2006 IEEE International Conference on
Multisensor Fusion and Integration for Intelligent Systems
September 3-6, 2006, Heidelberg, Germany

Multidimensional Localization of Multiple Sound Sources Using Blind Adaptive MIMO System Identification

Anthony Lombard, Herbert Buchner, and Walter Kellermann

Multimedia Communications and Signal Processing
University of Erlangen-Nuremberg
Cauerstr. 7, 91058 Erlangen, Germany
{lombard,buchner,wk}@LNT.de

Abstract— **The TDOA-based acoustic source localization approach is a powerful and widely-used method which can be applied for one source in several dimensions or several sources in one dimension. However the localization turns out to be more challenging when multiple sound sources should be localized in multiple dimensions, due to a spatial ambiguity phenomenon which requires to perform an intermediate step after the TDOA estimation and before the calculation of the geometrical source positions. In order to obtain the required set of TDOA estimates for the multidimensional localization of multiple sound sources, we apply a recently presented TDOA estimation method based on blind adaptive multiple-input-multiple-output (MIMO) system identification. We demonstrate that this localization method also provides valuable side information which allows us to resolve the spatial ambiguity without any prior knowledge about the source positions. Furthermore we show that the blind adaptive MIMO system identification allows a high spatial resolution. Experimental results for the localization of two sources in a two-dimensional plane show the effectiveness of the proposed scheme.**

I. INTRODUCTION

A popular approach in acoustic source localization is based on the estimation of time differences of arrival (TDOA) and consists of two separate steps. For each source, a relative temporal signal delay (i.e., the TDOA) is first estimated between each pair of microphones. In a second step the set of estimated TDOAs is used to calculate the position of each source in the three-dimensional space or in a two-dimensional plane. If the microphone array geometry is known, the second step becomes a purely geometrical problem. This two-step procedure makes it possible to localize each source in the near-field (where we are interested in the exact position of the source) as well as in the far-field (where we can only estimate directions of arrival (DOA), disregarding the range).

We first consider exemplarily the localization of two sources in the far-field, thereby estimating for each source a horizontal angle (i.e., the azimuth) and a vertical angle (i.e., the elevation) using two pairs of microphones (a horizontal and a vertical one). As a consequence a total of four angles should be calculated, each requiring one TDOA estimate (see Sect. IV). Two TDOA estimators should be used. With a first estimator measuring two TDOAs from the horizontal microphone pair, we can calculate the azimuths θ_1 and θ_2 of the first and the second source, respectively. Similarly we can obtain two elevations φ_1 and φ_2 from a second TDOA estimator measuring two TDOAs from the vertical microphone pair. However without additional information on the relative source positions, we cannot determine if the source one with azimuth θ_1 has the elevation φ_1 or the elevation φ_2. The same problem occurs of course for the second source, hence giving rise to a *spatial ambiguity* phenomenon, regardless of the TDOA estimation method used. Not limited to the case of two sources localized using two TDOA estimators, the spatial ambiguity problem will always occur when aiming at simultaneously localizing more than one source in a multidimensional space. In order to resolve the spatial ambiguity, an intermediate step should be performed after the TDOA estimation and before calculating the source positions using the microphone geometry.

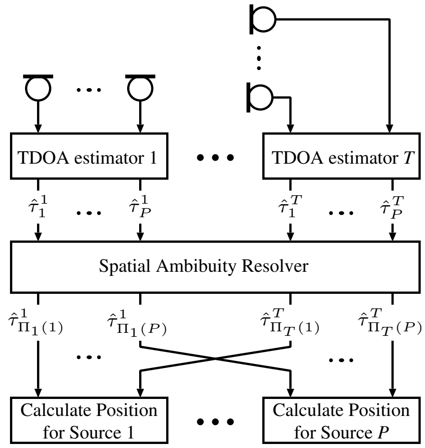

TDOA estimator 1
TDOA estimator T
$\hat{\tau}_1^1$ ... $\hat{\tau}_P^1$
$\hat{\tau}_1^T$ ... $\hat{\tau}_P^T$
Spatial Ambibuity Resolver
$\hat{\tau}_{\Pi_1(1)}^1$
$\hat{\tau}_{\Pi_1(P)}^1$
$\hat{\tau}_{\Pi_T(1)}^T$
$\hat{\tau}_{\Pi_T(P)}^T$
Calculate Position for Source 1
Calculate Position for Source P

Fig. 1. Architecture of a TDOA-based localization scheme for P sources in T dimensions.

In the general case of P sources and T dimensions illustrated by Fig. 1, T TDOA estimators are combined, each providing a set of P TDOAs to localize P sources. This results in a set of $T \times P$ estimates $\hat{\tau}_i^I, i = 1, \ldots, P$, $I = 1, \ldots, T$. The upper index denotes the TDOA estimator index and the lower one denotes the source index. We understand now the necessity to reorder the TDOA estimates so that the position of each source $i = 1, \ldots, P$ can be calculated using its corresponding set of TDOAs $\hat{\tau}_{\Pi_I(i)}^I, I = 1, \ldots, T$

1-4244-0567-X/06/$20.00 ©2006 IEEE.

where $\Pi_I(\cdot)$ is a permutation operator applied on the set of P TDOAs provided by the I^{th} TDOA estimator.

Note that the number T of TDOA estimators in Fig. 1 may be larger than the number of dimensions to be identified. It is actually possible to use more than three TDOA estimators to introduce some redundancy and improve the precision of the calculated source positions (see, e.g., [1] for the simpler case of only one source). In this paper we focus on the signal processing issues of source localization. For clarity we will therefore assume that no redundancy is exploited such that T corresponds to the number of dimensions to be identified but note that the framework illustrated in Fig. 1 and described throughout this paper can be applied to any number of TDOAs.

The key to obtain an accurate localizer is a robust TDOA estimator and an effective mechanism to solve the spatial ambiguity problem when no additional knowledge on the relative source position is available. In the following we will first consider a recently proposed TDOA estimation algorithm based on blind adaptive MIMO system identification (Sect. II). Since this algorithm can perform a simultaneous TDOA estimation for several sources [2] we can combine T such TDOA estimation algorithms to perform the T TDOA estimations in Fig. 1, making it possible to obtain the desired set of TDOAs $\hat{\tau}_i^I, i = 1, \ldots, P$, $I = 1, \ldots, T$. Moreover, we will show in Sect. III how side information provided by the blind adaptive MIMO system identification can be exploited to resolve the spatial ambiguity. We will then concentrate in Sect. IV on geometrical considerations for the multidimensional source localization problem using TDOAs, showing that the blind adaptive MIMO system identification is well suited to high-resolution spatial localization of sound sources. Finally, Sect. V gives some experimental results.

II. TDOA ESTIMATION USING BLIND ADAPTIVE MIMO SYSTEM IDENTIFICATION

A widely used and conceptually simple method to estimate a TDOA for one source from two sensor signals is to use the generalized cross-correlation function (GCC) [3]. However since this method is inherently based on a free-field propagation model its performance usually breaks down in reverberant environments. To address this reverberation problem, a completely different approach of TDOA estimation based on blind adaptive filtering was proposed in [4]. This so-called adaptive eigenvalue decomposition (AED) algorithm may be seen as a major advance in single source localization since this approach accounts for reverberation in its propagation model.

Motivated by the robustness of the above-mentioned approach based on adaptive single-input-multiple-output (SIMO) filtering for single source localization, a TDOA estimation approach for multiple sources maintaining the realistic reverberant propagation model was proposed in [2]. Based on *blind source separation* (BSS) techniques for convolutive mixtures, this method performs a blind adaptive MIMO system identification using the TRINICON framework [5] and can be considered as an extension of the AED to the simultaneous localization of multiple sources.

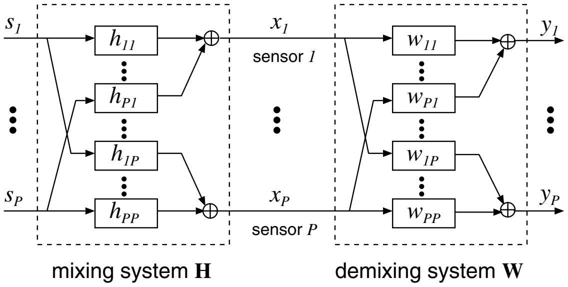


Fig. 2. Multiple-input-multiple-output (MIMO) model for BSS.

Fig. 2 shows the general acoustic setup for BSS. Due to the reverberation in the acoustic environment, the original source signals $s_q(n), q = 1, \ldots, P$ are filtered by a MIMO system $\mathbf{H}$ before they are picked up by the sensors. The recorded signals at the P microphones are denoted by $x_q(n), q=1, \ldots, P$. As the figure indicates, we assume here that the number of active sources is less or equal to the number of microphone signals (i.e., the P source signals in the figure might or might not all be active at the same time). Furthermore the sources are assumed to be mutually uncorrelated. In general this assumption holds for speech and audio signals. To separate the source signals $s_q(n), q = 1, \ldots, P$ without access to the mixing system $\mathbf{H}$, the BSS algorithm forces the output signals $y_q(n), q=1, \ldots, P$ to be statistically decoupled by suitably adapting the weights of the demixing system $\mathbf{W}$.

A. Matrix notation for convolutive mixtures

To express the algorithm for block processing of convolutive mixtures in a general way, we first formulate the convolution of the FIR demixing system of length L in the following convenient matrix form [6], [7]:

$$\mathbf{y}(m, j) = \mathbf{x}(m, j)\mathbf{W}(m), \tag{1}$$

where m denotes the block index over time, and $j = 0, \cdots, N-1$ is a time-shift index within a block of length N, and

$$\begin{aligned}
\mathbf{x}(m, j) &= [\mathbf{x}_1(m, j), \ldots, \mathbf{x}_P(m, j)], &(2)\\
\mathbf{y}(m, j) &= [\mathbf{y}_1(m, j), \ldots, \mathbf{y}_P(m, j)], &(3)\\
\mathbf{W}(m) &= \begin{bmatrix} \mathbf{W}_{11}(m) & \cdots & \mathbf{W}_{1P}(m) \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1}(m) & \cdots & \mathbf{W}_{PP}(m) \end{bmatrix}, &(4)\\
\mathbf{x}_p(m, j) &= [x_p(mL + j), \ldots, x_p(mL - 2L + 1 + j)], &(5)\\
\mathbf{y}_q(m, j) &= [y_q(mL + j), \ldots, y_q(mL - D + 1 + j)] &(6)\\
&= \sum_{p=1}^{P} \mathbf{x}_p(m, j)\mathbf{W}_{pq}(m). &(7)
\end{aligned}$$

In (6), D denotes the number of time lags taken into account to exploit the non-whiteness of the source signals as shown below. $\mathbf{W}_{pq}(m)$ denotes a $2L \times D$ Sylvester matrix that contains all L coefficients of the filter from the p^{th} sensor to the q^{th} output:

$$\mathbf{W}_{pq}(m) = \begin{bmatrix} w_{pq,0} & 0 & \cdots & 0 \\ w_{pq,1} & w_{pq,0} & \ddots & \vdots \\ \vdots & w_{pq,1} & \ddots & 0 \\ w_{pq,L-1} & \vdots & \ddots & w_{pq,0} \\ 0 & w_{pq,L-1} & \ddots & w_{pq,1} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & w_{pq,L-1} \\ 0 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix}. \quad (8)$$

B. Coefficient optimization

Based on the *natural gradient* [8] of the cost function in [5], the Second-Order-Statistics (SOS) realization of the generic TRINICON-based update rule reads:

$$\begin{aligned} \mathbf{W}(m) &= \mathbf{W}(m-1) - \mu\Delta\mathbf{W}(m), & (9) \\ \Delta\mathbf{W}(m) &= 2\sum_{i=0}^{\infty}\beta(i,m)\mathbf{W}(i) \\ &\quad \cdot \left\{\hat{\mathbf{R}}_{\mathbf{yy}}(i) - \hat{\mathbf{R}}_{\mathbf{ss}}(i)\right\}\hat{\mathbf{R}}_{\mathbf{ss}}^{-1}(i), & (10) \end{aligned}$$

where β is a window function with finite support that is normalized according to $\sum_{i=0}^{m}\beta(i,m) = 1$ allowing for online, offline, and block-online algorithms [7]. $\hat{\mathbf{R}}_{\mathbf{yy}}$ is a $PD \times PD$ correlation matrix estimated over a block of length N and containing correlation and cross-correlation sub-matrices between the BSS output signals for time lag $-D+1,\ldots,D-1$.

The BSS variant of this generic SOS natural gradient update also used for multiple TDOA estimation follows immediately by setting $\hat{\mathbf{R}}_{\mathbf{ss}}(i) = \mathrm{bdiag}_D\,\hat{\mathbf{R}}_{\mathbf{yy}}(i)$. Note that there are also efficient approximations of this algorithm with a reduced computational complexity allowing already real-time operation on a regular PC platform (see [9] for implementation details and parameterization).

C. TDOA estimation

For simplicity we focus here on the case of $P = 2$ sources. In this case the SOS-based BSS algorithm described above ideally converges to solutions fulfilling the following conditions:

$$\begin{aligned} h_{11}(n) * w_{12}(n) &= -h_{12}(n) * w_{22}(n), & (11) \\ h_{21}(n) * w_{11}(n) &= -h_{22}(n) * w_{21}(n), & (12) \end{aligned}$$

so that with a suitable coefficient initialization and broadband excitations the BSS algorithm performs a MIMO system identification of the mixing system $\mathbf{H}$ [10].

While performing a system identification the BSS algorithm provides the possibility to extract the location information of the sound sources from the unmixing system $\mathbf{W}$. The TDOA can actually be calculated after each coefficient update by identifying the direct propagation path between the sources and the microphones. We can use (11) to obtain the TDOA for source 1 and (12) to obtain the TDOA for source 2 [2]:

$$\begin{aligned} \hat{\tau}_1 &= \arg\max_n |w_{12}(n)| - \arg\max_n |w_{22}(n)|, & (13) \\ \hat{\tau}_2 &= \arg\max_n |w_{11}(n)| - \arg\max_n |w_{21}(n)|. & (14) \end{aligned}$$

Note that an exact estimation of the mixing system $\mathbf{H}$ is not necessary for all filter taps to perform a successful sound source localization since we only need to identify the filter tap corresponding to the direct propagation path, as can be seen in (13) and (14).

III. SOLVING THE SPATIAL AMBIGUITY

While easily applicable to the localization of one source in several dimensions or to the localization of multiple sources in one dimension, the two-step TDOA-based approach is not sufficient for the simultaneous localization of multiple sources in several dimensions because of the necessity to resolve the spatial ambiguity, regardless of the TDOA estimation method used (see Sect.I). For the problem at hand, the BSS algorithm described in Sect. II and used for the multiple TDOA estimation in Fig. 1 has the interesting feature of both simultaneously estimating TDOAs for several sound sources and unraveling the mixing system, thereby providing estimates of the original source signals at the BSS outputs.

Considering the notations of Fig. 1, we propose to resolve the spatial ambiguity by observing the cross-correlation between the P outputs of each of the T BSS algorithms used to estimate the TDOAs. For each processing block m we can calculate the $T(T-1) \times P^2$ cross-correlation coefficient estimates $\xi_{ij}^{IJ}(m), I, J = 1\ldots, T, I \neq J, i, j = 1,\ldots,P$ from short-time estimates of the cross-correlation $\hat{r}_{ij}^{IJ}$ as follows:

$$\begin{aligned} \hat{r}_{ij}^{IJ}(m,\kappa) &= \lambda_R \cdot \hat{r}_{ij}^{IJ}(m-1,\kappa) \\ &\quad + (1-\lambda_R)\cdot\hat{E}\left[y_i^I(m,n+\kappa)\cdot y_j^J(m,n)\right], & (15) \\ \hat{P}_i^I(m) &= \lambda_P \cdot \hat{P}_i^I(m-1) \\ &\quad + (1-\lambda_P)\cdot\hat{E}\left[\left(y_i^I(m,n)\right)^2\right], & (16) \end{aligned}$$

$$\xi_{ij}^{IJ}(m) = \sqrt{\frac{\sum_{\kappa=-\kappa_{\max}}^{\kappa_{\max}}(\hat{r}_{ij}^{IJ}(m,\kappa))^2}{\hat{P}_i^I(m)\cdot\hat{P}_j^J(m)}}, \quad (17)$$

where $\kappa = -\kappa_{\max},\ldots,\kappa_{\max}$ in (15) is the time-lag in the estimated correlation function $\hat{r}_{ij}^{IJ}(m,\kappa)$ between the i^{th} output from the I^{th} BSS algorithm y_i^I and the j^{th} output from the J^{th} BSS algorithm y_j^J. $\hat{P}_i^I(m)$ in (16) is the estimated signal power of y_i^I. The forgetting factors λ_R in (15) and λ_P in (16) are positive constants which should be chosen close to but smaller than one to allow a recursive smoothing in a block-by-block fashion of the correlation function estimates and of the signal power estimates respectively. In (15) and (16) n is the discrete-time index within one processing block and $\hat{E}(\cdot)$ is the expectation estimate operator over one processing block. Observing that $\hat{r}_{ji}^{JI}(m,\kappa) = \hat{r}_{ij}^{IJ}(m,-\kappa)$ in (15) and (17), we can easily establish the relation $\xi_{ji}^{JI}(m) = \xi_{ij}^{IJ}(m)$. As a consequence, only $T(T-1)/2 \times P^2$ cross-correlation

TABLE I

SPATIAL AMBIGUITY RESOLVER FOR $P = 2$ SOURCES IN $T = 2$ DIMENSIONS.

```
if $\left(\xi_{11}^{12}(m) > \alpha \cdot \xi_{12}^{12}(m)\right)$ and $\left(\xi_{22}^{12}(m) > \alpha \cdot \xi_{21}^{12}(m)\right)$
    $y_1^1$ is correlated with $y_1^2$ and $y_2^1$ is correlated with $y_2^2$.
    $\Rightarrow$ The TDOAs are already in the right order.
    $\begin{cases} position_1 = loc(\hat{\tau}_1^1, \hat{\tau}_1^2) \\ position_2 = loc(\hat{\tau}_2^1, \hat{\tau}_2^2) \end{cases}$
else if $\left(\xi_{12}^{12}(m) > \alpha \cdot \xi_{11}^{12}(m)\right)$ and $\left(\xi_{21}^{12}(m) > \alpha \cdot \xi_{22}^{12}(m)\right)$
    $y_1^1$ is correlated with $y_2^2$ and $y_2^1$ is correlated with $y_1^2$.
    $\Rightarrow$ The TDOAs are permuted.
    $\begin{cases} position_1 = loc(\hat{\tau}_1^1, \hat{\tau}_2^2) \\ position_2 = loc(\hat{\tau}_2^1, \hat{\tau}_1^2) \end{cases}$
else
    $\Rightarrow$ Undecided, keep the last decision.
end
```

coefficients $\xi_{ij}^{IJ}(m), I = 1 \ldots, T, J = I+1, \ldots, T, i, j = 1, \ldots, P$ need to be calculated.

For illustration we consider the example of $P = 2$ sources and $T = 2$ dimensions. In this case (15), (16) and (17) result in a set of four cross-correlation coefficients $\left(\xi_{11}^{12}(m), \xi_{12}^{12}(m), \xi_{21}^{12}(m), \xi_{22}^{12}(m)\right)$ at each block instant. These coefficients can serve directly to resolve the spatial ambiguity according to Table I.

In the procedure described in Table I, the constant $\alpha \geq 1$ is set to provide a security margin. A decision is made only if we have enough confidence, i.e., when both outputs of the first BSS algorithm are not correlated with the same output from the second BSS algorithm. We notice that the normalization of the cross-correlation coefficients using the BSS output power estimates in (17) makes the procedure robust against variations of the sound source power. A reasonable choice for the parameter α is to set it greater than but close to one, independently of the acoustical scenario under consideration.

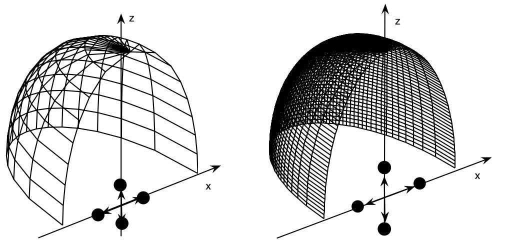
IV. SPATIAL RESOLUTION

In this section we focus on geometrical considerations, concentrating on the actual calculation of each source position independently and using a set of TDOAs estimated for each microphone pair, e.g., by the method described in Sect. II.

Independently of the TDOA estimation method used, the TDOAs are usually represented by integer numbers with maximum absolute value $\hat{\tau}_{\max} = \lceil d\, f_{\mathrm{s}}/c \rceil$ where d is the microphone spacing, f_{s} is the sampling rate and c is the sound velocity. Consequently the estimates of the potential source positions are restricted to a grid of discrete positions. Since with increasing $\hat{\tau}_{\max}$ the number of potential TDOA values for each microphone pair is also increased, we readily see that the density of this grid depends on the sampling rate and on the positions of the microphones.

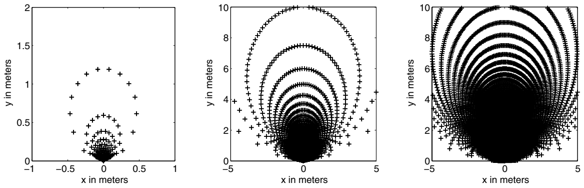
Given a TDOA estimate $\hat{\tau}$ expressed in samples between two microphones, a source can be localized in the far-field by calculating a DOA θ of the plane wave originating from the sound source, measured with respect to the normal of the microphone array axis:

$$\theta \quad = \quad \arcsin\left(\frac{c\,\hat{\tau}}{d\, f_{\mathrm{s}}}\right). \qquad (18)$$



(a) (b)

Fig. 3. Potential DOAs obtained from a four-sensor microphone array centered at the origin in the x-z-plane with spacing $d = 16$cm (left) and $d = 80$cm (right) at the sampling rate $f_{\mathrm{s}} = 16$kHz.



(a) (b) (c)

Fig. 4. Grid of potential positions obtained from a linear three-sensor microphone array centered at the origin point along the x-axis with different microphone spacings and sampling frequencies: (a) $d = 16$cm $f_{\mathrm{s}} = 16$kHz, (b) $d = 80$cm $f_{\mathrm{s}} = 16$kHz, (c) $d = 80$cm $f_{\mathrm{s}} = 48$kHz.

Fig. 3(a) shows the potential directions of arrival in a quarter of the space for a four-sensor array using two pairs placed orthogonal to each other, each with a microphone spacing $d = 16$cm and sampling rate $f_{\mathrm{s}} = 16$kHz. Since we are operating in the far-field the distance (range) from the source to the microphones is disregarded and two TDOA estimates using the two sensor pairs can directly be plugged into (18) to obtain the azimuth and the elevation of the source. Each curve intersection in Fig. 3(a) representing a possible DOA we see that the number of possible DOAs is limited. The spatial resolution can however be drastically improved by increasing the microphone spacing to, e.g., $d{=}80$cm as can be seen in Fig. 3(b).

In the near-field the information on the estimated source position $\hat{\mathbf{r}}_{\mathrm{s}}$ for a TDOA estimate $\hat{\tau}_{ij}$ between sensors i and j can be expressed as

$$c\hat{\tau}_{ij} = \|\hat{\mathbf{r}}_{\mathrm{s}} - \mathbf{r}_i\| - \|\hat{\mathbf{r}}_{\mathrm{s}} - \mathbf{r}_j\|. \qquad (19)$$

In a three-dimensional space such an equation describes a hyperboloïd. Given at least three TDOAs measured from correctly chosen microphone pairs we can obtain the three-dimensional source location estimate $\hat{\mathbf{r}}_{\mathrm{s}}$ at the intersection of three hyperboloïds. Note however that solving such a set of non-linear equations is not trivial. The existence of a closed-form solution is not guaranteed in any case. To illustrate the spatial resolution in the near-field we consider the two-dimensional grids of potential positions in Fig. 4. The grids were obtained from a linear array of three microphones, the center microphone being used twice, once to build a first microphone pair with the left microphone and once to build a second pair with the right microphone. Such

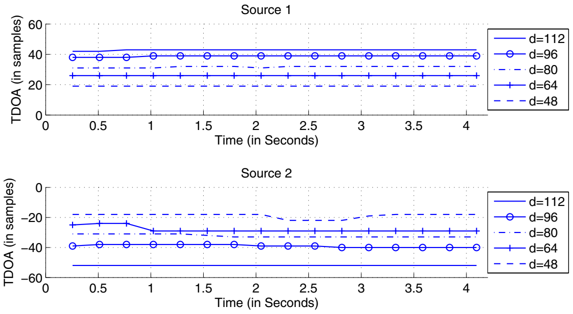


Fig. 5. Results of the TDOA estimation with large microphone spacings (given in cm) in a room with $T_{60} \approx$ 200ms.

an array has the advantage of having an exact closed-form solution. Like in the far-field case we see how increasing the microphone spacing and/or the sampling frequency can improve drastically the spatial resolution of a TDOA-based localizer.

When operating on broadband signals such as speech, the full potential of large microphone arrays can be exploited with the TDOA estimator based on system identification described in Sect. II because of its *broadband* formulation. Actually with other localizers the precision of TDOA estimation itself may be affected due to spatial aliasing if the microphone spacing is too large (i.e., $d > \lambda/2$). This ambiguity typically occurs with narrowband implementations, i.e., where the signal processing is carried out independently for each frequency bin. Moreover, to further improve the spatial resolution of the localizer at a low computational cost, fractional delays (as opposed to integer delays) can be obtained by performing a sinc interpolation [11] on the filters of the unmixing system $\mathbf{W}$ before performing the effective TDOA estimations (13) and (14), without increasing the sampling rate for the BSS operations.

V. EXPERIMENTAL RESULTS

In the following experiments the TDOA estimation was based on a block-online update procedure of the blind adaptive MIMO system identification [9] presented in Sect. II and conducted at the sampling rate f_s = 16kHz. Moreover fractional time delay estimates were obtained (Sect. IV) using an interpolation factor of three to improve the spatial resolution.

A. Robustness with large microphone spacings

We first performed experiments to verify the ability of the TDOA estimation algorithm described in Sect. II to operate on large microphone spacings. Two speech sources were positioned in front of a two-sensor array at $+60^o$ and -80^o relative to the microphone array axis. The experiments were conducted in an environment with reverberation time $T_{60} \approx$ 200ms.

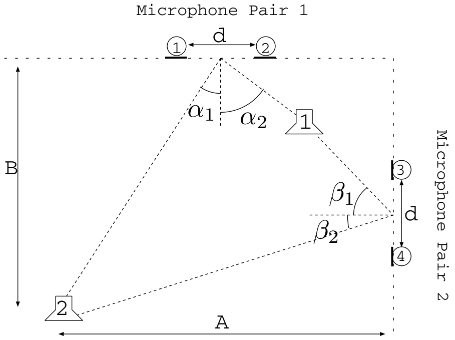


Fig. 6. Experimental setup with the position of the two sources and the two microphone pairs.

TABLE II
EXPERIMENTAL SETUP SPECIFICATIONS.

T_{60}	α_1	α_2	β_1	β_2	A	B
$\approx$ 250ms	$+25^o$	-50^o	-20^o	$+45^o$	490cm	450cm

Fig. 5 shows the TDOA estimation results for microphone spacings ranging from d = 48cm to d = 112cm. Note that all these microphone spacings are too large to avoid spatial aliasing for audio signals at high frequencies so that narrowband algorithms cannot be used here. For a microphone spacing of, e.g., d = 48cm the necessary condition to avoid spatial aliasing is to operate on signals with energy only in frequency regions below 350Hz. This condition is obviously not fulfilled for speech.

As expected, the TDOA estimates increase — in absolute value — with the size of the microphone array. Because of its *broadband* formulation the TDOA estimation based on MIMO system identification shows a robust behavior for a wide range of spacings d.

B. Two-dimensional localization of two speakers

To evaluate the ability of the BSS-based TDOA estimator not only to estimate a set of TDOAs but also to help solving the spatial ambiguity, experiments were conducted in a living room environment with reverberation time $T_{60} \approx$ 250ms. T = 2 TDOA estimators, each providing P = 2 TDOAs, were used with the experimental setup described in Fig. 6 and Table II. The two sound sources were to be localized in the plane spanned by the two sensor pairs. Since the localization of sources situated outside this plane would merely be reduced to finding their position projected on this plane, we assume in the following that both sensors and sources are in the same plane. The microphone spacing for both pairs of microphones was set to 80cm. Combined with an interpolation factor 3 increasing the effective sampling frequency of the TDOA estimation to 3×16 = 48kHz, such a TDOA-based localizer offers a reasonably good spatial resolution for a simultaneous source localization of two sound sources in a two-dimensional plane.

Fig. 7 shows the TDOA estimation results for the two microphone pairs. The two source signals were speech signals. At the beginning of the experiments only source 1 (see Fig. 6) was active and after 20 seconds source 2 started speaking while source 1 remained active. Fig. 7 confirms the robust TDOA estimation results already reported in [2].

Fig. 8 gives the results of the spatial ambiguity resolver implementing the procedure described in Table I using the output signals delivered by the MIMO system identification algorithm at a sampling rate of 16kHz. The upper graphs

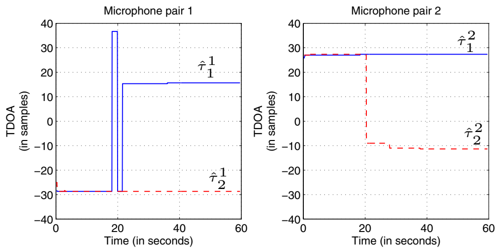


Fig. 7. TDOA estimation results.

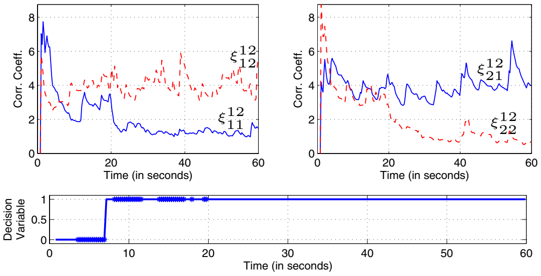


Fig. 8. Results of the spatial ambiguity resolver (in the lower graph, widened portions of line indicate undecided cases, i.e., the "else" condition in Table I).

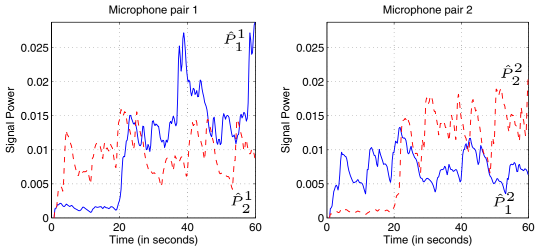


Fig. 9. Estimated signal powers of the BSS outputs.

show the cross-correlation coefficients (17) used in the "if" and "else if" statements of Table I. The lower graph represents the decision made. When this decision variable is set to one the "else if" condition in Table I is true and the TDOA estimates should be permuted. The decision variable was initially set to zero. The cross-correlation coefficients were calculated according to (17) on a block-by-block basis with block shift 4096 samples (the block processing delay of the TDOA estimation algorithm was 1024 samples) and block length 8192 samples. The forgetting factors used to estimate the correlation functions (15) and the signal powers (16) were set to $\lambda_R = 0.9$ and $\lambda_P = 0.9$ respectively. The maximum time-lag to compute the correlation function was $\kappa_{\max} = 1024$ samples. Finally the parameter α in Table I was set to $\alpha = 1.1$.

By observing the evolution of the cross-correlation coefficients in Fig. 8, we can see that the correct decision could be taken unambiguously when the two sources were active (i.e., after 20s). When only one source was active the procedure in Table I could not always lead to an unambiguous decision (see the widened portions of line in Fig. 8) but the resolution of the spatial ambiguity was not critical in this case since the active source was correctly localized in both channels of the TDOA estimators, as can be seen in Fig. 7. However since the TDOA estimators try to localize two sources, the two channels of each estimator might give different estimates when only one source is active. This happened e.g., after 19s in the first TDOA estimator (see Fig. 7). But note that the signal powers (16) shown in Fig. 9 can easily help in this case. When only one source is active we see actually that for each BSS algorithm (i.e., for each microphone pair) the active source is isolated in one output and tends to be suppressed in the other output, thus leading to a large difference in signal powers between the outputs of the BSS algorithms. We thus see clearly that during the first 20s of the experiments, only one source was active and was isolated in the second and first channel of the first and second BSS algorithms, respectively. The active source (which corresponds to source 1 in Fig. 6) can therefore be localized using the TDOA pair $(\hat{\tau}_2^1, \hat{\tau}_1^2)$.

VI. Conclusions

We presented a general TDOA-based framework for the simultaneous multidimensional localization of multiple sound sources. For the scheme to be successful we needed not only a good TDOA estimator but also a mechanism to solve the spatial ambiguity problem occurring when operating on several sources and also in several dimensions. We showed that the TDOA estimation based on blind adaptive MIMO filtering offered a well-suited solution to both issues and furthermore allowed us to obtain a satisfactory spatial resolution.

References

[1] J. Chen, Y. Huang, and J. Benesty. Time delay estimation. In Y. Huang and J. Benesty, editors, *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, pages 255–293. Kluwer Academic Publishers, Boston, 2004.

[2] H. Buchner, R. Aichner, J. Stenglein, H. Teutsch, and W. Kellermann. Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering. In *IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Philadelphia, PA, USA, Mar. 2005.

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

[4] J. Benesty. Adaptive eigenvalue decomposition algorithm for passive acoustic source localization. *J. Acoust. Soc. Am.*, 107:384–391, Jan 2000.

[5] H. Buchner, R. Aichner, and W. Kellermann. TRINICON: A versatile framework for multichannel blind signal processing. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 3, pages 889–892, Montreal, Canada, May 2004.

[6] H. Buchner, R. Aichner, and W. Kellermann. Blind source separation for convolutive mixtures: A unified treatment. In Y. Huang and J. Benesty, editors, *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, pages 255–293. Kluwer Academic Publishers, Boston, 2004.

[7] H. Buchner, R. Aichner, and W. Kellermann. A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics. *IEEE Trans. Speech Audio Processing*, 13(1):120–134, Jan. 2005.

[8] S.-I. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10:251–276, 1998.

[9] R. Aichner, H. Buchner, F. Yan, and W. Kellermann. A real-time blind source separation scheme and its application to reverberant and noisy environments. *Signal Processing*, 86(6):1260–1277, June 2006.

[10] H. Buchner, R. Aichner, and W. Kellermann. Relation between blind system identification and convolutive blind source separation. In *Proc. Joint Workshop on Hands-Free Communication and Microphone Arrays*, Piscataway, NJ, USA, Mar. 2005.

[11] T.I. Laakso, V. Valimaki, M. Karjalainen, and U.K. Laine. Splitting the unit delay. *IEEE Signal Processing Mag.*, 13:30–60, 1996.