A CLOSED-FORM METHOD FOR FINDING SOURCE LOCATIONS FROM MICROPHONE-ARRAY TIME-DELAY ESTIMATES

Michael S. Brandstein, John E. Adcock, and Harvey F. Silverman

Laboratory for Engineering Man/Machine Systems
Division of Engineering
Brown University
Providence, RI 02912

ABSTRACT

The *linear intersection* (LI) estimator, a closed-form method for the localization of source positions given only the sensor array time-delay estimate information, is presented. The array is constrained to be composed of 4-element sub-arrays configured in 2 centered orthogonal pairs. A bearing line in 3-space is estimated from each sub-array and potential source locations are found via closest intersection of bearing line pairs. The final location estimate is determined by a probabilistic weighting of these potential locations. The LI estimator is shown to be robust and accurate, to closely model the ML estimator, and to outperform a representative algorithm. The computational complexity of the LI estimator is suitable for use in real-time microphone-array applications.

1. INTRODUCTION

Microphone-array systems can be used to determine the positions of active talkers and can be electronically steered to provide spatially-selective speech acquisition. Since it is steered electronically, a microphone-array's directivity pattern can be updated rapidly to follow a moving talker or to switch between several alternating or simultaneous talkers. These features make microphone-arrays an attractive alternative to single microphone systems for hands-free speech acquisition, especially those involving multiple or moving sources.

The ability of microphone-array systems to determine talker location makes them attractive for use in multimedia teleconferencing systems where the location of the talker can be used not only for steering the directivity of the microphone-array, but also for pointing cameras or determining binaural cues for stereo imaging.

In microphone-array systems a directly observable signal characteristic is inter-sensor delay. Extensive literature exists on the topic of inter-sensor delay estimation and subsequent source location [1]. To achieve a high update rate in real-time situations, a practical source-location procedure must be computationally inexpensive. This requirement favors closed-form solutions [2, 3, 4, 5, 6] over search-based methods.

This work partially funded by DARPA/NSF Grant IRI-8901882, and NSF grants MIP-9314625 and MIP-9120843

2. SOURCE LOCALIZATION PROBLEM

The locationing problem addressed here may be stated as follows. There are N pairs of sensors m_{i1} and m_{i2} for $i \in [1, N]$. The ordered triplet (x,y,z) of spatial coordinates for the sensors will be denoted by $\mathbf{m}_{i1}$ and $\mathbf{m}_{i2}$, respectively. For each sensor pair, a time-difference of arrival (TDOA) estimate, τ_i, for a signal source located at s is available. The true TDOA associated with a source, s, and the i^{th} sensor-pair is given by:

$$\mathcal{T}(\{\mathbf{m}_{i1}, \mathbf{m}_{i2}\}, \mathbf{s}) = \frac{|\mathbf{s} - \mathbf{m}_{i1}| - |\mathbf{s} - \mathbf{m}_{i2}|}{c} \quad (1)$$

where c is the speed of propagation in the medium. In practice, τ_i is corrupted by noise and in general, $\tau_i \neq \mathcal{T}(\{\mathbf{m}_{i1}, \mathbf{m}_{i2}\}, \mathbf{s})$. In addition to the τ_i, a variance estimate, σ_i^2, associated with each TDOA is also assumed to be available as a byproduct of the time-delay estimation procedure. Given these N sensor-pair, TDOA-estimate combinations:

$$\{\mathbf{m}_{i1}, \mathbf{m}_{i2}\}, \tau_i, \sigma_i^2 \qquad \text{for } i = 1, \ldots, N$$

it is desired to estimate the source location s.

If the TDOA estimates are assumed to be independently corrupted by additive Gaussian noise, the Maximum Likelihood (ML) estimate $\hat{\mathbf{s}}_{ML}$ is given by the least-squares estimate [7]:

$$\hat{\mathbf{s}}_{ML} = \arg\min_{\mathbf{s}} \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \cdot [\tau_i - \mathcal{T}(\{\mathbf{m}_{i1}, \mathbf{m}_{i2}\}, \mathbf{s})]^2 \right) \quad (2)$$

However, since $\mathcal{T}(\{\mathbf{m}_{i1}, \mathbf{m}_{i2}\}, \mathbf{s})$ is a non-linear function of s, this error criteria is nonconvex and the solution of (2) requires burdensome, and frequently problematic, numerical search methods. For these reasons, it is worthwhile to develop a closed-form source-location

0-7803-2431-5/95 $4.00 © 1995 IEEE

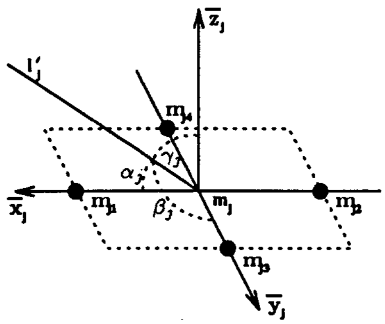

Figure 1: Quadruple sensor arrangement and local Cartesian coordinate system

estimator with performance features comparable to those of the less tractable ML estimate. The locationing procedure described below, termed the *linear intersection (LI)* method, satisfies both these conditions.

3. LI SOURCE LOCALIZATION ALGORITHM

For a specific sensor pair $\{\mathbf{m}_{i1}, \mathbf{m}_{i2}\}$ and their associated TDOA estimate, τ_i, the locus of potential source locations in 3-space forms one-half of a hyperboloid of two sheets. This hyperboloid is centered about the midpoint of $\mathbf{m}_{i1}$ and $\mathbf{m}_{i2}$ and has the directed line segment $\overline{\mathbf{m}_{i1}\mathbf{m}_{i2}}$ as its axis of symmetry. For sources with a large source-range to sensor-separation ratio, the hyperboloid may be well-approximated by the cone with vertex at the sensors' midpoint, having $\overline{\mathbf{m}_{i1}\mathbf{m}_{i2}}$ as a symmetry axis, and a constant direction angle relative to this axis. The direction angle, θ_i, for a sensor-pair, TDOA-estimate combination is given by:

$$\theta_i = \Theta(\{\mathbf{m}_{i1}, \mathbf{m}_{i2}\}, \tau_i) = \cos^{-1}\left(\frac{c \cdot \tau_i}{|\mathbf{m}_{i1} - \mathbf{m}_{i2}|}\right) \quad (3)$$

Now consider two pairs of sensors $\{\mathbf{m}_{j1}, \mathbf{m}_{j2}\}$ and $\{\mathbf{m}_{j3}, \mathbf{m}_{j4}\}$, where j is used to index the sets of sensor quadruples, along with their associated TDOA estimates, τ_{j12} and τ_{j34}, respectively. The sensors' placement positions are constrained to lie on the midpoints of a rectangle. A local Cartesian coordinate system is established with unit vectors defined as $\overline{\mathbf{x}_j} = \frac{\overline{\mathbf{m}_{j1}\mathbf{m}_{j2}}}{|\mathbf{m}_{j1} - \mathbf{m}_{j2}|}$, $\overline{\mathbf{y}_j} = \frac{\overline{\mathbf{m}_{j3}\mathbf{m}_{j4}}}{|\mathbf{m}_{j3} - \mathbf{m}_{j4}|}$, and $\overline{\mathbf{z}_j} = \overline{\mathbf{x}_j} \times \overline{\mathbf{y}_j}$ with the origin at the common midpoint of the two pairs, denoted by $\mathbf{m}_j$. This geometry is depicted in Figure 1. The first sensor-pair TDOA-estimate approximately determines a cone with constant direction angle, α_j, relative to the $\overline{\mathbf{x}_j}$ axis, as given by (3). The second specifies a cone with constant direction angle, β_j, relative to the $\overline{\mathbf{y}_j}$ axis. Each has a vertex at the local origin. If the potential source location is restricted to the positive-z half-space, the locus of potential source points common to these two cones is a line in 3-space. The remaining direction angle, γ_j, may be calculated from the identity

$$\cos^2\alpha_j + \cos^2\beta_j + \cos^2\gamma_j = 1$$

with $0 \leq \gamma_j \leq \frac{\pi}{2}$ and the line may be expressed in terms of the local coordinate system by the parametric equation

$$\mathbf{l}'_j = \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix} = r_j \begin{bmatrix} \cos\alpha_j \\ \cos\beta_j \\ \cos\gamma_j \end{bmatrix} = r_j \mathbf{a}'_j$$

where r_j is the range of a point on the line from the local origin at $\mathbf{m}_j$ and $\mathbf{a}'_j$ is the vector of direction cosines. The line $\mathbf{l}'_j$ may then be expressed in terms of the global Cartesian coordinate system via the appropriate translation and rotation. Namely,

$$\mathbf{l}_j = r_j \mathbf{R}_j \mathbf{a}'_j + \mathbf{m}_j$$

where $\mathbf{R}_j$ is the 3×3 rotation matrix from the j^{th} local coordinate system to the global coordinate system. Alternatively, if $\mathbf{a}_j$ represents the rotated direction cosine vector then

$$\mathbf{l}_j = r_j \mathbf{a}_j + \mathbf{m}_j$$

Given M sets of sensor quadruples and their corresponding bearing lines

$$\mathbf{l}_j = r_j \mathbf{a}_j + \mathbf{m}_j \qquad \text{for } j = 1, \ldots, M$$

the problem of estimating a specific source location remains. The approach taken here will be to calculate a number of potential source locations from the points of closest intersection for all pairs of bearing lines and use a weighted average of these locations to generate a final source-location estimate. More specifically, given two such bearing lines

$$\begin{aligned} \mathbf{l}_j &= r_j \mathbf{a}_j + \mathbf{m}_j \\ \mathbf{l}_k &= r_k \mathbf{a}_k + \mathbf{m}_k \end{aligned} \quad (4)$$

the shortest distance between the lines is measured along a line parallel to their common normal and is given by:

$$d_{jk} = \frac{|(\mathbf{a}_j \times \mathbf{a}_k) \cdot (\mathbf{m}_j - \mathbf{m}_k)|}{|\mathbf{a}_j \times \mathbf{a}_k|}$$

Accordingly, the point on $\mathbf{l}_j$ with closest intersection to $\mathbf{l}_k$ (denoted by $\mathbf{s}_{jk}$) and the point on $\mathbf{l}_k$ with closest intersection to $\mathbf{l}_j$ (denoted by $\mathbf{s}_{kj}$) may be found by first solving for the local ranges, r_j and r_k, and substituting these values into (4). The local ranges are found via solution of the overconstrained matrix equation:

$$\begin{bmatrix} \mathbf{a}_j & \vdots & -\mathbf{a}_k \end{bmatrix} \begin{bmatrix} r_j \\ r_k \end{bmatrix} = \begin{bmatrix} \mathbf{m}_k - \mathbf{m}_j + d_{jk} \cdot (\mathbf{a}_j \times \mathbf{a}_k) \end{bmatrix}$$

Each of the potential source locations is weighted based upon its probability conditioned on the observed set of $2M$ sensor-pair, TDOA-estimate combinations. The TDOA estimates are assumed to be normal distributions with mean given by the estimate itself. The weight associated with the potential source location, $\mathbf{s}_{jk}$, is calculated from:

$$W_{jk} = \prod_{l=1}^{M} P(\mathcal{T}(\{\mathbf{m}_{l1}, \mathbf{m}_{l2}\}, \mathbf{s}_{jk}), \tau_{l12}, \sigma_{l12}^2) \cdot P(\mathcal{T}(\{\mathbf{m}_{l3}, \mathbf{m}_{l4}\}, \mathbf{s}_{jk}), \tau_{l34}, \sigma_{l34}^2) \quad (5)$$

where $P(x, m, \sigma^2)$ is the value of a Gaussian distribution with mean m and variance σ^2 evaluated at x, i.e.

$$P(x, m, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp(\frac{-(x-m)^2}{2\sigma^2})$$

The final location estimate, which will be referred as the ***linear intersection estimate*** ($\hat{\mathbf{s}}_{LI}$), is then calculated as the weighted average of the potential source locations:

$$\hat{\mathbf{s}}_{LI} = \frac{\sum_{j=1}^{M} \sum_{k=1, k\neq j}^{M} W_{jk}\mathbf{s}_{jk}}{\sum_{j=1}^{M} \sum_{k=1, k\neq j}^{M} W_{jk}} \quad (6)$$

4. LOCATION ESTIMATOR EVALUATION

The ML estimate (2) is found via minimization of the sum-squared error of the differences between the observed TDOA and those of the hypothesized source. Because this error criterion is non-linear, (2) does not have a closed-form solution. For the location estimator presented here and those in the literature, the requirement of a closed-form solution necessitates the development of alternative error criteria. These alternative error criteria take several forms and vary in the degree to which they approximate the ML error criteria and perform in comparison to the ML estimator. A discussion of several of these closed-form estimators as well as a relative performance evaluation is presented in [6]. Smith and Abel found their estimation procedure, a linear least-squares approach termed the *spherical interpolation (SI)* method, to exhibit an RMS error superior to that of the estimators presented in [4] and [5].

As a means of evaluating the LI location estimator, the statistical characteristics of the LI and SI localization methods were compared through a series of Monte Carlo simulations modeled after those conducted in [6]. The experimental set-up, a nine-sensor orthogonal array with half-meter spacings and a source located at a range of 5 meters with equal direction angles, is depicted in Figure 2. The true TDOA values (1) were corrupted by additive white Gaussian noise. 100 trials were performed at each of 11 noise levels ranging from a standard deviation the equivalent of 10^{-3} meters to 10^{-1} meters when scaled by the propagation speed of sound in air ($c \approx 342\frac{\text{m}}{\text{s}}$). The LI method partitioned the array into 4 square sensor quadruples and required the evaluation of 8 TDOA estimates, one for each diagonal sensor-pair. The SI method required that all the TDOA values be relative to a reference sensor. The sensor at the origin was chosen for this purpose and the TDOA for each of the remaining 8 sensors relative to the reference were calculated. In addition to calculating the LI and SI estimates, the ML estimate (2) was computed via a quasi-Newton search method with the initial guess set equal to the true location (clearly this is not a practical algorithm since it requires prior knowledge of the actual source location).

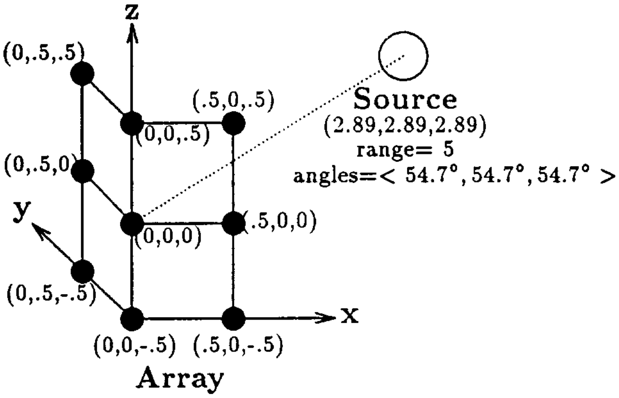

Figure 2: Experimental Set-Up: nine-element array (distances in meters)

Figures 3-5 summarize the results of these simulations. Figure 3 plots the sample bias for the estimated source bearing and range for each of the estimation methods as a function of the level of noise added to the true TDOA values. While each of the methods exhibits some degree of bias in the noisier trials, the situation is most extreme for the SI method. This tendency for the SI method to consistently bias its estimates towards the origin was noted by the authors of [6]. The LI method performs comparably to the ML estimate for all but the most extreme noise conditions. Figure 4 plots the sample standard deviations. For the standard deviation of the bearing estimates, a trend similar to the bearing bias is observed. The SI method's performance decays rapidly for noise levels above 10^{-2} meters. However, in terms of the range, each of the closed form estimators displays a smaller variance than the ML estimator at the higher noise conditions. This is a consequence of the estimator biases observed previously. Finally, Figure 5 shows the root-mean-square errors (RMSE).

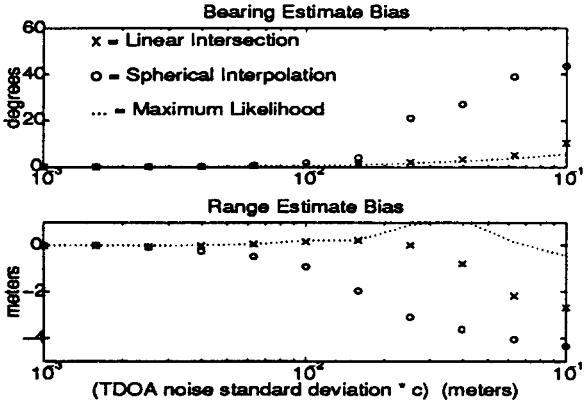

Figure 3: Location estimator sample bias

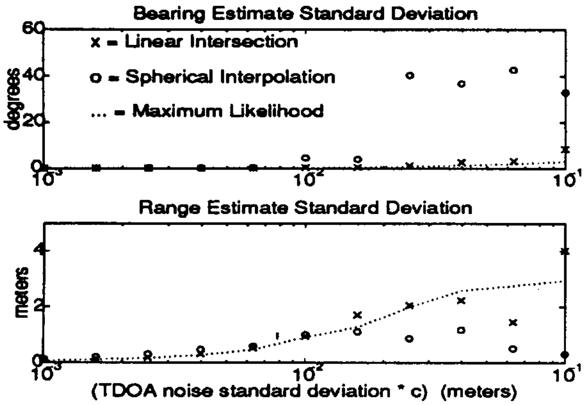

Figure 4: Location estimator sample σ

Once again, the LI method closely tracks the ML estimator in all but the most extreme condition while the SI method exhibits a marked performance decrease in both bearing and range for moderate and large noise levels.

Simulations performed over a broad range of source positions exhibit trends similar to those in Figures 3-5. The LI estimator is consistently less sensitive to noise conditions and possesses a significantly smaller bias in both its range and bearing estimates when compared to the SI estimator.

5. DISCUSSION

A closed-form method for the localization of source positions given only TDOA information has been presented. It was shown to be a robust and accurate estimator, closely modeling the ML estimator, and clearly outperforming a representative algorithm.

From an implementation standpoint, the constraint that the array be composed of rectangular 4-element sub-arrays is not problematic for typical room-oriented microphone-array applications. The *linear intersection* method has proven to be an effective source localization procedure when used in conjunction with a single 10-element planar-array in our laboratory and is easily

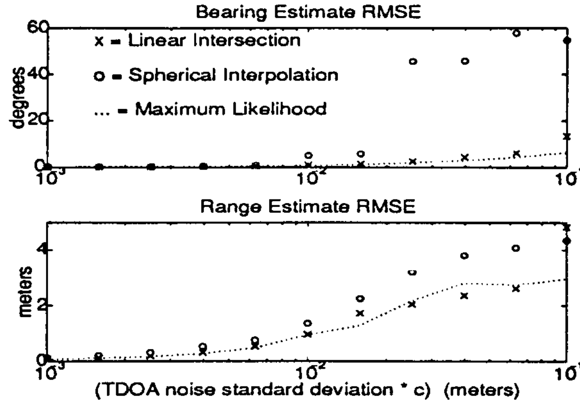

Figure 5: Location estimator root mean square error

extendible to real-time applications using more complex sensor arrangements. It is an advantage of the LI method that localization in 3-space can be performed with a 2-dimensional array. Also, since the LI method does not require the estimation of delays between sensors other than those in the local sub-array, the sub-arrays can be placed far apart and delay-estimation processing can be performed locally.

6. REFERENCES

[1] Special issue on time-delay estimation. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-29, Jun 1981.

[2] J. M. Delosme, M. Morf, and B. Friedlander. A linear equation approach to locating sources from time-difference-of-arrival measurements. In *Proceedings of ICASSP80*. IEEE, 1980.

[3] S. Reddi. An exact solution to range computation with time delay information for arbitrary array geometries. *IEEE Trans. Signal Processing*, 41(1):485–486, Jan 1993.

[4] H. Schau and A. Robinson. Passive source localization employing intersecting spherical surfaces from time-of-arrival differences. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-35(8):1223–1225, Aug 1987.

[5] R. Schmidt. A new approach to geometry of range difference location. *IEEE Trans. Aerosp. Electron.*, AES-8(6):821–835, Nov 1972.

[6] J. Smith and J. Abel. Closed-form least-squares source location estimation from range-difference measurements. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-35(12):1661–1669, Dec 1987.

[7] S. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, first edition, 1993.