# Robotic Sound-Source Localization and Tracking Using Interaural Time Difference and Cross-Correlation

John C. Murray, Harry Erwin and Stefan Wermter
Center for Hybrid Intelligent Systems
University of Sunderland,
Sunderland, SR6 0DD.
Email: {john.murray, harry.erwin, stefan.wermter}@sunderland.ac.uk

***Abstract*-Interaural Time Difference (ITD) is used in the mammalian auditory system to compute the angle of incidence of an acoustic sound-source on the horizontal plane. This paper describes how ITD can be incorporated into a robotic acoustic tracking system to enable the robot to locate and orient towards sound-sources within its environment. We describe a system compiled using cross-correlation and auditory cues that has a lot of potential for robot sound-source localization.**

***Keywords: Sound-Source Localization; Biologically Inspired Acoustics; Tracking; Bioinspired Robotics.***

## I. INTRODUCTION

With the increasing developments in robotics, robots are becoming more common in everyday surroundings [1]. However for humans to better interact with robots they have to be able to communicate in the same way and to be accepted by humans robots have to become more sociable [2, 3]. It is therefore evident that acoustics plays a large role in robotics of today and the future.

Tour guide robots are one example of such 'sociable' robots. These tour guides are able to move around their environment whilst avoiding obstacles and they also have the ability to interact with the people they encounter by answering questions and also taking instructions. One example of such a tour guide robot is PERSES which is being developed by Böhme et al. [1]. Sound-source localization is an important task for such robots in order to improve their speech understanding capabilities eventually.

Researchers are drawing on many different areas in acoustics ranging from engineering to biological systems. Previously robotic navigation uses mainly range sensors (such as sonar) and tracking has relied predominately on vision [16]. This modality is widely used as a means for locating objects within the scene; however, as with humans and most animals, our field of view is restricted to less than 180$^o$ as our eyes point forward. This restriction can be overcome in vision with the use of a conical mirror [12].

In the real world we also use our hearing which gives us a full 360$^o$ 'field of view'. This allows us to locate objects that may not be in our field of vision, i.e. obscured by objects or even located around a corner [8]. There are currently several acoustic tracking robots that have been developed. These acoustic robots however, differ in the principles they utilize in order to localize sound-sources. Many approaches use arrays of four or more microphones [14, 15, 8] and engineering principles including specialized digital signal processors (DSP). This paper describes our approach to the task of robotic acoustic tracking. Our aim is to develop the system as close to the mammalian acoustic system as possible, i.e. using two microphones to represent the human ears and taking advantage of the cues and principles that exist within the mammalian auditory cortex (AC) [13] in order to locate and track a sound-source within the robots environment.

Mammals are extremely efficient in the way in which they localize sound-sources with some animals reaching an accuracy of ±1$^o$ on the horizontal plane and ±5$^o$ with respect to elevation [7]. The areas of the mammalian AC that concern us are the regions that encode the various cues such as Interaural Time Difference (ITD) and Interaural Phase Difference (IPD). These cues are found to be encoded in lower brainstem regions such as the medial superior olive (MSO) [10]. However, higher order regions have also been found to encode location specific information [6], such as the Inferior Colliculus (IC) and the Planum Temporale (PT). These regions are also of interest to us, as it has been found that the PT specifically has properties that help towards maintaining an acoustic 'track' on an object.

## II. ACOUSTIC ROBOT LOCALIZATION

In this section we describe the general idea of robot localization, specifically sound-source detection and its purpose in robotics. Our research group is working towards building a robotic waiter system. Sound localization is therefore understandably required on such a system. For example, as shown in Fig. 1 a robot may be moving around the room when a person would like to gain the robot's attention. Logically this would be done via voice as this is the most common way of human communication, i.e. the guest saying something like "Robot". The robot in this case would therefore need to localize the person and orient itself in order to move up to them. Fig. 14 shows the current acoustic robotic system that we are using to test our architecture.

Sponsored by: EPSRC & OCF Plc

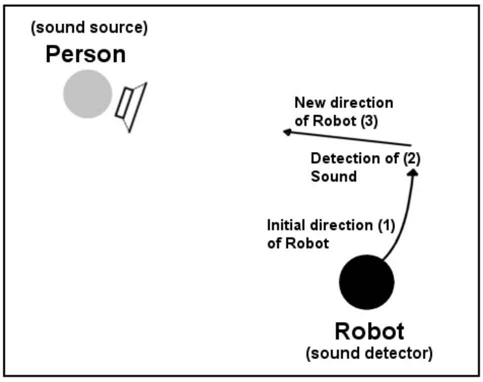


Figure 1. Robot sound localization scenario.

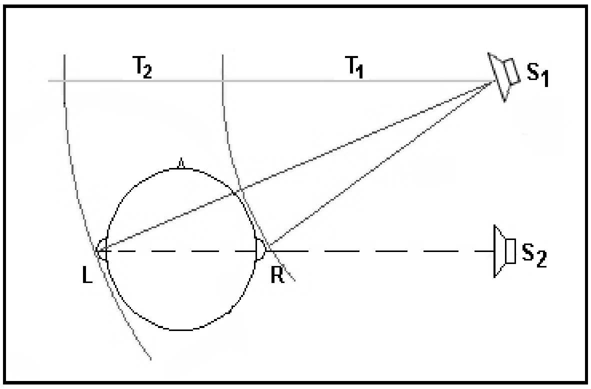


Figure 3. $T_2$ shows the ITD increment.

We are using an ActiveMedia PeopleBOT (shown in Fig. 11) as our base for the acoustic tracking robot. This robot contains a PC running the Linux Red Hat Operating System, a K6-2-500MHz processor and 128MB of RAM and sound card for the acoustic recordings. The robot operates on the local wireless network where it has direct access to our Beowulf cluster. The cluster consists of a 27 node dual Pentium 4, 2GHz set-up which is available to the robot on the network for the processing of any intensive data. This therefore provides additional computing power to increase the speed of computation. For the sound processing we have two omni-directional mini electrets microphones that are amplified using two LM386 audio amplifiers [17] and connected to the robot's sound card line-in port.

## III. METHOD FOR AZIMUTH CALCULATION

In order to locate a sound-source within the environment we need to calculate its azimuth with respect to the robot. The azimuth represents the angle from which the sound-source is located with respect to the robot's internal frame of reference with $0^{\circ}$ always being directly ahead of the robot irrespective of its direction as shown in Fig. 2. This is achieved by computing the time delay of arrival (TDOA) of the wave front at the two microphones, which in biological terms is the equivalent of the Interaural Time Difference cue used in the auditory cortex of the mammalian brain [13, 11]. ITD is the time it takes for the sound to arrive at the contralateral ear once the ipsilateral ear has detected the sound. Fig. 3 shows the ITD increment that is measured for two separately located sound sources.

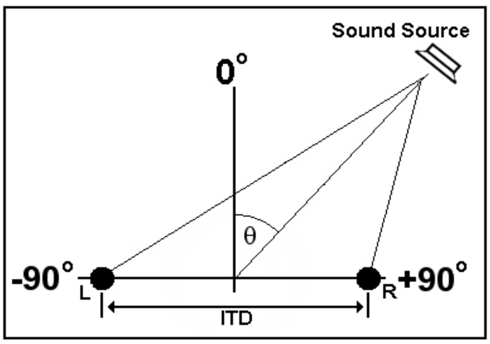


Figure 2. The angle calculated to determine the azimuth. Left of $0^{\circ}$ is negative, right of $0^{\circ}$ is positive.

In order to determine the angle of incidence of the received wave form, we have to be able to detect the Interaural Phase Difference (IPD), i.e. the lag of the wave at a specific point received at both microphones. When the first microphone detects the sound, we need to ensure that when we are calculating the TDOA, that we compute it between two identical points along the waveform in order to ensure we get an accurate measure of the ITD. To carry out this task on our robot the system records a sample of sound in a time interval (initially this was a one second slice). The stereo signal recorded at the microphones is then split into its left and right components. This is passed to the cross-correlation function which is used to compare the left and right channels for similarity i.e. where the signals are most matched, as the example in Figs. 4-7.

## IV. CROSS-CORRELATION

In order to calculate the azimuth of the sound-source we need to compute the point at which the independently received signals at the microphones *g(t)* and h*(t)* are at their maximum correlation i.e. the signals are at closest match when they are superimposed on each other. The cross-correlation function is defined in equation (1). The correlation function *Corr(g,h)(t)* will be at its maximum value at some point in time $(t_n)$ when the function *g(t)* is shifted in time across the function *h(t)*.

$$Corr(g,h)_j(t) \equiv \sum_{k=0}^{N-1} g_{j+k} h_k \qquad (1)$$

Cross-correlation in this instance is used to compare two vectors ***A*** and ***B*** (which contain the values for the signals *g(t)* and *h(t)* respectively) for similarity. Cross-correlation takes as its input the two vectors ***A*** and ***B*** which represent the audio signals recorded from the microphones. The signals are then slid across each other at all points to give a product vector ***C*** whose length is shown in equation (2).

$$\text{length}(\boldsymbol{C}) = (\text{length}(\boldsymbol{A}) + \text{length}(\boldsymbol{B}))-1. \qquad (2)$$

The maximum value in the returned vector ***C*** represents the position of maximum correlation between the two signals *g(t)* and *h(t)* with a time delay σ. As can be seen from the example in Fig 7 the maximum correlation value occurs at approximately position 20800 of vector ***C***. Within the plot of vector ***C*** we have two axes; the y-axis represents the product of all the values within the two vectors ***A*** and ***B*** at any given delay σ. The x-axis represents the current step within the cross-correlation algorithm i.e. the delay σ. If *g(t)* and *h(t)* were not delayed (i.e. the signal was at $0^{o}$) then the maximum value would occur at the mid point of vector ***C*** as when the two vectors are aligned the signal will be matched.

To compute the phase shift or delay of our two signals, we present the left and right channels as vectors ***A*** and ***B*** to the cross-correlation method. These vectors contain the amplitude values of the signals with each location within the vector representing a sample point along the waveform with the amplitude varying between ±1. The cross-correlation method then starts by offsetting vector ***B*** to the far left and zero padding, i.e. the first location of ***A*** is lined up with the last location of ***B***.

As is shown in Figs 4 – 6, as one channel is slid across the other, the correlation vector is created with the length as calculated in equation (2). As can be seen when the cross-correlation starts, i.e. vectors ***A*** and ***B*** are at their highest offset, the product of the vectors will be at their lowest as shown in the left of the plot in Fig. 7. This cross-correlation product will increase as the vectors draw closer to highest similarity and will then again decrease as the vectors are slid past maximum correlation.

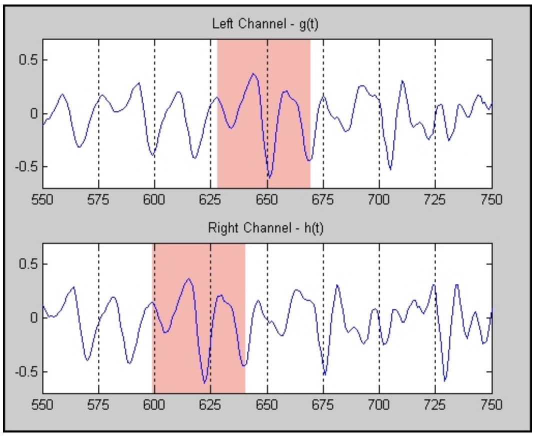


Figure 4. Beginning phase of the sliding window.

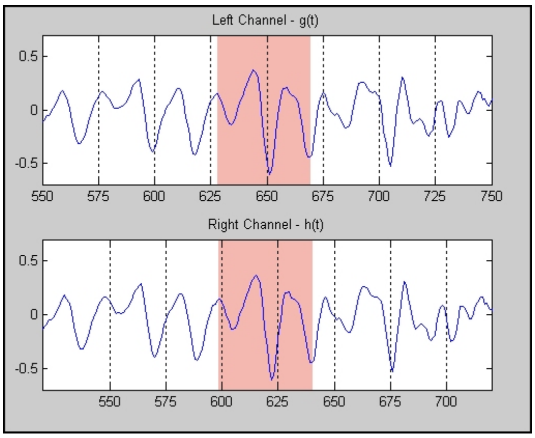


Figure 5. Middle (matched) phase of the sliding window.

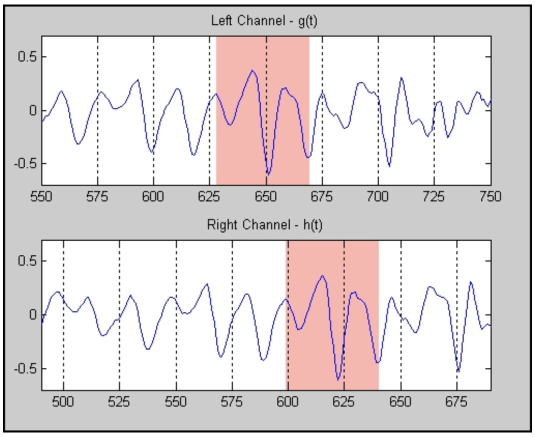


Figure 6. Final phase of the Sliding window.

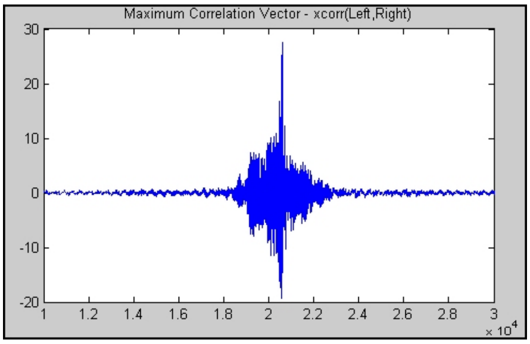


Figure 7. Shows the plot of cross-correlation product vs. vector location.

## V. ALGORITHM FOR LOCALIZATION USING CROSS-CORRELATION AND ITD

Azimuth is calculated from the cross-correlation method by detecting the delay (σ) offset of the highest correlation point in vector ***C***. The delay offset of the maximum correlation point is found by moving to the mid-point of ***C*** as this is $0^{o}$ as discussed previously. Then, counting the number of locations to the highest position gives the delay offset. This offset is either negative (less than the mid-point) or positive (higher than the mid-point). This delay is then used to calculate the TDOA within the system. To find the TDOA we first need to calculate several variables. The first variable we need to determine is the time increment between sampling.

$$\Delta = 1/44.1\times10^{3} = 2.2676\times10^{-5}s \qquad (3)$$

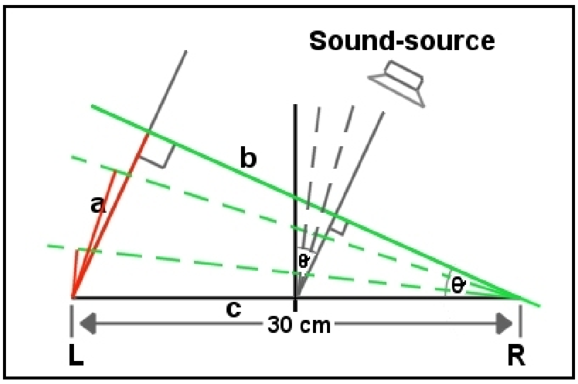


Figure 8. Calculating the angle of sound-source.

We assume that the sound arrives in parallel as shown by line 'b' in Fig. 8. In order to calculate the angle of incidence of the sound-source we need to determine the unknown variables of the triangle in Fig. 8. Using the basic trigometric functions of right angled triangles we can determine the angles Θ. From trigonometry we have the following equations:

$$Sin\Theta = \frac{a}{c},\ Cos\Theta = \frac{b}{c},\ Tan\Theta = \frac{c}{b} \qquad (4)$$

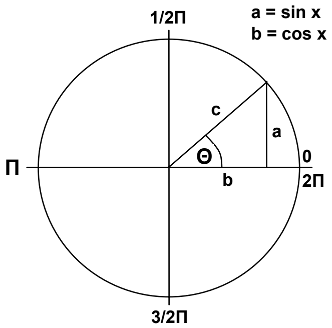


Figure 9. Geometric diagram of sin (x) and cos (x).

However, in order to determine the angle Θ shown in Fig. 8 (note that both angles labeled Θ are the same as again using trigonometry we know that the angles of a triangle must equal $180^o$). Before we can calculate the angle we need to find the length of at least two sides of the triangle. We already know that side 'c' is 0.30 meters as this is the distance between the robots microphones. The remaining sides to determine are 'a' and 'b' from these the side we can determine is 'a' which can be obtained from equation (6). To determine distance we use $V_{sound}$ x time. However, in order to determine the time for the sound to traverse line 'a' we need to use the following equations:

$$t = \Delta \times \sigma \qquad (5)$$

Were Δ = time between sound sampling, see equation (3), and σ = the number of delay samples returned from the cross-correlation function. Next, we determine the length of line 'a' by substituting equation (5) into the following:

$$length = t \times V_{sound} = (\Delta \times \sigma) \times V_{sound} \qquad (6)$$

Were speed of sound is taken to be, $v = 384m/s$ *at room temperature of* $24^o$ *@ sea level.* We now have the length of sides 'a' and 'c' and referring to equation (4) can see that we require the sine rule. However, we need to transpose this to find Θ.

$$Sin\Theta = \frac{a}{c} \therefore \Theta = Sin^{-1}\frac{a}{c} \qquad (7)$$

$$\Theta = Sin^{-1}\frac{(\Delta \times \sigma) \times V_{sound}}{c} \qquad (8)$$

In order to compute the azimuth we computed the cross-correlation of the waveform and determined the delay between the two channels and then using equation (5-8) we can calculate the azimuth of the sound-source.

## VI. ACOUSTIC ARCHITECTURE

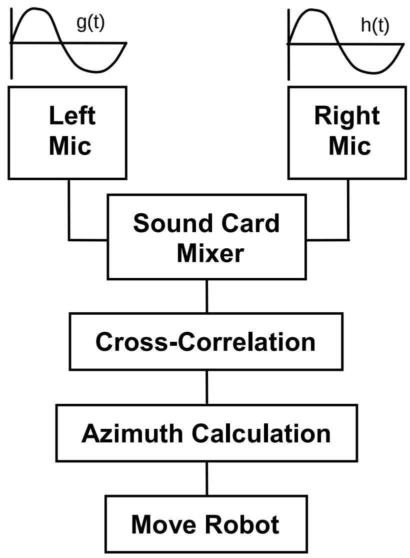


Figure 10. Basic Acoustic Architecture

Fig. 10 shows the basic acoustic system architecture that was initially developed for simple sound-source localization of single sounds. The signals *g(t)* and *h(t)* arrive at the two microphones and are recorded and preprocessed by the 'Sound Card Mixer' stage, from here the vectors ***A*** and ***B*** containing the processed signals *g(t)* and *h(t)* are passed to the 'Cross-Correlation' stage to determine the IPD of the two independent signals. Once the result has been obtained we calculate the azimuth and finally instruct the robot to move by the required angle determined in (8).

## VII. BASIC TRACKING SYSTEM

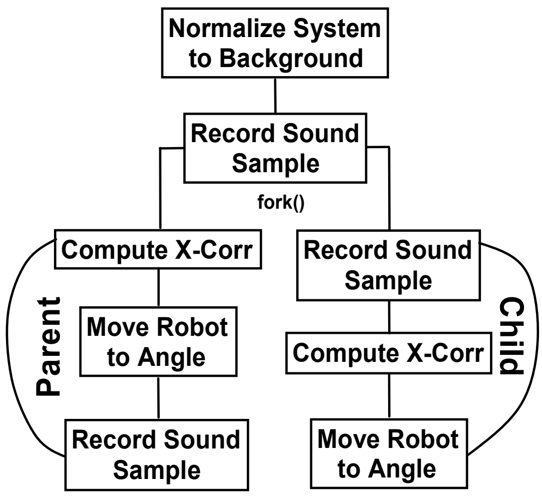


Figure 11. Basic Two Process Tracking System.

The second stage of our system was to enable the acoustic robot to 'track' that is follow a dynamic sound-source. If the sound of interest is moving around the environment then we need to be able to 'track' its movements. The mammalian auditory cortex, along with the rest of the mammalian brain is a fully parallel system; we therefore looked at several parallel processing methods before deciding on the one shown in Fig 11.

Here the system starts off with an initial single process normalizing the recording levels to the background level so that we do not to over saturate the system. From here the first recording is made (we have reduced the time slice from 1 second to 200ms to improve speed and efficiency). The system then performs a 'fork()' command and creates a child process.

The parent process uses the first recorded sound and computes the cross-correlation. During this processing time, the child process begins to record the second sound sample. As the parent process then calculates the angle and begins to move the robot to the desired angle the child process begins the cross-correlation from the sound sample it took. The use of semaphores is employed to ensure the two processes do not conflict with each other and to ensure that they run concurrently but with the specific stages offset.

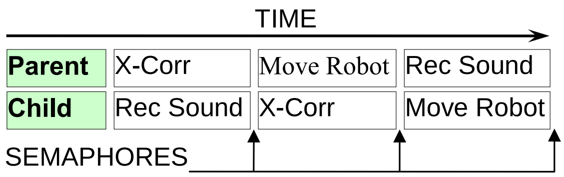


Figure 12. Time progression of multi process system.

Fig. 12 shows the parent and child processes with their order of execution of the functions enabling sound-source tracking. The arrows represent the semaphores that are used to ensure the processes remain in sync.

## VIII. EXPERIMENTATION - LOCALIZATION

The initial sets of experiments for the sound-source localization system were carried out on waveforms which had their phase manually altered using sound editing software; this enabled us to accurately measure and make changes to the phase delay and amplitude of the waveforms and therefore manipulate them so that we could effectively create any angles we wish for testing the system. The second sets of experiments were carried out using prerecorded sound files which were recorded via the two amplified microphones, and finally we tested the full localization system on the robot itself.

### A. *Simulated Waveforms*

To test our algorithm using Cross-Correlation to compute the TDOA of a waveform, we recorded the word 'LENA' (the name of our robot). This recording was made in mono, using a resolution of 8-Bits and a sampling rate of 44.1 KHz. This waveform was then converted into a stereo signal and three test cases were created to see if the delay of the signal could be calculated accurately using the cross-correlation method.

Firstly, the mono signal was added to both channels in a stereo file and both the phase delay and the amplitude of the signal were kept identical, this therefore simulated a sound originating from directly in front of the robot i.e. $0^{o}$ due to their being no phase shift in the waveform as the wave front arrives at both microphones at the same time and therefore having a TDOA of zero.

Secondly, the mono signal was again added to both channels. This time the phase delay and amplitude of a specific channel were altered. The phase of the LEFT channel was modified by creating a delay of 20ms and the amplitude was then decreased by 20%. This therefore gave the impression that the signal was originating from the right of the robot as the signal arrives at the right microphone first.

For the third test we swapped the channels over from 'test case 2', so that the sound-source location appeared to be originating from the left side of the robot as the RIGHT channel was delayed in its arrival.

Table I shows the details of each of the samples used to test our system giving details of delay and sampling. We used these preconfigured waveforms to see if our algorithm of cross-correlation gave the same results as using algebraic calculations. As we have preset the phase shift in our test samples (i.e. the real world variable) we can mathematically compute the number of delay samples that should be detected by cross-correlation.

From equation (3) we know that each sample $\Delta$ = $2.2676\text{x}10^{-5}$ seconds, therefore using (9) we can show the delay in samples:

$$\delta = \frac{\xi}{\Delta} \tag{9}$$

Where $\xi$ is the phase delay in seconds of the contralateral signal, $\Delta$ is the sample time and $\delta$ is the number of samples. Table II shows the results in number of samples that were obtained from both the cross-correlation function in equation (1) and also the mathematically calculated method in equation (9).

### B. *Real-World Prerecorded Waveforms*

On simulated waveforms the cross-correlation method and the mathematical calculations based on (9) returned the same results to 100% accuracy. This indicates that at this stage of our testing cross-correlation is a good and accurate measure of similarity for the localization of acoustic sound-sources.

With the next step of our experimentation we carried out the cross-correlation method on several real-world prerecorded sounds. Here the sample rate of the recorded sound was known. However unlike the manually altered sounds, the phase shift / delay time / lag was the unknown factor as this is what needed to be calculated in order to determine the azimuth. Preliminarily we conducted tests at 10 different angles using free field sound. The tests were conducted by a person standing at a range of 1.5 meters from the center of the microphones at the 10 measured angles. The angles recorded were –90°, –50°, -40°, -30°, 0°, +10°, +20°, +35°, +45°, +70°.

Knowing these angles we can calculate mathematically how many delay samples there should be with respect to the current angle. If we take the sine rule from equation (4) we know the angle Θ and the length of 'c' we can therefore transpose for 'a' giving:

$$a = Sin\Theta \times c \qquad (10)$$

once we know 'a' we can then calculate the delay σ by transposing equation (6):

$$\sigma = \frac{a}{(V_{sound} \times \Delta)} \qquad (11)$$

These results were then confirmed with the use of the cross-correlation method. Table III shows the comparisons between these two methods of cross-correlation and mathematical calculation. Each angle was recorded five times to test how resilient the system is to repeated recordings. The system seemed to provide results that were within an average accuracy of ±1.5°.

### C. Robot Testing

Finally, the remaining test for the sound-source localization architecture of Fig. 10 was to test on the robot itself. Here we positioned the robot in the center of the robot lab and had a person stand 1m away from the center point of the robot. We conducted 10 specific angle tests five times, then some random tests i.e. we moved to randomly chosen places to see if the robot would turn to face the speaker. The 10 tests conducted were -90°, -50°, -40°, -30°, 0°, +10°, +20°, +35°, +45°, +70°. Table IV shows the results of the tests on the robot, the five repetitions of the 10 individual angle tests were averaged in the table. The accuracy of the results has also been given, as can be seen, the smaller the angle to localize to the higher effect of the accuracy the error has, this is due to the error value having a larger effect over a smaller angle.

## IX. EXPERIMENTATION – BASIC TRACKING

The second sets of experiments were conducted to test the development of the sound-source 'tracking' system shown in Fig. 11. For this system the set of experiments were conducted on the robot only. In order to see if the system was functioning correctly the microphones have to be able to rotate in order to follow the sound-source.

To test the system we kept a distance of 1 meter from the center of the robot and walked around the robot in a circle to see if it would rotate with us. We found the system was not fully 'real-time' but lagged behind us by approximately 1-2 seconds due to the time taken to compute the cross-correlation and start up the motors to move the robot's wheels. At present we are running our tracking system on a single processor on board the robot itself. However, in further developments we aim to migrate our tracking architecture to our multiprocessor Beowulf cluster, which should speed up the system substantially.

TABLE I. MANUALLY CREATED WAVEFORMS

| Test | Phase Lag | Amplitude | Bit | KHz |
|---|---|---|---|---|
| Sample 1 | 0 | L, R = 100% | 8 | 44.1 |
| Sample 2 | L = 20ms | L = 80% | 8 | 44.1 |
| Sample 3 | R = 20ms | R = 80% | 8 | 44.1 |

TABLE II. NUMBER OF DELAY SAMPLES

| Test | Phase Lag | Calculated | Cross-correlation |
|---|---|---|---|
| Sample 1 | 0 | 0 | 0 |
| Sample 2 | L = 20ms | 960 | 960 |
| Sample 3 | R = 20ms | 960 | 960 |

TABLE III. CROSS-CORRELATION VS MATHEMATICAL CALCULATION (DELAY SAMPLES)

| Sample Angle | Cross-Correlation | Mathematically Calculated |
|---|---|---|
| -90 | 39 | 37 |
| -50 | 27 | 26.4 |
| -40 | 23 | 22 |
| -30 | 17 | 17 |
| 0 | 0 | 1 |
| 10 | 5 | 5.98 |
| 20 | 11 | 11.78 |
| 35 | 20 | 19.76 |
| 45 | 26 | 24.36 |
| 70 | 33 | 32.38 |

TABLE IV. FINAL ANGLE TESTS ON ROBOT

| Test | Actual Angle | Robot Position (Average) | Accuracy % |
|---|---|---|---|
| Test 1 | -90 | ±4 | 95.5 |
| Test 2 | -50 | ±2 | 96 |
| Test 3 | -40 | ±1 | 97.5 |
| Test 4 | -30 | ±0 | 100 |
| Test 5 | 0 | ±2 | 98 |
| Test 6 | +10 | ±2 | 80 |
| Test 7 | +20 | ±1 | 95 |
| Test 8 | +35 | ±2 | 94.3 |
| Test 9 | +45 | ±2 | 95.6 |
| Test 10 | +70 | ±3 | 95.7 |

TABLE V. RESULTS OF TEST ANGLES

| Test | Actual Angle | Calculated Azimuth | Accuracy % |
|---|---|---|---|
| Test 1 | -90 | -90 | 100 |
| Test 2 | -60 | -55 | 94.44 |
| Test 3 | -20 | -22 | 97.78 |
| Test 4 | 0 | 2 | 97.78 |
| Test 5 | +30 | +34 | 95.56 |
| Test 6 | +40 | +45 | 94.44 |
| Test 7 | +70 | +70 | 100 |
| Test 8 | +90 | +90 | 100 |

## X. RESULTS & DISCUSSION

Once we had conducted our experiments using cross-correlation to see if we could perform a similarity measure to calculate the delay of the left and right signals and ultimately determine the location of the sound-source, we then conducted some random angle tests to see how well the system performed. These results are shown in Table V with the calculated angle measured against the actual angle. Again each angle was recorded and tested five times to obtain an average and determine accuracy and robustness.

As can be seen from the results presented in this paper, it is clear that using cross-correlation and ITD is an effective method of sound-source localization, the testing performed wielded some promising results, as can be seen from the above tables this system is capable of locating a sound-source within the environment to an average accuracy of $\pm 1.5^{o}$ which is comparable to its mammalian counterpart as is shown by Jens Blauert [18] to be $0.9^{o}$ – $1.5^{o}$ for speech.

The first and second stages (localization and tracking) of our acoustic system have now been prototyped. The system developed using cross-correlation is able to locate sound-sources on the azimuthal plane with only a small time delay of approximately 1 second (on localization tasks). The system is also capable of tracking a sound-source within the environment, again with a time lag of 1-2 seconds. The system will undergo further development in order to increase the speed and accuracy of the localization and tracking. Below are future additions to the robotic system to enable it to carry out sound localization and tracking to a higher standard (including system enhancements).

In order to approach the functionality of the mammalian auditory cortex this system is required to be able to separate an incoming sound stream into its independent components. A common scenario is the 'cocktail party effect' [5, 9]. This involves more than one person talking at the same time but still being able to understand what the person we are focused on is saying. The next stage of our system is to be able to localize 'specific' sounds of interest. This means that the robot would only orient itself to a sound it deemed interesting.

We have carried out the bulk of our experiments within our robotics lab which has an ambient background level of 52db. This is fairly low and therefore, the next stage in our experimentation is to cover a range of background levels and see how well the current system can still detect azimuth. One problem however that still exists at present is what we term the 'mirror effect'; this is where the system has difficulty in determining whether the sound is originating from in front of or behind the robot, this is shown in Fig. 13. This ambiguity also occurs in the mammalian auditory system [7]. In our robotic system, this phenomena is due to the current microphones being fully isotropic, that is, they receive with the same strength from their entire circumference and are not directional as in the mammalian ear. This problem is partially overcome in the mammalian system by the use of non-isotropic pinnae as this makes changes to the spectral patterns of the received sound. Combining this with the other available cues [7] overcomes this.

One way that we can help to overcome this effect on our robot is by allowing it to turn several degrees left (this method is also used by mammalian's) and taking another reading, therefore changing the interaural intensity difference (IID) of the sound received at the microphones. From this change we can then easily determine if the sound-source is in front of or behind the robot. Another way to tackle this is from an engineering perspective and incorporating more microphones [8]. However this moves away from our goal of being biologically inspired.

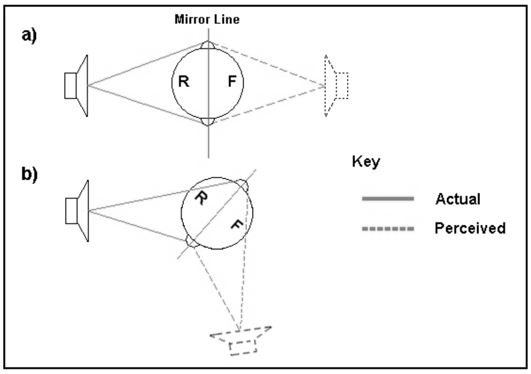


Figure 13. Example of the 'mirror effect' caused by purely isotropic microphones.

## XI. CONCLUSIONS

In this paper we have shown a robotic system for sound-source localization and basic tracking along the horizontal plane using cross-correlation to provide a similarity measure of the sound signals *g(t)* and *h(t)* arriving at the two microphones. This is the first stage in creating our biologically inspired acoustic tracking robot. Showing how the cues that are available in the auditory cortex of the mammal can be incorporated into a robotic system to provide a similarly efficient model of sound-source localization. This paper has shown how combining cross-correlation with TDOA of the sound at the microphones does in fact enable us to locate the azimuth of the sound-source within the environment with acceptable accuracy.

This paper also shows how creating a multiprocessing model of our sound-source localization architecture can be used to create a basic acoustic sound-source tracking system. The use of cross-correlation has proved a successful method for the similarity measure of two sound signals (as a basis for further development). As can be seen from our results we are able to localize a sound-source to within an average of $\pm 1.5^{o}$. This is an accurate result which demonstrates the potential of cross-correlation as a basis for robotic sound-source localization.

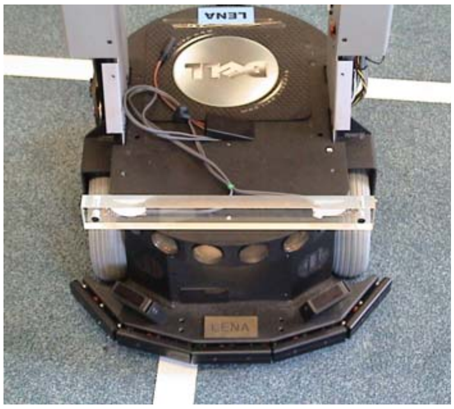


Figure 14. Picture of the PeopleBOT used in the department with ears attached.

## ACKNOWLEDGMENTS

The authors would like to thank EPSRC and OCF plc. for their funding of this research. We would also like to thank Dr. Cornelius Weber for his help with refining some of the algorithms.

## REFERENCES

[1] Hans-Joachim Böhme, Torsten Wilhelm, Jürgen Key, Carsten Schauer, Christof Schröter, Horst-Michael Groβ, Torsten Hempel. An approach to multi-modal human-machine interaction for intelligent service robots. Robotics and Autonomous Systems, Vol 44(1) 2003, pages 83-96.

[2] Cynthia Breazeal. Towards Sociable Robots. Robotics and Autonomous Systems, Vol. 42(3-4) 2003, pages 167-175.

[3] Cynthia Breazeal and Brian Scassellati. Robots that imitate humans. TRENDS in Cognitive Sciences, Vol. 6(11) 2002, pages 481-487.

[4] Terrence Fong, Illah Nourbakhsh, Kerstin Dautenhahn. A survey of socially interactive robots. Robotics and Autonomous Systems, Vol. 42(3-4) 2003, pages 143-166.

[5] Mark Girolami. A nonlinear model of the binural cocktail party effect. Neurocomputing, Vol 22(1-3) 1998, pages 201-215.

[6] Timothy D. Griffiths and Jason D. Warren. The planum temporale as a computational hub. TRENDS in Neurosciences, Vol. 25(7) 2002, pages 348-353.

[7] Harold L. Hawkins, Teresa A McMullen, Arthur N. Popper, Richard R. Fay. Auditory Computation. Springer Handbook of Auditory Research. Springer 1996, pages 334-336.

[8] Jie Huang, Tadawute Supaongprapa, Ikutaka Terakura, Fuming Wang, Noboru Ohnishi, Noboru Sugie. A model-based sound localization system and its application to robot navigation. Robotics and Autonomous System, Vol. 27(4) 1999, pages 199-209.

[9] Gil-Jin Jang, Te-Won Lee, and Yung-Hwan Oh. Blind seperation of single channel mixture using ICA basis functions. In Proceedings of 3rd International Conference on ICA and BSS (ICA2001), San Diego, CA, USA. pages 595-600.

[10] Philip X. Joris, Philip H. Smith and Tom C. T. Yin. Coincidence Detection in the Auditory System: 50 Years after Jeffress. Neuron, Vol. 21(6) 1998, pages 1235-1238.

[11] Jack B. Kelly and Dennis P. Phillips. Coding of interaural time differences of transients in auditory cortex of *Rattus norvegicus*: Implications for the evolution of mammalian sound localization. Hearing Research, Vol 55(1) 1991, pages 39-44.

[12] Pedro Lima, Andrea Bonarini, Carlos Machado, Fabio Marchese, Carlos Marques, Femando Ribeiro, Domenico Sorrenti. Omni-directional catadioptric vision for soccer robots. Robotics and Autonomous System, Vol 36(2-3) 2001, pages 87-102.

[13] David McAlpine and Benedikt Grothe. Sound localization and delay lines – do mammals fit the model?. TRENDS in Neuroscience. Vol. 26(7) 2003, pages 347-350.

[14] Jean-Marc Valian, et al. Robust Sound Source Localization Using a Microphone Array on a Mobile Robot. In Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robotics and Systems, pages 1228 – 1233.

[15] Qing Hua Wang, Teodor Ivanov and Parham Aarabi. Acoustic robot navigation using distributed microphone arrays. Information Fusion, In Press, Corrected Proof.

[16] S. Wermter, C. Weber, M. Elshaw, C. Panchev, H. Erwin, F. Pulvermüller. Towards Multimodal Neural Robot Learning. Robotics and Autonomous Systems Journal, in press, 2004.

[17] National Semiconductor, www.national.com LM386, http://www.national.com/ds.cgi/LM/LM386.pdf

[18] Jens Blauert. Spatial Hearing – The Psychophysics of Human Sound Localization. 1997, Table 2.1, page 39.