RELIABILITY CRITERIA EVALUATION FOR TDOA ESTIMATES IN A VARIETY OF REAL ENVIRONMENTS

Dirk Bechler and Kristian Kroschel

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

ABSTRACT

Time delay estimates in microphone pairs are commonly used for acoustic sound source localization. In real environments, the reliability of time delay estimates quickly degrades for increasing noise or room reverberation levels. This work presents two computationally non-demanding criteria to determine the reliability of time delay estimates. The criteria are evaluated for real data in a variety of acoustic environments with different reverberation and noise levels. The results show the generality of the proposed reliability indicators in delivering information of credibility of the current time delay estimate for many typical room environments.

1. INTRODUCTION

The task 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 a more and more important area of research nowadays [1].

The technique of choice in most passive acoustic sound source localization systems using a microphone array is a two-step procedure. First, *Time Delays Of Arrival* (*TDOAs*) in microphone pairs are estimated. Second, these TDOAs are used together with the array geometry to determine the position of the sound source. The most common technique to estimate the TDOAs is the *Generalized Cross Correlation* (*GCC*) method [2]. While being 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 noise or room reverberation rise above minimal levels [3]. Therefore, reliability criteria are desirable to determine the confidence of every single TDOA estimate.

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

In this work, the reliability criteria presented in [4] are evaluated for real data in a variety of environments with different noise and reverberation levels to test their generality.

2. TIME DELAY ESTIMATION

2.1. 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 $s(t)$, emitted by a remote sound source in a reverberant and noisy environment, can be modeled mathematically as

$$\begin{aligned} x_i(t) &= h_i(t) * s(t) + n_i(t) \\ x_j(t) &= h_j(t) * s(t - \tau_{ij}) + n_j(t)\,, \end{aligned} \tag{1}$$

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 $n_j(t)$.

2.2. TDOA Estimation with GCC Method

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

$$\hat{\tau}_{ij} = \operatorname*{argmax}_{\tau} R_{ij}^{(g)}(\tau)\,. \tag{2}$$

In real environments, the GCC function $R_{ij}^{(g)}(\tau)$ with *Phase Transform* (*PHAT*) weighting has shown the best performance [5]. It is defined as

$$R_{ij}^{(g)}(\tau) = \int_{-\infty}^{+\infty} \frac{X_i(\omega) X_j(\omega)^*}{|X_i(\omega) X_j(\omega)^*|} e^{j\omega\tau} d\omega\,. \tag{3}$$

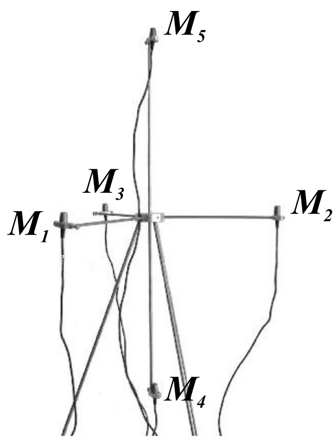

Fig. 1. Microphone array

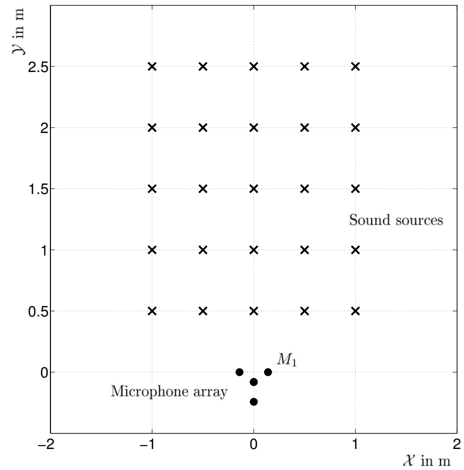

Fig. 2. Microphone and sound source positions

3. RELIABILITY CRITERIA FOR TDOA ESTIMATES

Although the GCC approach seems to be convenient, its application in real acoustic environments is only of limited use. Even in mildly reverberant or noisy rooms, the TDOA estimation error rises strongly, delivering unreliable time delays and hence non-credible sound source positions. Consequently, reliability indicators are required allowing to evaluate the credibility of TDOA estimates. In [4] it was shown that two properties of the GCC function can be used in a typical office room environment to evaluate the reliability of the current TDOA estimate, namely the value of the maximum peak and the ratio between the first and second largest peak in the GCC function. In this work, these reliability criteria are evaluated for different room acoustics, i.e. a variety of reverberation times and noise levels.

4. EXPERIMENTAL SETUP

A microphone array of 5 omni-directional electret condenser microphones in an equilateral double-tetrahedron geometry with a side length of $D = 28$ cm as shown in Fig. 1 was used for data recording. Recordings were carried out in three different office rooms with typical environmental office noise, coming from fans, mechanical equipment, etc. Utterances from 6 speakers (3 male and 3 female) of different German sentences (altogether 3840 words) were played back by a loudspeaker. In all three rooms, the loudspeaker was placed on the same grid with 25 different positions shown in Fig. 2. The height of the reference microphone M_1 and the sound sources was 1.5 m. The sampling frequency was f_s = 16 kHz. The recorded speech signals were analyzed in frames of 32 ms to assure quasi-stationarity. 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 $\mathrm{T_{dec}} = 1.5$ samples

$$f_s \cdot |\hat{\tau}_{ij} - \tau_{ij}| \begin{cases} \leq \mathrm{T_{dec}} & : \ \text{correct} \\ > \mathrm{T_{dec}} & : \ \text{false.} \end{cases} \tag{4}$$

To determine the relationship between the GCC criteria and the TDOA reliability, the TDOA estimates were divided for every criterion into 10 intervals, whose borders are given in Tab. 1.

For the evaluation of the reliability criteria with respect to different reverberation times, recordings were conducted in three rooms with different room characteristics. The recording conditions are detailed in Tab. 2. The 60 dB reverberation times RT_{60} in the rooms were experimentally determined by means of Schroeder's backward integration method [6]. With reverberation time values between 190 and 480 ms, a large bandwidth of varying room characteristics of typical office rooms is covered.

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

	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	$0.225 < m \leq 0.250$	$2.000 < r \leq 2.250$
Interval 8	$0.250 < m \leq 0.300$	$2.250 < r \leq 2.500$
Interval 9	$0.300 < m \leq 0.400$	$2.500 < r \leq 3.000$
Interval 10	$m > 0.400$	$r > 3.000$

Table 2. Room characteristics for the evaluation of the reverberation time

	Size	Mean SNR	RT_{60}
Room 1	3 m x 4 m x 3 m	30.4 dB	190 ms
Room 2	5 m x 5 m x 3 m	30.1 dB	350 ms
Room 3	8 m x 6 m x 3 m	29.8 dB	480 ms

Table 3. Mean SNR values for the evaluation of the noise influence in Room 2

Room 2	Mean SNR
Session 1	30.1 dB
Session 2	22.1 dB
Session 3	15.3 dB
Session 4	6.3 dB
Session 5	-1.1 dB

The influence of environmental office noise on the reliability criteria was analyzed in Room 2 for a variety of SNR values adjusted with different loudspeaker volume levels for the speech signals. The mean SNR values of the conducted five sessions are summarized in Tab. 3.

5. RESULTS AND DISCUSSION

5.1. Reverberation influence on reliability criteria

The total percentage of correct TDOA estimates for the recordings in all three rooms with different reverberation is given in Tab. 4. As expected, the number of correct estimates decreases with increasing room reverberation levels.

The relationship between the GCC reliability criteria and the TDOA credibility is shown by the interpolated curves in Figs. 3 and 4. The markers of the supporting points are positioned at the centroids of the corresponding intervals given by Tab. 1. As can be clearly seen, the maximum peak as well as the ratio between the first and second peak in the GCC function allow a very confident judgement about the reliability of the current TDOA estimate in every room. Low criteria values mean low reliability, whereas for high values of the criteria highly reliable estimates are delivered.

Table 4. Total percentage of correct TDOA estimates for different reverberation times RT_{60}

Room number	1	2	3
RT_{60}	190 ms	350 ms	480 ms
Correct TDOAs	91.51%	81.27%	77.96%

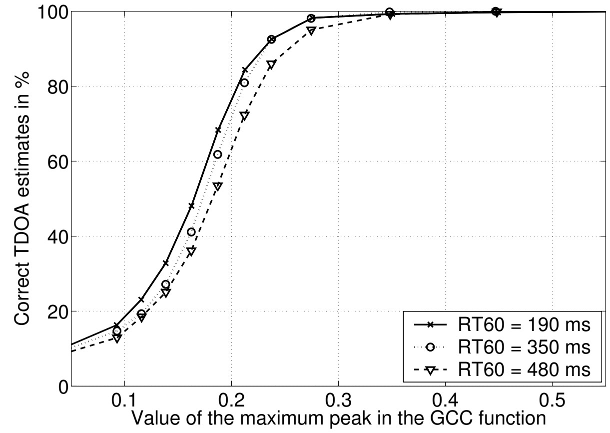

Fig. 3. Reverberation influence on maximum peak criterion

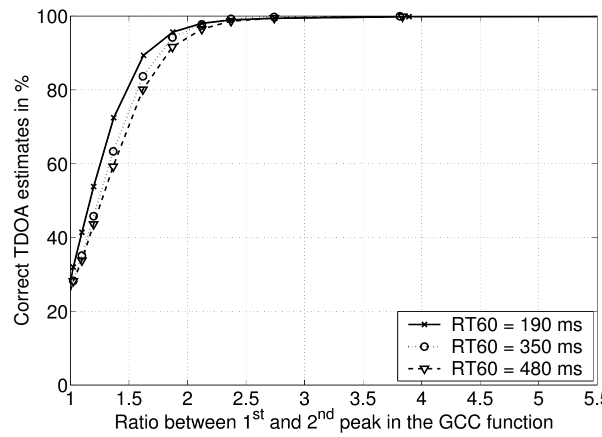

Fig. 4. Reverberation influence on ratio criterion

Comparing the progression of the curves for the 3 different rooms in Figs. 3 and 4, the differences of the total percentage of correct TDOA estimates of Tab. 4 are approximately represented in the gap between the curves. Environments with lower reverberation show steeper rises for slightly smaller criterion values, but generally speaking, the reliability judgement of TDOA estimates works convincingly well in the investigated range of reverberation times between 190 and 480 ms.

5.2. Noise influence on reliability criteria

Table 5 shows the noise influence on the total percentage of correct TDOA estimates for recordings in Room 2 with different SNR levels. The percentage of correct estimates falls significantly with decreasing SNR values to only 25.10% for a mean SNR of -1.1 dB.

Figures 5 and 6 illustrate the results for the reliability criteria for these recordings. Compared with the ratio criterion, the maximum peak criterion shows better perfor-

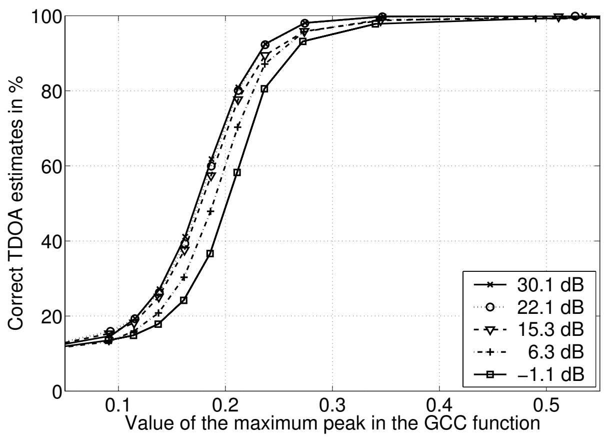

Fig. 5. Noise influence on maximum peak criterion

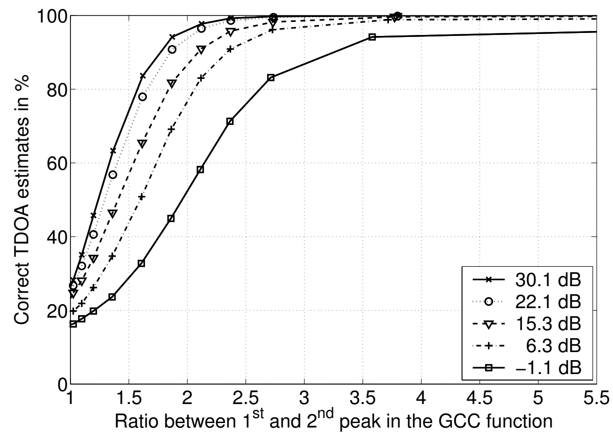

Fig. 6. Noise influence on ratio criterion

Table 5. Total percentage of correct TDOA estimates for recordings with different mean SNR values in Room 2

Mean SNR values				
30.1 dB	22.1 dB	15.3 dB	6.3 dB	-1.1 dB
Correct TDOAs				
81.27%	70.74%	57.10%	40.26%	25.10%

mance. In Fig. 5 the curves with high SNRs up to a mean SNR of 15.3 dB are very close. For lower SNR values, the gaps between the curves increase but are far below the differences of total percentages of correct estimates of Tab. 5. For the ratio criterion in contrast, the gaps of the curves in Fig. 6 correspond for mean criterion values roughly to the difference of correct estimate percentages. But even for an SNR of -1.1 dB with a total percentage of correct TDOAs of only 25.10%, the criterion can be efficiently used for a reliability judgement reaching for high criterion values confident TDOA estimates.

6. CONCLUSIONS

Accurate time delay estimates in microphone pairs are the basis for robust acoustic 3D sound source localization with microphone arrays. Even low levels of reverberation and background noise become rapidly detrimental to GCC-based TDOA estimation methods. This work presents two computationally non-demanding reliability criteria, namely the value of the maximum peak and the ratio of the values of the first and second largest peak in the GCC function. These criteria allow a very efficient evaluation of the credibility of every single TDOA estimate and hence the rejection of false TDOA estimates. Even if the TDOA reliability assigned to a certain criterion value is dependent on the reverberation and noise level, the general progression of the dependence of TDOA reliability and criterion values remains constant for a variety of environments with different reverberation and noise levels: small criterion values signify low TDOA credibility, whereas high values result in highly reliable TDOA estimates. The simplicity of the proposed criteria makes them highly attractive for real-time systems.

7. 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*, vol. 24, no. 4, pp. 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)*, Adelaine, Australia, April 1994, pp. II:261–264.

[4] D. Bechler and K. Kroschel, "Three different reliability criteria for time delay estimates," in *12th European Signal Processing Conference EUSIPCO*, Wien, Austria, September 2004.

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

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