Optimal Microphone Placement for Source Localization using Time Delay Estimation

E.A.P. Habets and P.C.W. Sommen
Eindhoven University of Technology, Dept. of Electrical Engineering,
P.O.Box 513, 5600 MB Eindhoven, The Netherlands
E.A.P.Habets@tue.nl, P.C.W.Sommen@tue.nl

Abstract— **A commonly used method for Acoustic Source Localization is based on Time Delay Estimation (TDE). In general discrete TDE algorithms are used which result in discrete Time Difference of Arrival (TDOA) estimates, i.e. range differences. The Position Location of an acoustic source can be calculated using a set of TDOAs. The accuracy of the position greatly depends upon the accuracy of the estimated TDOAs. In this paper it is shown that discrete TDOAs result in a non-uniformly distributed spatial resolution across the room furthermore it is shown how the sample frequency and the microphone distance effect the spatial resolution. Finally the effect upon spatial resolution for different microphone setups is demonstrated.**

Keywords—**Source localization, Time Delay Estimation, Optimal Microphone Placement.**

I. Introduction

In the last decade many research efforts have been devoted to microphone array processing techniques. Existing array systems have been used in a number of applications. These include teleconferencing, speech recognition, speaker identification, speech acquisition in an automobile environment, sound capture in reverberant enclosures, large room recording-conferencing, acoustic surveillance and hearing aids devices.

An essential requirement of these sensor-array systems is the ability to locate and track an audio source. For audio-based applications, an accurate fix on the primary talker, as well as knowledge of any interfering talkers or coherent noise sources, is necessary to effectively enhance a given source while simultaneously attenuating the undesired ones. Location data may be used as a guide for discriminating individual speakers in a multi-source scenario. With this information available, it would then be possible to automatically focus on and track a given source. Additionally, the speaker location estimates can be applied to steer a camera or series of cameras in a video-conferencing system. In this regard, the automated localization information eliminates the need for one or more human camera operators.

A commonly used method for Acoustic Source Localization is based on Time Delay Estimation (TDE) [1]. In general discrete TDE algorithms are used which result in discrete Time Difference of Arrival (TDOA) estimates, i.e. range differences. The Position Location of an acoustic source can be calculated using a set of TDOAs. Since the position is obtained in two dependent steps the solution is suboptimal, the amount of data is decreased in the first step which makes it very efficient, however the accuracy of the position greatly depends upon the accuracy of the estimated TDOAs.

In section II a general solution for estimating TDOAs is presented which holds for certain types of room impulse responses. The results are used in section III to calculate the position of the acoustic source. Better insight is obtained by looking at the properties of the obtained TDOAs in section III-A. Simulation results for different microphone setups are presented in section IV. In section V the conclusions are presented.

II. Time Delay Estimation

A widely used discrete signal model for TDE problems is the following. Let $x_i[n],\ i = 1, 2$, denote the i-th microphone signal, then:

$$x_i[n] = s[n] * h_i[\mathbf{r}_s, n] + \nu_i[n], \tag{1}$$

where $*$ denotes convolution and $h_i[\mathbf{r}_s, n]$ is the acoustic impulse response between the source $s[n]$ at position $\mathbf{r}_s$ and the i-th microphone. The additive noise signals, $\nu_1[n]$ and $\nu_2[n]$, might be correlated in case of directional noise, e.g. from a ceiling fan or overhead projector. The TDOA, denoted by D_{21} is caused by the difference in propagation time of the direct waves between the source and the spatially separated microphones. To estimate the TDOA we calculate the *difference room impulse response* (dRIR) using a LMS adaptive filter as shown in figure 1. Assuming the acoustic impulse responses are linear-time-invariant,

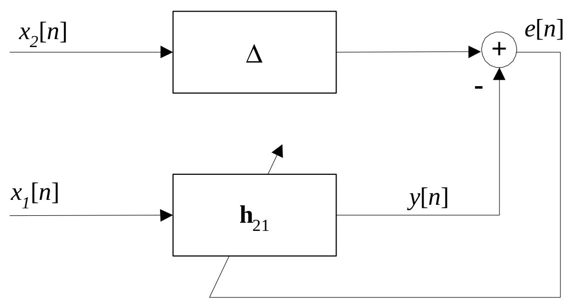


Fig. 1. LMSTDE.

the dRIR can be denoted as,

$$h_{21}[\mathbf{r}_s] = h_1^{-1}[\mathbf{r}_s] * h_2[\mathbf{r}_s] * \delta[-\triangle]. \quad (2)$$

The delay $\triangle$ is defined as $L/2$, where L is the length of the modelling filter h_{21}. This delay makes it possible to estimate both positive and negative TDOAs. In an anechoic environment the acoustic impulse responses can be represented using pure delays. The TDOA can now be calculated using,

$$\hat{D}_{21} = \arg\max_k h_{21}[\mathbf{r}_s, k] - \triangle,\ k = 0, \ldots, L-1. \quad (3)$$

This LMS TDE algorithms can be seen as an adaptive method to estimate the cross-correlation between $x_1[n]$ and $x_2[n]$ using the Roth Impulse Response weighting [1]. Other adaptive filter implementations may be used depending on the application, desired performance or complexity. It should be noticed that above method only works correct in case the amount of energy concentrated in both direct waves are dominant with respect to the reflections. This assumption could be violated in case the source-microphone distance is large, the maximum distance depends upon the position of the source and the amount of reverberation that is present.

The inner microphone distance (R_{21}) and the sample rate (f_s) that are used result in an upper bound for the maximum TDOA that can be measured. This upper bound can be used to determine an efficient filter length for h_{21},

$$\text{Number of filter taps} = \lceil 2 \cdot f_s \cdot \frac{R_{21}}{c} \rceil, \quad (4)$$

where c denotes the speed of sound in meters per second.

III. Position Location

Once the relative TDOA have been obtained they can be used to calculate the location of the source. The problem can be stated mathematically in the following way:
A microphone array of $N + 1$ microphones is located at,

$$\mathbf{r_i} \triangleq (x_i, y_i, z_i)^T, \quad i = 0, \ldots, N \quad (5)$$

in Cartesian coordinates. The acoustic source is located at $\mathbf{r}_s \triangleq (x_s, y_s, z_s)^T$. The distance between the source and the i-th microphone is denoted by,

$$R_{is} \triangleq ||\mathbf{r}_i - \mathbf{r}_s||_2 \quad (6)$$

$$R_{is} = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2 + (z_i - z_s)^2}. \quad (7)$$

The distance difference between microphones i and j from the source is given by,

$$d_{ij} \triangleq R_{is} - R_{js}, \quad i, j = 0, \ldots, N. \quad (8)$$

The difference is usually termed as the range difference and is proportional to the time delay of arrival with the speed of sound,

$$d_{ij} = c \cdot D_{ij}. \quad (9)$$

The localization problem then is to estimate $\mathbf{r}_s$ given the set of $\{\mathbf{r}_i\}$ and $\{\hat{D}_{ij}\}$. A huge variety of algorithms can be found in literature to solve this set of highly non-linear equations [2], [3], [4], [5]. In some methods, like the spherical-interpolation and -intersection, all TDOA measurements are relative to one microphone , which limits the number of microphone setups. We search for the position of the source $\mathbf{r}_s$, without any constraints with respect to the TDOA measurements, by minimizing the following cost function,

$$J(\tilde{\mathbf{r}}_s) = \sum_{\text{all desired i,j}} (D_{ij}(\tilde{\mathbf{r}}_s) - \hat{D}_{ij})^2. \quad (10)$$

The position of the source can now be found using,

$$\hat{\mathbf{r}}_s = \arg\min_{\tilde{\mathbf{r}}_s} J(\tilde{\mathbf{r}}_s). \quad (11)$$

During the simulation $\hat{\mathbf{r}}_s$ is estimated using a non-linear gradient decent algorithm. In figure 2 two contour plots for $J(\mathbf{r}_s = (300, 200))$ are shown for two different microphone setups. The positions of the microphones are denoted by stars at the bottom of each figure. From this figures we may conclude that the cost function is smooth and has a strong global minimum. The cost function for the microphone setup used in figure 2b is conditioned better compared to the setup used in figure 2a, which additionally results in a smaller sensitivity to noise. In both cases a minimum set (2) of TDOA is used to estimate the 2D position of the source. Evaluations for other source and microphone positions undermine these conclusions.

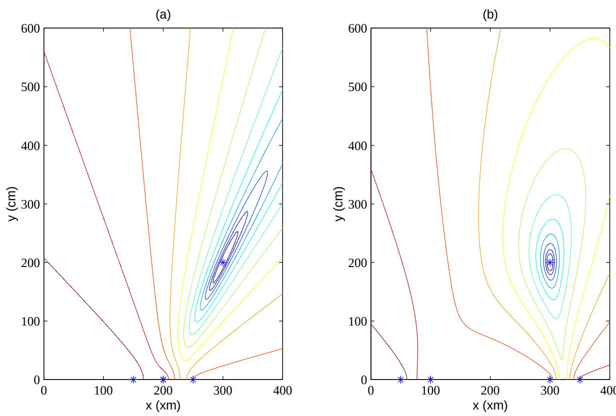


Fig. 2. Contour plots for $J(\mathbf{r}_s = (300, 200))$ in dB for different microphone setups.

A. Spatial Resolution

The accuracy of the position location depends on the on the accuracy of the the measured TDOAs, i.e. the range differences, that are obtained. In general a discrete TDE algorithm will be used, resulting in a discrete TDOA and range difference. The TDOA resolution can be increased using higher sample rates or by interpolating adjacent filter coefficients of the modelling filter h_{21}. In theory the TDOA and hence range difference can be obtained with infinite precision. Increasing the resolution using interpolation is however limited by the SNR of the signal. There is however an other possibility to increase the spatial resolution, namely increasing the inner microphone distance. By increasing the inner microphone distance the range difference (d_{ij}) increases for almost all source positions in the room. This effect is illustrated in figure 3. Here discrete range differences were calculated across the room using different microphone distances, each gray level represents a different range difference value. The results show a clear increase in spatial resolution for increasing inner microphone distances. The intersection at $y = 200$ cm and $y = 400$ cm for $R_{21} = 50$ cm is shown in figure 4a. In this figure it's clearly visible that the spatial resolution decreases for increasing source-microphone array distance. The intersection at $R_s = 200$ cm, the distance between the center of both microphones and the source, for $R_{21} = 50$ cm is shown in figure 4b. The spatial resolution is decreasing for larger off-broadside angles θ. It should however be noticed that accurate TDE for large inner microphone distances becomes harder due to the non-minimum phaseness of the RIRs and that the maximum distance is often limited due to the radiation pattern of the source.

B. Optimal Microphone Placement

Sensor placement is the one parameter of the system over which there is a considerable degree of freedom at installation time and no further control afterwards. While algorithms can be modified on the fly based on gathered statistics, and the room characteristics should be assumed to be unchangeable (any acoustic characteristic modifications required for proper installation are a major inconvenience and defeat the intent of microphone array use), microphone sensor placement may be chosen within a reasonable set of constraints. The problem of optimal microphone placement is translated in an analytic expression that can be used to find an optimal microphone setup.

A general expression must be derived that defines performance as a function of microphone placement:

$$PF(\mathcal{G}, M, \mathbf{r}_s, \hat{\mathbf{r}}_s) \tag{12}$$

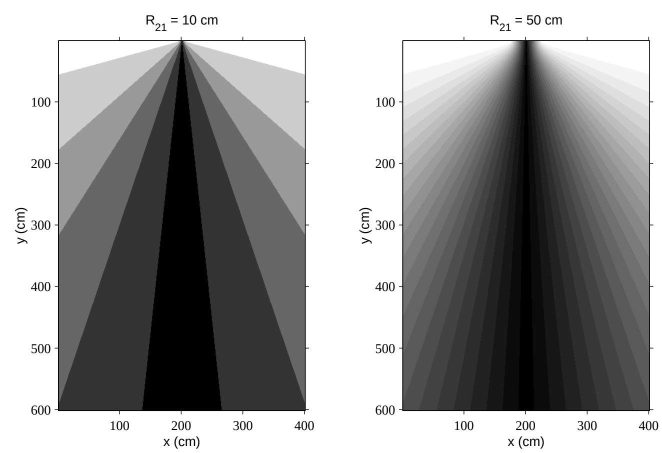


Fig. 3. Range differences at 16 kHz for different inner microphone distances.

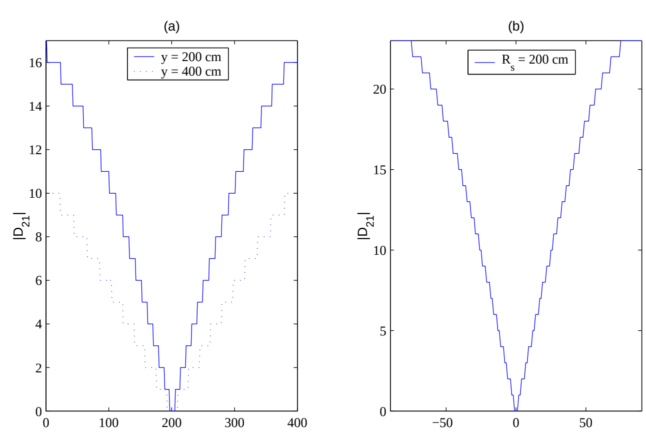


Fig. 4. For $R_{12} = 50$ cm (a) Range differences at $y = 200$ cm, and $y = 400$ cm, (b) Range differences at $R_s = 200$ cm.

where $\mathcal{G}$ defines the enclosure geometry and environment, M is an $N \times 3$ matrix of microphone coordinates, $\mathbf{r}_s$ is the single source location, and $\hat{\mathbf{r}}_s$ is the estimate of the source location. As a total performance score the average value over an acoustic space $\mathcal{A}$ could be used:

$$\overline{PF(\mathcal{G}, M, \mathbf{r}_s, \hat{\mathbf{r}}_s)} = \mathrm{E}[PF(\mathcal{G}, M, \mathbf{r}_s, \hat{\mathbf{r}}_s)] = \iint_{\mathcal{A}} p(\hat{\mathbf{r}}_s|\mathbf{r}_s)\, d\hat{\mathbf{r}}_s\, p(\mathbf{r}_s)\, d\mathbf{r}_s. \quad (13)$$

The three product terms in the integral in (13) must be expressed analytically in order to evaluate the expression.

The term $p(\mathbf{r}_s)$ represents the spatial probability distribution of the acoustic source, i.e. how likely is a source to be at a particular location. If the sound source is human talker, considerable information is available regarding the pattern of its behavior. It is probable that a person will talk from either a sitting or standing position, and that he/she is likely to use a chair etc. Although a highly accurate probabilistic source location model is difficult to obtain and varies with the application, a good heuristic model may be defined based on assumed human behavior.

The term $p(\hat{\mathbf{r}}_s|\mathbf{r}_s)$ defines the source estimation distribution that describes the accuracy of the source locator. An ideal locator would produce a distribution

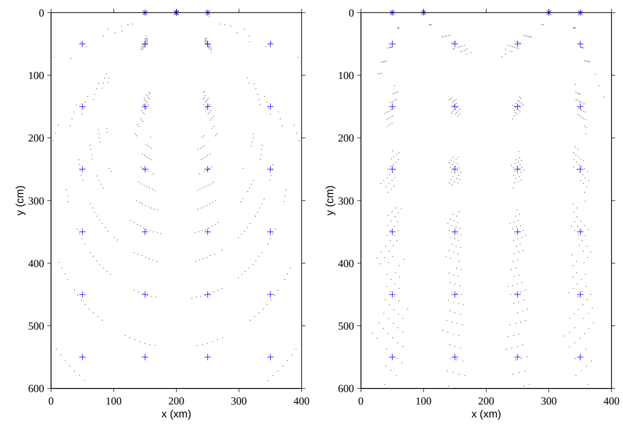
$$p(\hat{\mathbf{r}}_s|\mathbf{r}_s) = \delta(\mathbf{r}_s - \hat{\mathbf{r}}_s). \quad (14)$$

Which implies that the source position is estimated perfectly all the time at any given place. In practice, source location produces errors due to TDOA errors and non-uniform spatial resolution. Additionally the position location algorithm, which uses these TDOA estimates, and the microphone setup will finally determine the position error. Finding an exact analytic expression for this term is unlikely. Using some basic simplifications it is however possible to produce a good estimate depending on the algorithm that is used.

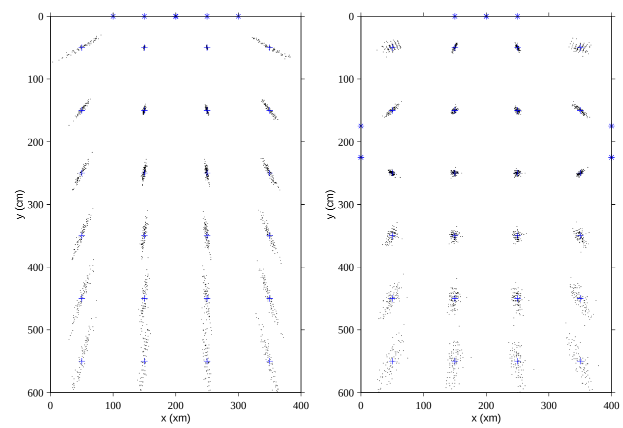
Once all quantities in (13) are defined, it remains to maximize the Performance Function over the constrained of allowable microphone positions $\{\mathcal{M}\}$:

$$M_{\text{best}}(\mathcal{G}) = \max_{M \epsilon \{\mathcal{M}\}} \overline{PF(\mathcal{G}, M)} \quad (15)$$

Although (13) is almost certain to not have a closed form solution, it may be evaluated numerically, or at least evaluated over a representative set of locations. Numerical methods may then be applied to maximize (15).



(a)2 TDOAs (b)2 TDOAs



(c)4 TDOAs (d)4 TDOAs

Fig. 5. Position Location for various source positions.

IV. Simulation Results

During the simulations in this section the discrete TDOAs are estimated with an accuracy of 1/44100 seconds. The continuous TDOA values are distorted with a white gaussian noise ($\sigma^2 = 0.5/44100$). The points in figure 5 represent the estimated source position. The sources used in this experiment are denoted by a plus-sign. Four different setups where used to demonstrate the effect of different microphone setups. All TDOAs used in 5a and 5c where relative to the center microphone. While estimating the position of the source using such setup the position can be calculated using spherical algorithms, which can be very efficient. The arbitrary microphone setup used in figure 5b and 5d is much more flexible. Compared to the setups a and c with the same number of TDOA there is a clear increase in spatial resolution. From these results one can clearly see the effect upon the increase of spatial resolution and the sensitivity of the noise as was already mentioned in section III.

V. Conclusions

The information presented in this paper gives a better insight with respect to the accuracy of the position location given a set of TDOAs. The performance of a source localization system can be increased significant using proper microphone array setups. A general expression is presented which can be used to find an optimal setup. The results presented here clearly show the advantage of arbitrary placed microphones in which not all TDOA are measured relative to one microphone.

References

[1] C.H. Knapp and G.C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-24, no. 4, pp. 320–327, August 1976.

[2] H.C. Schau and A.Z. Robinson, "Source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 8, pp. 1223–1225, August 1987.

[3] J. Smith and J. Abel, "The spherical interpolation method for source localization," *IEEE Journal of Oceanic Engineering*, vol. OE-12, no. 10, pp. 246–252, December 1987.

[4] Y. Chan and K. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Transactions. Signal Processing*, vol. 2, no. 8, pp. 1905–1915, August 1994.

[5] J. Benesty G.W. Elko Huang, Y.A. and R.M. Mersereati, "Real-time passive source localization: a practical linear-correction least-squares approach," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 943–956, November 2001.