ACOUSTIC SOURCE LOCALIZATION IN DISTRIBUTED SENSOR NETWORKS

Thibaut Ajdler[*◇], *Igor Kozintsev*[◇], *Rainer Lienhart*[◇], *Martin Vetterli*[*]

[◇]Intel Labs, Intel Corporation, Santa Clara, USA
[*]Ecole Polytechnique Federale de Lausanne (EPFL), LCAV, Lausanne, Switzerland

ABSTRACT

This paper studies the problem of sound source localization in a distributed wireless sensor network formed by mobile general purpose computing and communication devices with audio I/O capabilities. In contrast to well understood localization methods based on dedicated microphone arrays, in our setting sound localization is performed using a sparse array of arbitrary placed sensors (in a typical scenario, localization is performed by several laptops/PDAs co-located in a room). Therefore any far-field assumptions are no longer valid in this situation. Additionally, localization algorithm's performance is affected by uncertainties in sensor position and errors in A/D synchronization. The proposed source localization algorithm consists of two steps. In the first step, time differences of arrivals (TDOAs) are estimated for the microphone pairs, and in the second step the maximum likelihood (ML) estimation for the source position is performed. We evaluate the Cramer-Rao bound (CRB) on the variance of the location estimation and compare it with simulations and experimental results. We also discuss the effects of distributed array geometry and errors in sensor positions on the performance of the localization algorithm. The performances of the system are likely to be limited by errors in sensor locations and increase when the microphones have a large aperture with respect to the source.

1. INTRODUCTION

Arrays of audio sensors and actuators (microphones and loudspeakers) along with array processing algorithms (beamforming, speaker tracking, 3D audio) offer a rich set of new features for emerging multimedia applications. In the past, audio array processing required expensive dedicated sensor arrays, multi-channel I/O cards and high-throughput computing systems due to the requirement to process all channels on a single machine. Recent advances in mobile computing and communication technologies, however, suggest a novel and very attractive platform for implementing these algorithms. Students in classrooms and co-workers at meetings are nowadays accompanied by several mobile computing and communication devices with audio and video I/O capabilities onboard such as laptops, PDA's, and tablets. In addition, high-speed wireless network connections, like IEEE 802.11a/b/g, are available to network those devices. Such ad-hoc sensor/actuator networks can in the future replace dedicated arrays. However, several technical and theoretical problems have to be addressed in order for this to happen. In a particular case of audio processing we need to adapt the existing algorithms that are based on certain assumptions about existing underlying hardware structure to new features of distributed arrays: arbitrary position of sensors, possibility to form sparse arrays over large areas, uncertainty in exact sensor locations.

In this paper, we address the problem of a single sound source localization in a distributed sensor environment. We propose a method and a statistical model for the localization problem and derive the performance bounds. Our experimental results demonstrate that the sound source localization can be efficiently implemented using a distributed sensor network. Many existing techniques aiming at solving the problem of source localization make certain assumptions regarding the location of the source (e.g. far field assumption) or the spacing between the microphones. For example, most beamforming-based approaches [1] require the distance between microphones to be at least a half of the wavelength of the emitted sound. Trying to localize a source emitting at a higher frequency is often difficult because of grating lobes in beam patterns of microphone arrays [1]. Another popular assumption used in sound localization literature is the one of the far-field geometry [2]. This assumption is justified in practice by small apertures of microphone arrays and allows approximating sound by a plane wave. In our scenario, both of these assumptions are no longer valid since microphones can have a large spacing between them and the aperture of the distributed array may be considerable (compared to the distance to the sound source).

The rest of the paper is organized as follows. Section 2 formulates the problem of sound source localization using array of microphones, describes our approach to solving this problem and proposes a statistical model and corresponding Cramer-Rao bound on the performance of the localization. In Section 2.5 we discuss the effect of distributed array geometry and sound source position on performance of localization algorithm. In Section 2.6 we investigate the effect of sensor position errors on the sound source localization performance. Experimental results are presented in Section 3 and the conclusions are drawn in Section 4.

2. SYSTEM DESCRIPTION

In order to solve our problem of source localization, we use a two steps technique. The first step determines the TDOAs different pairs of microphones. The second step finds the ML estimation for the source location from the knowledge of the TDOAs and the microphone positions. Once the TDOA of a microphone pair is obtained, we expect the source to be located on an hyperboloid whose foci are the two microphones positions. Ideally, with the knowledge of 4 TDOAs a unique solution can be found. In a practical situation, with ambient noise and reverberation the 4 hyperboloids will never intersect in a single point. We will then need to find the point that has the ML to be the intersection point and the source position. As considered in the literature [3, 4], we assume in this work the measurement error of the times of arrivals to be zero mean Gaussian and independent for each microphone.

2.1 Time Difference of Arrival: estimation and modeling

In this work we consider the case where there is only one active source (at each moment of time) captured by M synchronized microphones. The steps taken to calculate the TDOA for one microphones pair are outlined in Figure 1.

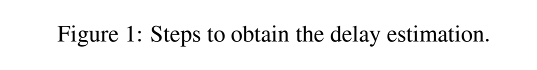


Figure 1: Steps to obtain the delay estimation.

Preprocessing

Initially, energy based silence detection is performed to check

Thibaut Ajdler is with the Laboratory of Audio Visual Communications (LCAV), Ecole Polytechnique Federale de Lausanne, 1015 Lausanne, Switzerland. E-mail:thibaut.ajdler@epfl.ch. The work was performed while he was an intern at Intel Labs, Intel Corporation, Santa Clara, USA.

wether the signal is present. When the energy over a window is below a threshold this channel is rejected and muted. Next, the input signal is preemphasized using a two-tap high-pass filter ($[1, -.95]$) [5]. This high pass filtering has been demonstrated to be able to attenuate the low frequency part of the signal where the TDOA estimation is less reliable due to the presence of low frequency noise (e.g. computer fan).

GCC-PHAT

The TDOA between microphone i and microphone j originating from the source is defined as $TDOA_{ij}$. The TDOA is defined as:

$$TDOA_{ij}(s) = \frac{|s-m_i| - |s-m_j|}{c} \quad (1)$$

with s being the source position, m_i the position of the i^{th} microphone and c the speed of sound. In order to measure TDOAs for the microphone pairs, we use the Generalized-Cross-Correlation method described in [6]. We choose the GCC-PHAT method that reduces the degradation due to reverberation [6]. This method finds the maximum of the normalized cross-correlation between the two input signals. The analysis window used in our experiments has duration of $60ms$ with an overlap of $20ms$.

Interpolation

To increase the precision of the TDOA estimation and to be able to get better approximation for sound source location, we perform interpolation of the normalized cross-correlation before finding the maximum [7] using a windowed sinc filter. This allows us to work with sub-sample precision in TDOA estimation.

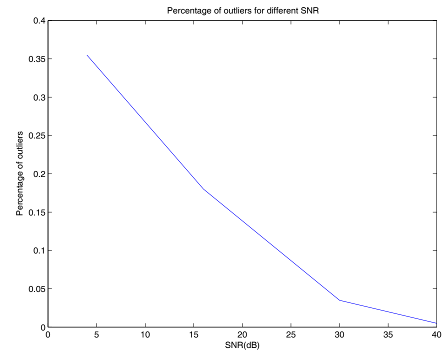
Outliers Rejection

Using the technique explained so far, we estimated TDOAs from recorded measurements for SNR ranging between $40dB$ and $4dB$. With lower SNR values, two different effects were observed. Firstly, the number of TDOA estimates that significantly differ from the true value (outliers) increases. We define outliers as TDOA estimates that are more than 1.5 samples away from the correct value [5]. By rejecting these outliers the dataset becomes smaller and source localization has to be performed on fewer windows. In Figure 2(a), we plot the percentage of outliers in the experiments as a function of SNR. For high SNR, the percentage is almost zero while for smaller SNR values the percentage can significant (above 35 percent). Secondly, we see that even in case of low SNR values, the standard deviation of remaining TDOA estimates remains relatively small (well under one sample). Figure 2(b) suggests that the standard deviation of the TDOA estimates (excluding outliers) varies from around .05 to .2 samples for a wide range of SNR values. Approximately 200 non-overlapping windows were analyzed and an histogram of the non-outliers delay estimations provided by one microphones pair is shown in Figure 3. These values were obtained using an interpolation factor of 32. The histogram suggests that Gaussian distribution is a good candidate to model the noise in TDOA estimation. In practice, we detect outliers by rejection of peaks that are not in the range of physically possible positions. Also, when tracking a source, another solution to remove outliers could be to only keep consistent peaks detected in a certain interval around the preceding estimation. This is based on the idea that in time, if the source is not moving very fast, the new estimated TDOA will still be in the neighborhood of the preceding TDOA. In this paper, however, we do not discuss the tracking of moving acoustic sources.

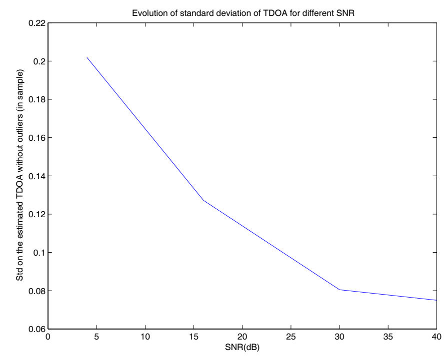
As conclusion, we see that TDOA estimations remain reliable even for low SNR when the outliers are removed. At the same time Gaussian noise model leads us to the following TDOA estimation approximation:

$$\tau = t + \eta \quad (2)$$

with η being a zero-mean white Gaussian noise, $\eta = N(0, \sigma)$, t being the exact TDOA and τ being the measured TDOA.



(a)



(b)

Figure 2: (a) Percentage of outliers in TDOA estimation as a function of SNR. (b) Standard deviation of the non outliers for increasing SNR.

2.2 Maximum Likelihood Estimation

Consider all the TDOA between the different microphones and a reference microphone. Let i represent the index going over these $M-1$ different microphone pairs. Define Δ as the vector of all the exactly calculated TDOAs, Γ the vector containing the noisy measurements and Σ the covariance matrix of the noise vector.

$$\Delta = [t_1, t_2, \ldots t_{M-1}],\ \Gamma = [\tau_1, \tau_2, \ldots \tau_{M-1}].$$

The likelihood function of Γ given s is

$$p(\Gamma; s) = (2\pi)^{-(M-1)/2} |\Sigma|^{-1/2} e^{-\frac{1}{2}(\Gamma-\Delta)^T \Sigma^{-1} (\Gamma-\Delta)} \quad (3)$$

We assume the estimation error of the times of arrival to be zero mean Gaussian and independent for each microphone. If we call $\frac{\sigma_i^2}{2}$ the variance of this error for the i^{th} microphone and $\frac{\sigma_0^2}{2}$ this variance for the reference microphone, the matrix Σ can be written as follows:

$$\Sigma = \sigma_0^2 \begin{pmatrix} \frac{1}{2}(1+\frac{\sigma_i^2}{\sigma_0^2}) & \frac{1}{2} & \cdots & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}(1+\frac{\sigma_i^2}{\sigma_0^2}) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{1}{2} \\ \frac{1}{2} & \cdots & \frac{1}{2} & \frac{1}{2}(1+\frac{\sigma_i^2}{\sigma_0^2}) \end{pmatrix}. \quad (4)$$

We diagonalize the matrix Σ,

$$\Sigma = U \Lambda V^T = U \Lambda^{1/2} \Lambda^{1/2} V^T,$$

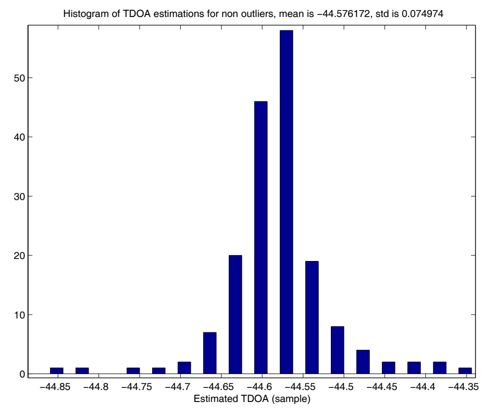


Figure 3: Histogram for the interpolated TDOA estimation.

with Λ being a diagonal matrix. Using the new eigenvector bases, we represent the vector$(\Gamma-\Delta)^T V \Lambda^{-\frac{1}{2}}$ by $(\Gamma'-\Delta')^T$. It can then be shown that the ML estimate $\hat{s_{ML}}$ minimizes the following function:

$$f_{ML}(s) \quad = \quad \sum_{i=1}^{M-1} \frac{(\tau_i' - t_i'(s))^2}{2} \tag{5}$$

We have:

$$\hat{s_{ML}} \quad = \quad \arg_s \min\left[f_{ML}(s)\right]. \tag{6}$$

This function can be minimized using standard numerical optimization methods. We use Levenberg-Marquardt (L-M) method to solve the problem. One known difficulty in applying this method is the requirement to find a good initial guess for the algorithm. In our experiments we exploit the closed-form solution derived using the spherical interpolation method as in [8]. Despite the fact that the accuracy of the initial guess is quite sensitive to sensor noise and uncertainties in microphone positions, we found the spherical interpolation approximation to be a good starting point.

2.3 Theoretical Cramer-Rao bound

The CRB gives a lower bound on the variance of any unbiased estimate. In this section, we derive the CRB for the estimates of the position of the source. Let T be the vector of all the exactly calculated TDOAs, and Γ the vector containing the noisy measurements. The variance of any unbiased estimator of Φ is bounded:

$$E[(\hat{\Phi}-\Phi)(\hat{\Phi}-\Phi)^T] > J^{-1}(\Phi), \tag{7}$$

where $J(\Phi)$ is called the Fischer's information matrix. It is given by

$$J = E\left\{\left[\frac{\partial \log p(\Gamma;\Phi)}{\partial \Phi}\right]\left[\frac{\partial \log p(\Gamma;\Phi)}{\partial \Phi}\right]^T\right\} \tag{8}$$

The likelihood function can be written as

$$p(\Gamma;\Phi) = (2\pi)^{-(M-1)/2}|\Sigma|^{-1/2}e^{-\frac{1}{2}(\Gamma-\Delta)^T\Sigma^{-1}(\Gamma-\Delta)} \tag{9}$$

with Φ corresponding to the three coordinates of the source position. Using the generalized chain rule, we obtain that:

$$J = \left[\frac{\partial \Delta(\Phi)}{\partial \Phi}\right]^T \Sigma^{-1}\left[\frac{\partial \Delta(\Phi)}{\partial \Phi}\right] \tag{10}$$

As shown in (7), in order to calculate the minimal variance of the estimates of the position of the source, we need to calculate the trace of the inverse of the Jacobian matrix. Each diagonal element of the inverse Jacobian matrix corresponds to the minimal variance of one coordinate of the source position, e.g. J_{11}^{-1} correspond to the minimal variance of the x-coordinate of the source position. The total variance on the estimation of the source position is the trace of the inverse Jacobian matrix. If one considers all σ_i to be equal, it can be seen from (4) and (10) that the bound on the variance of the measurements is directly proportional to the variance of the noise.

2.4 Monte Carlo simulations

We performed Monte Carlo simulations to study the performance of the Levenberg-Marquardt algorithm and compare it to the CRB. We performed 2000 source localization estimations using different noise realizations. In the simulations the positions of the microphones followed a normal distribution with standard deviation of 2 meters. The results of the simulations are illustrated in Figure 4. In Figure 4, one can see that for noise deviation ranging from 10^{-8}

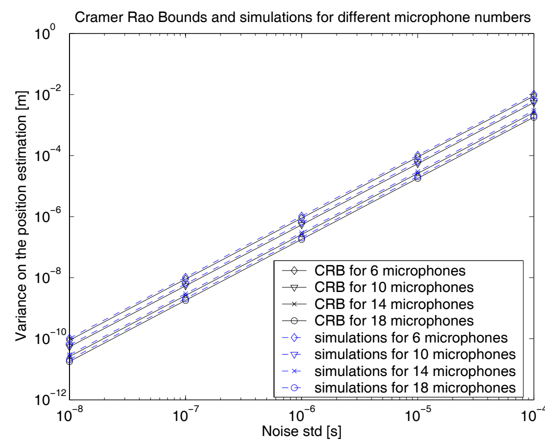


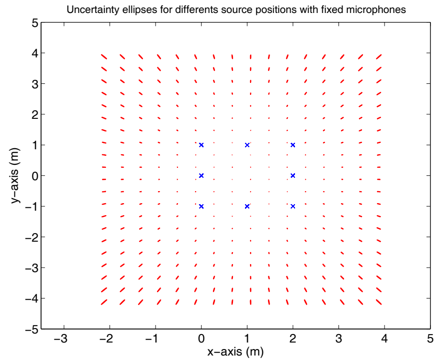
Figure 4: CRB and Monte Carlo simulations for different number of microphones. The theoretical bounds are shown in solid lines and the simulations are shown in dashed lines.

to 10^{-3} s, the performance of the algorithm follows a linear pattern parallel to the theoretical CRB. Therefore, the variance of the position estimates obtained with L-M method is proportional to the variance of the noise.

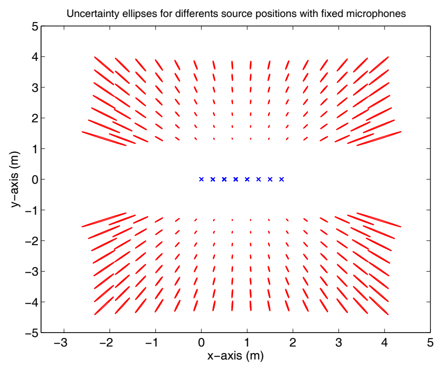
2.5 Effect of array geometry

To study the effect of relative geometry of sensor array and source, we show the uncertainty ellipses (obtained using the CRB for a given microphone array geometry) in Figure 5. For simplicity we present the graphs in 2 dimensions. Figure 5(a) shows a typical distributed sensor setup where the microphones are located with a large spacing in a rectangular shape and Figure 5(b) presents the same number of microphones along a line as it is usually the case in the microphone arrays setup. We first study the case where the microphones are located along a line as in Figure 5(b). For a source in front of the array at small distance, the aperture formed by the microphones with respect to the source is large. The uncertainty ellipse is small and can be explained by Figure 6(a). It shows hyperboles with their domain of uncertainty corresponding to TDOAs measured by 2 microphone pairs (pairs $(m1,m3)$ and $(m2,m3)$). The hyperboles intersect each other in a small area what explains the small uncertainty[1]. When the source is located on the far side

[1] Remark that in the specific case of an omnidirectional linear microphone array, two symmetrical source positions are possible. Using directional microphones or adding a microphone outside of the line would solve this uncertainty. Also, the same analysis applies in three dimensions using hyperboloids.



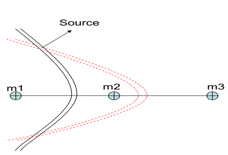
(a)



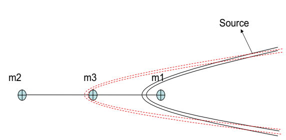
(b)

Figure 5: (a) Uncertainty ellipses in the case of a large distributed microphone array. (b) Uncertainty ellipses in the case of a line microphone array.

of the microphone array, Figure 5(b) shows a large increase in the uncertainty. The aperture becomes very small and in that case the hyperboles intersect each other in large areas as can be seen in Figure 6(b). In the case of distributed sensors shown in Figure 5(a), for a source located inside the rectangular shape, the uncertainty is very small because the aperture of the microphones with respect to the source is very large. The uncertainty ellipses only increases when the source gets far outside of the domain formed by the microphones, that is when the aperture becomes smaller.



(a)



(b)

Figure 6: hyperboles with their domain of uncertainty corresponding to TDOAs measured by 2 microphone pairs (pairs $(m1, m3)$ and $(m2, m3)$.(a) Intersection of hyperboles in small area. (b) Intersection of hyperboles in large area.

2.6 Effect of calibration error

An additional source of errors leading to errors in source localization estimation is the result of imprecise knowledge of microphone positions. The positions are either obtained manually by hand measurement or automatically by using auto-calibration schemes. In [4], results suggest that the localization error of the sensors positions remains of the order of a couple centimeters. To study the effect of these imprecisions, we performed 10000 simulations where white Gaussian noise of different standard deviation was added to the exact microphone positions (we used 8 microphones located at the 8 vertices of a cube). We varied the size of the side of the cube ranging from 3 m to .3 m. As can be seen from the Figure 7, the effect of the sensor location noise dependents on the spacing of the microphones. For the largest cube of side 3 m, the variance on the sound source location estimation is very small even for sizeable errors (on the order of 10 cm) in the positions of the microphones. When the side of the cube becomes smaller, however, the variance on the sound source position estimate becomes larger. With a side of .3 m, a standard deviation of 1 cm leads to a variance of approximatively 1 cm in the estimated source position. For higher errors on the positions in this cube, the algorithm does not converge anymore. This suggests that the uncertainty in the microphone positions can lead to high errors in the position estimation. Luckily, placing the microphones further away from each other allows us to keep small variance in the estimation. In the experiments in Section 2.1, we

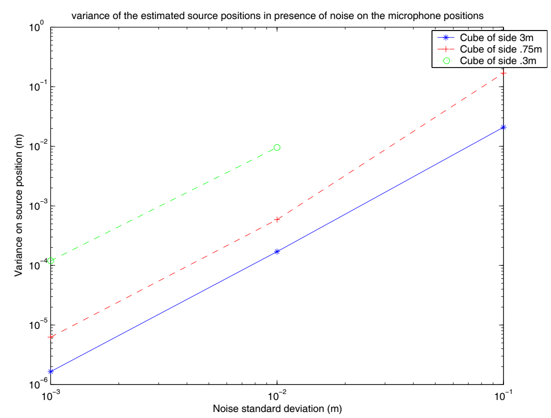


Figure 7: Variance on the source localization in the presence of errors in the microphone positions.

demonstrated that the TDOA measurements have a standard deviation of around .1 sample due to the sensor noise, therefore it can be seen on Figure 4 that the variance given by the CRB would have a value of about 10^{-5} m. As seen in Figure 7, the variances on the source estimation are much higher. Therefore, inaccurate measurement of the sensor positions is expected to be the dominant cause of errors for the sound source localization. Remark that in this setup, increasing the number of microphones helps reducing the variance on the source position estimation. Simulations were carried on using microphones positions generated with a standard deviation of 1 m and .6 m. On these positions a noise of 5 cm standard deviation was added. The simulation results are shown in Fig. 8. The results show that using more microphones will result in an increase of the precision of the microphone position estimation. In the present case, multiplying the number of microphones by a factor 6 leads to a reduction by a factor 10 on the variance of the position estimation.

3. EXPERIMENTS

To test the performance of the algorithm, we used a loudspeaker to simulate the sound source in a room of dimensions of about $3*5*3$ m. Some computers were present in the room adding noise. The microphones used were *Behringer ECM* 800 spaced by a distance of about 40 cm. Figure 9 shows the histogram of the estimates of the *x*-coordinate of the source. The estimated position remains nearly constant. The standard deviation of the measured source coordinate is on the order of 2 cm. The total variance on the source position is on the order of 10^{-2} m. Compared to the theoretical

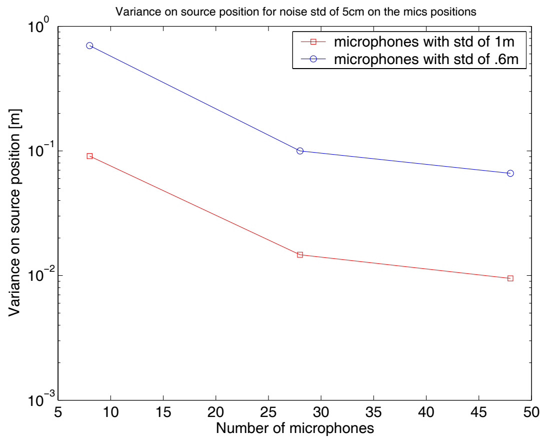


Figure 8: Variance on the source position in the presence of noise of standard deviation of 5 cm on the microphone positions in function of the number of microphones.

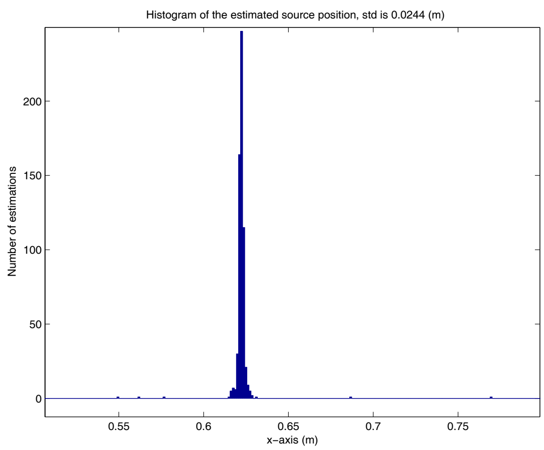


Figure 9: Histogram of the estimations of 1 coordinate of the source obtained using real measurements.

results presented in Figure 4, this variance appears to be high. We think that this result is due to inevitable errors in the measurement of the sensor positions. As our setup was arranged with distances between microphones on the order of approximatively .4 m, it is very likely that an error of the order of 1 cm was made during the calibration. By taking this error in consideration, the experimental variance matches well the simulation results. The precision provided by our source localization algorithm is on the order of several *cm* in typical environments and is surely sufficient for most target applications.

4. CONCLUSION

We presented an algorithm for a single sound source localization using a distributed microphone array of acoustic sensors. In our method we modify a conventional TDOA estimation procedure to make it robust to noise. We investigated the dependence of the performance of the proposed algorithm on different relative geometry of the array and sound source, and on uncertainties in sensor positions. Our theoretical and experimental results indicate that a large aperture sparse arrays offer high precision (on the order of several *cm* in typical environments) of source localization. Our results suggest that arrays of general purpose computing platforms with on-board audio I/O devices represent an attractive alternative to conventional dedicated microphone arrays.

REFERENCES

[1] B. D. Van Veen and K. M. Buckley, ”Beamforming: A Versatile Approach to Spatial Filtering,” *IEEE Acoust., Speech, Sig. Process. Magazine*, pp. 4-24, Apr. 1988.

[2] J. DiBiase, H. Silverman, and M. Brandstein, ”Robust localization in reverberant rooms,” *In M. Brandstein and D. Ward, editors, Microphone Arrays*, chapter 8, pp. 157 180. Springer, 2001.

[3] Y.T.Chan, K.C.Ho, ”A Simple and Efficient Estimator for Hyperbolic Location”, *IEEE Trans. on Signal Processing*, Vol.42, No.8, Aug. 1994.

[4] V. Raykar, I. Kozintsev, R. Lienhart, ”Position Calibration of Audio Sensors and Actuators in a Distributed Computing Platform,” *Submitted to IEEE Trans. on Acoustics, Speech and Signal Processing*, Aug. 2003.

[5] D. Bechler , K. Kroschel, ”Confidence Scoring of Time Difference of Arrival Estimation for Speaker Localization with Microphone Arrays, ” 13. Konferenz ”Elektronische Sprachsignalverarbeitung (ESSV)”, Sept. 2002, Dresden, Germany.

[6] C. H. Knapp and G. C. Carter, ”The generalized correlation method for estimation of time delay,” *IEEE Trans. on Acoustics, Speech and Signal Processing*,vol. 24(4), pp. 320-327,1976.

[7] M. Omologo, P.Svaizer, ”Acoustic source location in noisy and reverberant environment using csp analysis,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97), pp. 921-924, April 1996.

[8] J.Smith, J.Abel, ”Closed-form least-squares source location estimation from range-difference measurements,” *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 35(12),pp. 1661-1669, 1987.