A COMPARISON OF ALGORITHMS FOR SOUND LOCALIZATION

Pedro Julián, Andreas G. Andreou, Larry Riddle, Shihab Shamma*, Gert Cauwenberghs*

Johns Hopkins University
Electrical and Computer Engineering Dept.
3400 North Charles St., Baltimore, MD 21218, USA

ABSTRACT

In this paper, we compare the performance of four algorithms for sound localization: one-bit correlation, one-bit correlation derivative, and two methods inspired from biology, namely, gradient flow and sterausis. We employ real-data recorded from four microphones to compare the localization performance.

1. INTRODUCTION

We present a comparison of four algorithms for sound localization using four microphones and data experimentally recorded in a natural environment. We consider the situation in which the sensors are passive; in our case, a pair of microphones to sense the signal and estimate the source position. Two of the algorithms employ the classical approach of cross-correlation [1]; the other two are bio-inspired: spatial-temporal gradients techniques [2] and the stereausis network architecture proposed in [3]. The comparison study presented in this paper is aimed at a micropower sound localizer in CMOS technology. The companion paper [4], discusses the implementation and testing of a micropower binary cross-correlation architecture.

2. SETUP AND DATA COLLECTION

The localization setup under consideration consists of four microphones, as shown in Fig. 1, with an effective distance $L = 15.87cm$. We are assuming that the sound source is far away from the microphones ($L << L_s$), and is also limited in frequency from $20Hz$ to $300Hz$. To compare the algorithms in a natural environment a series of experiments were performed in an open field with a speaker set at a distance of approximately 18 m from the microphones. For each angular location of the speaker, 30 seconds of data were emitted and recorded simultaneously by the four microphones. This experiment was repeated for nineteen angles between 0° and 180° in steps of 10°. As all algorithms were designed to produce an estimation after one second, for every angle we obtained a set of 30 readings of time delay.

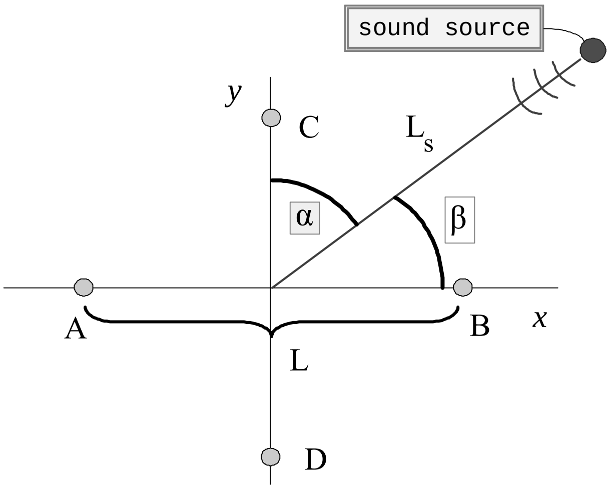

Figure 1: Microphones setup to measure the bearing angle.

3. DESCRIPTION OF THE DIFFERENT METHODS

In this section, we describe four algorithms employed to estimate the bearing angle. Two of them are algorithms previously reported in the literature, based on correlation [1] and spatial gradients techniques [2]. The other two are new approaches: the first one is based on a modification of the correlation approach, and the second is based on a neuromorphic approach [3].

3.1. Correlation Approach

This is a standard approach that has been extensively reported in the literature [1], [5]. Let us consider one pair of microphones and assume that the signals $x_A(t)$, $x_B(t)$ entering that pair of microphones are described by:

$$\begin{array}{l} x_A(t) = s(t) + n_A(t) \\ x_B(t) = s(t-D) + n_B(t) \end{array} \quad (1)$$

where $s(\cdot)$ is the signal emitted by the source, $n_A(\cdot)$ and $n_B(\cdot)$ are uncorrelated noise signals and D is the time delay between microphones. Under the assumption that the source is far away, the wave arriving at the microphones can be considered as plane, and then the following relation

*P. Julián is with CONICET (Consejo Nacional de Investigaciones Cientificas y Técnicas), Av. Rivadavia 1917, 1033 Cap. Fed., Argentina; and on leave from the Departamento de Ingenieria Eléctrica y Computadoras, Universidad Nacional del Sur, Av. Alem 1253, 8000 Bahia Blanca, Argentina. E-mail: pjulian@ieee.org

†S. Shamma is with the Department of Electrical Engineering, University of Maryland at College Park

holds:

$$D = L/c\cos(\beta) = D_{\max}\cos(\beta), \tag{2}$$

where $c = 345m/s$ is the speed of sound in air at ambient temperature and $D_{\max} = 460\mu s$ is the maximum delay. Considering that n_A and n_B are uncorrelated, the correlation between signals x_A and x_B is given by:

$$R_{x_A x_B}(\tau) = \int_{-\infty}^{\infty} s(t)\, s(t - D + \tau)\, dt$$

This function will exhibit a maximum at $\tau = D$. Therefore, one way to estimate the time delay is to generate the correlation function numerically and calculate the time where the maximum is achieved. In practice, the signal is sampled at a certain frequency $f_s = 1/T_s$ and the correlation is approximated using a discrete time sum:

$$\tilde{R}_{x_A x_B}(iT_s) = \sum\nolimits_{k=0}^{K} x_A(kT_s)\, x_B((k-i)T_s), \tag{3}$$

where K is such that $K \cdot T_s$ is the time window under consideration. Operation (3) can be implemented in a digital fashion after quantization of the signals. Using experimental data, we found that a one bit quantization leads to accurate estimations, as we will show later. From a hardware prospective, coding the signal with just one bit produces a dramatic reduction in the density and complexity of the design. The associated structure is composed of a number of stages

$$y(i) = \sum\nolimits_{k=0}^{K} x_A(k)\, x_B(k-i), \tag{4}$$

where i is an index to the stage number. As was explained in [4], a sampling frequency of $200KHz$ permits to estimate angles in the range $\{\alpha \in [0, 50] \cup [+130, +180]\}$, with an accuracy of one degree. This choice of sampling frequency implies that every discrete time delay is $T_s = 5\mu s$. It also implies that the maximum possible delay –corresponding to an angle $\beta = 90°$– is $D_{\max} = 460\mu s$, so that it is necessary to implement 92 stages. Accordingly, index i in (4) ranges from 0 to 91. From a hardware viewpoint, the digital implementation of (4) requires shift registers to generate the delayed versions of x_B, a counter implementing the correlation operation and finally one block to determine where the maximum has occurred.

A drawback to this approach is that once the signal is quantized with one bit, the information corresponding to the time delay between signals is encoded in the changes of state from zero to one, and viceversa. No information is contained in those parts of the signal where there are no state changes. However, every stage (4) is counting all the time at the frequency clock, regardless of the input values. As the frequency of the clock is much higher than the frequency of the signal, this architecture will dissipate power at a much higher rate than what is actually necessary. This observation motivated the approach presented in the following sub-section. An additional disadvantage of this approach is the need to calculate the occurrence of the maximum of (4), which would require the implementation of additional circuitry (a winner-takes-all circuit or an equivalent digital circuit).

3.2. Correlation Derivative Approach

As we said, the maximum of the correlation occurs when the delay produced by the shift register chain coincides with the relative delay between signals. Mathematically, detecting the maximum of the correlation function is equivalent to detecting the zero-crossing of its derivative when the second derivative is negative. This methodology has several advantages that we will describe now. If we consider (4) and calculate the discrete difference between adjacent elements, we get for every stage

$$\begin{gathered}\Delta y(i) := y(i) - y(i-1) = \\ \sum\nolimits_{k=0}^{l} x_A(k)\left[x_B(k-i) - x_B(k-(i-1))\right]\end{gathered} \tag{5}$$

Careful observation of (5) reveals an UP/DOWN counter, which counts up when $x_A(k) = 1$, $x_B(k-i) = 1$ and $x_B(k-(i-1)) = 0$, and counts down when $x_A(k) = 1$, $x_B(k-i) = 0$ and $x_B(k-(i-1)) = 1$. In this case, the count is only updated when one of the two signals changes its state, and it is idle the rest of the time. This mode of operation reduces the circuit activity and therefore the power consumption, and also reduces the size of the counters. In addition, to obtain the value of the delay it is just necessary to read the position of the stage where the zero-crossing occurred, eliminating the need for searching the maximum of the outputs.

3.3. The Stereausis Approach

This approach is inspired in the stereausis network described in [3] and uses two cochlea channels to pre-process the microphones input signals. Indeed, the sound from the left and right microphones are fed to two cochlea channels. All outputs are quantized to 1 bit and the outputs of every stage of one channel are digitally correlated with the outputs of the other channel. In this way, a spatial arrangement of elements results, which can be associated to an image C, whose (i, j) element $C_{i,j} \geq 0$ represents the correlation between the output of the i-th element of the left cochlea and the output of the j-th element of the right cochlea (see Fig. 2). When the left and right signals are equal, the resulting image C will have a high density of nonzero elements along the main diagonal. However, if there is a delay in one of the signals, the image C will show a shift of the main diagonal towards one of the sides. The network that we used in the simulations consists of a 32–stages cochlea with cut-off frequencies between $252Hz$ and $618Hz$. Notice that as a delay of τ seconds is equivalent to a phase shift of $\phi(f) = 2\pi f\tau$, the higher the frequency the more noticeable the unbalance of the image with respect to the main diagonal (see [3]). The indication of time delay is calculated by measuring the unbalance of the image C with respect to the main diagonal. This is done by computing the difference between the sum of upper diagonal elements and lower diagonal elements, i.e.,

$$\Upsilon = \sum_{i<j} C_{i,j} - \sum_{i>j} C_{i,j}.$$

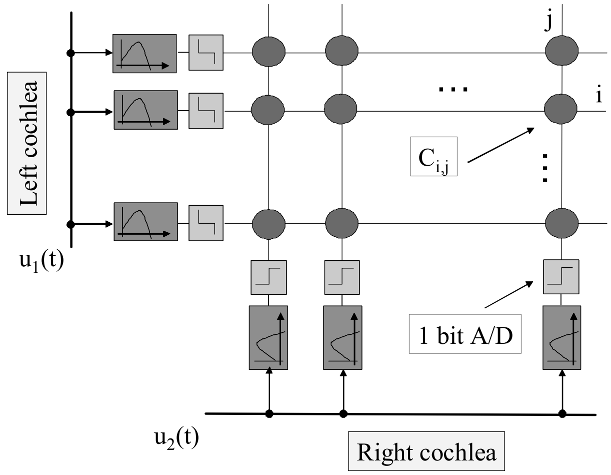

Figure 2: Architecture for estimation based on the stereausis approach.

3.4. Spatial Gradients Approach

In this approach, the signals recorded by the microphones are interpreted as samples of a sound field $s(\cdot)$ and the bearing angle is estimated using first order derivatives ([2]). This algorithm, in contrast to the previous cases, takes full advantage of the four microphones for the time delay estimation. For the present situation let us consider the position of the microphones with respect to the center of the array. We will assume that for any given location $\mathbf{r}$ in the plane, where $\mathbf{r} \in \mathbf{R}^2$, the magnitude $\tau(\mathbf{r})$ represents the time delay between the wavefront of the sound wave at $\mathbf{r}$ and the wavefront of the sound wave at the center of the array. Using this definition and a Taylor series, we can express the field $s(t+\tau(\mathbf{r}))$, $s : \mathbf{R}^1 \mapsto \mathbf{R}^1$ in a neighborhood of the origin as

$$s(t+\tau(\mathbf{r})) = s(t) + \tau(\mathbf{r}) \frac{d}{dt} s(t) + \frac{1}{2}\tau(\mathbf{r})^2 \frac{d^2}{dt} s(t) + O\left(\tau(\mathbf{r})^3\right)$$

To first order, and after geometric considerations, it can be easily seen that

$$\begin{gathered} x_A(t) \approx s(t) + \tau_2 \dot{s}\ , \ x_D(t) \approx s(t) + \tau_1 \dot{s} \\ x_B(t) \approx s(t) - \tau_2 \dot{s}\ , \ x_C(t) \approx s(t) - \tau_1 \dot{s} \end{gathered}$$

where $\tau_1 = \frac{1}{2}\frac{L}{c}\cos(\alpha)$, $\tau_2 = \frac{1}{2}\frac{L}{c}\cos(\beta)$ are the delays with respect to the coordinate axes. Then, a simple manipulation of the variables leads to

$$\begin{gathered} s(t) = \frac{1}{4}\left(x_A(t) + x_B(t) + x_C(t) + x_D(t)\right) \\ \tau_1 \dot{s} = \frac{1}{2}\left(x_D(t) - x_C(t)\right), \tau_2 \dot{s} = \frac{1}{2}\left(x_A(t) - x_B(t)\right) \end{gathered} \qquad (6)$$

If we sample the signals with a sampling time $T_s = 1/f_s$, and assume that $ds(t)/dt$ at $t = kT_s$ can be adequately measured by filtering $s(kT_s)$, then (6) is a standard least squares problem and τ_1, τ_2 can be obtained independently after collecting $N+1$ samples.[1] This approach heavily relies on the accuracy of the signals measurement, especially to calculate the derivative with precision. Due to this, in this case the amplitude cannot be quantized. In practice, the original signal was used with the original sampling rate of 2048 samples per second, and the derivative was calculated using finite differences.

Table 1: Accuracy of the algorithms (STD) in degrees

	Corr.	Ster.	Spatial Gr.
STD	1.18	1.47	0.87

4. COMPARISON OF RESULTS

Based on the collected data, we used the mean to define the transfer curve time delay–angle, and the standard deviation to quantify the precision. As was shown in [4], the time delay variation corresponding to a change of one degree at an angle β^* is

$$\Delta D|_{1^0} = D_{\max} \sin(\beta^*)\,\pi/180. \qquad (7)$$

In the present case, it is useful to quantify the error in degrees. This requires a conversion of the measurement from seconds to degrees. Accordingly, if the reading of a certain time delay has a standard deviation of σ_T, then the standard deviation in degrees is given by

$$\sigma_D = \frac{\sigma_T}{D_{\max}\sin(\beta^*)\,\pi/180}$$

The correlation and correlation derivative approach give indistinguishable results, therefore, in what follows we are only referring to the latter approach. Figure 3 shows the mean value of the output corresponding to the three algorithms in the range $\{\alpha \in [0,50] \cup [+130,+180]\}$; Fig. 4 shows the standard deviation corresponding to the range $\{\alpha \in [0,90]\}^2$. From this figure, it can be seen that whenever the pair of microphones $A-B$ is used, the precision of the estimation deteriorates in the range $\{\alpha \in [50,130]\}$. However, the other pair of microphones can be used in this range to obtain the same accuracy. Accordingly, Table 1 summarizes the average standard deviation of the three algorithms in the range $\{\alpha \in [0,50]\}$.

5. CONCLUSIONS

We have compared four different algorithms for sound localization. Two of the algorithms were previously reported in the literature, and the other two were developed specifically for this application. The spatial gradients method shows the best accuracy results, but from an implementation viewpoint it requires a sampled data analog architecture able to solve adaptively an LMS problem. The stereausis based approach shows acceptable results but requires a two dimensional array of correlators in addition to the two

[1] Similar results can be obtained using adaptive algorithms.

[2] This plot is symmetric with respect to $\alpha = 90°$, so for the sake of clarity only the portion between 0° and 90° is shown.

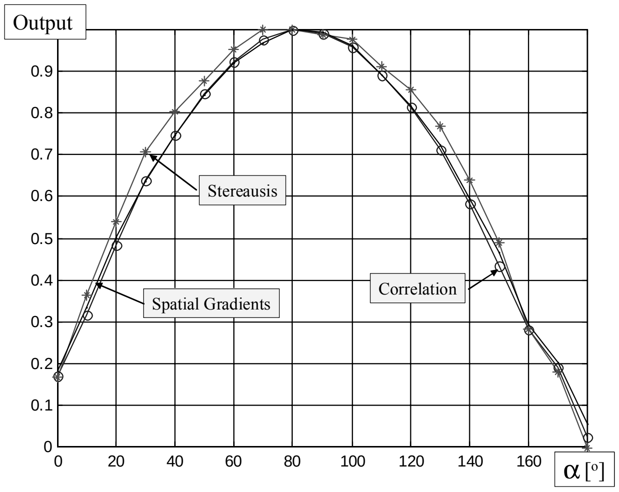

Figure 3: Mean value of the output for the three algorithms in the range of interest given by $\{\alpha \in [0, 180]\}$.

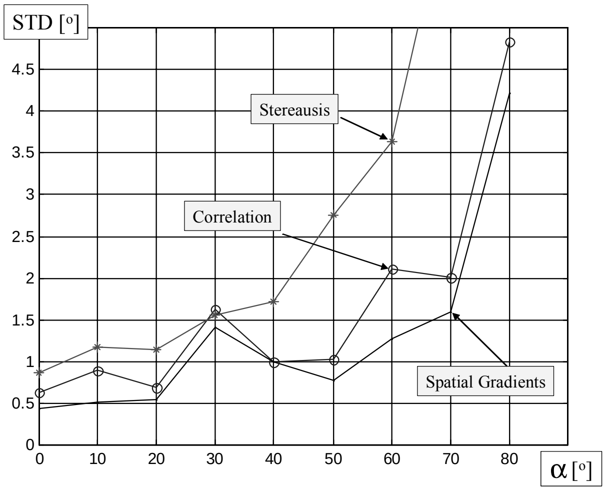

Figure 4: Standard deviation in degrees for the three algorithms in the range of interest given by $\{\alpha \in [0, 90]\}$.

cochlea filter channels. Finally, the correlator derivative approach shows an accuracy very close to the spatial gradients approach, but it offers a very convenient architecture, evidenced not only by its simplicity but also by the associated low power consumption due to the low temporal activity. Experimental results of an integrated circuit that implements this approach can be found in the companion paper [4].

6. REFERENCES

[1] G. C. Carter, "Time delay estimation for passive sonar signal processing," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-29, pp. 463–470, June 1981.

[2] G. Cauwenberghs, M. Stanacevic, and G. Zweig, "Blind broadband source localization and separation in miniature sensor arrays," in *Proc. of the IEEE Int. Symp. on Circuits and Syst. (ISCAS)*, vol. III, pp. 193–196, 2001.

[3] S. Shamma, N. Shen, and P. Gopalaswamy, "Stereausis: Binaural processing without neural delays," *J. Acoust. Soc. Am.*, vol. 86, pp. 989–1006, 1989.

[4] P. Julian, A. Andreou, P. Mandolesi, and D. Goldberg, "A low-power CMOS integrated circuit for bearing estimation." to appear in Proc. of the IEEE Int. Symp. on Circuits and Syst. (ISCAS), 2003.

[5] G. C. Carter, "Coherence and time delay estimation," *Proc. IEEE*, vol. 75, pp. 236–255, February 1987.