THREE DIFFERENT RELIABILITY CRITERIA FOR TIME DELAY ESTIMATES

Dirk Bechler and Kristian Kroschel

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

ABSTRACT

Time Difference Of Arrival (*TDOA*) estimates are often used for passive acoustic single sound source localization with microphone arrays. Commonly, the *Generalized Cross-Correlation* method is applied for TDOA estimation. In real environments the reliability of TDOA estimates is degraded due to noise and especially to room reverberation. This work presents three reliability criteria allowing convincingly the evaluation of the confidence of every single TDOA estimate. The criteria are evaluated and compared to each other for real data recorded in a noisy and reverberant office room environment. The use of the reliability criteria permits a robust sound source localization even in adverse acoustic environments with increased reverberation times.

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. 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]. Former studies tried to increase the overall confidence of the TDOA estimates with only little improvement [4]. In this work, three reliability criteria are presented and compared to each other allowing a confidence scoring of every single estimate and thus permitting robust sound source localization even for increased reverberation times.

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 $s(t)$, coming from a remote sound source in a reverberant and noisy environment, can be modeled mathematically as

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

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* (*GCC*) method [2]. The relative time delay τ_{ij} is estimated as the time lag with the global maximum peak in the GCC function $R_{ij}^{(g)}(\tau)$:

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

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

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

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 [5]. The PHAT weighting function is defined as

$$\psi_{ij}^{\mathrm{PHAT}}(\omega) = \frac{1}{|X_i(\omega) X_j(\omega)^*|}\,, \tag{4}$$

which can be regarded as a whitening filter.

3. RELIABILITY CRITERIA FOR TDOA ESTIMATES

Although this 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. For this study, three potential reliability criteria are regarded and compared to each other.

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

The first criterion uses psychoacoustical knowledge and is based on the precedence effect [6]. In closed reverberant rooms, the human auditory system uses for acoustic source localization this important psychoacoustical feature. To suppress the echos coming from reflections of the sound waves on walls, floor and ceiling, humans take advantage of the so called *precedence effect* or *law of the first wave-front*. To help the listener in estimating the position of a sound source, the human auditory system focuses on the direct sound wave, i.e. the first wave-front also called *onset*, ignoring the information contained in reflective, later arriving wave-fronts within a certain inhibition time. This inhibition time depends on the type of signal [7]. In case of speech the inhibition time is about 30 to 50 ms.

The two other criteria are properties of the GCC function, 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. With the study of the maximum peak value in the GCC function, we expect that the higher the value of this peak, the higher is the probability that the TDOA estimate is correct. Likewise we wish that the higher the largest peak dominates the second largest one, the higher is the probability of a confident TDOA estimate.

4. 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. To evaluate the confidence criteria, real experiments were carried out in an office room of 5 m x 5 m x 3 m. 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 15 different positions in the office room with typical environmental noise (SNR $\approx$ 20 dB) coming from fans, mechanical equipment, etc. and relatively strong reverberations (reverberation time $T_{60} \approx 360$ ms). The height of the microphone array and the sound sources was 1.5 m. For the *x*- and *y*-coordinates of the sound source positions see Fig. 1.

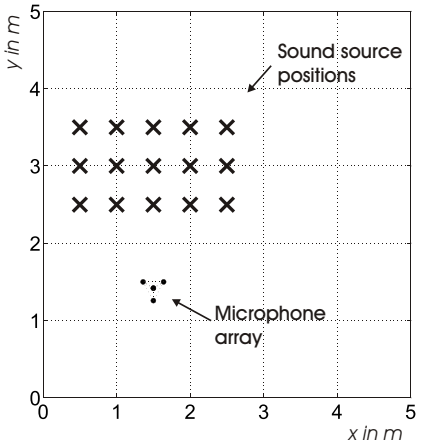

Figure 1: Microphone and sound source positions

4.1 Signal Processing

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 $T_{dec} = 1.5$ samples

$$f_s \cdot |\hat{\tau}_{ij} - \tau_{ij}| \begin{cases} \leq T_{dec} & : \text{ correct} \\ > T_{dec} & : \text{ false.} \end{cases} \quad (5)$$

4.2 Onset Detector

To analyze the influences of the precedence effect on the sound source localization using microphone arrays, TDOA estimates need not to be calculated for the complete signal, but only for those signal segments corresponding to the onsets. Therefore, a robust onset detector has to be implemented.

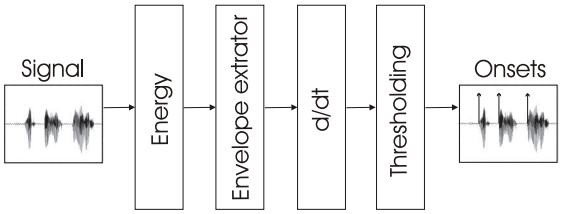

Figure 2: Block diagram of the onset detector

Figure 2 shows the block diagram of the applied onset detector. After the segmentation of the input signal in analysis frames of 32 ms, the energy per frame is calculated. Then the envelope of the energy signal is extracted. This envelope is differentiated and by means of thresholding, the positions of the onsets are determined. With an onset detection rate of 100% and a false alarm rate of only 2.82% high robustness is achieved, delivering reliably those analysis frames containing an onset.

5. RESULTS AND DISCUSSION

5.1 Evaluation of the Reliability Criteria

5.1.1 Precedence Effect Criterion

To evaluate the influence of the precedence effect on the percentage of correct TDOA estimates compared with the one for all TDOA estimates during speech activity, the analysis windows containing an onset have to be determined. Table 1 shows the results of this comparison. The absolute

	Correct TDOA estimates in %
All frames with speech activity	*72.97%*
Frames with onset	*93.37%*

Table 1: Percentage of correct TDOA estimates for all analysis frames and for precedence effect frames with onsets

performance gain of 20.4% leads to highly reliable TDOA estimates with an average percentage of correct estimates of 93.37% for analysis frames containing an onset. Hence, in consideration of the precedence effect, the confidence of TDOA estimates can be significantly increased. It has to be mentioned, that the number of frames containing an onset is restricted. Only 2.42% of all frames with speech activity are frames with onsets. Thus the precedence effect is limited to the initialization of a sound source track. As an accurate initialization is crucial for defining a certain region of interest for consecutive estimates, the precedence effect can be very helpful to assure robust source localization and tracking.

5.1.2 GCC Criteria

To determine the relationship between the GCC criteria and the TDOA reliability, the TDOA estimates were divided for every criterion into 8 intervals. The interval borders are extracted from the respective histograms for the maximum peak and the ratio criterion values of all analysis frames (Figs. 3-4). The interval limits for every confidence criterion were chosen such that every interval contains a similar number of TDOA estimates. Table 2 details the interval borders.

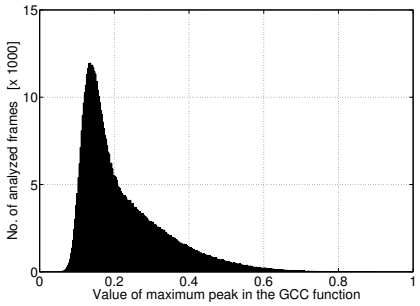

Figure 3: Histogram for the maximum peak criterion values of all analysis frames

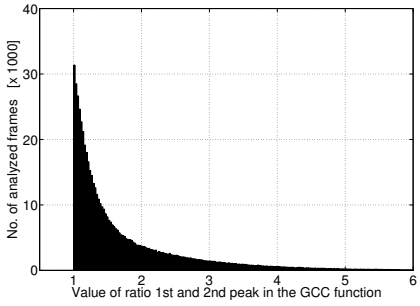

Figure 4: Histogram for the ratio criterion values of all analysis frames

As can be clearly seen in Fig. 5, 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 of only 26.30% {31.57%} for the maximum peak {ratio} criterion in interval 1, whereas for high values of the criteria the confidence increases to around 97%, delivering highly reliable estimates. Consequently these two properties of the GCC function can be used to detect outliers and

	Maximum peak m	**Ratio** r
Interval 1	$m \leq 0.129$	$r \leq 1.08$
Interval 2	$0.129 < m \leq 0.146$	$1.08 < r \leq 1.23$
Interval 3	$0.146 < m \leq 0.163$	$1.23 < r \leq 1.40$
Interval 4	$0.163 < m \leq 0.195$	$1.40 < r \leq 1.55$
Interval 5	$0.195 < m \leq 0.236$	$1.55 < r \leq 2.31$
Interval 6	$0.236 < m \leq 0.247$	$2.31 < r \leq 3.10$
Interval 7	$0.247 < m \leq 0.283$	$3.10 < r \leq 5.20$
Interval 8	$m > 0.283$	$r > 5.20$

Table 2: Interval borders of the reliability criteria values maximum peak (m) and ratio (r)

to suppress real environment influences such as noise and room reverberation considerably. With the confidence criteria, a trade-off has to be made between a high number of estimates, which is necessary for a continuous target tracking, and a high percentage of correct TDOA estimates, which is crucial for robust source localization.

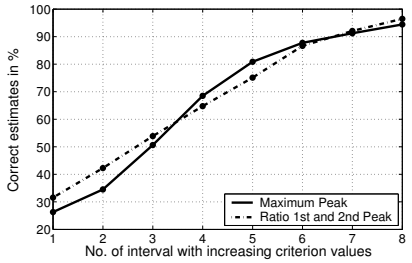

Figure 5: Correct estimate percentage per interval for increasing values of the maximum peak and ratio criterion

5.2 Comparison of the Reliability Criteria

5.2.1 Precedence Effect Criterion vs. GCC Criteria

In comparison to Figs. 3-4, the histograms for the reliability criteria of the analysis frames containing an onset tend to higher values of the maximum peak, as well as to the ratio between 1^{st} and 2^{nd} peak in the GCC function (Figs. 6-7). Hence, there is only a small number of frames with onsets having low criteria values.

To study the relation between the GCC criteria values and the TDOA reliability for precedence effect frames, the

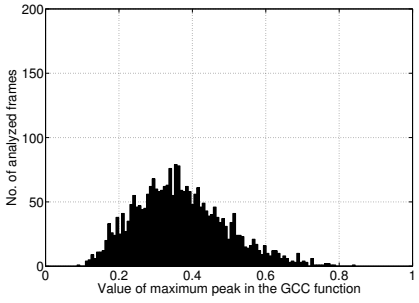

Figure 6: Histogram for the maximum peak criterion values of analysis frames with an onset

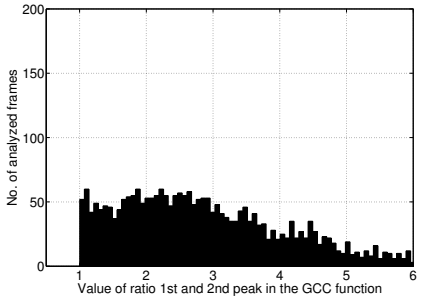

Figure 7: Histogram for the ratio criterion values of analysis frames with an onset

frames having an onset were analyzed separately in the same interval borders given in Tab. 2. The curves in Fig. 8 show explicitly, that also for analysis frames with onsets corresponding to the first wave-front, the percentage of correct TDOA estimates depends strongly on the value of the maximum peak and the ratio between the 1^{st} and the 2^{nd}peak in the GCC function, respectively. Nevertheless the percentages of correct TDOA estimates are significantly increased compared to Fig. 5, especially for intervals with low and medium criteria values. As an example, with a percentage of correct estimates of 43.75% {55.74%} in interval 1, the absolute gain is 17.45% {24.17%} for the maximum peak {ratio} criterion.

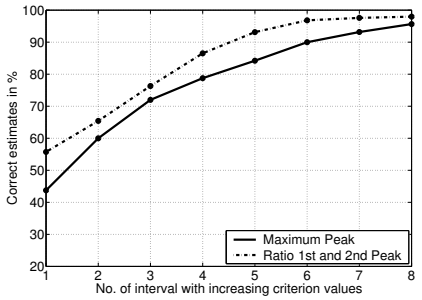

Figure 8: Percentage of correct TDOA estimates per interval with increasing criterion values for onset frames

As mentioned before, the number of estimates in the regions of low and medium criterion values is very small. Hence, when the two features of the GCC function for robust localization of sound sources are already used, the improvements contributed by the precedence effect are limited.

5.2.2 Comparison between the GCC Criteria

Figure 9 shows the comparison of the two criteria which are properties of the GCC function. In a scatter plot the ratio between the 1^{st} and the 2^{nd} peak is plotted over the value of the maximum peak of the GCC function for all frames with speech activity. In general, analysis frames with an increased ratio value tend to have a higher value of the maximum peak in the GCC function, too. Nevertheless, the large width of this scatter plot in direction of the maximum peak values (x-axis) as well as in direction of the ratio values (y-axis) signifies a great number of frames having either a large maximum peak value and a low ratio value or a low maximum peak value and a large ratio value. For these frames the correlation between the GCC criteria is limited and its combination

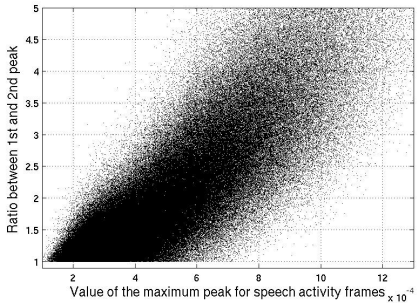

Figure 9: Percentage of correct TDOA estimates per interval with increasing criterion values for onset frames

will increase the reliability of the TDOA estimate even more.

6. CONCLUSIONS

With the proposed reliability criteria, the confidence of every single TDOA estimate in a microphone pair can be measured allowing the effective suppression of noise influences and room reverberation. It was shown that the precedence effect is for the most part covered by the criterion of the maximum peak and the criterion of the ratio between the 1^{st} and the 2^{nd} peak of the GCC function. In contrast, there is only a limited correlation between the two GCC criteria. Consequently, an additional gain in TDOA reliability can be achieved by combining the maximum peak with the ratio criterion. This is implemented successfully in a real-time 3D acoustic sound source tracker. This localization system shows robustness in the noisy and reverberant office environment and can track with ease a moving speaker.

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] A. Stéphenne and B. Champagne. Cepstral prefiltering for time delay estimation in reverberant environments. In *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pages 3055–3058. Detroit, USA, May 1995.

[5] J.H. DiBiase et al. Robust localization in reverberant rooms. In M. Brandstein and D. Ward, editors, *Microphone Arrays*, chapter 7, pages 131–154. Springer, 2001.

[6] *Spatial Hearing - Revised Edition*. The MIT Press, Cambridge, MA, 1996.

[7] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman. The precedence effect. *Journal of the Acoustical Society of America*, 106(4):1633–1654, October 1999.