DESKTOP MIC ARRAY FOR TELECONFERENCING

Peter L. Chu

PictureTel Corporation, MS 635
222 Rosewood Dr.
Danvers, MA 01923, USA
chu@pictel.com

ABSTRACT

Reducing the noise and reverberance in sound pickup has been a problem ever since the microphone was invented. Elegant solutions using multiple microphones in an array are a current hotbed of research [1] [2] [3] [4]. Unfortunately, because of the computational / monetary cost of these approaches, they have not been widely implemented in products.

In this paper, an automatically steered mic array, which works by taking linear combinations of two dipole microphones, is presented whose cost is low enough to have been implemented in a videoconferencing product. The array is positioned centrally on the conference table and provides very reasonable pickup for people speaking within a 7 foot radius, adequate for most conferencing situations. While simple in structure, the array provides a large increase in convenience and performance compared to the common method of laying out multiple cardioid microphones on the table, where each participant must be within the pickup angle / range of a cardioid microphone.

1. DESKTOP ARRAY STRUCTURE

The desktop array consists of two dipole microphones mounted perpendicularly to each other, as close to each other and as close to the table as possible. The main beams of the microphones are parallel to to the table top surface. The two dipole microphone outputs go to the left and right channels of a stereo A/D converter, whose output, in turn, goes to the DSP chip. A block diagram of the structure is shown in figure 1. Assuming the source is in the far field, as a function of angle the response of the dipole microphone to the source is

$$D(\theta) = A\ cos(\theta) \qquad (1)$$

where A is the sound pressure level of the source at the dipole and θ is the angle between the source and the on-axis angle of the dipole. Examination of (1) shows that the dipole has a response of 1 when $\theta = 0$, -1 when $\theta = 180$ degrees, and 0 when $\theta = +/-$ 90 degrees. The dipole decreases isotropic noise and reverberance by 4.8 dB compared to an omnidirectional microphone, assuming the source is on-axis. The more commonly used cardioid or unidirectional microphone has response

$$C(\theta) = A\ (.5 + .5cos(\theta)). \qquad (2)$$

The cardioid pattern has a response of 1 when $\theta = 0$, 0 when $\theta = 180$ degrees, and .5 when $\theta = +/-$ 90 degrees. Both cardioid and dipole directivity patterns reduce isotropic noise and reverberance by equal amounts. However, if the noise source is predominantly overhead, as is the case for air conditioning vents, the dipole with its main beam parallel to the tabletop surface will do a better job than the cardioid in attenuating the vent noise because of its response null in the vertical axis. On the other hand, because of strong reflections from surfaces directly opposite the person speaking, the cardioid, with its null in the opposite direction of its main axis, often sounds slightly less reverberant than the dipole. Overall, weighing the advantages and disadvantages of both, the two patterns are fairly equal choices for microphone pickup.

Assuming a fixed frame of reference for the two perpendicularly mounted microphones, the response for microphone A is

$$D_A(\theta) = A\ cos(\theta) \qquad (3)$$

and the response for microphone B is

$$D_B(\theta) = A\ sin(\theta). \qquad (4)$$

Adding D_A, the signal of microphone A, to D_B, the signal from microphone B, and scaling by $\sqrt{2}$,

$$D_C(\theta) = \frac{1}{\sqrt{2}}(D_A(\theta) + D_B(\theta)), \qquad (5)$$

0-7803-2431-5/95 $4.00 © 1995 IEEE

yields

$$D_C(\theta) = cos(\theta - 45°) \quad (6)$$

which is simply a dipole microphone pattern shifted 45 degrees relative to the main axis of microphone A. Similarly, subtracting D_B from D_A and scaling by $\sqrt{2}$ yields

$$D_D(\theta) = cos(\theta + 45°) \quad (7)$$

which is a dipole microphone pattern shifted -45 degrees relative to the main axis of microphone A. Therefore, by taking the sum and difference of the two dipole signals and scaling appropriately, easily done in the DSP, it is possible to derive two additional dipole patterns oriented halfway in angle between the two original patterns. The four pickup patterns defined, $D_A(\theta)$, $D_B(\theta)$, $D_C(\theta)$, and $D_D(\theta)$ (shown in figure 2) adequately cover a full 360 degrees of arc, since a source halfway between two beams is down only .688 dB from maximal on-axis response. In fact, *any* arbitrary angle of rotation of the dipole pattern can be achieved by taking the appropriately weighted linear combination of the two dipole micropone signals.

2. DESKTOP MIC ARRAY BEAM SELECTION

The algorithm which chooses which of the four beam patterns to use in picking up the source should be insensitive to constant background noise from air vents and reverberant energy. Computational simplicity is also a major concern.

The steps in the algorithm for beam selection will now be outlined.

1. Bandpass Filtering- The left and right channels from the stereo A/D converter are fed into two separate but identical FIR bandpass filters which let through frequencies in the 1-4 kHz region (the sampling rate of the system is 16 kHz). The bandpass filtering gets rid of much of the lower and higher frequency background noise. The speech signal below 1 kHz tends to be more reverberant than higher frequencies so is less useful for finding the source direction. For the left channel, the bandpassed output is

$$l_b(n) = \sum_{k=0}^{k<L} l(n-k)h(k), \quad (8)$$

and for the right channel,

$$r_b(n) = \sum_{k=0}^{k<L} r(n-k)h(k). \quad (9)$$

2. Decimation by Four- To reduce computations involved in the FIR bandpass filter in the previous step by a factor of four, the outputs of the bandpass filters are decimated by four. While aliasing is introduced in this process, the aliasing has little effect on later calculations in which energy will be measured. For the left channel,

$$L_b(m) = l_b(4m) \quad (10)$$

and for the right channel,

$$R_b(m) = r_b(4m). \quad (11)$$

3. Formation of Four Beams- Signals from the four dipole patterns are derived by taking the appropriately scaled sums and differences of the two bandpass, subsampled signals. The absolute value is taken of the samples, so that

$$A_1(m) = |L_b(m)| \quad (12)$$

$$A_2(m) = |R_b(m)| \quad (13)$$

$$A_3(m) = |\frac{1}{\sqrt{2}}(L_b(m) + R_b(m))| \quad (14)$$

$$A_4(m) = |\frac{1}{\sqrt{2}}(L_b(m) - R_b(m))| \quad (15)$$

4. Average Level found in 20 msec. Blocks- The terms $A_i(m)$, $i = 1, 2, 3, 4$ are averaged in 20 millisecond blocks.

5. Background Noise Level Estimate- Over the last 2 seconds, the minimum 20 millisecond block level, derived in step 4, is found for each of the 4 beam patterns. This value is averaged against previously found minima in previous 2 second intervals of time. The result is a somewhat biased estimate of the background noise level due to vents, fans, etc. A different background noise level estimate results for each of the 4 beam patterns.

6. Background Noise Level Subtraction- The background noise estimate is subtracted from the terms in step 3. If the result is less than zero the term is set to zero. For $i = 1, 2, 3, 4$ and N_i defined as the noise estimate for dipole pattern i,

$$B_i(m) = A_i(m) - N_i \quad (16)$$

under the condition that if $B_i(m) < 0$, then $B_i(m) = 0$. The purpose of this subtraction is to eliminate the influence of background noise on beam selection.

7. Short Term Integrator- The samples from step 6 are next fed to a short time integrator, to provide some smoothing of isolated peaks. For $i = 1, 2, 3, 4$,

$$C_i(m) = .25B_i(m) + .75C_i(m) \quad (17)$$

8. Running Peak - To mitigate the effects of reverberant energy on beam selection, a running peak value is developed for each beam. The philosophy is that the peak value of a signal will be proportional to the direct path energy while the decaying tails of the signal will have a larger portion due to reverberant energy. For $i = 1, 2, 3, 4$

$$\text{if } D_i(m) > C_i(m), D_i(m) = C_i(m) \tag{18}$$

$$\text{else } D_i(m) = .996 D_i(m) \tag{19}$$

9. Sum of Running Peak and Beam Selection - Over a 20 millisecond frame, the sum of the values of $D_i(m)$ for $i = 1, 2, 3, 4$ are found, and the index i that produces the largest sum is the dipole pattern which is chosen as maximizing the source pickup quality. Making decisions every 20 milliseconds has been found to lead to no noticeable degradation in performance. In fact, the beam selection algorithm has been found to yield high quality sound pickup even for the case of multiple people talking simultaneously.

3. DAISY CHAINING DESKTOP ARRAYS

It has been found by experiment that a single mic desktop array picks up people well in a 7 foot radius circle about the mic desktop array. Two mic desktop arrays may be used by simply adding the left channel of the first desktop array to the left channel of the second desktop array and adding the right channel of the first desktop array to the right channel of the second desktop array and then feeding the resultant summed left and right channel signals to the stereo A/D converter. Each desktop array will have a beam active. The beam selection algorithm for a single mic desktop array works well for multiple mic desktop arrays. The addition of a second mic desktop array increases the noise and reverberance by 3 dB because the added presence of a second beam. The effect of the 3 dB worsening of the signal-to-noise is to reduce the radius of coverage of each desktop array to 5 feet. Thus, the use of multiple desktop arrays does not increase the total area of coverage but merely serves to alter the shape of the area of coverage. In the case of one mic desktop array vs. two mic desktop arrays, the pickup area changes from a single circle of radius 7 feet to two circles of radius 5 feet. The area of the two smaller circles equals that of the single large circle.

4. ACOUSTIC ECHO CANCELLATION

The acoustic echo canceller duplicates the room transfer function between loudspeaker and microphone, filters the loudspeaker signal with this transfer function, and subtracts the result from the microphone signal. Via this procecedure, the component of the loudspeaker signal is eliminated from the microphone signal, with no effect on other components of the microphone signal. The loudspeaker-to-microphone transfer function changes drastically for different desktop array beams, so therefore, the loudspeaker-to-mic transfer function appropriate to the currently chosen desktop array beam must be used for echo cancellation. The acoustic echo canceller could store four sets loudspeaker-to-mic filter coefficients (which would have to be continually updated due to the changing nature of the acoustic paths). Alternatively, two echo cancellers, one for the left channel signal and one for the right channel signal of the stereo A/D converter, could be used, with the echo canceller outputs being summed or subtracted together to produce the four beam patterns. As yet another alternative, one echo canceller could be used with only two sets of loudspeaker-to-mic coefficients stored, those corresponding to the two base component dipole microphones. When needed, the two missing sets of loudspeaker-to-mic filter coefficients could be derived from these two stored sets by invoking the same operations needed to generate the two shifted dipole beams from the two component dipole beams, i.e., either add or subtract filter taps from the two sets of stored loudspeaker-to-mic filter coefficients, and then scale the resulting sum or difference by .7071 to derive the tap values for the missing loudspeaker-to-mic filter coefficient set. By examining the adaptive tap values found for the previous beam choice and the current (different) beam choice, one could easily derive the adaptive filter tap values for the two base component adaptive filters (two equations in two unknowns).

5. REFERENCES

[1] 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.

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

[3] 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.

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

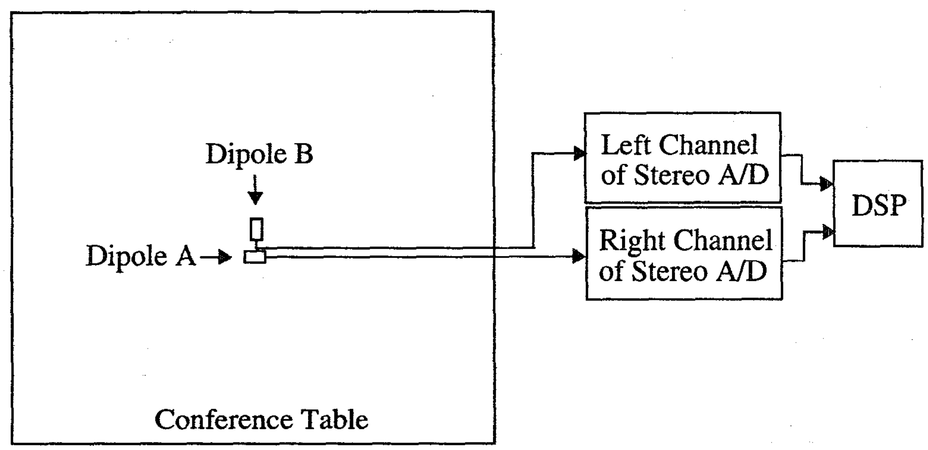

Figure 1. Schematic of the desktop mic array structure, overhead view, looking down onto the conference table.

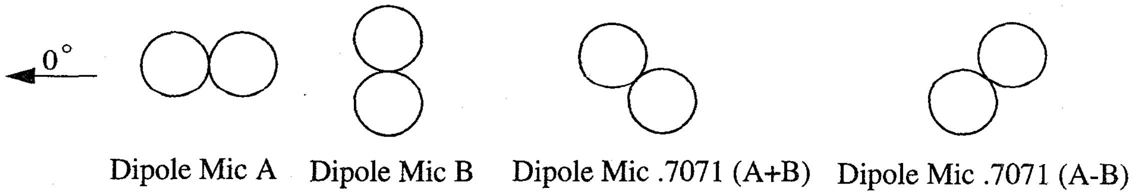

Figure 2. 4 dipole pickup patterns