Robustness of Acoustic Multiple-Source Localization in Adverse Environments

Robert Aichner, Herbert Buchner, Stefan Wehr, and Walter Kellermann
Multimedia Communications and Signal Processing,
University of Erlangen-Nuremberg,
Cauerstr. 7, 91058 Erlangen, Germany
{aichner, buchner, wehr, wk}@LNT.de

Abstract

In this paper the robustness of a previously introduced localization algorithm for multiple acoustic sources is investigated and compared to two popular single-source localization algorithms. To show the versatility of the proposed method, experiments are conducted in various environments with different reverberation times and background noises. In addition, moving speakers are used to show the tracking capability. The application to binaural hearing aids shows the applicability of the algorithm to realistic adverse environments such as, e.g., a cafeteria.

1 Introduction

In the literature on passive acoustic source localization the most common approach is the estimation of time differences of arrival (TDOA) which comprises two steps. First the temporal signal delays between different pairs of microphones (TDOA) are estimated. In a second step the position in the three-dimensional space or in a two-dimensional plane is calculated using the TDOA estimates. Under the assumption that the microphone positions are known a-priori the second step reduces to a purely geometrical problem. In the following we address the TDOA estimation for single-source localization and our recently proposed multiple-source localization and investigate their behaviour in noisy and reverberant environments.

2 TDOA estimation in reverberant and noisy environments

The most widely used and conceptually simple method to estimate the TDOA is to use the generalized cross-correlation function (GCC) [1]. The basic principle of this technique consists of the maximization of the inverse Fourier transformation of a weighted cross-power spectral density, i.e.,

$$\tau = \arg\max_{\tau} \mathcal{F}^{-1}\left\{\frac{S_{x_1x_2}(f)}{|S_{x_1x_2}(f)|}\right\} \qquad (1)$$

where τ is the TDOA between the sensor signals x_1 and x_2 and $S_{x_1x_2}(f)$ denotes the cross-power spectral density (PSD) at frequency f. The normalization by the magnitude of the PSD in (1) is commonly referred to as the phase transform (PHAT) technique. The underlying model for the GCC is based on free-field wave propagation and, thus, in reverberant environments the performance degrades.

To address the reverberation problem, a completely different approach to TDOA estimation based on blind adaptive filtering was proposed in [2]. This so-called adaptive eigenvalue decomposition (AED) algorithm blindly identifies the impulse responses h_1 and h_2 (assumed to be FIR) between a source s and the two microphones and thus, this approach is inherently based on a dispersive propagation model (Fig. 1a). By minimizing the mean square of the output signal

$$e(n) = s(n) * (h_1(n) * w_1(n) + h_2(n) * w_2(n)) \qquad (2)$$

This work was in parts supported by a grant from the European Union FP6, Project 004171 HEARCOM.

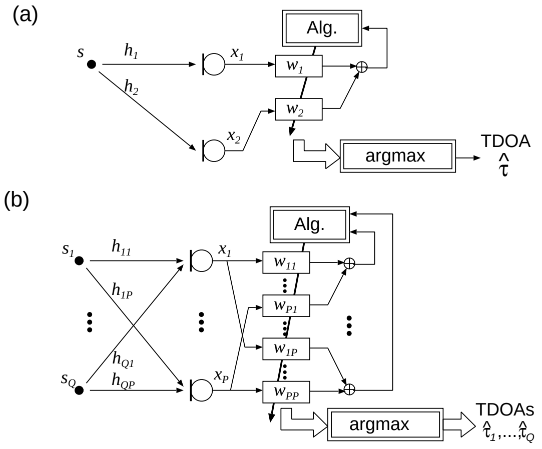

Fig. 1. TDOA estimation via (a) SIMO and (b) MIMO model-based blind adaptive filtering.

ideally the relation $w_1(n) * h_1(n) = -w_2(n) * h_2(n)$ is obtained which leads to a single-input multiple-output (SIMO) system identification. It can be shown that the impulse responses can be identified up to a scalar constant c resulting in $w_1(n) = c \cdot \hat{h}_2(n)$ and $w_2(n) = c \cdot \hat{h}_1(n)$. The TDOA can then be calculated as

$$\tau = \arg\max_n |w_1(n)| - \arg\max_n |w_2(n)|. \quad (3)$$

In [3] the adaptive SIMO filtering approach for single source localization was extended to blind adaptive multiple-input multiple-output (MIMO) filtering for simultaneous localization of multiple sources. Fig. 1b shows the corresponding MIMO-based structure. For the case of two sources and two microphones it was shown in [3] that by applying the TRINICON framework presented in [4] it is possible to calculate the TDOAs for both sources from the estimated FIR filters $w_{pq}(n)$, $p, q \in \{1, 2\}$ by evaluating

$$\tau_1 = \arg\max_n |w_{12}(n)| - \arg\max_n |w_{22}(n)| \quad (4)$$

$$\tau_2 = \arg\max_n |w_{11}(n)| - \arg\max_n |w_{21}(n)| \quad (5)$$

Here, it is assumed that the number of *simultaneously active* sources does not exceed the number of microphones. Furthermore, it is assumed that the sources are mutually uncorrelated which generally holds for speech and audio signals. The adaptation algorithm used for the FIR filters $w_{pq}(n)$ is a broadband blind source separation algorithm derived from the framework in [4]. The adaptation algorithm and its real-time implementation is explained in detail in [5] where also a pseudo-code is included. It should be pointed out that due to the fact that the algorithm is based on time-domain optimization utilizing the broadband signal model, this approach aims at blind MIMO identification and thus the localization performance is not affected by spatial aliasing. Therefore no constraints are put on the sensor geometry and large spacings allowing for a good spatial resolution of the localization results can be chosen.

In literature also localization experiments based on narrowband blind source separation algorithms have been reported (e.g., [6]). Due to the narrowband approach these methods have to adhere to the spatial sampling theorem and thus the sensor geometry is limited to small spacings. Moreover, these narrowband algorithms rely on the far-field approximation.

It was shown in [5] that the broadband adaptation algorithm is very robust against reverberation and background noise. Therefore, we will investigate the effect of reverberation and background noise on the localization performance in the next section. Moreover, the localization performance is investigated in applications where the sensor geometry is only known approximately such as, e.g., with binaural hearing aids. There, additionally head shadowing effects occur introducing further signal delay.

3 Localization experiments in adverse environments

In the experiments, the two point sources which should be localized were emulated by loudspeakers. The speech signals have been recorded with two microphones with a sensor spacing of 0.8 m at a sampling frequency $f_s = 16$ kHz in different scenarios. Moreover, in Sect. 3.2 one experiment with a moving speaker is presented (for more experiments with moving speakers see [3]). The block size for the GCC-PHAT algorithm (1) has been chosen to 1024 samples and for the other algorithms to 2048 samples. For the MIMO adaptation algorithm a block-online update procedure using $K = 8$ blocks and $j_{\max} = 5$

iterations has been used as described in [5]. The blockshift for the GCC algorithm was 512 samples and for the other algorithms 1024 samples. The filter length for the AED and MIMO algorithms is 1024 taps.

3.1 Reverberant environments

The localization results for a lecture room with a reverberation time of $T_{60} \approx 1\,\mathrm{sec}$ and a living room with $T_{60} \approx 250\,\mathrm{msec}$ are shown in Fig. 2 and 3, respectively. In the beginning only one male source is active and after approximately 3 sec a second female speaker starts speaking. The speakers were

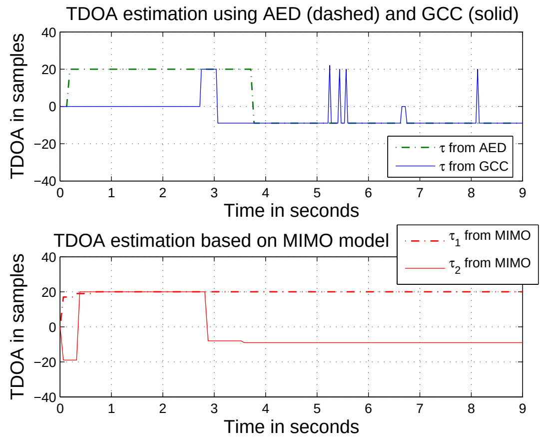

Fig. 2. Results for a living room with $T_{60} \approx 250\,\mathrm{msec}$.

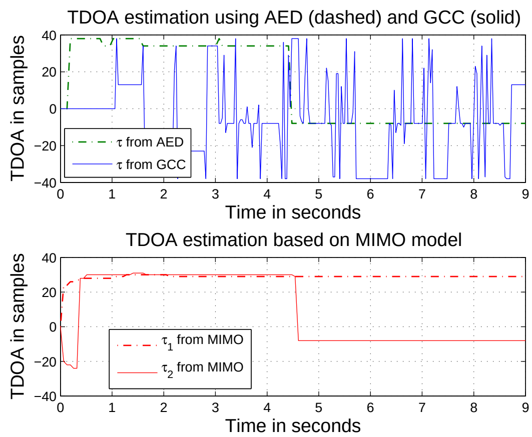

Fig. 3. Results for a lecture room with $T_{60} \approx 1\,\mathrm{sec}$.

positioned relative to the microphone array axis at $+35°$ and $-15°$ for the living room and $+50°$ and $-15°$ for the lecture room. The distance between speakers and microphones was approx. 4 m. It can be seen that for moderate reverberation also the GCC works reasonably well for localizing one source (Fig. 2). Large reverberation, however, needs a convolutive model and thus only the AED and the MIMO localization provide good results (Fig. 3). It can be observed that the MIMO approach accurately tracks the two sources which are simultaneously active after approx. 3 seconds whereas the AED switches to the second source and the GCC jumps between the TDOA of both sources.

3.2 Noisy environments

To assess the robustness against background noise separate noise scenarios have been recorded in a city center to capture various kinds of complex real-life noise sources and to allow for different SNR levels. Speech signals convolved with impulse responses recorded in a quiet public space using the same array have been added to the noise. In the recordings reflections from buildings have been observed which ideally would again require a convolutive model. In public spaces, however, usually the background noise is the dominating problem. The speakers of equal average

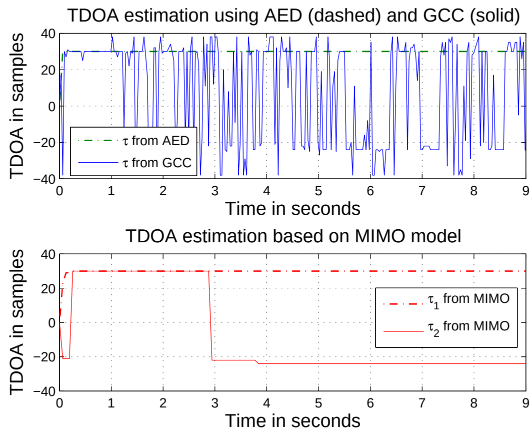

Fig. 4. Experimental results for city center with SNR=-5dB.

power were positioned at $+50°$ and $-40°$ at a distance of 5 m to the array. The background noise contained mainly noise originating from street traffic and pedestrians walking by. For the results in Fig. 4, the SNR was chosen to $-5\,\mathrm{dB}$. The results show that both, the AED and the MIMO algorithm are very robust against background noise. The AED algorithm continuously localizes source 2 whereas the GCC algorithm oscillates between the TDOAs of both sources and exhibits also several outliers due to the background noise.

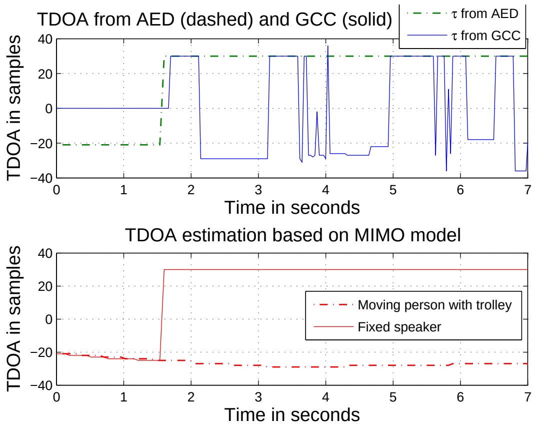

Fig. 5. City center with one fixed and one moving speaker and additional background noise.

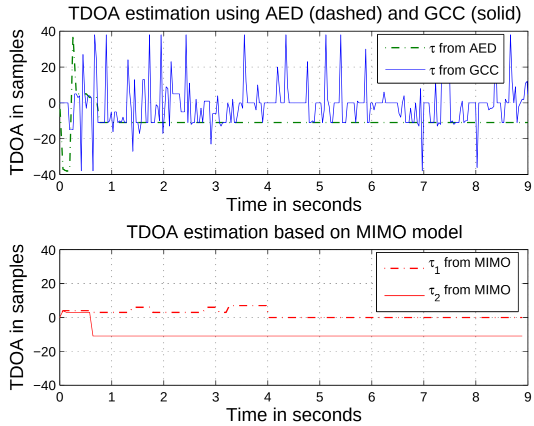

Fig. 6. Results for binaural hearing aids in cafeteria with $T_{60} \approx 2$ sec and SNR=5dB.

In a further experiment a second person was walking by with a trolley suitcase and after approx. 1.5 sec the speaker from the fixed position at $+50°$ started speaking. Moreover, again background noise was present. In Fig. 5 it can be seen that the AED algorithm localizes the fixed speaker very well, whereas the GCC algorithm again switches between the speaker and the person with the trolley. The MIMO-based algorithm can accurately track both, the speaker and the moving person with the trolley and thus shows the effectiveness of the algorithm to time-variant scenarios.

3.3 Application to binaural hearing aids

One application of localization algorithms is the estimation of the position of the desired source for steering a beamformer in multiple-microphone hearing aids. Recently also blind source separation algorithms have been applied to binaural hearing aids as they can adapt without position estimates. However, for a distinction between target source and interfering sources it is desirable to estimate the direction of arrival of all sources from the adapted filters of the blind source separation algorithm. Hearing aids have to work in adverse environments such as cafeterias where usually large reverberation together with a high background noise level is encountered. An additional difficulty is given by shadowing effects which are caused by the head, especially if binaural algorithms are considered. In Fig. 6, the localization results in a cafeteria with background babble noise with SNR=5 dB are presented. A male speaker arrives from $90°$ and after approx. 3 sec a second female speaker at $0°$ is active. Both TDOAs can be accurately estimated by the MIMO algorithm. The value of τ_2 adapts first to background noise effects and after 4 sec locates the second speaker. The GCC exhibits again many outliers and the AED algorithm accurately estimates the TDOA of the male speaker.

4 Conclusions

In this paper we have shown the robustness of a previously proposed acoustic multiple source localization algorithm with respect to reverberation, background noise, and shadowing effects caused by objects placed between the sensors.

References

[1] C.H. Knapp and G.C. Carter, “The generalized correlation method for estimation of time delay,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, pp. 320–327, Aug. 1976.

[2] J. Benesty, “Adaptive eigenvalue decomposition for passive acoustic source localization,” *J. Acoust. Soc. Am.*, vol. 107, pp. 384–391, Jan. 2000.

[3] H. Buchner, R. Aichner, J. Stenglein, H. Teutsch, and W. Kellermann, “Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering,” in *Proc. ICASSP*, Mar. 2005, vol. 3, pp. 97–100.

[4] H. Buchner, R. Aichner, and W. Kellermann, “TRINICON: A versatile framework for multichannel blind signal processing,” in *Proc. ICASSP*, Montreal, Canada, May 2004, vol. 3, pp. 889–892.

[5] R. Aichner, H. Buchner, F. Yan, and W. Kellermann, “A real-time blind source separation scheme and its application to reverberant and noisy acoustic environments,” *Signal Processing*, 2006, to appear.

[6] H. Sawada, R. Mukai, S. Araki, and S. Makino, “Direction of arrival estimation for multiple source signals using independent component analysis,” in *Proc. ISSPA*, Paris, France, 2003.