ACOUSTIC LOCALIZATION BY INTERAURAL LEVEL DIFFERENCE

Stanley T. Birchfield and Rajitha Gangishetty

Dept. of Electrical and Computer Engineering, Clemson University, Clemson, SC 29634
{stb, rgangis}@clemson.edu

ABSTRACT

Interaural level difference (ILD) is an important cue for acoustic localization. Although its behavior has been studied extensively in natural systems, it remains an untapped resource for computer-based systems. We investigate the possibility of using ILD for acoustic localization, deriving constraints on the location of a sound source given the relative energy level of the signals received by two microphones. We then present an algorithm for computing the sound source location by combining likelihood functions, one for each microphone pair. Experimental results show that accurate acoustic localization can be achieved using ILD alone.

1. INTRODUCTION

When a sound is emitted in an environment, the signals received by the different microphones are generally shifted versions of one another, depending upon the relative locations of the source and receivers. Techniques for acoustic localization have, to date, relied exclusively upon this cue, known either as interaural time difference (ITD) or interaural phase difference (IPD). For example, the methods of time-delay estimation (TDE) [1, 2, 3], beamforming [4, 5, 6], hemisphere sampling [7], and accumulated correlation [8, 9] are different ways of utilizing the relative shifts in the signals received by the microphones to determine the location of the sound source. A significant amount of research has also been conducted to discover prefilters to make such computations robust to noise [10, 11, 12, 13].

Another important cue exists which has received little or no attention in the signal processing community. The signals received by the microphones not only differ in their relative time shift but also in their intensity level, with a microphone closer to the sound source receiving a higher-intensity signal than that received by a farther microphone. This cue, known as interaural level difference (ILD), forms the basis of the "intensity-difference theory" of directional hearing, which is the oldest theory of directional hearing going back over 100 years [14]. Although we now know, unlike the early proponents, that ILD is not the *only* cue

for acoustic localization, extensive psychoacoustic and psychophysical experiments have shown ILD to be an important cue used by the human localization system [14, 15]. Despite its importance in nature, however, including the localization systems of animals such as owls [16], no technique utilizing ILD has yet been proposed for computer-based systems.

In this paper we present a preliminary investigation into the possibility of using ILD for acoustic localization. We derive a model for computing the likelihood that the sound source is placed in a particular location using only the relative energies received by the microphones, without any information as to their relative phase. From this formulation, an algorithm is proposed to compute the sound source location using multiple microphones. Experiments demonstrate the algorithm's ability to localize accurately a sound source in a reverberant environment, and highlight some issues regarding ILD.

2. ILD FORMULATION

Suppose we have N microphones and a source signal $s(t)$ propagating through a generic free space with noise. According to the so-called inverse-square-law, the signal received by the ith microphone can be modeled as

$$x_i(t) = s(t)/d_i + \xi_i(t),$$

where d_i is the distance from the source to the ith microphone, and $\xi_i(t)$ is additive white Gaussian noise. To focus our attention upon the ILD cue, this formula ignores the relative time shift between the signals that is so important for ITD.

Let us assume that the sound source is audible and in a fixed location during the time interval $[0, W]$, where W is the window size. Then the energy received by the ith microphone can be obtained by integrating the square of the signal over this time interval:

$$
\begin{aligned}
E_i &= \int_0^W x_i^2(t)\,dt = \int_0^W [s^2(t)/d_i^2 + \xi_i^2(t)]\,dt \\
&= \frac{1}{d_i^2} \int_0^W s^2(t)\,dt + \int_0^W \xi_i^2(t)\,dt,
\end{aligned}
$$

because the integration of the cross-term is zero if $\xi_i(t)$ is uncorrelated and zero-mean. From this equation the name of the inverse-square-law is apparent: the received energy decreases as the inverse of the square of the distance to the source.

Given two microphones, the above equation leads to a simple relationship between the energies and distances:

$$E_1 d_1^2 = E_2 d_2^2 + \eta, \qquad (1)$$

where $\eta = \int_0^W [\xi_1^2(t) - \xi_2^2(t)] \, dt$ is a zero-mean random variable if the variance of $\xi_i(t)$ is constant.

Let (x_i, y_i) be the coordinates of the ith microphone, and let (x, y) be the coordinates of the sound source. To simplify the analysis we will assume a planar world throughout. Then $d_i^2 = (x - x_i)^2 + (y - y_i)^2$. Substituting this expression into Eq. (1) yields, after some algebraic manipulation, the following quadratic equation in x and y:

$$\begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} c_e & 0 & -c_x \\ 0 & c_e & -c_y \\ -c_x & -c_y & c \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \eta, \qquad (2)$$

where

$$\begin{aligned}
c_e &= E_1 - E_2 \\
c_x &= E_1 x_1 - E_2 x_2 \\
c_y &= E_1 y_1 - E_2 y_2 \\
c &= E_1(x_1^2 + y_1^2) - E_2(x_2^2 + y_2^2).
\end{aligned}$$

With $\eta = 0$, this equation describes the locus of points where a source emitting a sound will cause the two microphones to receive signals with energies of E_1 and E_2, respectively. This equation holds regardless of the overall energy of the original signal, which can be seen by dividing the entire equation by E_2 to obtain an equivalent expression only in terms of the energy ratio $\Delta_E = E_1 / E_2$.

Homogeneous coordinates are used in Eq. (2) to capture all possible cases with a single expression. One such case occurs when the received energies are not identical, i.e., $E_1 \neq E_2$, in which the equation can be written in a more familiar form

$$\left(x - \frac{c_x}{c_e} \right)^2 + \left(y - \frac{c_y}{c_e} \right)^2 = \frac{E_1 E_2 d_{12}^2}{c_e^2} + \eta',$$

where $d_{12} = (x_1 - x_2)^2 + (y_1 - y_2)^2$ is the squared distance between the two microphones, and $\eta' = \eta / c_e$. According to this expression the sound source is constrained to lie on a circle centered at $(c_x/c_e, c_y/c_e)$ with a radius of $d_{12}\sqrt{E_1 E_2}/c_e$, ignoring noise. In 3D, of course, the circle becomes a sphere.

Another case arises when $E_1 = E_2$, in which the equation reduces to
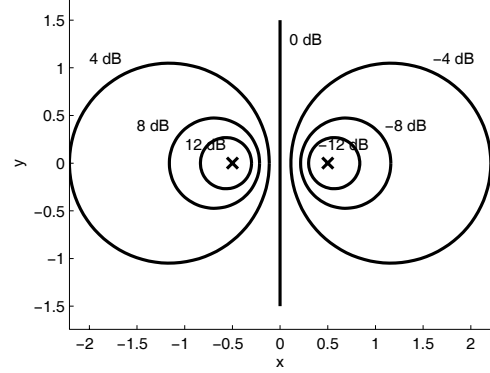
$$2c_x x + 2c_y y = c + \eta,$$

Fig. 1. Isocontours of Eq. (2) for different values of $10 \log \Delta_E$. The sound source lies on a circle (sphere) unless the two energies are equal, in which case it lies on a line (plane, the *mid-sagittal plane*) between the microphones. Here microphones 1 and 2 are located at $(-0.5, 0)$ and $(0.5, 0)$, respectively.

which is the equation of the line passing halfway between the microphones and perpendicular to the line joining them (i.e., the perpendicular bisector). In 3D, the line becomes a plane.

This line for $E_1 = E_2$, and the circles for $E_1 \neq E_2$, are evident in the isocontours of the quadratic equation displayed in Figure 1. The shape of these isocontours correspond qualitatively with those measured in the ILD localization system of owls (see the figures in [16]).

3. ILD LOCALIZATION

As seen in the previous section, with only two microphones ILD is not able to pinpoint the sound source location. Instead the source is constrained to lie on a curve (or surface in 3D), similar to the "cone of confusion" [15]. Ignoring noise, all sources emanating from a point on this curve yield an identical interaural level difference.

Our approach to solving this ambiguity is to employ multiple microphone pairs. Each microphone pair determines a different curve in the environment, so the intersection of these curves yields the sound source location. However, instead of computing this intersection directly using a closed form or least squares solution, we choose instead to use probabilistic sampling. That is, we select a number of candidate locations in the space and, for each of these locations, compute the likelihood that the sound source is located there. This total likelihood is computed as the sum of the likelihoods using each microphone pair. Assuming that the microphone pairs yield independent measurements, this technique is equivalent to computing the joint probability by multiplying the individual probabilities using the
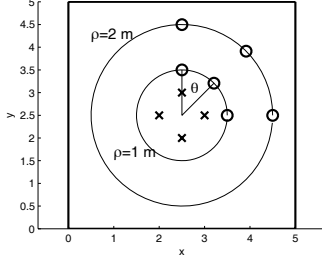
Fig. 2. The simulated room, with four microphones (×) and six sound source locations (○).

sum of log likelihoods. This simple approach to sensor fusion has been used successfully in ITD acoustic localization [8].

An issue remains to be solved, namely computing the likelihood at an arbitrary candidate location given a curve (circle or line) computed for a microphone pair. We solve this problem by noting that, for any given candidate location $(\tilde{x}, \tilde{y})$, the expected value for Δ_E can be calculated as the ratio of the squares of the distances to the two microphones:

$$\tilde{\Delta}_E = \frac{(\tilde{x} - x_2)^2 + (\tilde{y} - y_2)^2}{(\tilde{x} - x_1)^2 + (\tilde{y} - y_1)^2}.$$

This result is obtained by substituting $(\tilde{x}, \tilde{y})$ for (x, y) in Eq. (2), setting $\eta = 0$, and solving for Δ_E. Using this expression, $\tilde{\Delta}_E$ for all the candidate locations is computed once off-line. Then, at run time, the likelihood that the sound source is at a candidate location is computed by treating $10 \log \Delta_E$ as a Gaussian random variable with mean $10 \log \tilde{\Delta}_E$ and variance σ_e^2.

4. EXPERIMENTAL RESULTS

We tested the ILD algorithm in a 5 m × 5 m simulated room, with four microphones arranged in a square so that opposing microphones were separated by 1 m, as shown in Figure 2. A sound file of a male voice counting from one to ten (16-bit, 44.1 kHz) was played at a location in the room and captured by the microphones, using the image method [17] with linear interpolation between samples and up to sixth order reflections for the four walls. For these experiments, the entire 2.5-second utterance was treated as a single audio frame.

We calculated the error $|10 \log \Delta_E - 10 \log \tilde{\Delta}_E| = 10 |\log(\Delta_E / \tilde{\Delta}_E)|$ using the two horizontal microphones for different source locations and values of the reflection coefficient β. The results are shown in Figure 3. The accuracy of Δ_E estimation is highly dependent upon the sound source location and the amount of reverberation. At higher reverberations, or in positions where the reverberations are
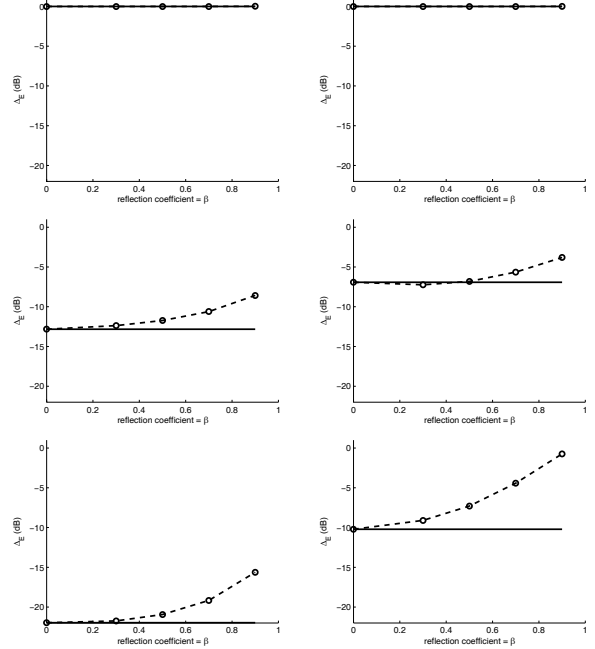

Fig. 3. The results of Δ_E estimation using the two horizontal microphones for the six sound source locations. From top to bottom, $\theta = 0, 45, 90$ degrees; from left to right $\rho = 1, 2$ m. The solid line is ground truth.

asymmetric, the accuracy of the estimation decreases significantly. Not only does the error increase, but there appears to be a systematic bias in the estimation.

Nevertheless, the ILD algorithm is able to localize a sound source accurately, as seen in Figure 4. For this experiment we placed the sound source in one of the more challenging positions and set reverberation to the maximum value of the previous tests ($\theta = 45$ degrees, $\rho = 2$ m, $\beta = 0.9$). The variance σ_e^2 was computed from the error in Figure 3 to be 9.6. Shown in the figure are the contour lines and mesh plots of the individual and combined likelihood functions. The algorithm computes the bearing angle to the sound source with 0.0 degree error. However, the algorithm exhibits a bias toward locations far from the microphones and hence is unable to estimate distance accurately.

5. CONCLUSION

Interaural level difference (ILD) is an important cue for acoustic localization in natural systems. We have investigated the possibility of using ILD in computer-based systems. Equations were derived that constrain the location of a sound source based upon received energy levels of two microphones, and an algorithm for computing the location using multiple microphone pairs has been proposed. Experiments in reverberant environments demonstrate the algorithm's ability to
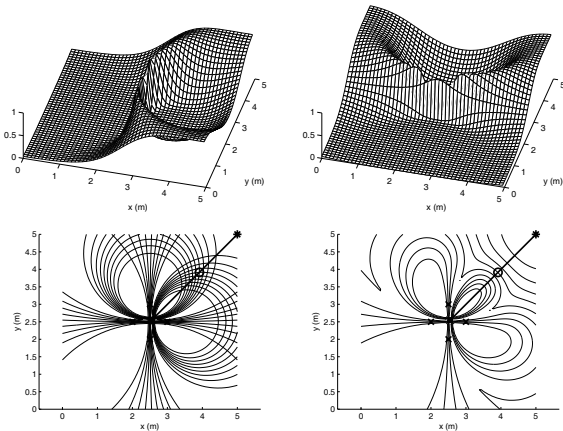
Fig. 4. TOP: The likelihood function computed by the horizontal (left) and vertical (right) microphone pairs. BOTTOM: Contour plots of the two likelihood functions overlayed (left), and contour plot of combined likelihood function (right). Also shown are the microphones (x), the true sound source location (○), the peak of the combined function (⋆), and the computed bearing angle to the peak (solid line).

compute exact bearing angle in a reverberant environment, thus validating the utility of the cue. This paper, however, only scratches the surface of possibilities with ILD, leaving many questions unanswered, such as the bias of the algorithm toward distant locations.

6. REFERENCES

[1] Michael S. Brandstein and Harvey F. Silverman, "Practical methodology for speech source localization with microphone arrays," *Computer Speech and Language*, vol. 11, no. 2, pp. 91–126, 1997.

[2] Piergiorgio Svaizer, Marco Matassoni, and Maurizio Omologo, "Acoustic source location in a three-dimensional space using crosspower spectrum phase," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1997, vol. 1, pp. 231–234.

[3] J. Vermaak and A. Blake, "Nonlinear filtering for speaker tracking in noisy and reverberant environments," in *ICASSP*, 2001.

[4] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *Journal of the Acoustical Society of America*, vol. 78, no. 5, 1985.

[5] R. Duraiswami, D. Zotkin, and L. Davis, "Active speech source localization by a dual coarse-to-fine search," in *ICASSP*, 2001.

[6] Darren B. Ward and Robert C. Williamson, "Particle filter beamforming for acoustic source localization in a reverberant environment," in *ICASSP*, 2002.

[7] Stanley T. Birchfield and Daniel K. Gillmor, "Acoustic source direction by hemisphere sampling," in *ICASSP*, 2001.

[8] Stanley T. Birchfield and Daniel K. Gillmor, "Fast Bayesian acoustic localization," in *ICASSP*, 2002.

[9] Stanley T. Birchfield, "A unifying framework for acoustic localization," in *Proceedings of the 12th European Signal Processing Conference (EUSIPCO)*, 2004.

[10] Charles H. Knapp and G. Clifford Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[11] Michael S. Brandstein, "Time-delay estimation of reverberated speech exploiting harmonic structure," *Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2914–2919, 1999.

[12] Maurizio Omologo and Piergiorgio Svaizer, "Use of the crosspower-spectrum phase in acoustic event location," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 3, 1997.

[13] Alex Stéphenne and Benoît Champagne, "Cepstral prefiltering for time delay estimation in reverberant environments," in *ICASSP*, 1995, vol. 5, pp. 3055–3058.

[14] Jens Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, Cambridge, Massachusetts: The MIT Press, 2001.

[15] William A. Yost, *Fundamentals of Hearing: An Introduction*, San Diego: Academic Press, 2000.

[16] M. S. Brainard, Eric I. Knudsen, and Steven D. Esterly, "Neural derivation of sound source location: Resolution of spatial ambiguities in binaural cues," *Journal of the Acoustical Society of America*, vol. 91, no. 2, pp. 1015–1027, 1992.

[17] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," *Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, 1979.