Proceedings of the 42nd IEEE
Conference on Decision and Control
Maui, Hawaii USA, December 2003

A Biomimetic Apparatus for Sound-source Localization

Amir A. Handzel, Sean B. Andersson, Martha Gebremichael and P.S. Krishnaprasad
Institute for Systems Research and Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742, USA
{handzel,sanderss,krishna}@isr.umd.edu

***Abstract*— Sound source localization systems are typically made of free field microphone arrays. We recently proposed the design of a localizing system based on the principles observed in nature, where directional acoustic sensing evolved to rely on diffraction about the head with only two ears. Localization is performed using the resultant frequency dependent acoustic phase and intensity differences between the two ears. These interaural functions can be computed analytically by modeling the head as a sphere. Now we report the first successful implementation of these ideas in an artificial head with two antipodally placed microphones. In order to determine source direction, we define a suitable metric associated to interaural functions. The true source direction is given by the global minimum of the metric between a measurement and a table of theoretical functions. The system was tested with a broadband source and was found to perform well even in a non-ideal reverberant laboratory environment. For comparison, we also calculated the source direction using the standard cross-correlation algorithm which is based on interaural time delay. Our approach performs more accurately and allows to use briefer signal durations. A key motivation for this work is to devise effective means to guide robotic navigation in environments with acoustic sources. The apparatus described in this paper is installed on a mobile robot in the laboratory.**

I. Introduction

Directional sensing of sound enables the localization of its source in space. More broadly, it can aid in the separation of signals from multiple sources and in their identification. Designed systems typically comprise free field sensor arrays for extraction of directional information. Most use differences in time of arrival between combinations of pairs of microphones. Applications include the localization and tracking of speakers in conference rooms and improved hearing aids having directional sensitivity; see [3] for a comprehensive overview. Several groups installed free field microphone rigs on mobile robots to endow them with localizing capability [4], [9], which also use differences in time of arrival between microphone pairs. A recent robotic localizing device was based on a pair of free field microphones which were rendered directional by means of reflectors placed around the microphones [8]. This augmented the usual time difference information with intensity difference.

In nature, directional acoustic sensing evolved to rely on diffraction about the head with only two sensors — the ears. The impinging sound waves are modified by the head in a frequency and direction dependent way. Additional complex filtering is performed by the external ears (*pinnae*). The cochlea decomposes the sound pressure signal into frequency bands. The brain then uses interaural differences in phase (IPD) and intensity level (ILD) in the various frequency bands to infer the location of a source [1], [6].

Inspired by human sound localization, we recently suggested building artifical systems operating by similar principles [5]. An artificial head with microphones placed on its surface would provide the required directional acoustic information. The sound pressure at each microphone can be computed analytically by modelling the head as a sphere [2]. A suitable algorithm would then be used to extract the direction of the source from the received sound. Here we report a first successful implementation of such a system for localization in the horizontal plane. In Section II we briefly review the derivation of the pressure function measured on the surface of a scattering sphere, followed by Section III in which we prescribe the computations to extract the direction of a source from a pair of sound signals. Section IV contains a description of the experimental set-up. Results are presented in Section V followed by a brief conclusion.

II. Spherical scattering

The acoustic pressure can be described by small fluctuations in a scalar field called the velocity potential Π, such that $\mathbf{v} = -\nabla\Pi$ where $\mathbf{v}$ is the velocity of a point mass of the fluid medium. Π is a scalar function that serves as a dynamic variable from which all relevant fluctuating quantities are derived, including pressure: $p = \rho\partial_t\Pi$, where ρ is the mass density of air. The velocity potential is governed by the wave equation:

$$0 = \frac{1}{c^2}\frac{\partial^2\Pi}{\partial t^2} - \triangle\Pi \tag{1}$$

where $\triangle$ is the Laplacian and $c = 344$m/s is the speed of sound.

A solution to the wave equation separates in time and space, i.e. $\Pi = T(t)\psi(\mathbf{r})$, giving time harmonic behaviour, $T(t) = e^{-i\omega t}$ while the spatial variables satisfy the Helmholtz Equation:

$$\triangle\psi + k^2\psi = 0, \tag{2}$$

where $k = \omega/c$ is the wave number corresponding to angular frequency ω. A general solution of (2) is expressed through separation of variables in spherical coordinates, such that each variable is governed by an ODE: $\psi(r,\theta,\phi) = R(r)\Theta(\theta)\Phi(\phi)$. The coordinate system is determined naturally by the geometry of the apparatus. We designate a

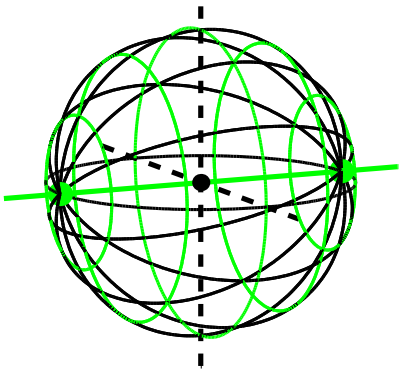

Fig. 1. Coordinates of spherical head. The microphones (grey) are located at the two poles on the horizontal plane The polar angle θ is measured along meridians (black) connecting the poles; the elevation (i.e. azimuthal) angle ϕ is measured along latitude circles (grey) which are lines of constant θ.

"north" pole as the position of the left microphone, and the antipodally positioned right one is assigned the "south" pole; see Figure 1. The *polar* angle (θ) is measured along meridians between the north and south poles, and is constant along circles of latitude relative to the two poles. The elevation angle ϕ marks the position on latitude circles, in other words, it marks the meridian between the poles relative to the horizontal. For the present problem we set the latitude of the equator to $\theta = 0^{\circ}$, so the microphones are located at $\theta = \pm 90^{\circ}$.

The impinging wave, Π^0, and scattered wave, Π^s, are expanded in spherical harmonics and the position of the source and measurement points. Thanks to the orthonormality of spherical harmonics, the boundary conditions can be imposed independently for each component [2]. For an acoustically hard sphere, the normal velocity of the total wave vanishes on its surface (Neumann conditions):

$$\frac{\partial}{\partial \mathbf{n}}(\Pi^0 + \Pi^s)|_d = 0 \tag{3}$$

where d is the radius of the sphere. Taking the sum of impinging and scattered waves as an expansion series, the total velocity potential is obtained for a general measurement point $\mathbf{r} = (r, \theta, \phi)$ in space. The field does not depend on the separate absolute angular positions of the source and sensor but on the angle between them. Then if the source lies at position $\mathbf{r_0} = (r_0, 0, 0)$, the spatial part of the potential is

$$\psi_{r=d} = C \sum_n (2n+1) P_n(\cos\theta) \frac{h_n(kr_0)}{h_n'(kd)} \tag{4}$$

where C is an aggregate coefficient, P_n are the Legendre functions, and h_n are the spherical Hänkel functions; prime denotes derivation with respect to the function's argument. The nature of the solution is such, that except for the near-field (i.e. very close to the head), the pressure on the head is insensitive to the source distance r_0. We therefore limit ourselves to finding the *direction* to the source.

The symmetry of the problem governs its solution and properties. The scattering configuration has cylindrical symmetry, because the total pressure depends only on the polar angle. The same pressure will, therefore, be measured on each circle of latitude (Fig. 1). The sound source, the pair of antipodal sensors and the center of the sphere, always lie on a plane whose orientation in space is determined by the angle ϕ. *The present work is restricted to localization in the horizontal plane*. Each circle of latitude intersects the plane at two points: one in the front half-plane and the other in the back, and we encounter only this discrete ambiguity ("front-back confusion"). The localization algorithms determine source direction up-to front or back hemisphere.

III. Localization algorithm

A. The IPD-ILD algorithm

The measured sound pressure is a complex response to the excitation by a source:

$$p = Ae^{i\alpha - i\omega t} \tag{5}$$

where α is the part of the phase containing spatial information. With pressure measured at the right (R) and left (L) microphones, we define the Interaural Level Difference and Interaural Phase Difference:

$$\mathrm{ILD} = \log A_L - \log A_R \qquad \mathrm{IPD} = \alpha_L - \alpha_R. \tag{6}$$

The solution of the wave equation involves separation to independent frequency components. Consequently, both ILD(ω) and IPD(ω) are functions of frequency ω. We consider the ILD-IPD plane as a basic feature space in which localization is performed. For every source direction and frequency there is a point in the ILD-IPD plane. Since ILD and IPD depend smoothly on frequency, every broadband sound source generates a whole curve $\sigma(\omega)$ in this plane which is its specific *signature* depending on the source location in space.

The picture can be summarized as follows: A source at position $\mathbf{r_0}$ emits sound which is mapped through the scattering process, S, to a pair of sound pressure measurements, i.e. a pair of smooth complex functions of some frequency interval Ω. Extracting the binaural, i.e. relative, phase and intensity, reduces them to a pair of Real functions:

$$\begin{array}{ccccc} \mathbb{R}^3 & \xrightarrow{S} & C_{\mathbb{C}}(\Omega) \times C_{\mathbb{C}}(\Omega) & \xrightarrow{I} & C_{\mathbb{R}}(\Omega) \times C_{\mathbb{R}}(\Omega) \\ \mathbf{r_0} & \overset{S}{\longmapsto} & (p_L, p_R) & \overset{I}{\longmapsto} & (\mathrm{ILD}, \mathrm{IPD}) \end{array} \tag{7}$$

The task is to prescribe a localization operator that would, in effect, invert the above to recover the source direction, up to the intrinsic symmetry of the problem. We do so by defining the squared L^2 norm distance between the measured interaural functions (ILD(ω),IPD(ω)) and the theoretical functions ($\mathrm{IPD}(\theta, \omega), \mathrm{ILD}(\theta, \omega)$). The theoretical functions are stored in a table. After suitable normalization of the metric, we

pick the angle whose interaural functions are closest to those measured. The metric for IPD is:

$$\begin{aligned}\mathbf{D}_2^{\mathrm{IPD}}(\theta) &\equiv \| \operatorname{IPD}(\theta,\omega) - \operatorname{IPD}(\omega) \|_2^2 \\ &= \textstyle\sum_\omega \left(\operatorname{IPD}(\theta,\omega) - \operatorname{IPD}(\omega)\right)^2\end{aligned} \tag{8}$$

and similarly for ILD. We normalize each of the two metrics with respect to its maximal value (over θ):

$$\mathbf{D} \longrightarrow \frac{1}{M}\mathbf{D} \qquad M = \max_{(\theta)} \mathbf{D}(\theta), \tag{9}$$

and combine the normalized metrics for ILD and IPD to produce a combined distance function:

$$\mathbf{D}_2^{\mathrm{Comb}} = \mathbf{D}_2^{\mathrm{IPD}} + \mathbf{D}_2^{\mathrm{ILD}}. \tag{10}$$

The chosen source direction is the one for which the combined metric is at minimum.

In [5], a modification of the metric (10) is given, based on the *variation* of IPD and ILD under head movements. This modification contains information to determine the angle ϕ, i.e. elevation of a source *off* the horizontal plane. Experimental results based on this idea will be presented in a future publication.

B. The ITD algorithm

The standard algorithm in use is based on the interaural time delay (ITD) to determine the angle to the sound source. The source is assumed to be far enough so that the impinging wavefront is planar. The algorithm relies on a common ray approximation [7] of the propagating wave and its path difference to the two microphones as depicted in Figure 2. For a source at polar angle θ, after the sound impinges upon

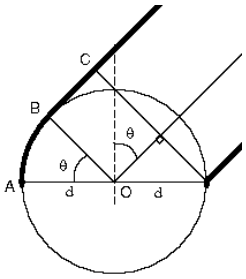

Fig. 2. Interaural time delay

the first microphone it must travel along line segment BC and then along arc $\angle AOB$ before arriving at the second microphone. This extra path, a, is

$$a = d\sin\theta + d\theta \tag{11}$$

where d is the radius of the sphere. If the path difference and the head radius are known then the angle to the source can be determined by solving equation (11) for θ. As in the IPD-ILD algorithm, there is an inherent front-to-back symmetry and the algorithm can only localize up to this symmetry.

To determine the path difference a, we measure the ITD as follows. Let $l[k]$ and $r[k]$ denote the sequences of N digitized samples of the sound pressure levels at the left and right microphones respectively. The cross-correlation between the two sequences is defined by

$$\mathrm{C}_{lr}[m] = \frac{1}{N}\sum_{n=0}^{N-|m|-1} l[n]\, r[n+m] \tag{12}$$

with m ranging from $-m_{max}$ to m_{max} where m_{max} corresponds to the extra distance traveled by the sound when the sound source is at $\theta = 90^{\mathrm{o}}$. This is given by

$$m_{max} = \text{floor}\left(\frac{f_s d}{c}\left[\frac{\pi}{2}+1\right]\right) \tag{13}$$

Here f_s is the sample rate and floor$(\cdot)$ is a function returning the closest integer no larger than the argument. Define m^* by

$$m^* = \arg\max_m \mathrm{C}_{lr}[m]. \tag{14}$$

This in turn yields the ITD from

$$\mathrm{ITD} = \frac{m^*}{f_s} \tag{15}$$

Finally we calculate the distance a as $a = c\,\mathrm{ITD}$ and retrieve θ from equation (11).

IV. Experiment and data analysis

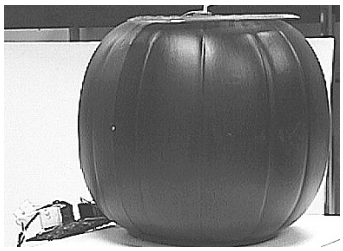

Fig. 3. Experimental setup: side view

The experiments were performed in our laboratory, a reverberant environment with various low-level background noise sources, primarily computer fans. A pair of Knowles FG-3329 microphones were mounted antipodally on an approximately spherical plastic "head" (Figure 3) with flat top and bottom. Its minimum radius, between top and bottom, is 9.5 cm; its maximum, in the horizontal plane, is 11.3

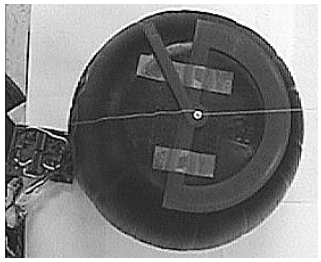

Fig. 4. Experimental setup: top view

cm. The head was filled with styrofoam to dampen internal resonances. A Cambridge Soundworks multimedia speaker, model SPS52, was placed along a semi-circular arc of fixed radius from the center of the head. The experiment was conducted at positions over the full hemispheric range of directions, from -90° to 90°, at 2.5° increments, measured with a protractor attached to the top of the head (Figure 4).

A broadband signal $y(t)$ comprising a sum of pure tones at every 43Hz from 43 Hz to 11 kHz was presented through the speaker. The amplitude was chosen in each case so as to make the output as large as possible without saturating the electronics. The output of each microphone was anti-aliased by passing it through a sixth-order low-pass Bessel filter with a cutoff frequency of approximately 11 kHz. This signal was then run through amplifiers with gains selected to equalize the background levels on the two channels. The outputs of the amplifiers were connected to an analog-to-digital (A/D) converter and the signals sampled at $f_s = 22,095$ Hz with 16-bit resolution. The A/D converter used was a built-in sound card on an Advantech PCM-5862 single board computer. The converter alternated between sampling the left and right channels, introducing a frequency dependent phase difference between the two channels given by

$$\Delta\phi_{sampling}(f) = \frac{2\pi f}{f_s} \tag{16}$$

where f_s is the sample rate. Ten 0.5s samples were recorded for each source direction.

To determine the IPD and ILD, each 512 point (23 ms) portion of the left and right channel data was passed through a fast Fourier transform, giving the amplitudes and phases of both signals. The IPD and ILD at each frequency were calculated (6) and the IPD corrected using equation (16). The IPD and ILD curves were smoothed using a nine-point moving average to produce the measured signature curve.

Theoretical signatures were calculated for all source locations. Since the head is not precisely spherical, an effective radius was determined by matching the theoretical IPD function to recorded data of one source direction. It was found to equal the horizontal head radius of 11.3cm. When calculating the distance between a measured and theoretical curve using equation (10), we included only those frequencies whose amplitudes exceeded the background noise threshold as determined by experiment.

For the ITD algorithm the same 23ms portions of the left and right channel data were cross-correlated using equation (12), m^* found from equation (14), and the ITD determined using equation (15). As shown in the Results Section, in order to improve the performance we repeated the ITD calculation for time segments of 115ms. For the calculation of the direction to the sound source, an effective head radius of 12cm was fit to the data. Due to the discretization of the signal at the given sampling rate, the ITD algorithm is able to localize only to a discrete set of directions. Moreover, because of the $\sin\theta$ factor in (11), this discretization is not linear, having smaller increments around the front direction ($\theta = 0^{\circ}$) and increasing to nearly 7° at lateral directions.

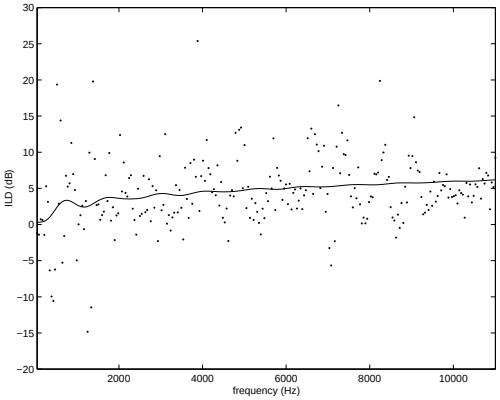

Fig. 5. ILD for source at $\theta = 17.5^{\circ}$

V. Results

First, we compare the experimental data to the theoretical curves. In Figure 5 we show a scatter plot of the measured ILD for one 23 ms sample from a source direction of 17.5° together with the theoretical curve. The measured ILD is widely scattered but shares the same trend as the theoretical curve. In Figure 6 we show the effect of the smoothing operation. In Figure 7 we show the measured IPD and the theoretical curve for the same data set. The smoothed version (Figure 8) is quite close to the theoretical curve.

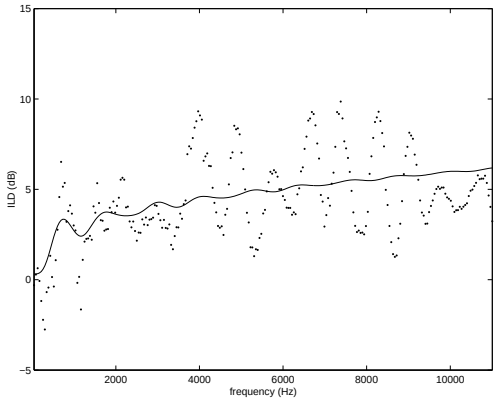

Fig. 6. Smoothed ILD for source at $\theta = 17.5^{\circ}$

The distance functions calculated from these data is remarkably sharp. In Figure 9 we show the metric for two adjacent sources at 15.0° and 17.5°. For each source location, there is a deep and sharp valley and a clear minimum. Most importantly, the two sources are clearly distinguishable from one another on both curves. The picture becomes somewhat less sharp as we move away from the centerline. In Figure 10 we show the distance functions for sources located at 72.5°

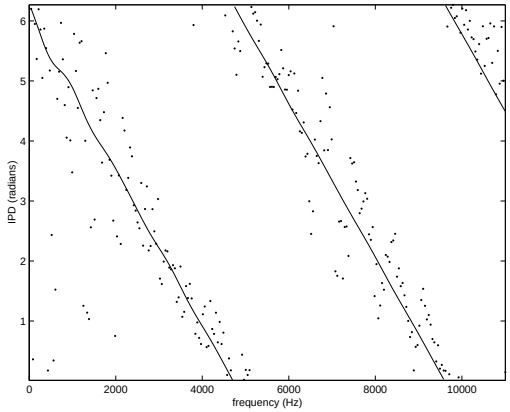

Fig. 7. IPD for source at $\theta = 17.5^{\circ}$

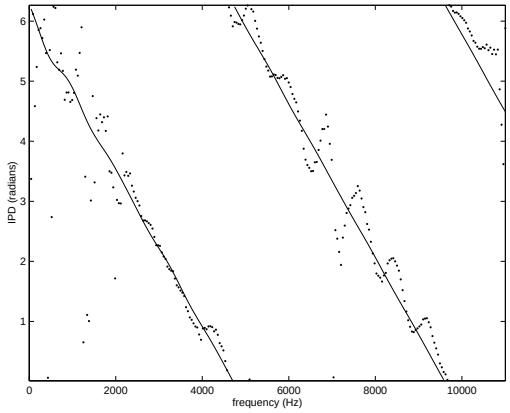

Fig. 8. Smoothed IPD for source at $\theta = 17.5^{\circ}$

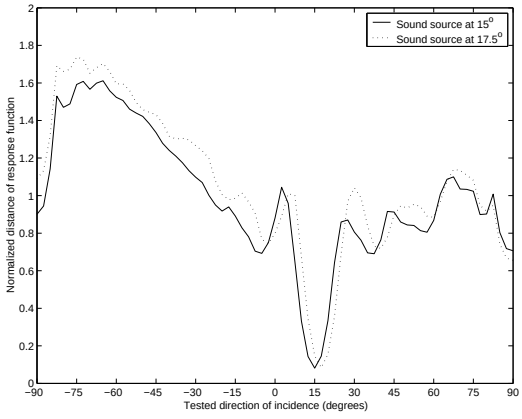

Fig. 9. Distance functions for source at 15° and 17.5°

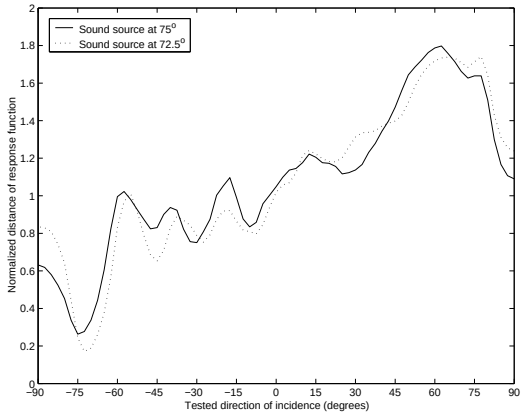

Fig. 10. Distance functions for source at 72.5° and 75°

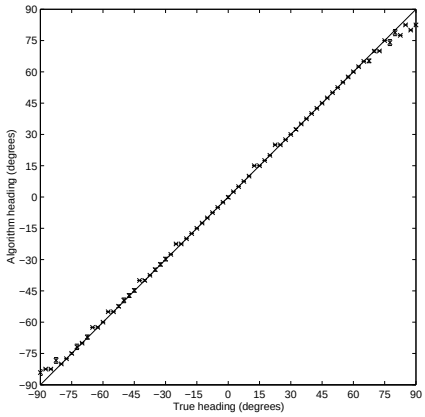

Fig. 11. Localization performance of the IPD/ILD algorithm

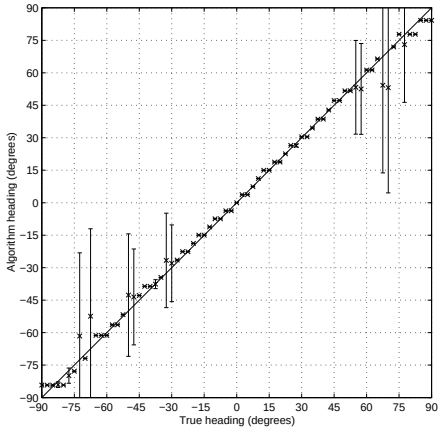

Fig. 12. Localization performance of the ITD algorithm

and 75°. Although the valleys are not as sharp as for more "frontal" source pairs, the two are still distinguishable. From these metric functions we expect high localization acuity near centerline and with less accuracy as we move laterally.

In order to obtain a statistically meaningful picture of the localization performance, we picked 100 data sets of 23ms at each source angle and applied the IPD-ILD algorithm to each. Figure 11 shows the sample mean and variance of localization as a function of the true source direction over the complete range of angles. The variance of the algorithm output is so small that it is either zero or barely noticable. The performance is excellent, except for a slight bias towards the

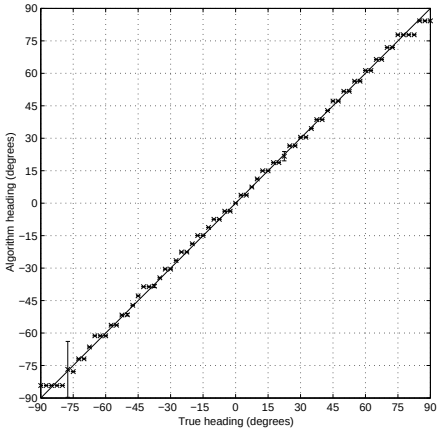

Fig. 13. Localization performance of the ITD algorithm, extended samples)

front for sources in the lateral directions. A possible cause is the existence of the "bright spot" — pronounced constructive diffraction in the front scatter direction — which substantially diminishes IPD realtive to other source directions.

We now turn to the standard ITD algorithm. In Figure 12 we show the sample means and variances for the estimates from the same data as for ILD-IPD. The ITD algorithm is also quite accurate out to $\pm 30^{\circ}$. Interestingly, for source directions more lateral than $\pm 30^{\circ}$ the algorithm occasionally produces estimates that are vastly wrong. These explain the large variance seen for some source directions. As cross-correlation is sensitive to signal length, we repeated the procedure for signals five times longer than the original 23 ms. In Figure 13 we show the results based on these 116ms (2560 points) samples. With the longer sample length, the ITD algorithm performs well, no longer producing wildly inaccurate estimates. It is accurate out to $\pm 75^{\circ}$, where the inherent "graininess" prevents better performance.

VI. Concluding remarks

We have presented experimental results for a novel sound source localization algorithm based on measuring the IPD and ILD between a pair of microphones mounted on the surface of an acoustically hard sphere. These experiments show that for broadband signals, the method is remarkably accurate over the entire range of source directions. We note that the performance was not degraded by the fact that we relied on a theoretical model of a spherical head, even though the real artificial head was only approximately spherical.

In calibrating the apparatus, we computed the effective radius for the spherical model, based on a small data set and found it to be equal to the actual radius in the horizontal plane. This indicates that the scattering around this plane was dominant in effecting the measured interaural functions.

We compared this biomimetic algorithm with the standard cross-correlation algorithm, which relies on ITD, and showed that under the same conditions the IPD-ILD algorithm outperforms ITD. However, at the cost of using longer data samples, the ITD algorithm can perform nearly as well.

The IPD-ILD algorithm is also beneficial in terms of the discretization of the space of source directions. Using the ITD algorithm, this discretization is forced by the choice of sample rate and it is not uniform along the various directions. In applications which allow one to do so, angular acuity can be increased by choosing a larger head or increasing the sample rate. The IPD-ILD algorithm, in contrast, is not limited by discretization precision; one need only calculate the corresponding theoretical signatures.

Initial experiments in using the localization algorithm of this paper for robot navigation have been successful.

VII. Acknowledgements

This research was supported in part by the NSF Learning and Intelligent Systems Initiative Grant CMS9720334, by the ARO ODDR&E MURI01 Program Grant No. DAAD19-01-1-0465 to the Center for Communicating Networked Control Systems (through Boston University) and by the ONR ODDR&E MURI97 Program Grant No. N000149710501EE to the Center for Auditory and Acoustics Research. S. Andersson was also supported by a fellowship from the ARCS Foundation.

VIII. References

[1] Jens Blauert. *Spatial Hearing*. MIT Press, revised edition, 1997.

[2] J.J. Bowman, T.B.A. Senior, and P.L.E. Uslenghi. *Electromagnetic and Acoustic Scattering by Simple Shapes*. North-Holland, 1969.

[3] Michael Brandstein and Darren Ward, editors. *Microphone Arrays*. Digital Signal Processing. Springer-Verlag, 2002.

[4] Huang Jie *et al.* A model-based sound localization system and its application to robot navigation. *Robotics and Autonomous Systems*, 27(4):199–209, 1999.

[5] Amir A. Handzel and P.S. Krishnaprasad. Biomimetic sound-source localization. *IEEE Sensors Journal*, 2(6):607–616, 2002.

[6] William M. Hartmann. How we localize sound. *Physics Today*, pages 24–29, November 1999.

[7] George F. Kuhn. Model for the interaural time difference in the azimuthal plane. *J. Acoustical Society America*, 62(1):157–167, July 1977.

[8] Lorenzo Natale, Giorgio Metta, and Giulio Sandini. Development of auditory-evoked reflexes: Visuo-acoustic cues integration in a binocular head. *Robotics and Autonomous Systems*, 39:87–106, 2002.

[9] Juyang Weng and Kamen Y. Guentchev. Three-dimensional sound localization from a compact non-coplanar array of microphones using tree-based learning. *J. Acoustical Society America*, 110(1):310–323, July 2001.