicating the performances of representative small and large aperture arrays at various points in a large focal area. From these results, it is clear that uniformity over a large focal area is an advantage of large-aperture arrays.

1. INTRODUCTION

Obtaining high-quality remote speech data is an increasingly important goal for such audio-visual applications as teleconferencing and recording of meetings. In contrast to small office or desktop applications, these situations require that sources be allowed to move freely over a large fraction of a room. This large area is often called the system's *focal area.* Acceptable performance over a wide focal area calls for one of three solutions: 1) each talker (source) wears a head or body mounted microphone with wired or wireless connections to a decision program or human operator; 2) a program or human operator directs a parabolic-reflector or *shotgun* microphone at a targeted speaker; or 3) a fixed, microphone-array system focuses automatically on one or more simultaneous sources. This paper explores the design of microphone arrays with particular emphasis on their uniformity of response and spatial resolution over large focal areas.

Earlier work has shown through simulation [1, 2] and measurement [3, 4] that arrays which come close to surrounding possible sources - *large-aperture* arrays - have better spatial resolution near a central focus than spatially compact arrays. Here we extend that work to a wider range of practical array configurations and examine the uniformity of their behavior over large focal areas. Measured data for a 256 microphone, large-aperture array is compared to simulations, showing that real systems can achieve substantial uniformity even though overall performance may be limited by source radiation patterns and reverberations that are not usually modeled in simple simulations.

In Section 2, the *beampattern* [5] is introduced as a measure both for the simulations and the measurements. Delay-and-sum beamforming is used because of its simplicity and robustness in a variety of acoustic environments. Simulated results are shown first because, as a practical matter, spreading microphones around a room for large numbers of microphones is a painful exercise! Thus, in Section 3, we indicate what performance tradeoffs make sense, albeit with an idealized model. Results from four arrays subtending from 15 to 360 degrees at the center of the focal area are presented. In Section 4, we present data from our large-array system of 256 microphones. This system has been fully described in [6, 7]. Our experimental system was a U-shaped array of 256 microphones placed on the sides of a 5Mx6M room. Validation measurements include spatial resolution and SNR response over the entire usable focal area of this array.

2. PERFORMANCE MEASURE AND ENVIRONMENT

The beampattern of an array is widely used as an observable measure of an array's performance when aimed at a particular location. The beampattern shows the energy of the array's output as a function of source location $\vec{x}_s$ with a fixed aiming point $\vec{x}_a$. For a spherically uniform, sinusoidal source, $e^{j\omega t}$, the transfer function of the acoustic signal to the array electrical output for a weighted delay-and-sum beamformer is

$$B(\vec{x}_s, \vec{x}_a, \omega) = \sum_{m=1}^{256} \frac{W_m}{|\vec{x}_m - \vec{x}_s|} . e^{j\omega \frac{|\vec{x}_m - \vec{x}_a| - |\vec{x}_m - \vec{x}_s|}{c}}, \tag{1}$$

where c is the speed of sound and W_m is a set of weights. The beampattern depends upon both the distance from each microphone to the fixed aiming location and the distance from each microphone to the sources location. As the source is moved throughout the environment, the individual microphones receive signals that are inversely proportional to the distance from source to microphone. In this work we have used a uniform weighting with $W_m = W$ chosen to make $|B| = 0dB$ at $\vec{x}_s = \vec{x}_a$. In order to observe the performance over a range of frequencies in the simulations, we use a uniformly weighted sum of the squared magnitudes of B over the range $[1000Hz, 4000Hz]$ in 100Hz steps, or

$$Q(\vec{x}_s, \vec{x}_a) \equiv 10\ log \left[\sum_{m=10}^{40} |B(\vec{x}_s, \vec{x}_a, 200\pi m)|^2 \right] \tag{2}$$

Four arrays of 256 omnidirectional microphones, but with different geometries, are simulated in Section 3. These are shown in Figure 1. In all experiments, the z-direction is ignored. Each of the arrays has the same z size, 0.67M. All actual microphone locations were placed randomly with the restriction that the random points were on a 3cmx3cm grid.

- **Array I:** Planar, small aperture – 0.67M wide. Due to its small aperture much of the focal area is in the far-field. We

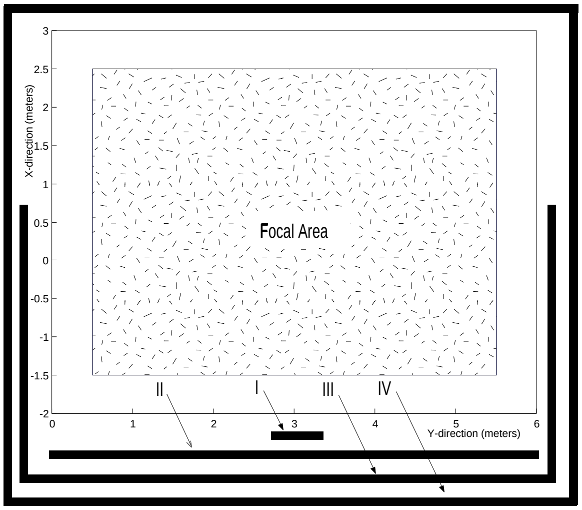

Fig. 1. Top View of the Room Showing the Placement of the Four Arrays and their Common Focal Area

would expect the array's beampattern to have only angular discrimination with little range selectivity.

- **Array II:** Planar , "medium aperture" – 6.0M wide. This array subtends 80^o for a source at the center of the focal area. We expect this configuration to perform better than the smaller array since a larger proportion of the focal area will be in the near field, enhancing range selectivity.
- **Array III:** U-shaped (on three walls), large aperture – Angle subtended $= 186^o$.
- **Array IV:** Fully enclosed, large-aperture array.

3. RESULTS FROM SIMULATIONS

Figure 2 shows the beampatterns produced by placing three aiming locations within the focal area of Array I. The beamformer isolates sources poorly, having inadequate range discrimination. The maximum responses do not even occur when the source is at the aiming points. While the peak response is in the correct direction, the only attenuation with distance results from the inverse distance factor in (Equation 1). There is insufficient variation in the source-to-microphone distances to produce significant attenuation from phase mismatch. The rapid accumulation of phase mismatch with a change of source position is critical to any satisfactory spatial discrimination.

Figures 3, 4 and 5 present similar data for the other array geometries.

We see that the medium aperture, planar array, Array II, does not exhibit uniformity in its beampattern. Although the peak responses occur near the aiming points, there are still some far-field effects that stretch the contours. There is poor range discrimination in the directions of these elongations. Aiming points near the array are less affected by this phenomenon since they are in the near-field of the array.

For the U-shaped Array III, the four evaluation aiming points were selected to be as separated as possible, but easily measurable (*i.e.*, no interference from cabinets and furniture). As seen in Figure 4, Array III has better performance than either planar array. The sides of the U substantially improve selectivity in the X-direction and to a lesser extent in the Y-direction. While still

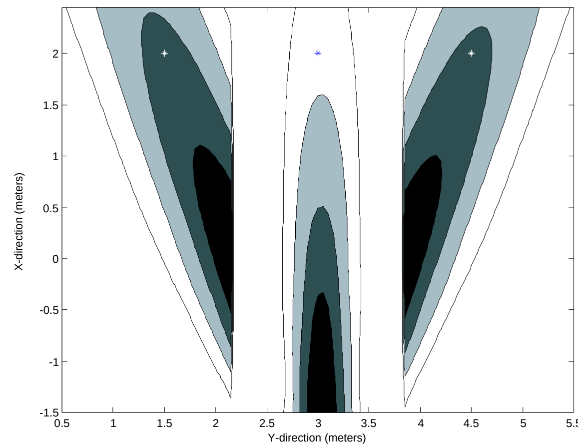

Fig. 2. Contour Plot from Simulation for Array I for Three Aiming points Indicated by Plus Signs. Contours are for -3, -6, -9, and -12dB below the Maximum Response

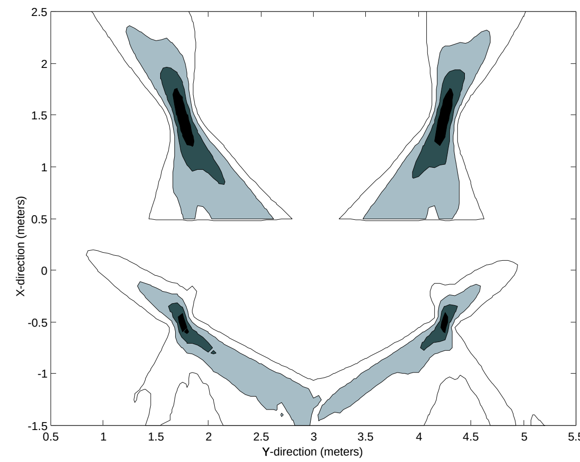

Fig. 3. Contour Plot from Simulation for Array II for Four Aiming Points in the Focal Area. Contours are for -3, -6, -9, and -12dB below the Maximum Response.

partially elongated, beampatterns are substantially nearer to approximate circular symmetry.

Twenty different aiming points were used to illustrate the effectiveness of Array IV, the one that entirely surrounds the focal area. The highest gain, dark-color regions surround the aiming points tightly. There is little asymmetry in the contour shapes. The elongation of the -12dB contours near the edge of the room is caused by proximity to individual microphones. Perhaps this implies that the definition of the focal area – 0.5M from each of the surrounding walls – is too large for applications with the most demanding uniformity requirements.

4. MEASURED RESULTS

The room in which the Huge Microphone Array (HMA) is located is 8.4M on each side with a ceiling height of 3 meters. Walls and floor are hard acoustic materials and the room has a long (450ms) reverberation time. Some partitions divide the room, so that

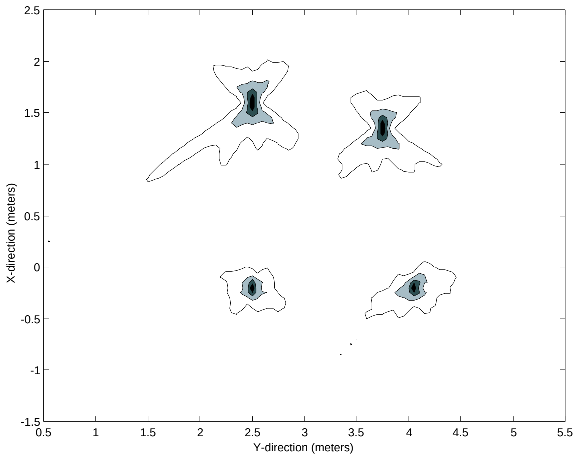

Fig. 4. Contour Plot from Simulation for Array III for Four Aiming Points in the Focal Area. Contours are for -3, -6, -9, and -12dB below the Maximum Response

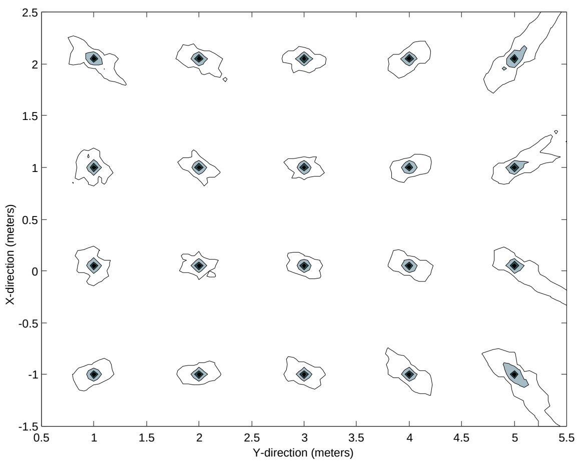

Fig. 5. Contour Plot from Simulation for Array IV for 16 Aiming Points in the Focal Area. Contours are for -3, -6, -9, and -12dB below the Maximum Response

the 256-microphone array is placed on three adjacent sides of a 5Mx6M rectangle. Array III uses exactly this set of microphone locations and so the results in Figure 4 are the point of comparison with our measurements. Further descriptions of the room and testing environment are given in [3] and [6].

The source used for these measurements is a 5cm diameter, long-throw, full-range speaker in a sealed enclosure, the forward response of which is known to vary less than 3dB over the frequency range of 1-4kHz. It was mounted on a tripod and moved at a fixed height as needed. A continuous swept sinewave changing linearly with time from 1kHz to 4kHz in $\frac{1}{7}s$ was used as the signal. Measurements of energy were taken over roughly 35 sweeps or about 5s. As the chirp started its sweep at 1kHz, and room background noise from fans and computers is predominantly low frequency, a linear phase FIR highpass filter with cutoff at about 400Hz was used to reduce the background effects in the measurements.

In Figure 6 a contour plot of the beampattern is shown for the

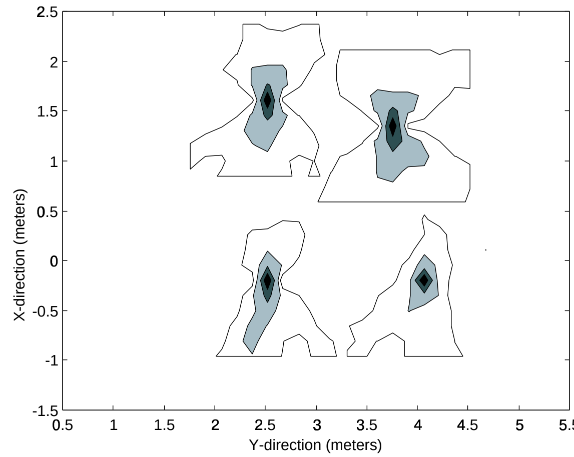

Fig. 6. Measured Contour Plot for Array III for 4 Aiming Points in the Focal Area. Contours are for -3, -6, -9, and -12dB below the Maximum Response

same four locations shown for the simulation, Figure 4. The quantization of the measured data is about three times coarser than in the simulations ($\sim$ 15cm vs 5cm). The shapes of the measured volume-selective peaks in the top 6dB are quite uniform and are similar in size to those in the simulation. (Remember the U-shaped array does not show the same uniformity over the focal area that is seen in the fully enclosed array, Array IV.) While the contour lines at -9 and -12dB show the same directionality in both simulation and measurement, the measured contours show significant spreading. This is not surprising. The ideal assumptions of the simulation ignore 1) the reverberations of this highly-reverberent room, 2) the spectral coloration and very non-spherical radiation pattern of the speaker, 3) the variations in the gains of the microphones, 4) the slight errors in aiming, and even with high-pass filtering, 5) the residual effects from acoustic background noise. All these factors tend to spread the beampattern function.

In Figures 7 and 8 we show the results of measurements made over most of the focal area to characterize the uniformity of amplitude response for tracked sources. Figure 7 shows a contour plot of the variation in the signal energy measured as the source was placed at a succession of aiming points on a uniform coarse mesh across the focal area. Because of the choice of weighting coefficient W in equation 1, the ideal array output should be precisely constant. The observed variation of $\pm$ 2dB is quite good considering the effects mentioned above that are not included in either the simulation or the calculation of weights.

Because of the renormalization of amplitude by the weighting coefficient, the amplitude response is not an adequate measure of the effective sensitivity of the array. One has to consider how the signal-to-noise ratio varies from point to point. To determine the noise, a measurement was made at each aiming point in the same mesh with the source turned off. In Figure 8, we show changes in the SNR as the source moved over the same large focal area. The mean SNR for these measurements was 30.3 dB, and the figure shows a total variation of only 4dB across the entire area. The darkest contour represents a 32dB SNR. The general bias of slightly higher SNR in the direction of increasing Y is the result of a discrete source of acoustic noise from computer fans near the the

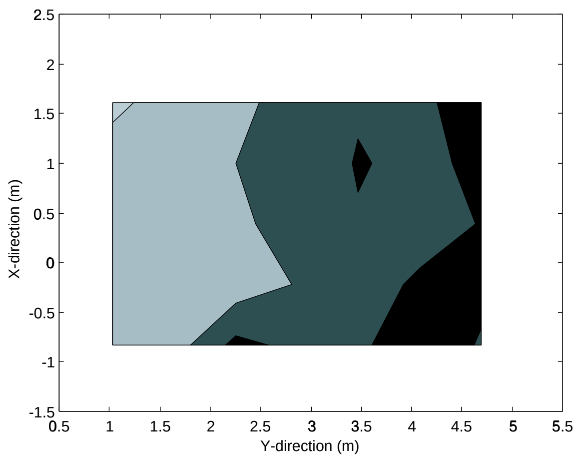

Fig. 7. Measured Contour Plot for Array III for Peak Signal Response over a Large Focal Area. Contours are for -1, -2, -3, and -4dB below the Maximum Response

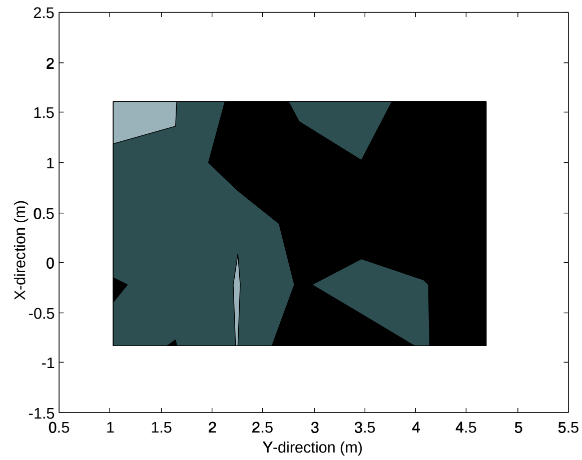

Fig. 8. Measured Contour Plot for Array III for SNR over a Large Focal Area. Contours are for -1, -2, -3, and -4dB below the Maximum Response

upper left corner. This is better than expected performance, given all the complicating factors mentioned above.

5. CONCLUSIONS

We have investigated the performance of microphone arrays with small, medium and large apertures over a wide focal area. It is clear that small and medium aperture planar arrays suffer from an inability to select volumetrically due to poor range discrimination. Simulations show that ideally a large-aperture array that totally encloses potential sources will provide the uniform response and high spatial selectivity needed for demanding applications. Measurements of a large U-shaped array confirm the conclusions from these simulations. However, the measured results are somewhat different from those of the simulation primarily due to the non-spherical radiation pattern of the source transducer and the reverberation patterns of the room. Improved uniform performance in a real environment might be achieved by taking the transfer function for the source angular dispersion into account as a source moves about the focal area and changes orientation. Future work will extend measurements to arrays that totally surround the focal area and will try to exploit source dispersion as well.

6. REFERENCES

[1] H. F. Silverman and W. R. Patterson III. Visualizing the performance of large-aperture microphone arrays. In *Proceedings of ICASSP-1999*, pages 969–972, Phoenix, AZ, March 1999.

[2] D. Rabinkin. *Optimum Placement of Sensors for Large Microphone Arrays*. PhD thesis, Rutgers University, Piscataway, NJ, May 1998.

[3] H. F. Silverman, W. R. Patterson III, and J. M. Sachar. Early results for a large-aperture microphone array system. In *Proceedings of SAM2000)*, pages 207–211, Boston, MA, March 1999.

[4] H. F. Silverman, W. R. Patterson III, and J. M. Sachar. First measurements of a large-aperture microphone array system for remote audio acquisition. In *Proceedings of IEEE International Conference on Multimedia and Expo*, New York, NY, July/August 2000. Session TP0, Paper 5.

[5] Don H. Johnson and Dan E. Dudgeon. *Array Signal Processing: Concepts and Techniques*. Prentice-Hall, Englewood Cliffs, NJ, 1993.

[6] H. F. Silverman, W. R. Patterson III, and J. L. Flanagan. The huge microphone array (HMA)- Part I. *IEEE Transactions on Concurrency*, 6(4):36–46, October-December 1998.

[7] H. F. Silverman, W. R. Patterson III, and J. L. Flanagan. The huge microphone array (HMA) - Part II. *IEEE Transactions on Concurrency*, 7(1):32–47, January-March 1999.