ACTIVE BINAURAL SOUND LOCALIZATION

Greg Reid, Evangelos Mikos
Department of Computer Science
York University
North York, Ontario, CANADA M3J 1P3
e-mail: eem@cs.yorku.ca
URL: http://www.cs.yorku.ca/~eem

ABSTRACT

Estimating the direction of arrival of sound in three dimensional space is typically performed by generalized time-delay processing on a set of signals from an array of omnidirectional microphones. This requires specialized multichannel A/D hardware. Our work is motivated by the desire to only use standard two-channel audio A/D, which is commonly available on personal computers. To estimate direction of arrival of sound, we change the pose of the microphones by mounting them on a computer-controlled pan-and-tilt unit. In this paper, we describe two approaches. The first uses two omnidirectional microphones on a fixed baseline, which has two rotational degrees of freedom. The second uses a directional microphone with two rotational degrees of freedom. We show algorithms for estimation of sound direction of arrival using these configurations. Preliminary experimental results demonstrate the feasibility of the approach.

1 INTRODUCTION

Most previous work on sound source localization has used arrays of omnidirectional microphones and generalized time-delay estimation techniques [4]. This approach requires special purpose multichannel A/D hardware, generates a lot of signal data, and requires intensive computation. In our work, we pose the question of how well sound direction of arrival can be determined through the use of two, possibly directional, microphones. Cues that we use are either the interaural time or interaural intensity difference. Sounds from a pair of microphones can easily be captured on standard A/D hardware. To resolve ambiguities, and to obtain the best possible attenuation of background noise (interference), our microphones are mounted on computer-controlled pan-and-tilt units (PTU), and their orientation in space can be changed at will (Figure 1). The computer controlled Pan-Tilt Unit we use is made by Directed Perception Inc. (Model PTU-46-17.5).

Figure 1: Top: An active directional microphone or 'ear' mounted on a computer controlled pan-tilt unit. Bottom: An active pair of omnidirectional microphones.
The use of active microphones achieves with physics what microphone arrays must achieve with massive computation. We call our approach “Active Audition”, the auditory equivalent of “Active Vision” [5], in which cameras are mounted on PTUs and verging stereo is used for tracking visual targets. In our work we investigate the computational principles underlying the above approach. Our main objective is to develop and evaluate the performance of algorithms for source direction determination using an active audition system.

2 PREVIOUS WORK

In human auditory perception, it is believed that there are three basic cues from which most (if not all) sound localization is derived [7, 1].

- **ITD** - Interaural Time Differences are phase changes or time delays in the signals caused by the path length difference between the two ears. It is the primary horizontal cue for humans in lower frequencies (below 1KHz).

- **IID (or ILD)** - Interaural Intensity (or Level) Differences, which are due to the directionality of the human ear and the shadowing effect of the head between the ears. This is the primary horizontal cue in higher frequencies (above 4KHz), which correspond to wavelengths smaller than the size of the ear.

- **Spectral Cues** - these are an effect of a human’s presence in the sound field. The spectral characteristics of a perceived sound are affected by the presence of our outer ears, head and torso. For humans these cues extend our perception into the vertical plane.

Estimating the direction of a sound source from signals received at two directional microphones has been addressed and tested in simulation mode only in [2]. In that work, the microphones are fixed in space, both pointing forward with a slight difference in the elevation of the maximum gain direction. The central problem addressed is how to represent the nonlinear mapping from signal features to source direction, which is solved by an artificial neural network. The inputs to the neural network are measures for the ITD and the IID at distinct frequencies. The measure for the ITD is the phase difference at the two microphones. The measure for the IID is the intensity ratio (in dB) at the two microphones at distinct frequencies. The conclusion from that work is that for the realistic case of noisy input during training, the accuracy of localization is tolerable only in the central region of the training space, namely for a source near zero azimuth and elevation.

Figure 2: Two different poses of the baseline, b1 and b2, yield two solution cones, which intersect at two lines denoted by S, representing possible directions of arrival.

3 METHODOLOGY AND ALGORITHMS

3.1 An active omnidirectional microphone pair

The intuition behind this method is the following. A single time delay measurement from a single position of the microphone baseline constrains the source direction to be on a right circular cone, which has its vertex at a fixed reference point (for example the midpoint of the baseline), and its axis of symmetry is the baseline itself. A single rotation of the baseline about a horizontal or vertical axis through its midpoint yields another cone on which the source direction should lie. Figure 2 shows the concept.

For a single baseline position, the solution cone is defined as follows. Its vertex is the reference point (the midpoint of the baseline), its axis of symmetry is the unit vector along the baseline, and its angle $\alpha$ between its axis and any line of the cone that contains its vertex is given by the following equation:

$$\sin \alpha = cn/fd$$  \hspace{1cm} (1)

where $c$ is the speed of sound, $n$ is the ITD measured in samples, $f$ is the sampling frequency, and $d$ is the length of the baseline.

Solving for the source direction is a geometric problem of finding the intersection between two cones, or, more generally, of finding the source direction that satisfies a least squares criterion, in case time delay measurements from more than the minimum number of baselines are obtained.

First we consider our unknown source direction as a unit vector $s$ with its start at the reference point and pointing towards the sound source. This vector is the unique solution and is independent of the orientation of the baseline. The following constraint on $s$ then applies for a particular orientation $i$ of the baseline $b_i$, and a direction of arrival at angle $\alpha_i$ with respect to baseline (unit) vector $b_i$:

$$s \cdot b_i = \cos \alpha_i$$  \hspace{1cm} (2)
or equivalently,
\[
    s_x b_{ix} + s_y b_{iy} + s_z b_{iz} = \cos \alpha_i
\]  
(3)

Quantities \(b_{ix}, b_{iy}, b_{iz}\) are the cartesian coordinates of a unit vector with azimuth and elevation given by \((\theta_i, \phi_i)\) respectively. Azimuth and elevation are set by the user by controlling the motors of the pan-and-tilt unit. Angle \(\alpha_i\) represents the direction of arrival with respect to the selected baseline pose, and it is computed from the measured ITD of the microphones. Therefore equation 3 is a linear equation with three unknowns, \(s_x, s_y, s_z\). The minimum number of equations needed (obtained by using different baseline poses), is two, since vector \(s\) is a unit vector, yielding the implicit constraint \(s_x^2 + s_y^2 + s_z^2 = 1\). With three or more such equations (generated by three or more orientations of the baseline) we can solve for \(s\), which is the direction of the sound source, through a linear least squares approach.

### 3.2 An active directional microphone

Directional microphones are microphones which exhibit an intensity response that varies with the sound wave’s direction of arrival. The directivity pattern represents the intensity of the received sound signal as a function of the direction of arrival to the microphone. These patterns are typically cardioid in shape [3], and are parameterized by frequency. Figure 3 shows the measured 2D directivity pattern for our directional microphone, which is equipped with a parabolic reflector (Electret condenser microphone SONY Model ECM-PB1IC). The side lobes of the pattern are not reliable and depend on the acoustic properties of the surrounding space and the position and orientation of the microphone in it. The main lobe within \(\pm 45^\circ\) of the direction of maximum response has been experimentally determined to be stable with respect to changes in the environment surrounding the microphone.

In our work, we assume a sound source which has fixed position in space, and is persistent over time. This allows us enough time to change the pose of the microphones and take more than one measurement. We need enough measurements so that we can fit the pattern shown in figure 3 to them and estimate the direction of arrival of sound as the orientation corresponding to the peak of the fitted pattern. The problem thus is split into two subproblems. First, select a number of measurements that are guaranteed to correspond to the main lobe of the pattern and fairly high above the values at the tails. Second, fit the pattern to them and estimate the direction of arrival of sound.

The selection problem is simplified by assuming that the direction of arrival of the source is between -70 and 70 degrees. In this range there is practically no side-lobe present and the directivity pattern is unimodal. By applying a small number of steps (2-3) of the golden section search in one dimension [6], we collect sufficient measurements on the main lobe and high enough above the tails to do the fitting.

To solve the fitting problem with the measurements extracted above, we view the 2D directivity pattern as a function \(f(\phi)\), where \(\phi\) is the direction of arrival. Furthermore, assume that the set of measurements extracted is \(I = \{I_1, I_2, \ldots, I_N\}\), corresponding to azimuth \(\phi_1, \phi_2, \ldots, \phi_N\) respectively. The problem of fitting \(f(\phi)\) to the set \(I\) involves estimating a scale \(s\) and a translation \(t\) that will induce the best fit of \(f\) to \(I\). For a perfect fit at translation \(t\), the following will be true:

\[
    \frac{f(\phi_1 - t)}{I_1} = \frac{f(\phi_2 - t)}{I_2} = \ldots = \frac{f(\phi_N - t)}{I_N} = s
\]  
(4)

Therefore, we can formulate the estimation of translation \(t\) (and implicitly the scale factor \(s\)) as an one dimensional search for the minimum of the following function of \(t\).

\[
    \min_t \sum_{i \neq j} \left( \frac{f(\phi_i - t)}{I_i} - \frac{f(\phi_j - t)}{I_j} \right)^2
\]  
(5)

This method requires only that the directivity pattern of the directional microphone be known a priori, i.e. measured during a calibration stage. Extension of this approach to 3D sound direction determination involves search in two dimensions (in both azimuth and elevation).

### 4 EXPERIMENTAL RESULTS

We experimented with the two configurations described earlier, and we now present preliminary results of the estimation of the direction of arrival of sound from a single sound source at a single frequency (3KHz). The
sound source was placed at approximately 3m from the microphones. In the omnidirectional microphone pair configuration, the baseline was 25cm.

Figure 4: The graph shows a number of sound directions of arrival and the quality of the associated estimates using the omnidirectional microphone pair configuration. The diamond represents the true direction of arrival, and the cross represents the standard deviation of the azimuth (horizontal axis) and elevation (vertical axis) of 25 estimates for each true direction of arrival.

The diagrams show the true directions of arrival, and the standard deviation of the estimates of azimuth and elevation for each direction of arrival (for the omnidirectional pair configuration) and the standard deviation of the azimuth at zero elevation (for the directional-omni microphone pair configuration). The estimation of both azimuth and elevation using the directional microphone is currently being investigated.

5 DISCUSSION

We presented two different configurations of microphones mounted on a pan-and-tilt unit and methods for estimating direction of arrival of sound from a persistent distant source. The omnidirectional pair configuration was able to determine direction of arrival with an approximate standard deviation of ±4° in azimuth and elevation. The directional microphone was able to determine direction of arrival with an approximate standard deviation of ±5° in azimuth. These results demonstrate that the proposed methods are a promising alternative to the use of microphone arrays, using off-the-shelf and low-cost consumer audio equipment.

We are currently working on the following topics: (a) tracking of a slowly moving sound source. The availability of a prediction of the direction of arrival allows us to optimize the pose of the microphones to minimize error; (b) direction of arrival estimation for impulsive sources by exploiting the dependence of the directivity pattern on frequency; (c) 3D (azimuth and elevation) direction of arrival estimation with the directional microphone.

References


