SUPERDIRECTIVE MICROPHONE ARRAY FOR A SET-TOP VIDEOCONFERENCING SYSTEM

Peter L. Chu

PictureTel Corporation, MS 635
100 Minuteman Road
Andover, MA 01801-1031, USA
chu@pictel.com

ABSTRACT

In set-top videoconferencing, the complete videoconferencing system fits unobtrusively on top of the television. The microphone sound pickup system is one of the most important functional blocks with constraints of small size, high performance, and low cost. Persons speaking several feet away from the system must be picked up satisfactorily while noise generated internally in the system by the cooling fan and hard drive, and noise generated externally from air conditioning and nearby computers must be attenuated.

In this paper, a three microphone superdirective array is described which meets these constraints. An analog highpass and lowpass filter are used to merge two of the microphone signals to form a single channel, so that a single stereo A/D converter is required to process the three microphone signals. The microphone signals are then linearly combined so as to maximize the signal-to-noise ratio, resulting in nulls steered toward nearby objectionable noise sources.

1. INTRODUCTION

Superdirectivity is a decades-old technique which has been used for radio frequency and sonar applications. It is the best known technique for maximizing SNR (signal-to-noise ratio) and directivity given a array of a limited number of microphones and limited physical space. For this reason its application to hearing aids has been recently explored by Kates [1], [2] with good results. Algorithmic details of superdirectivity are discussed by Cox et. al. [4]. Hudson [3] has an excellent tutorial on the mathematics involved in the procedure.

Similarly, superdirectivity is a natural choice for set-top videoconferencing, where the complete system must fit unobtrusively on top of the television, greatly constraining the physical size of the mic array. The microphone sound pickup system is one of the most important functional blocks which should provide high performance at low cost. Persons speaking several feet away from the system must be picked up satisfactorily while noise generated internally in the system by the cooling fan and hard drive, and noise generated externally from air conditioning and nearby computers must be attenuated.

In this paper, a three microphone superdirective array is described which meets these constraints. An analog highpass and lowpass filter are used to merge two of the microphone signals to form a single channel, so that a single stereo A/D converter is required to process the three microphone signals. The microphone signals are then linearly combined so as to maximize the signal-to-noise ratio, resulting in nulls steered toward nearby objectionable noise sources.

2. DELAY-AND-SUM VS. SUPERDIRECTIVITY

A popular microphone array design methodology is the delay-and-sum technique. It has been investigated by many researchers [5], [6], [7], [8]. The delay-and-sum approach seeks to maximize signal energy as opposed to SNR. Array geometries usually consist of microphone elements lined up broadside to the source direction. The microphone signals are appropriately delayed and summed together so as to maximize signal strength in the source direction. The strengths of this approach lie in its very weak sensitivity to microphone variations (eliminating the need for calibration), minimal loss in performance over a wide steering angle, and convenient control of the beam shape over these angles. Practically speaking, the main weakness of the approach lies in the large physical sizes needed to achieve large array gain (more than a wavelength).

Alternatively, the superdirective approach has the potential to yield twice as much SNR gain in dB than the delay-and-sum approach [9] in isotropic acoustic noise fields (assuming the self-noise of the microphones is small in comparison) for the same number of microphones. For example, the commonly used hypercardioid microphone, which may be viewed as a two element superdirectional array, has an SNR gain of 6 dB relative to a single omnidirectional microphone while the delay-and-sum approach for a two microphone array would yield a maximum gain of 3 dB. Typically, the hypercardioid capsule is less than an inch in dimension, while the delay-and-sum array would require a significant fraction of a wavelength spacing between the two microphones to achieve any significant gain.

The disadvantages of superdirectivity lie in the necessary careful characterization of microphone element response (calibration) and loss of performance when wide steering ranges are required. Maximum SNR gain occurs for end-fire arrays, and steering away from the end-fire direction entails substantial SNR loss. Using array configurations other than end-fire can reduce the relative loss in performance for different directions, but results in a significant loss compared to the end-fire array for the end-fire direction given a fixed number of microphone elements.

3. SUPERDIRECTIVE ALGORITHM DESCRIPTION

3.1. CALCULATING THE TAP WEIGHTS

Assume we have N microphone signals, $m_i(t)$, $i = 0, ..., N-1$. Each real microphone signal is fed to its own FFT-based polyphase filterbank, producing complex bandpassed signals $m_{ki}(t), i = 0, ..., N-1, k = 0, ..., K-1$ where K is the total number of bands. For our implementation, 256 bands span an 8 kHz audio bandwidth, with each band having a bandwidth of 31.25 Hz. For each band k the dot product of the the microphone signals and weights are taken to produce the superdirective output for that band,

$$s_k(t) = \mathbf{a}_k^T \mathbf{m}_k(t). \tag{1}$$

For each band, k, the vector of N complex weights $\mathbf{a}_k$ is chosen so as to maximize the SNR of $s_k(t)$. The noise statistics are specified by the crosscorrelation of the noise between the microphones, specified by the complex NxN matrix $\mathbf{Q}_k$, and the signal is specified by the magnitude and phase relationships between the microphones in the vector $\mathbf{d}_k$. Usually d_{ik} for one of the microphones is assumed 1, and so that all other microphones' magnitude and phase are specified relative to that reference microphone. For every band k, the same i th microphone signal is assumed 1, and the resulting tap weights will normalize the final output so it equals the signal at the i th microphone, but with attenuated noise and reverberance. The solution for the optimal weights for band k is:

$$\mathbf{a}_k = \frac{\mathbf{Q}_k^{-1}\mathbf{d}_k}{\mathbf{d}_k^* \mathbf{Q}_k^{-1}\mathbf{d}_k} \tag{2}$$

Intuitively, the solution of (2) may be viewed as a two step process of first decorrelating the noise components between the microphones and secondly applying a matched filter based the microphone signal statistics to maximize the SNR [3]. The N weights of the vector $\mathbf{a}_k$ found in (2) are used to compute $s_k(t)$ via (1), and the resulting bandpass signals $s_k(t)$ are then combined via a synthesis bandpass filterbank producing a real fullband signal which is the superdirective array output.

If noise crosscorrelation statistics are updated continously while the microphone array is in use, an automatic noise nulling capability results. The depth of the null will depend on the reverberance of the noise at the microphone array and the proximity of the noise source direction to the signal source direction.

The signal statistics may be gathered in an anechoic chamber with a white noise signal source placed in the far field in the desired look direction of the array, i.e., the end-fire direction for an end-fire array.

3.2. MICROPHONE MERGING VIA ANALOG FILTERS

While the procedure of section 3.1 may be used to find optimal weights for any array geometry of microphones, the geometry which gives maximum increase in SNR for a single direction is the endfire array.

In the case of an endfire array of two microphones, if the spacing between the microphones is too small, the circuit self-noise of the microphones (in electret micophones this noise is from the FET transistor preamplifier embedded in the microphone capsule) limits the acoustic noise reduction possible. Alternatively if the spacing of the microphones is too large, spatial aliasing of nulls and peaks in the directivity pattern occurs, limiting the attenuation of acoustic noise. Experimently, we have found the optimal spacing is approximately a quarter wavelength. Therefore, ideally, the physical distance between the microphones should decrease with increasing frequency. Using economical analog low pass and high pass filters, this distance vs. frequency characteristic may be achieved via the technique illustrated in figure 1. The rear-most microphone is 4 inches from the front microphone which is a quarter wavelength at 846 Hz, while the middle microphone is 1 inch from the front microphone which is a quarter wavelength at 3385 Hz. The crossover frequency of the analog filters is 2000 Hz.At frequencies below crossover the array appears to be two microphones with 4 inch spacing while at higher frequencies it appears to be two microphones with 1 inch spacing. While the design is an endfire three microphone array, only a single (expensive) stereo A/D converter is required since the rear two microphone signals are effectively merged via economical analog highpass and lowpass filters into a single signal. Clearly the design philosophy could be extended to more microphones and more filters to cover an even wider audio frequency range. This design has been implemented in a PictureTel set-top videoconferencing product.

3.3. PERFORMANCE OF THE THREE ELEMENT SUPERDIRECTIVE ARRAY

In the set-top videoconferencing system where the array is used, hard drive noise and fan noise from the unit itself are objectionable. The superdirective array nulls these noise sources to inaudibility. Neglecting these local noise sources, in typical office environments, the total amount of noise reduction relative to a single cardioid microphone is about 3-4 dB. Unfortunately, the subjective SNR gain is not as great because, along with the noise, reverberant signal energy is attenuated, and our ears integrate the direct and reverberant energy of a source together in the perception of loudness.

In general, the speech sounds clearer, less reverberant with the superdirective array compared to the single cardioid microphone.

4. RING OF MICS SUPERDIRECTIVE ARRAY

By having a ring of four dipole microphones as shown in figure 2, it is possible to achieve a two element endfire superdirective array for eight directions, effectively allowing pickup of audio sources originating from any direction in the plane of the microphones with negligible loss in performance. Such a topology would make a cost effective superdirective table-top mounted array where the conferencing participants would be seated around the array.

Refering to figure 2, sources in the North or South directions have microphones 1 and 3 forming a two element endfire array. Sources in the East or West directions have microphones 2 and 4 forming an endfire array. Sources that are in a Northeast or Southwest direction have the virtual dipole consisting of microphone 1 summed with microphone 2, and the virtual dipole of microphone 3 summed with microphone 4 forming a two element endfire array. Finally, sources in the Northwest or Southeast direction have virtual dipoles of microphone 4 summed with microphone 3 and microphone 1 summed with microphone 2 to form an endfire array. Thus, for eight directions a two element endfire array exists using the four microphone topology. This array has been implemented in hardware with very good test results.

5. SUMMARY

In this paper, physically small, very economical superdirective array structures have been discussed which provide significant gain over single microphone elements. A three element superdirective array has been implemented in a commercial set-top videoconferencing system with good results.

Although superdirectivity has been around for decades, its application to audio pickup problems has not been popular. When the audio microphone application demands small physical size, the acoustics DSP engineer should give supedirectivity techniques careful consideration.

6. REFERENCES

[1] J. Kates, "Superdirective Arrays for Hearing Aids", J. Acoust. Soc. Am. 94(4), pp. 1930-1933.

[2] J. Kates, "An Evaluation of Hearing-Aid Processing", 1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York.

[3] J. E. Hudson, Adaptive Array Principles, pp. 59-69, copyright 1981, New York: Peter Peregrinus for IEE.

[4] H. Cox et. al., "Practical Supergain", IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-34, pp. 393-398, June 1986.

[5] J. L. Flanagan, J. D. Johnston, R. Zahn, G. W. Elko, "Computer-steered Microphone Arrays for Sound Transduction in Large Rooms", J. Acoust. Soc. Am. 78(5), November 1985, pp. 1508-1518.

[6] M. M. Goodwin, G. W. Elko, " Constant Beamwidth Beamforming", ICASSP-93, pp. I-169-I-172.

[7] J. L. Flanagan, D. A. Berkley, G. W. Elko, J. E. West, and M. M. Sondhi, "Autodirective microphone systems", Acustica 73(2), 1991, pp. 58-71.

[8] Yves Grenier, "A Microphone Array for Car Environments", Speech Communication 12(1), March, 1993, pp. 25-39.

[9] G. W. Elko, personal communication, June 1996.

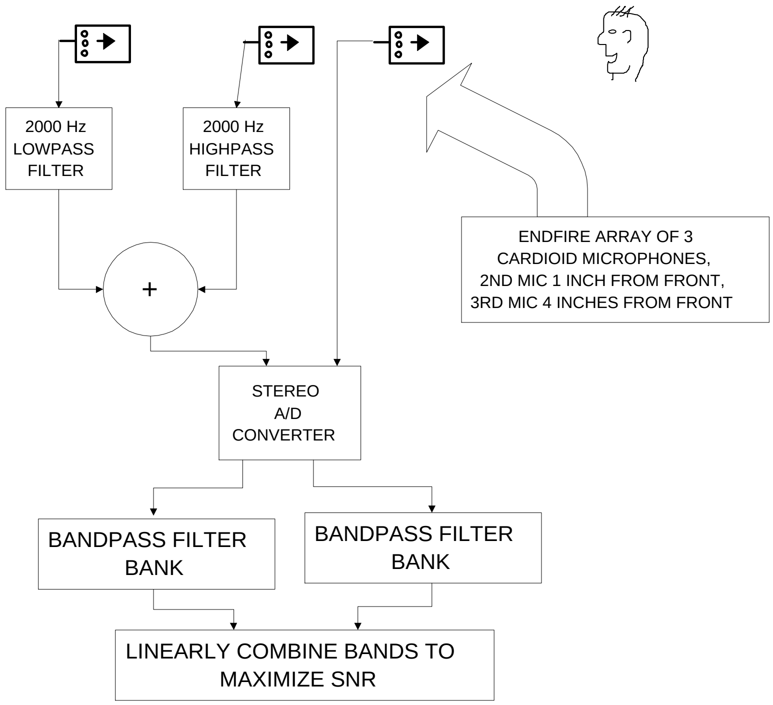

Figure 1: Three microphone superdirective array using a stereo A/D

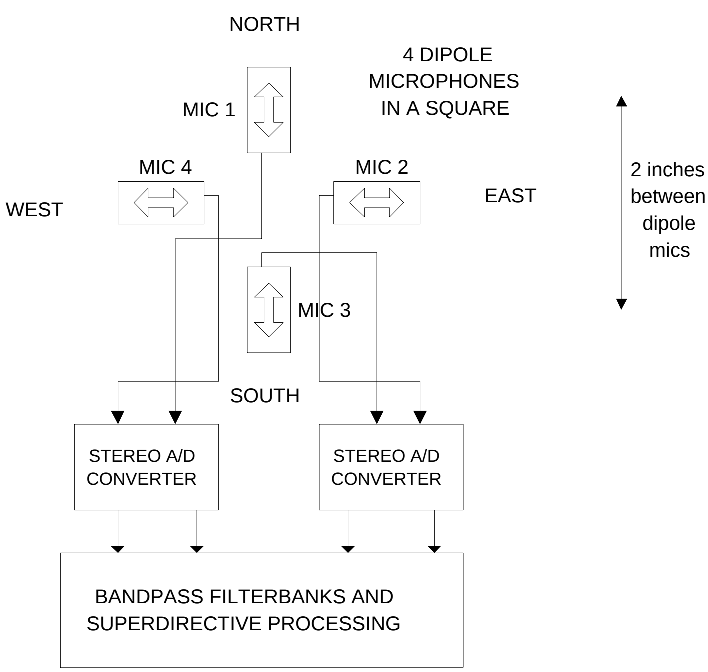

Figure 2: Ring of dipoles superdirective array, top view