Improved Time Delay Estimation Exploiting Redundancy in Microphone Arrays

Dirk Bechler and Kristian Kroschel

Institut für Nachrichtentechnik
Universität Karlsruhe, Kaiserstr. 12
D-76128 Karlsruhe, Germany
{bechler,kroschel}@int.uni-karlsruhe.de

Abstract: In general, a set of *Time Difference Of Arrival* (*TDOA*) estimates is used for 3D passive acoustic sound source localization with microphone arrays. In real environments the reliability of TDOA estimates is degraded due to noise and room reverberation. To determine the set of TDOAs, standard methods don't take advantage of all possible microphone pairs in a microphone array, using only one set of independent time delay estimates. In this work, the redundant information lying in the remaining dependent time delays is used to improve the robustness of the TDOA estimation. Experimental results for real data recorded in a noisy and reverberant office room are presented. The proposed methods show an absolute decrease of the percentage of false TDOA estimates of about 8 %, enhancing significantly the TDOA estimate reliability.

1 Introduction

The need of acoustic sound source localization is of interest in many technical systems. While acoustic surveillance and teleconferencing systems are traditional applications, the integration of acoustic perception into humanoid robots becomes nowadays a more and more important area of research [1].

The technique of choice in most passive acoustic sound source localization systems using a microphone array is a two-step procedure. First, the *Time Delays Of Arrival* (*TDOA*) in microphone pairs of the sensor array are estimated. In a second step, these TDOAs are used together with the microphone array geometry to determine the position of the active sound source. Thereby, usually not every possible time delay in the sensor array but only one set of independent TDOA estimates is used, neglecting the redundant information lying in the residual dependent TDOAs. The most common technique to estimate the TDOAs is the Generalized Cross Correlation (GCC) method [2]. While computationally very efficient, this method has big problems in realistic acoustic environments. The reliability of the TDOA estimates and consequently the robustness of the localization suffer severely if the room reverberations rise above minimal levels [3].

Using confidence criteria for the TDOA estimates found in [4, 5], this work presents methods to determine the most confident independent TDOAs out of all possible ones to increase the robustness of the acoustic source localization system.

2 Problem Definition

With N microphones, $\binom{N}{2} = \frac{N(N-1)}{2}$ possible microphone pairs can be generated, but only $N-1$ pairs are independent. As an example, the case with $N = 3$ microphones as shown in Fig. 1 is regarded. The relation between the TDOA τ_{ij} in the microphone pair M_iM_j with the Cartesian coordinates of M_i $(\mathcal{X}_i, \mathcal{Y}_i, \mathcal{Z}_i)$ and the sound source position $(\mathcal{X}_s, \mathcal{Y}_s, \mathcal{Z}_s)$ can be written as

$$\begin{aligned}\tau_{ij} \;=\;& \frac{1}{c}\Big(\sqrt{(\mathcal{X}_s-\mathcal{X}_i)^2+(\mathcal{Y}_s-\mathcal{Y}_i)^2+(\mathcal{Z}_s-\mathcal{Z}_i)^2}\;- \\ & \sqrt{(\mathcal{X}_s-\mathcal{X}_j)^2+(\mathcal{Y}_s-\mathcal{Y}_j)^2+(\mathcal{Z}_s-\mathcal{Z}_j)^2}\Big)\,, \qquad (1)\end{aligned}$$

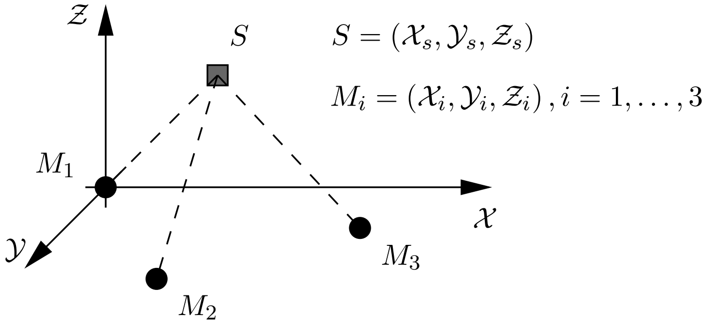

$\mathcal{Z}$
S
$S = (\mathcal{X}_s, \mathcal{Y}_s, \mathcal{Z}_s)$
$M_i = (\mathcal{X}_i, \mathcal{Y}_i, \mathcal{Z}_i), i = 1, \ldots, 3$
M_1
$\mathcal{X}$
$\mathcal{Y}$
M_3
M_2

Figure 1 - Example of a configuration with 3 microphones $M_i, i = 1, \ldots 3$ and a sound source S

where c is the velocity of sound. Only 2 out of the 3 possible TDOAs τ_{12}, τ_{13} and τ_{23} are independent, as the following dependence can be easily deduced from (1).

$$\tau_{23} = \tau_{13} - \tau_{12} \,. \tag{2}$$

To come from the TDOAs of a microphone array to the 3D localization of a sound source , the exact solution necessitates solving a set of 3 independent highly non-linear equations from the type given in (1) for the 3 unknown Cartesian coordinates of the sound source $\mathcal{X}_s$, $\mathcal{Y}_s$ and $\mathcal{Z}_s$. Hence, at least 4 microphones are required. This exact solution can be computationally very demanding and in the presence of measurement errors unambiguous. Fortunately, there is an extensive class of sub-optimal, closed-form location estimators which approximate sufficiently the exact solution to the non-linear problem and which are computationally undemanding [6, 7]. As little disadvantage, these methods need an additional microphone, i.e. at minimum $N = 5$ sensors with 4 independent microphone pairs for the 3D localization. In real environments, noise and especially reverberation originating from reflections of sound waves can cause wrong TDOA estimates and therewith, wrong source positions. This suggests to use the redundant information in the dependent time delays to enhance the confidence of the TDOA estimates. As an example, for $N = 5$, there are $\binom{N}{2} = 10$ possible TDOAs and $N-1 = 4$ independent time delays. The idea is now to exploit the redundancy lying in the $10-4 = 6$ additional dependent TDOAs. Usually, no information about the error distribution or the reliability of a single TDOA estimate is available. That is why up to now, attempts to use the redundancy in a microphone array got no improvement [8]. State of the art is to choose one reference microphone in the array with N sensors and to estimate the $N-1$ independent time delays in the remaining microphones referred to the reference microphone.

In this work, reliability criteria for the TDOA estimates found in [4, 5] and detailed in Sect. 4 are used to determine the $N-1$ most confident TDOAs out of the $\binom{N}{2}$ possible time delays to diminish the percentage of false TDOA estimates in order to increase the robustness of the acoustic source localization system.

3 Time Delay Estimation

3.1 Signal Model

For a given pair of spatially separated microphones M_i and M_j, the recorded sensor signals $x_i(t)$ and $x_j(t)$ for a signal $s(t)$, coming from a remote sound source in a reverberant and noisy environment, can be modeled mathematically as

$$\begin{aligned} x_i(t) &= h_i(t) * s(t) + n_i(t) \\ x_j(t) &= h_j(t) * s(t-\tau_{ij}) + n_j(t) \,, \end{aligned} \tag{3}$$

where τ_{ij} represents the relative time delay of arrival to be determined, $*$ signifies the convolution operator, $h_i(t)$ is the acoustic impulse response between the sound source and the i^{th} microphone and the additive term $n_i(t)$ summarizes the channel noise in the microphone system as well as environmental noise for the i^{th} sensor. The noise term $n_i(t)$ is assumed to be uncorrelated with $s(t)$ and $n_j(t)$.

3.2 TDOA Estimation with GCC Method

The most popular approach for determining the TDOAs is the *Generalized Cross Correlation* (*GCC*) method [2]. The relative time delay τ_{ij} is estimated as the time lag with the global maximum peak in the GCC function $R_{ij}^{(g)}(\tau)$:

$$\hat{\tau}_{ij} = \underset{\tau}{\text{argmax}}\ R_{ij}^{(g)}(\tau)\,. \tag{4}$$

This GCC function $R_{ij}^{(g)}(\tau)$ is defined as

$$R_{ij}^{(g)}(\tau) = \int_{-\infty}^{+\infty} \psi_{ij}(\omega) X_i(\omega) X_j(\omega)^* e^{j\omega\tau} d\omega\,. \tag{5}$$

The weighting function $\psi_{ij}(\omega)$ intends to decrease noise and reverberation influence and tries to emphasize the GCC peak at the true TDOA τ_{ij}. For real environments, the *Phase Transform* (*PHAT*) technique has shown the best performance [9]. The PHAT weighting function is defined as

$$\psi_{ij}^{\text{PHAT}}(\omega) = \frac{1}{|X_i(\omega) X_j(\omega)^*|}\,. \tag{6}$$

4 Reliability Criteria for TDOA Estimates

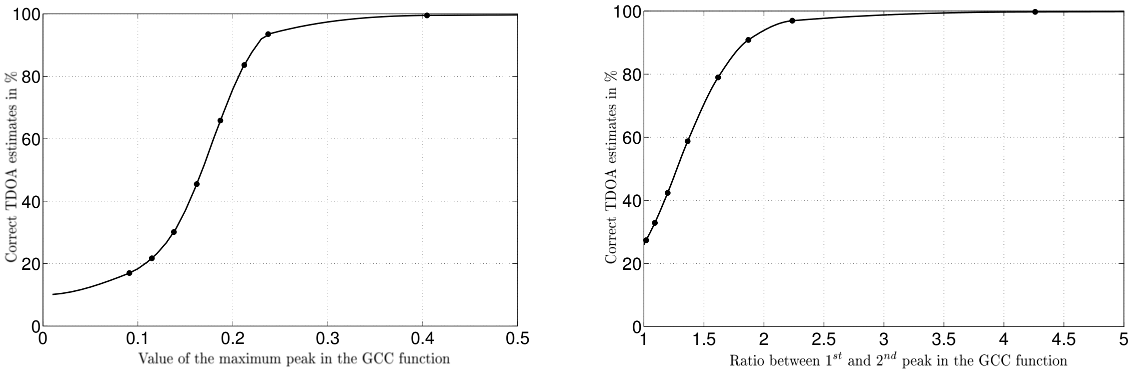
Although the GCC approach seems to be practical, its application in real acoustic environments is only of limited use. Even in mildly reverberant rooms, the TDOA estimation error rate rises strongly, delivering unreliable time delays and hence non-confident sound source locations. Therefore, reliability indicators are required allowing to evaluate the confidence of every single TDOA estimate.

Two properties of the GCC function can be used to evaluate the reliability of every single TDOA estimate, namely the value of the maximum peak and the ratio of the values of the 1^{st} and 2^{nd} largest peak in the GCC function [4, 5]. For this analysis, real data with stationary sources at different positions were recorded with the environmental setup described in Sect. 5. To determine the relationship between the GCC criteria and the TDOA reliability, the TDOA estimates were divided for every criterion into 8 intervals, which borders are given in Tab. 1. As can be clearly seen in the interpolated curves in Fig. 2, the maximum peak, as well as the ratio between 1^{st} and 2^{nd} peak in the GCC function allow very convincingly a judgment about the reliability of the current TDOA estimate. Low criteria values mean low reliability, whereas for high values of the criteria highly reliable estimates are delivered.

Table 1 - Interval borders of the reliability criteria values maximum peak (m) and ratio (r)

	Maximum peak m	**Ratio** r
Interval 1	$m \leq 0.100$	$r \leq 1.075$
Interval 2	$0.100 < m \leq 0.125$	$1.075 < r \leq 1.150$
Interval 3	$0.125 < m \leq 0.150$	$1.150 < r \leq 1.250$
Interval 4	$0.150 < m \leq 0.175$	$1.250 < r \leq 1.500$
Interval 5	$0.175 < m \leq 0.200$	$1.500 < r \leq 1.750$
Interval 6	$0.200 < m \leq 0.225$	$1.750 < r \leq 2.000$
Interval 7	$0.225 < m \leq 0.250$	$2.000 < r \leq 2.500$
Interval 8	$m > 0.250$	$r > 2.500$

Consequently, these two properties of the GCC function can not only be used to detect outliers and to suppress real environment influences such as noise and room reverberation, but also to determine per analysis frame the most confident independent TDOAs out of all possible ones to increase the percentage of correct TDOA estimates.



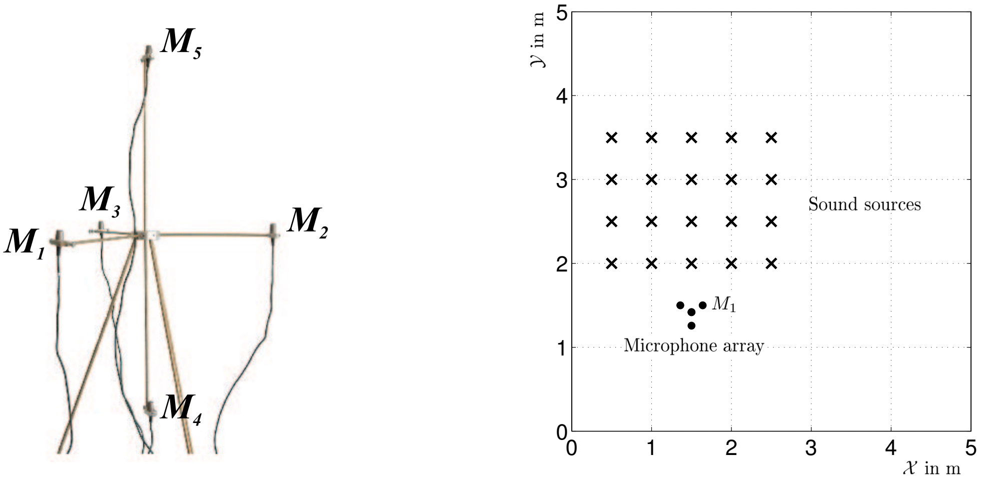
(a) Maximum peak reliability criterion

(b) Ratio reliability criterion

Figure 2 - Confidence criteria of TDOA estimates

5 EXPERIMENTAL SETUP

For data recording, a microphone array of 5 omni-directional electret condenser microphones in an equilateral double-tetrahedron geometry with a side length of $D = 28$ cm was used (Fig. 3(a)). To evaluate the confidence criteria, real experiments were carried out in a typical office room measuring 5 m x 5 m x 3 m. The level of reverberation in the room was experimentally determined by means of Schroeder's backward integration method [10]. Measuring the 60 dB decay period of the sound pressure level after the source signal is switched off for a number of loudspeaker and microphone positions provided a reverberation time RT$_{60} = 0.36$ s. The level of noise coming from fans, mechanical equipment, etc. leads to an SNR of 15 dB. Different utterances of German sentences (altogether 3840 words) from 6 speakers (3 male and 3 female) were played back by a loudspeaker. The loudspeaker was placed in 25 different positions in the office room. The height of the microphone array and the sound sources was 1.5 m. For the $\mathcal{X}$- and $\mathcal{Y}$-coordinates of the sound source positions see Fig. 3(b).



(a) Microphone array

(b) Microphone and sound source positions

Figure 3 - Experimental setup

The sampling frequency was $f_s = 16$ kHz. The recorded speech signals were analyzed in frames of 32 ms to assure quasi-stationarity. For this data segmentation a Hamming window with a 50% overlap was applied. A TDOA estimation in the microphone pair M_iM_j is deemed correct if the product of the sampling frequency f_s and the term $|\hat{\tau}_{ij} - \tau_{ij}|$, i.e. the absolute value of the difference of the estimated and the real TDOA value of the sound source is less than a decision threshold of $\mathrm{T}_{\mathrm{dec}} = 1.5$ samples

$$f_s \cdot |\hat{\tau}_{ij} - \tau_{ij}| \begin{cases} \leq \mathrm{T}_{\mathrm{dec}} & : \text{ correct} \\ > \mathrm{T}_{\mathrm{dec}} & : \text{ false.} \end{cases} \tag{7}$$

6 Results and Discussion

As mentioned in Sect. 2, the use of efficient 3D closed-form estimators necessitate 5 microphones with 4 independent microphone pairs. In this section, the reliability criteria for the TDOA estimates described in Sect. 4 are used to determine the 4 *most confident* TDOAs out of the $\binom{5}{2} = 10$ possible ones in the microphone array in Fig. 3(a). The percentage of false TDOA estimates is evaluated with this pre-selection of TDOA estimates and compared to the conventional method, where only 4 independent time delays referred to the reference microphone M_1 are determined exploiting no redundant information lying in the 6 remaining dependent time delays.
Three possible methods of this TDOA estimate pre-selection are examined:

1. *Maximum peak pre-selection:* According to the maximum peak criterion, the 4 independent TDOAs with the highest value for its maximum peak in the GCC function are taken out of the 10 possible time delays.

2. *Ratio pre-selection:* According to the ratio criterion, the 4 independent TDOAs with the highest value for its ratio between the 1^{st} and the 2^{nd} peak in the GCC function are taken out of the 10 possible time delays.

3. *Pre-selection by combining maximum peak and ratio criterion:* For every microphone pair, the estimation reliability for the actual analysis frame according to the maximum peak and the ratio criterion are determined by means of the curves in Fig. 2. These two reliabilities are summed up and divided by 2 to get a combined confidence information. The 4 independent TDOAs with the highest combined confidence values are taken out of the 10 possible time delays.

The percentages of false TDOA estimates as well as the percentages of replaced TDOAs for the different pre-selection methods compared to the conventional approach are summarized in Tab. 2.

Table 2 - Percentage of false TDOA estimates for the conventional method and for the TDOAs with pre-selection, the absolute and relative improvement and the rate of replaced TDOAs

	Conventional	**Maximum peak**	**Ratio**	**Combining maximum peak and ratio**
Percentage of false TDOAs	27.90 %	19.95 %	19.88 %	19.71 %
Absolute improvement	-	7.95 %	8.02 %	8.19 %
Relative improvement	-	28.49 %	28.75 %	29.35 %
Replaced TDOAs in %	-	64.82 %	64.50 %	64.99 %

It is clearly visible that exploiting the redundancy in the proposed way leads to a significant decrease of false TDOA estimates. With an absolute improvement of about 8 % and a relative improvement of about 29 %, the pre-selection methods deliver an important reliability enhancement of time delay estimates and hence a more robust sound source localization. The convincing results justify the 2.5 times higher computational load of the methods exploiting redundancy. In comparison with the conventional method, the percentages of replaced TDOA estimates of about 65 % in the 4 chosen microphone pairs per analysis frame show a considerable exchange of TDOA estimates for the pre-selection methods. Comparing the three pre-selection methods, no big performance differences can be noticed. The ratio pre-selection slightly outperforms the maximum pre-selection. With the combination of both methods, only a small additional improvement can be achieved, denoting a relatively high correlation of the maximum peak and the ratio reliability criterion.

7 Conclusions

Accurate time delay estimates in microphone pairs are the basis for robust acoustic 3D sound source localization with microphone arrays. This work presents methods to improve the time delay estimation exploiting redundancy in a microphone array. By means of reliability criteria, the most confident independent microphone pairs are chosen out of all possible ones leading to a significant absolute decrease of the percentage of false TDOA estimates of about 8 %.

The proposed methods are implemented successfully in a real-time acoustic 3D tracker with an Extended Kalman filter post-processing unit smoothing the initially estimated source trajectory [11]. The system shows robustness to noise and reverberation and performs well in a real office room environment: a moving speaker can be tracked without problems.

8 Acknowledgments

This work is part of the research project Sonderforschungsbereich (SFB) No. 588 *"Humanoid robots"* at the University of Karlsruhe in Germany. The SFB is supported by the Deutsche Forschungsgemeinschaft (DFG).

REFERENCES

[1] K. Nakadai, H.G. Okuno, and H. Kitano. Active audition based humanoid audition system an its evaluation: Localization, separation and recognition of simultaneous speeches. In *Humanoids*, Karlsruhe, Germany, 2003.

[2] C. H. Knapp and G. C. Carter. The generalized correlation method for estimation of time delay. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 24(4):320–327, August 1976.

[3] S. Bédard, B. Champagne, and A. Stéphenne. Effects of room reverberation on time-delay estimation performance. In *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pages II:261–264, Adelaine, Australia, April 1994.

[4] D. Bechler and K. Kroschel. Confidence scoring of time difference of arrival estimation for speaker localization with microphone arrays. In *13. Konferenz Elektronische Sprachsignalverarbeitung ESSV*, September 2002.

[5] D. Bechler and K. Kroschel. Reliability measurement of time difference of arrival estimations for multiple sound source localization. In *17th Annual Meeting of the IAR*, November 2002.

[6] Y. Huang, J. Benesty, and G. W. Elko. Passive acoustic source localization for video camera steering. In *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pages 909–912, June 2000.

[7] Y. Huang, J. Benesty, G. W. Elko, and R. M. Mersereau. Real-time passive source localization: A practical linear-correction least-squares approach. *IEEE Trans. Speech and Audio Processing*, 9(8):943–956, November 2001.

[8] F. Berthommier and J. Leber. Speaker localisation with four microphones: study of the t-shaped configuration. In *12th Annual Meeting of the IAR*, 1997.

[9] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein. *Microphone Arrays*, chapter Robust Localization in Reverberant Rooms. Springer, 2001.

[10] M. R. Schroeder. New method of measuring reverberation time. *Journal of the Acoustical Society of America*, 37:409–412, 1965.

[11] D. Bechler and K. Kroschel. System for robust 3d speaker tracking using microphone array measurements. In *IEEE/RSJ International Conference on Intelligent Robots and Systems IROS*, Sendai, Japan, September 2004.