IEEE copyright notice

Disambiguation of TDOA Estimation for Multiple Sources in Reverberant Environments

Jan Scheuing and Bin Yang, *Senior Member, IEEE*

***Abstract*—This paper presents a novel approach to estimate the time difference of arrival (TDOA) for multiple sources in reverberant environments. It resolves ambiguities in TDOA estimation caused by multipath propagation and multiple sources. By exploiting two TDOA constraints, the raster condition and the zero cyclic sum condition, we are able to identify and reject the echo path TDOAs and to assign the direct path TDOAs correctly to different sources. For the latter purpose, an efficient algorithm for the synthesis of approximately consistent TDOA graphs has been developed. A real experiment demonstrates the superior performance of our algorithms.**

***Index Terms*—Disambiguation of TDOA estimation in multipath multisource environments (DATEMM), multiple sources, raster matching, reverberant environments, synthesis of consistent graphs, time difference of arrival (TDOA) ambiguity, TDOA estimation, zero cyclic sum matching.**

I. Introduction

THE ESTIMATION of time difference of arrival (TDOA) from the signals of a microphone array plays an important role for many applications like acoustic source localization and beamforming [1], [2]. Two approaches are well known for this task: generalized cross-correlation (GCC) and blind estimation of the room impulse response. In the former case, the TDOA estimate is the peak position in the cross-correlation between two microphone signals [3], [4]. In the latter case, the room impulse responses are estimated by an adaptive eigenvalue decomposition [5]. Both approaches have been approved in many single source scenarios. However, the TDOA estimation remains a difficult problem for multiple sources in reverberant environments because the multipath propagation, the presence of multiple sources, and periodic signals make the TDOA estimation ambiguous.

One idea to combat this problem is to extend the single source impulse response technique [5] to the multiple source case by splitting the multi-input multi-output system to several single-input multi-output systems. Usually, this is achieved either under the ideal assumption that each source is exclusively active during some time intervals [6]–[8] or by a blind source separation [9], [10]. Another completely different idea is to scan the volume of interest for possible sources by maximizing, e.g., the steered response power (SRP) [11].

In this paper, we present a novel approach for disambiguation of TDOA estimation in multipath multisource environments (DATEMM) [12], [13]. It is based on a fairly simple observation of two TDOA constraints implying information redundancy. By applying these constraints to TDOA estimates derived from, e.g., GCC, the ambiguity of TDOA estimation can be significantly reduced. The first constraint is that the extremum positions of a cross-correlation between two microphone signals appear in a well-defined distance which can be predicted from the extremum positions of the corresponding autocorrelations of the microphone signals. Under ideal conditions, combining the cross-correlation with the autocorrelations will uniquely identify the desired direct path TDOA and reject all ambiguous cross-correlation extrema caused by echo paths. The second TDOA constraint is the zero cyclic sum of TDOAs over any number of microphones as long as the TDOAs originate from the same sources and the same propagation paths. This provides an useful mean to assign TDOA estimates to different sources.

In Section II, we formulate the signal model and the TDOA estimation problem. Then we analyze different TDOA ambiguities in Section III and show the basic ideas of DATEMM in Section IV. In Section V, we present an algorithm exploiting the information redundancy contained in the autocorrelation of the microphone signals. By using a so-called raster matching approach, we show how to identify and reject the echo path TDOAs. Section VI formulates the problem of combining TDOA estimates of different microphone pairs in the framework of consistent TDOA graph. Based on the zero cyclic sum constraint, we search for groups of matching TDOAs by a synthesis of consistent TDOA graphs. We present an efficient synthesis algorithm based on consistent triples. Section VII describes a real experiment to locate multiple sources in reverberant environments. It demonstrates the effectiveness and localization accuracy of our algorithms.

II. Signal Model

We assume N acoustic sources and M microphones. We also assume a linear channel from source a to microphone k containing a total number of $\Lambda_{a,k}$ propagation paths. Neglecting noise and assuming omnidirectional characteristics of sources

Manuscript received August 06, 2007; revised July 22, 2008. Current version published October 17, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hiroshi Sawada.

J. Scheuing was with the Chair of System Theory and Signal Processing, University of Stuttgart, 70550 Stuttgart, Germany. He is now with Bosch Engineering GmbH, 74232 Abstatt, Germany (e-mail: jan.scheuing@LSS.uni-stuttgart.de).

B. Yang is with the Chair of System Theory and Signal Processing, University of Stuttgart, 70550 Stuttgart, Germany (e-mail: bin.yang@LSS.uni-stuttgart.de).

Digital Object Identifier 10.1109/TASL.2008.2004533

and microphones, the discrete-time signal of the kth microphone is given by

$$x_k(n) = \sum_{a=1}^{N} \sum_{\mu=0}^{\Lambda_{a,k}-1} h_{a,k,\mu} s_a(n - \tau_{a,k,\mu}). \quad (1)$$

$s_a(n)$ is the signal of source a. $h_{a,k,\mu}$ and $\tau_{a,k,\mu}$ are the amplitude and (integer) delay of path μ between source a and microphone k, respectively. All delays $\tau_{a,k,\mu}$ are sorted in ascending order, i.e., $\tau_{a,k,\mu} > \tau_{a,k,\nu}$ for $\mu > \nu$. The TDOA between path μ of microphone k and path ν of microphone l for source a is

$$n_{a,kl,\mu\nu} = \tau_{a,k,\mu} - \tau_{a,l,\nu}. \quad (2)$$

We assume that the line-of-sight propagation condition is satisfied for all pairs of source and microphone. Hence, all direct paths exist and are denoted by the path index $\mu = 0$; otherwise, a localization would be hardly possible.

The goal of *TDOA estimation* is to estimate a *source TDOA vector*

$$\mathbf{n}_a = [n_{a,12,00}, n_{a,13,00}, \ldots, n_{a,M-1M,00}]^T \quad (3)$$

of length $\binom{M}{2}$ for each source a subject to four requirements.

- All TDOAs in $\mathbf{n}_a$ should originate from direct paths only.
- All TDOAs in $\mathbf{n}_a$ should originate from the same source.
- The vector $\mathbf{n}_a$ should be as complete as possible (few missing elements).
- The TDOA estimation should be computationally as efficient as possible.

While the last two requirements represent soft wishes, the first two requirements are mandatory because otherwise we would obtain a wrong source position estimation. Unfortunately, a number of reasons make the TDOA estimation ambiguous and difficult. Below, three different types of TDOA ambiguity are analyzed using simple scenarios [12]. For notational convenience, we drop the index a or μ in (1) if we consider only one source or one path.

III. TDOA Ambiguities

A. Ambiguity Due to Periodic Signals

The first ambiguity is well known. Speech signals contain voiced segments which show a high periodicity. The same also happens for many natural and machine sounds. The periodic extrema in the autocorrelation of the source signals will also appear in the cross-correlation of the microphone signals, even for a single source without multipath propagation. GCC [4] have been proposed to combat this problem.

B. Multipath Ambiguity

The second TDOA ambiguity is caused by the multipath propagation. For a single source signal $(N = 1)$ propagating on Λ_k paths to microphone k, we obtain the signal

$$x_k(n) = \sum_{\mu=0}^{\Lambda_k-1} h_{k,\mu} s(n - \tau_{k,\mu}). \quad (4)$$

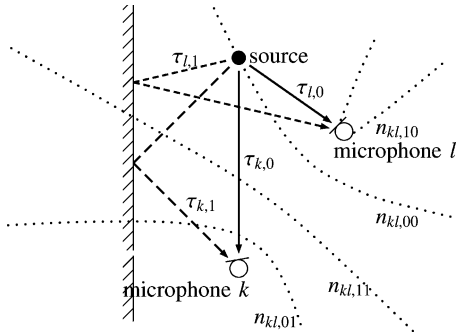

Fig. 1. Assuming a direct path (solid) and an echo path (dashed) from the source to each of the two microphones, four TDOA values corresponding to four hyperbola (dotted) are possible. Only the hyperbola of the direct path TDOA $n_{kl,00}$ passes the source location.

If $s(n)$ is zero mean and white, the cross-correlation $r_{kl}(n) = \mathrm{E}[x_k(m+n)x_l(m)]$ between the two microphone signals $x_k(n)$ and $x_l(n)$ will show $\Lambda_k\Lambda_l$ extrema at the TDOA positions

$$n_{kl,\mu\nu} = \tau_{k,\mu} - \tau_{l,\nu},\ \mu \in \{0, \ldots, \Lambda_k - 1\}, \nu \in \{0, \ldots, \Lambda_l - 1\}. \quad (5)$$

We are only interested in the *direct path TDOA* $n_{kl,00} = \tau_{k,0} - \tau_{l,0}$. All other $\Lambda_k\Lambda_l - 1$ TDOA values involve at least one echo path and are called *echo path TDOAs*. They correspond to wrong hyperbola of possible source location as shown in Fig. 1. The problem is how to determine which of the $\Lambda_k\Lambda_l$ extrema in the cross-correlation $r_{kl}(n)$ represents the desired direct path TDOA.

C. Multiple Source Ambiguity

The third TDOA ambiguity is due to multiple sources. Assuming a direct path propagation of N source signals, the kth microphone signal is

$$x_k(n) = \sum_{a=1}^{N} h_{a,k} s_a(n - \tau_{a,k}). \quad (6)$$

Here, $\tau_{a,k}$ denotes the direct path delay from source a to microphone k. If all source signals are zero mean, white, and uncorrelated, the cross-correlation $r_{kl}(n)$ will show N extrema at the TDOA positions

$$n_{a,kl} = \tau_{a,k} - \tau_{a,l}. \quad (7)$$

The difficulty is to assign them correctly to the N sources such that TDOAs of the same source are grouped together as in (3). By considering $L \leq \binom{M}{2}$ microphone pairs where each pair contributes N extrema, there are N^L different possibilities to construct *one* length-L TDOA vector whose each element can take N possible TDOA values. Any erroneous combination of TDOAs will likely cause a phantom source; see Fig. 2.

In practice, all three types of ambiguity occur simultaneously, making the TDOA estimation and grouping even more difficult. Fig. 3 shows the generalized cross-correlation PHAT [4] of two microphone signals as two speakers talked simultaneously in a medium-reverberant room; see Section VII for more details about the experiment. Since the GCC shows many peaks, it is

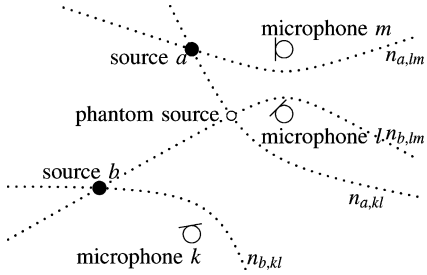

Fig. 2. Combination of TDOAs originating from different sources causes a phantom source.

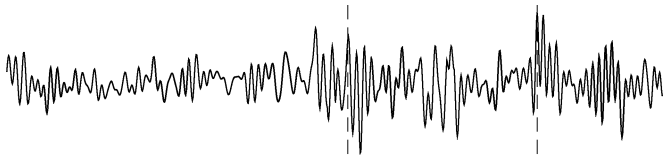

Fig. 3. Generalized cross-correlation of two microphone signals for two sources in a reverberant room.

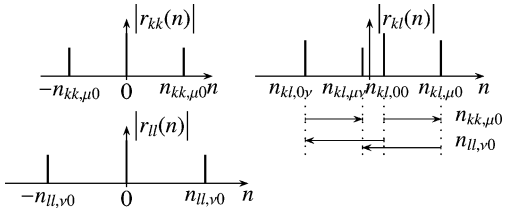

Fig. 4. Relationship between extremum positions in auto- and cross-correlation.

not trivial to estimate the two direct path TDOAs (indicated by dashed lines) and to assign them correctly to both sources.

IV. Principles of TDOA Disambiguation

A. Raster Condition

Below, we present a novel approach DATEMM [12] to resolve these TDOA ambiguities. It is based on two simple observations. The first one is the relationship between the extremum positions in the cross-correlation and autocorrelation of the microphone signals. For simplicity, we consider the single source and two path scenario in Fig. 1 again. Fig. 4 shows the four extremum positions in the cross-correlation $r_{kl}(n)$. It also shows the extremum positions of the two autocorrelations $r_{kk}(n) = \mathrm{E}[x_k(m+n)x_k(m)]$ and $r_{ll}(n) = \mathrm{E}[x_l(m+n)x_l(m)]$. Obviously, the cross-correlation extrema appear in well-defined distances which can be predicted by the extremum positions in the autocorrelations.

Let

$$\begin{aligned} x_k(n) &= h_{k,0}s(n-\tau_{k,0}) + h_{k,\mu}s(n-\tau_{k,\mu}) \\ x_l(n) &= h_{l,0}s(n-\tau_{l,0}) + h_{l,\nu}s(n-\tau_{l,\nu}) \end{aligned} \tag{8}$$

be two microphone signals. If $s(n)$ is zero mean and white, the autocorrelations $r_{kk}(n)$ and $r_{ll}(n)$ show, in addition to the zero-lag extrema $r_{kk}(0)$ and $r_{ll}(0)$, four other extrema at the positions $n_{kk,\mu 0} = \tau_{k,\mu} - \tau_{k,0}$, $n_{kk,0\mu} = -n_{kk,\mu 0}$, and $n_{ll,\nu 0} = \tau_{l,\nu} - \tau_{l,0}$, $n_{ll,0\nu} = -n_{ll,\nu 0}$. They coincide with the differences of the cross-correlation extremum positions

$$\begin{aligned} n_{kk,\mu 0} &= \tau_{k,\mu} - \tau_{k,0} = (\tau_{k,\mu} - \tau_{l,\eta}) - (\tau_{k,0} - \tau_{l,\eta}) \\ &= n_{kl,\mu\eta} - n_{kl,0\eta} > 0 \\ n_{ll,\nu 0} &= \tau_{l,\nu} - \tau_{l,0} = (\tau_{k,\eta} - \tau_{l,0}) - (\tau_{k,\eta} - \tau_{l,\nu}) \\ &= n_{kl,\eta 0} - n_{kl,\eta\nu} > 0 \end{aligned} \tag{9}$$

for any direct or echo path η. This condition is referred to as the *raster condition*. Since the direct path always has the shortest delay, $n_{kk,\mu 0}$, and $n_{ll,\nu 0}$ in (9) are positive. This implies for the first sensor k that the cross-correlation extremum $n_{kl,\mu\eta}$ of the echo path μ is always right to the extremum $n_{kl,0\eta}$ of the direct path 0. In contrast, for the second sensor l, the extremum $n_{kl,\eta\nu}$ of the echo path ν is left to the extremum $n_{kl,\eta 0}$ of the direct path 0. In Fig. 4, the relationships in (9) are illustrated by arrows below $r_{kl}(n)$. The arrows have a length equal to $n_{kk,\mu 0}$ or $n_{ll,\nu 0}$. The arrow direction is defined in such a way that the arrow $n_{kk,\mu 0}$ of the first sensor k points from left to right and the arrow $n_{ll,\nu 0}$ of the second sensor l points from right to left, respectively. Combined with the previous observation of "echo path extremum is right/left to direct path extremum for senor k/l," we conclude that *each arrow points from the direct path extremum to the echo path extremum* for the corresponding sensor. Clearly, the direct path TDOA $n_{kl,00}$ is that extremum in $r_{kl}(n)$ which shows only arrow tails and no arrowheads. This *raster matching* approach combines the extremum positions of both auto- and cross-correlations and enables us to identify the desired direct path TDOA $n_{kl,00}$ even in a reverberant environment.

B. Zero Cyclic Sum Condition

The second important observation is as follows. For each subset of microphones $\{k,l,\ldots,m,o\} \subseteq \{1,\ldots,M\}$ and the same number of corresponding direct or echo paths $\mu,\nu,\ldots,\eta,\kappa$, the following *zero cyclic sum condition*

$$\begin{aligned} &n_{a,kl,\mu\nu} + \ldots + n_{a,mo,\eta\kappa} + n_{a,ok,\kappa\mu} \\ &\quad = (\tau_{a,k,\mu} - \tau_{a,l,\nu}) + \ldots + (\tau_{a,m,\eta} - \tau_{a,o,\kappa}) \\ &\qquad + (\tau_{a,o,\kappa} - \tau_{a,k,\mu}) \\ &\quad = 0 \end{aligned} \tag{10}$$

always holds for TDOAs originating from the same source [14]. In (10), each path delay $\tau_{a,k,\mu}$ occurs two times with opposite signs. If the cyclic sum is not zero, either 1) different paths for the same microphone are used or 2) different sources are involved. This *zero cyclic sum matching* allows us to group TDOA estimates of different microphone pairs according to their sources and hence to avoid phantom sources like in Fig. 2.

In DATEMM, we use two additional criteria for TDOA disambiguation.

- The direct path amplitudes $h_{a,k,0}$ in (1) are always positive. Hence, we only search for the maxima instead of extrema in the cross-correlation $r_{kl}(n)$ [15].
- Due to the triangular inequality, each direct path TDOA between two microphones can never exceed in magnitude

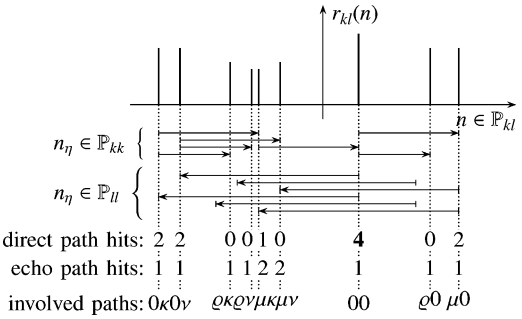

Fig. 5. Number of direct path and echo path hits. The desired direct path TDOA $n_{kl,00}$ has the highest number of direct path hits (arrow tails) 4.

the distance between the two microphones divided by the speed of sound. Any TDOA estimate beyond this geometrical upper bound is discarded.

V. Raster Matching

In the following, we assume that a set of TDOA estimates $\mathbb{P}_{kl}$ has been computed for each microphone pair (k, l) by GCC. In the ideal case, $\mathbb{P}_{kl}$ contains all N direct path TDOAs of N sources and no echo path TDOAs. In reality, estimation errors will cause both *false detection* (echo path TDOA accepted) and *miss detection* (direct path TDOA rejected). While a false detection is typically caused by echo paths, a miss detection of a source is usually due to the weak amplitude of that source signal. In general, a miss detection is more critical than a false detection in our application. While a rejected direct path TDOA is lost for ever, echo path TDOAs in $\mathbb{P}_{kl}$ can still be identified by DATEMM. Hence, we strongly recommend an overestimation $|\mathbb{P}_{kl}| > N$ for the initial TDOA estimation.

In addition, we also compute the autocorrelations of all microphone signals. The positive positions of the strongest autocorrelation extrema of the microphone signal $x_k(n)$ are collected in the set $\mathbb{P}_{kk}$. They are used to identify the echo path TDOAs in $\mathbb{P}_{kl}$. In a first attempt, the raster condition in (9) motivates a search for all pairs of TDOAs (n_μ, n_ν) whose difference matches an autocorrelation extremum position

$$n_\mu - n_\nu = n_\eta \quad \text{with} \quad n_\mu, n_\nu \in \mathbb{P}_{kl},\ n_\eta \in (\mathbb{P}_{kk} \cup \mathbb{P}_{ll}). \tag{11}$$

If such a pair has been found, that TDOA assigned to an arrowhead can be rejected immediately as Fig. 4 illustrates. Unfortunately, this hard decision would not work in practice due to several reasons.

First, the raster condition (9) is necessary but not sufficient. It is theoretically possible that two TDOA estimates from $\mathbb{P}_{kl}$ caused by different paths or different sources match one of the autocorrelation extremum positions. If we erroneously reject a direct path TDOA too early based on the matching of only one pair of TDOAs, this direct path TDOA is lost in all future steps. In order to prevent this from happening, we propose to count all direct path hits (arrow tails) and echo path hits (arrowheads). Fig. 5 shows the number of direct path and echo path hits in a scenario with one direct and two echo paths between one source and each of the two microphones k, l. In this case, the desired direct path TDOA $n_{kl,00}$ is that cross-correlation maximum position with the highest direct path hits 4.

Second, the raster condition (9) is not satisfied exactly due to estimation errors and quantized time delays. Instead of a perfect raster match, we tolerate an approximate raster match

$$|n_\eta - |n_\mu - n_\nu|| < 0.5\Gamma_{\mathrm{TWRM}},\ n_\eta \in (\mathbb{P}_{kk} \cup \mathbb{P}_{ll}),\ n_\mu, n_\nu \in \mathbb{P}_{kl} \tag{12}$$

where the so called *tolerance width of raster match* (TWRM) Γ_{TWRM} is typically in the order of a few samples. In addition, we introduce a quality value $q(n_\mu)$ for each TDOA estimate $n_\mu \in \mathbb{P}_{kl}$. In the case of GCC for TDOA estimation, it is defined by

$$\begin{aligned} q(n_\mu) = r_{kl}(n_\mu) &+ \sum_{n_\eta \in \mathbb{P}'_{kk}} \operatorname{sign}(n_\nu - n_\mu)\, |r_{kk}(n_\eta)| \\ &\times \Gamma_{\mathrm{TFRM}}(n_\eta - |n_\mu - n_\nu|) \\ &+ \sum_{n_\eta \in \mathbb{P}'_{ll}} \operatorname{sign}(n_\mu - n_\nu)\, |r_{ll}(n_\eta)| \\ &\times \Gamma_{\mathrm{TFRM}}(n_\eta - |n_\mu - n_\nu|) \end{aligned} \tag{13}$$

with $n_\nu \in \mathbb{P}_{kl}$. For each TDOA estimate $n_\mu \in \mathbb{P}_{kl}$, its initial quality value is the positive cross-correlation amplitude $r_{kl}(n_\mu)$. It is then increased or decreased during the subsequent raster matching. The sets

$$\begin{aligned} \mathbb{P}'_{kk} &= \{n_\eta \in \mathbb{P}_{kk} \,|\, |n_\eta - |n_\mu - n_\nu|| < 0.5\Gamma_{\mathrm{TWRM}}\} \\ \mathbb{P}'_{ll} &= \{n_\eta \in \mathbb{P}_{ll} \,|\, |n_\eta - |n_\mu - n_\nu|| < 0.5\Gamma_{\mathrm{TWRM}}\} \end{aligned}$$

contain those autocorrelation extremum positions from $\mathbb{P}_{kk}$ and $\mathbb{P}_{ll}$ which match a pair of cross-correlation TDOA estimates (n_μ, n_ν) in the sense of (12). $\Gamma_{\mathrm{TFRM}}(n)$ is a nonnegative symmetric function with the width Γ_{TWRM}. It is called the *tolerance function of raster match* (TFRM) and assigns a high/low score to a good/bad raster match. One simple example is the triangular function

$$\Gamma_{\mathrm{TFRM}}(n) = \begin{cases} 1 - \frac{|n|}{0.5\Gamma_{\mathrm{TWRM}}} & \text{if } |n| < 0.5\Gamma_{\mathrm{TWRM}} \\ 0 & \text{if } |n| \geq 0.5\Gamma_{\mathrm{TWRM}} \end{cases}. \tag{14}$$

The sign function $\operatorname{sign}(n)$ in (13) adds/subtracts a weighted magnitude of the involved autocorrelation extremum $r_{kk}(n_\eta)$ or $r_{ll}(n_\eta)$ to/from the quality value $q(n_\mu)$ if n_μ is likely a direct/echo path TDOA. This can be easily seen from Fig. 4 whether n_μ corresponds to an arrow tail ($n_\mu < n_\nu$ for $\mathbb{P}_{kk}$ or $n_\mu > n_\nu$ for $\mathbb{P}_{ll}$) or an arrowhead.

The final decision about the TDOA estimate n_μ is based on the final quality value of $q(n_\mu)$

$$n_\mu \text{ is viewed as } \begin{cases} \text{a direct path TDOA, if } q(n_\mu) > t_{kl} \\ \text{an echo path TDOA, otherwise.} \end{cases} \tag{15}$$

All direct path TDOA estimates are collected in a reduced set $\mathbb{P}'_{kl}$. We used the threshold $t_{kl} = \min_{n_\nu \in \mathbb{P}_{kl}} r_{kl}(n_\nu)$ in (15). This choice is intentionally conservative in order to ensure that no direct path TDOAs are rejected at this early step. Echo path TDOAs which are still contained in $\mathbb{P}'_{kl}$ can be detected by using the zero cyclic sum condition in the next section.

VI. Consistent TDOA Graphs

Starting from the sets $\mathbb{P}'_{kl}$ of direct path TDOA estimates determined in the previous section, we now apply the zero cyclic

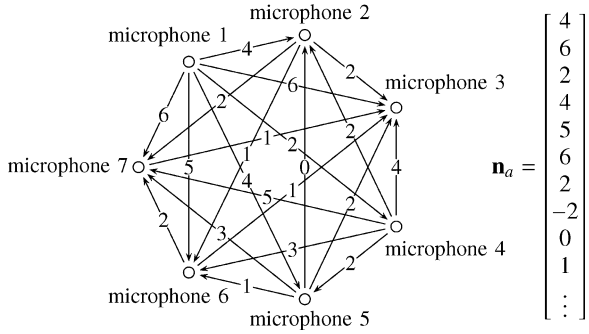

Fig. 6. Fully connected consistent TDOA graph with seven nodes and the related source TDOA vector $\mathbf{n}_a$.

sum condition (10) to examine which of the TDOA estimates of different microphone pairs belong together to the same source. We study this combination problem in the framework of consistent graphs.

A. TDOA Graph

As shown in Fig. 6, the content of a source TDOA vector $\mathbf{n}_a$ defined in (3) can be visualized by a weighted directed graph. It is called a *TDOA graph*. Each node represents a microphone and each directed edge between two nodes has a weight corresponding to the TDOA value between these two microphones. The edge direction is given by the order of microphones in the cross-correlation. A change of the edge direction has the effect of a sign change of its weight. For simplicity, we use integer TDOA values for illustration. However, the concept of TDOA graphs applies to real-valued weights as well.

A fully connected TDOA graph corresponds to a source TDOA vector of length $\binom{M}{2}$. For *incomplete graphs*, some vector elements are missing. The aim of this section is to compose TDOA graphs with the highest number of nodes and the maximum degree of connections from the sets of TDOA estimates $\mathbb{P}'_{kl}$.

A TDOA graph consisting of exact TDOA values is always *consistent* in the sense that the sum of all edge weights along any closed path is zero according to the zero cyclic sum condition (10). This is very similar to Kirchhoff's second law for electrical circuits (voltage graphs). The difference is that we are interested in the synthesis instead of analysis of consistent graphs. In the graph theory, a closed path with a zero sum of weights is sometimes called a *zero-cost cycle*. Thus, a consistent graph only contains zero-cost cycles. Unfortunately, the problem of synthesis of consistent graphs has never been addressed in the literature to our knowledge. This is the reason why we have to develop efficient synthesis algorithms by ourselves.

B. Consistency Check of a TDOA Graph

Below, we first analyze the complexity of different strategies to check the consistency of a given TDOA graph. We assume a fully connected graph with M nodes and $\binom{M}{2}$ edges. For simplicity, we count each addition or comparison as one operation.

In analogy to electrical voltage and potential, we can define a *time potential* at each node as its TDOA value with respect to any reference node. Let the time potential of the reference node be zero. The time potentials of the remaining $M-1$ nodes are then determined by their TDOAs relative to the reference node. In order to check the consistency of the graph, we only need to compare the remaining $\binom{M-1}{2}$ edge weights between these $M-1$ nodes with the corresponding potential differences. This leads to

$$C_{\text{potential}} = 2 \cdot \binom{M-1}{2} = (M-1)(M-2) \qquad (16)$$

operations. It can be shown that this is also the minimum complexity of consistency check.

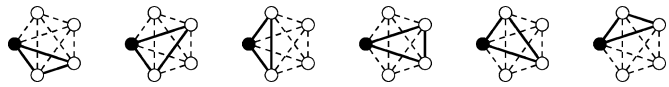

Fig. 7. Six independent triples sharing a common reference node •.

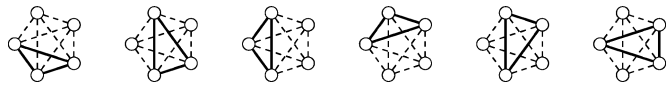

Fig. 8. Six independent triples without a common reference node.

Alternatively, we can check the consistency of $\binom{M-1}{2}$ *independent triples*; see Fig. 7. A *triple* is a three-node subgraph. Since the consistency check of a triple requires two operations, the total complexity is $C_{\text{triple}} = 2 \cdot \binom{M-1}{2} = C_{\text{potential}}$. The advantage of this approach over the time potential one is that there is no need to define a reference node. Instead of the six independent triples in Fig. 7 sharing a common reference node, we can also check the six independent triples in Fig. 8 without a reference node. It can be shown that the remaining $\binom{M}{3} - \binom{M-1}{2} = \binom{M-1}{3}$ dependent triples need not to be studied further since their consistency follows immediately from that of the independent triples.

C. Strategies of Graph Synthesis

Given the sets $\mathbb{P}'_{kl}$ of direct path TDOA estimates for the microphone pair (k,l), there exist different strategies to synthesize a TDOA graph. The brute force approach tries all possible combinations of TDOA values for $\binom{M}{2}$ edges. Since a TDOA graph can be incomplete, we consider $|\mathbb{P}'_{kl}|+1$ possibilities for the edge (k,l) $|\mathbb{P}'_{kl}|$ different edge weights and the case of a missing edge. The total number of possible TDOA graphs is the product of $|\mathbb{P}'_{kl}|+1$ for all $\binom{M}{2}$ edges with $1 \le k < l \le M$

$$G_{\text{brute-force}} = \prod_{k=1}^{M-1} \prod_{l=k+1}^{M} \left(|\mathbb{P}'_{kl}|+1\right). \qquad (17)$$

For $M = 8$ microphones and $|\mathbb{P}'_{kl}| = 7$ TDOA estimates for each microphone pair, the number of TDOA graphs to be checked for consistency is $G_{\text{brute-force}} = (7+1)^{\binom{M}{2}} = 8^{28} \approx 2 \cdot 10^{25}$. This is unacceptable for real-time applications.

One possibility to reduce the complexity is the use of the time potential approach introduced in the previous subsection. This requires the choice of a common reference node for *all* sources. Obviously, a reference node should be connected to all other

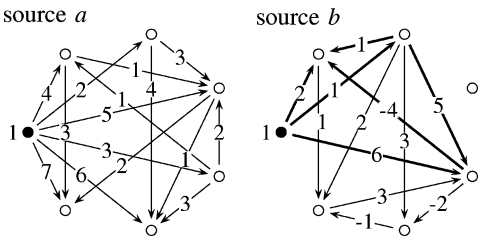

Fig. 9. Incomplete TDOA graphs make a choice of a common reference node for all sources difficult.

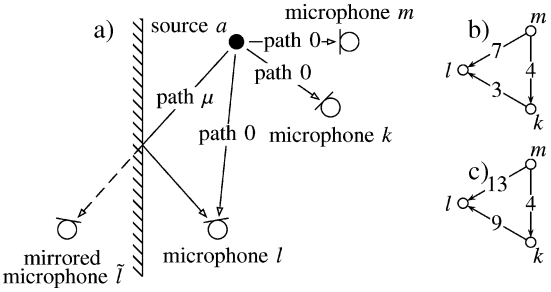

Fig. 10. Both the microphone l and the mirrored microphone $\tilde{l}$ cause a consistent TDOA graph in b) and c).

nodes in order to determine their time potential. This is, however, difficult in practice since TDOA graphs are often incomplete. Due to miss detection of TDOA, different edges in TDOA graphs of different sources are missing as illustrated in Fig. 9 for source a and b. For source a, node 1 is a good reference node since it is connected to all other nodes. The complete TDOA graph as shown can thus be synthesized. In contrast, node 1 is a bad reference node for source b since it is only connected to three nodes. Hence, only the bold subgraph containing four nodes can be synthesized; the other TDOA estimates are wasted. The situation becomes even worse if we choose any other reference node than node 1. For this reason, we do not follow the time potential approach. Instead, we propose a synthesis algorithm based on consistent triples.

D. Consistent Triples

A *TDOA triple* involves three nodes and three edges. It is consistent if its cyclic sum of edge weights is zero. Note that besides the desired direct path TDOA triples, other combinations of TDOAs can also form a consistent triple. There are two reasons for this phenomenon of *false consistency*. First, the zero cyclic sum condition (10) is necessary but not sufficient for TDOAs originating from a common source. Theoretically, scenarios are possible where TDOAs of different sources a, b, and c satisfy

$$n_{a,kl,00} + n_{b,lm,00} + n_{c,mk,00} = 0. \tag{18}$$

Of course, the probability of this occurrence is small for randomly placed sources.

The second, more critical situation is the *mirrored microphone* as shown in Fig. 10. The microphone l is close to a wall. When we model sound propagation and reflection by acoustic rays like the image source method [16], a reflecting wall has the same effect on a microphone signal as a corresponding mirrored microphone $\tilde{l}$. Clearly, both the direct path graph b) and the graph c) in Fig. 10 containing two echo path TDOAs $n_{a,kl,0\mu}$ and $n_{a,lm,\mu 0}$ are consistent because of

$$n_{a,kl,0\mu} + n_{a,lm,\mu 0} + n_{a,mk,00} = 0 \quad \text{for any } \mu \geq 0. \tag{19}$$

This false consistency cannot be identified at this stage. We will see in Section VII that the residual error in the source position estimation will help us to resolve this ambiguity.

Another problem is that a TDOA triple is never exactly consistent in practice because TDOA estimates are quantized and noisy. As a consequence, we look for *approximately consistent* TDOA triples and graphs whose cyclic sum of TDOAs is approximately zero. In analogy to the tolerance function of raster match $\Gamma_{\mathrm{TFRM}}(n)$ in Section V, we also introduce a nonnegative, symmetric, and smoothly decreasing *tolerance function of triple match* (TFTM) $\Gamma_{\mathrm{TFTM}}(n)$ to take the approximate consistency into account. The width of $\Gamma_{\mathrm{TFTM}}(n)$ is characterized by the parameter *tolerance width of triple match* (TWTM) Γ_{TWTM}. As for $\Gamma_{\mathrm{TFRM}}(n)$ in (14), the choice of $\Gamma_{\mathrm{TFTM}}(n)$ is up to the user.

TABLE I
PROCEDURE OF THE SYNTHESIS ALGORITHM

S1:	Find all approximately consistent TDOA triples.
S2:	Determine a quality value for each consistent TDOA triple.
S3:	Use the highest-quality TDOA triple as the initial triple.
S4:	Extend the initial triple to a consistent quadruple for each of the remaining fourth nodes.
S5:	Combine all consistent quadruples with a common initial triple to a star graph.
S6:	Connect as much fourth nodes of the star graph as possible.
S7:	Return this TDOA graph.
S8:	From the unused TDOA triples, select the highest-quality one as the initial triple for the next TDOA graph and go back to S4.

E. Efficient Synthesis Algorithm

Below, we present an efficient algorithm for the synthesis of approximately consistent TDOA graphs based on consistent triples. The starting point is the sets of TDOA estimates $\mathbb{P}'_{kl}$ for all microphone pairs (k,l). We assume that each TDOA estimate $n_\mu \in \mathbb{P}'_{kl}$ has a corresponding quality value $q(n_\mu)$.

Table I illustrates the basic procedure of our synthesis algorithm. In the first step S1, we search for all approximately consistent TDOA triples. For each microphone triple (k,l,m), let $\mathbb{T}_{klm}$ denote the set of approximately consistent TDOA triples

$$(n_{kl,\mu}, n_{lm,\nu}, n_{mk,\eta}) \quad \text{with} \quad |n_{kl,\mu} + n_{lm,\nu} + n_{mk,\eta}| < 0.5\Gamma_{\mathrm{TWTM}}. \tag{20}$$

The total number of TDOA triples to be checked is

$$\sum_{k=1}^{M-2} \sum_{l=k+1}^{M-1} \sum_{m=l+1}^{M} |\mathbb{P}'_{kl}| \cdot |\mathbb{P}'_{lm}| \cdot |\mathbb{P}'_{mk}|. \tag{21}$$

This number can be further reduced if the TDOA sets are stored as sorted lists. Since the number of consistent TDOA triples is much smaller than the total number of possible TDOA triples

$$|\mathbb{T}_{klm}| \ll |\mathbb{P}'_{kl}| \cdot |\mathbb{P}'_{lm}| \cdot |\mathbb{P}'_{mk}| \tag{22}$$

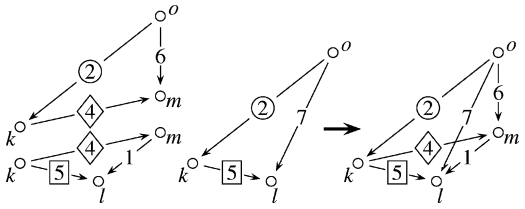

Fig. 11. Three consistent triples with pairwise common edge weights are combined to a consistent quadruple.

the complexity of our synthesis algorithm is significantly reduced.

In step S2, we compute a quality value

$$q_{\mathrm{T}}(n_{kl,\mu}, n_{lm,\nu}, n_{mk,\eta}) = \Gamma_{\mathrm{TFTM}}(n_{kl,\mu} + n_{lm,\nu} + n_{mk,\eta}) \cdot (q(n_{kl,\mu}) + q(n_{lm,\nu}) + q(n_{mk,\eta})) \quad (23)$$

for each consistent TDOA triple from $\mathbb{T}_{klm}$. It takes both the preciseness of the zero cyclic sum match and the quality values of the TDOA estimates into account. The larger the value is, the higher the quality of the TDOA triple

Then we choose that TDOA triple $(n_{kl,\mu_1}, n_{lm,\nu_1}, n_{mk,\eta_1})$ with the highest triple quality as the initial triple for our synthesis algorithm at step S3. For each of the remaining "fourth" microphones $o \in \{1, \ldots, M\} \setminus \{k, l, m\}$, we try to extend the initial TDOA triple to a consistent TDOA quadruple involving the four microphones k, l, m, o at step S4. We search for at least two other TDOA triples in the new sets $\mathbb{T}_{lmo}$, $\mathbb{T}_{mok}$, and $\mathbb{T}_{okl}$ with pairwise common edge weights. If, for example, the triples $(n_{mo,\sigma_2}, n_{ok,\varrho_2}, n_{km,\eta_2}) \in \mathbb{T}_{mok}$ and $(n_{ok,\varrho_3}, n_{kl,\mu_3}, n_{lo,\kappa_3}) \in \mathbb{T}_{okl}$ have common edge weights

$$n_{kl,\mu_1} = n_{kl,\mu_3}, \quad n_{mk,\eta_1} = -n_{km,\eta_2}, \quad n_{ok,\varrho_2} = n_{ok,\varrho_3}$$

we build a fully connected consistent TDOA quadruple by combining these three triples; see Fig. 11.

We repeat the synthesis of quadruples for all $M-3$ fourth nodes. Those fourth nodes $o, p, \ldots$ for which a complete consistent quadruple has been successfully composed are collected in a new set $\mathbb{K}$ with $|\mathbb{K}| \leq M-3$. The consistent quadruples with the common initial triple (k, l, m) form a consistent but not fully connected *star graph* at step S5. Fig. 12 shows such a star graph for $|\mathbb{K}| = 2$ on the left-hand side. The missing $\binom{|\mathbb{K}|}{2}$ edges among the $|\mathbb{K}|$ fourth nodes can be completed by triples at step S6 which have two edges in common with the star graph. One such completing triple from $\mathbb{T}_{mop}$ is shown in Fig. 12. Clearly, any other matching triple from $\mathbb{T}_{lop}$ or $\mathbb{T}_{kop}$ can also be used for this purpose.

Remember that, due to consistent echo path TDOA triples like in Fig. 10(c), different star graphs for the same initial triple are possible. Fig. 13 illustrates this phenomenon for $M = 6$ nodes. Starting with the boldface initial triple, four quadruples have been synthesized for the remaining three nodes o, p, q. While the nodes p and q each produce only one quadruple, the first two quadruples caused by the fourth node o may originate from the true microphone and its mirror. At this position, we are not able to decide which one is the correct one. Hence, we accept all four quadruples and combine them to two star graphs as shown in Fig. 13. Then we complete these star graphs by looking

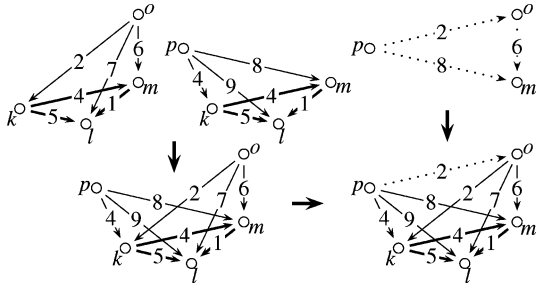

Fig. 12. Combining quadruples with a common initial triple (bold line) results in a star graph. Then the star graph is completed by matching triples (dotted line).

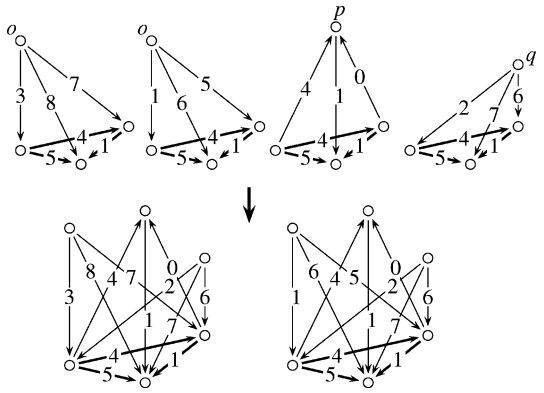

Fig. 13. When combining quadruples to a star graph, echo path triples result in different star graphs.

for triples which connect the nodes o, p, q. Each processed star graph returns a final TDOA graph which hopefully combines all direct path TDOAs from the same source. All triples used in a synthesized TDOA graph σ are summarized in a set $\mathbb{G}_\sigma$ with $|\mathbb{G}_\sigma| \leq \binom{M}{3}$.

One important feature of our synthesis algorithm is that no further consistency check is necessary during the synthesis process. Starting from (approximately) consistent triples, our synthesis algorithm guarantees by construction that each closed path and each subgraph of the resulting TDOA graph are (approximately) consistent.

F. Multiple Sources

After the synthesis of one or several TDOA graphs, we have to initialize the search for a new graph. In order to avoid the synthesis of identical graphs, those TDOA triples which have already been used in existing TDOA graphs are not considered as initial triples. Among the remaining consistent TDOA triples, we again select that with the highest quality as the initial triple and precede as before. The complete synthesis algorithm is terminated, if each triple has been used in a TDOA graph or if the remaining triples cannot be combined even to quadruples. These isolated triples are rejected since a three-dimensional source localization requires at least four microphones.

In practice, the number of TDOA graphs returned by the above described synthesis algorithm is larger than the true number of sources, mainly due to echo path triples. Typically, TDOA graphs corresponding to true source positions are highly connected while erroneous graphs caused by echo path triples

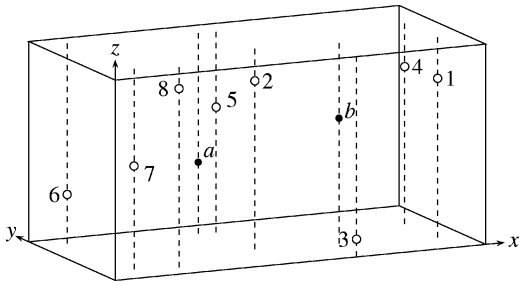

Fig. 14. Position of microphones (∘) and sources (•) in the lab.

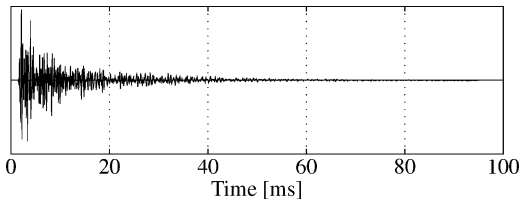

Fig. 15. Room impulse response measured in the lab.

have a small number of nodes and edges. This motivates the introduction of the *connectivity* w for each synthesized graph. It measures the degree of connection of the graph and how good the zero cyclic sum condition is satisfied for each valid triple

$$w = \sum_{(n_{kl,\mu}, n_{lm,\nu}, n_{mk,\eta}) \in \mathbb{G}_\sigma} \Gamma_{\mathrm{TFTM}}(n_{kl,\mu} + n_{lm,\nu} + n_{mk,\eta}). \tag{24}$$

The maximum value of w is $\binom{M}{3} \cdot \max_n \Gamma_{\mathrm{TFTM}}(n)$. The larger w is, the higher the connectivity of the graph. Our experiments show a significant gap in w between correct and erroneous TDOA graphs in most cases. Hence, the number of high-connectivity graphs can be used to estimate the number of sources if it is unknown.

VII. Experimental Results

A. Localization System

We evaluated our proposed algorithms for TDOA estimation in a real-time demonstration system for multiple speaker localization. Our small rectangular acoustic lab is shown in Fig. 14. It has the size $4\,\mathrm{m} \times 2\,\mathrm{m} \times 2\,\mathrm{m}$. The floor and the wooden walls are covered by a thin carpet. The ceiling is an acrylic glass. Fig. 15 shows a measured room impulse response. The reverberation time is $T_{60} \approx 300$ ms. $N = 2$ speech signals are played back from two loudspeakers at the position

$$\mathbf{p}_a = [1.67, 1.66, 0.71]^T, \quad \mathbf{p}_b = [2.72, 0.65, 1.25]^T \tag{25}$$

in meters. Due to fans and illumination, there is a weak background noise. The signal-to-noise ratio (SNR) is roughly 50 dB. $M = 8$ capacitive microphones are randomly placed at the positions

$$\begin{aligned}&(3.71, 0.49, 1.59)^T,\ (1.86, 0.75, 1.68)^T,\ (2.61, 0.01, 0.17)^T,\\ &(3.69, 1.21, 1.57)^T,\ (1.85, 1.63, 1.25)^T,\ (0.30, 1.76, 0.49)^T,\\ &(0.39, 0.40, 1.03)^T,\ (0.82, 0.29, 1.78)^T\end{aligned} \tag{26}$$

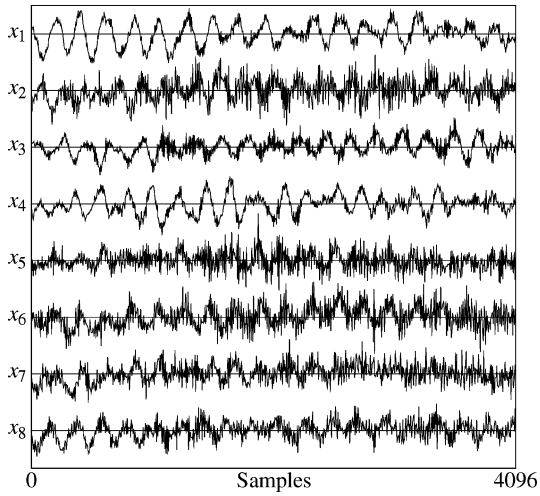

Fig. 16. One block of eight microphone signals.

as shown in Fig. 14 to record the speech signals. The microphone signals are sampled at 96 kHz and processed by a Linux PC (kernel 2.6, dual-core CPU at 2.8 Hz). This high sampling rate corresponds to a fine range quantization of 3.6 mm at the sound speed of 343 m/s.

B. TDOA Estimation of a Single Signal Block

Our experiment is based on a block length of 4096 samples. This corresponds to a speech frame of approximately 43 ms. Fig. 16 shows one block of eight microphone signals. During this time, the source signals contain a fricative [f] of one speaker and a diphthong [au] of the other speaker.

For each signal block, we calculated $M = 8$ autocorrelations and $\binom{M}{2} = 28$ generalized cross-correlations (GCC-PHAT) [4]. For each GCC $r_{kl}(n)$, we selected the 15 strongest maxima and stored their positions in $\mathbb{P}_{kl}$. For each autocorrelation $r_{kk}(n)$, the positions of the four strongest extrema were collected in $\mathbb{P}_{kk}$. Then we applied the raster matching as described in Section V to $\mathbb{P}_{kl}$. This reduces the total number of TDOA estimates for all microphone pairs from $28 \cdot 15 = 420$ to 148. The exact number of desired TDOAs for two sources is $28 \cdot 2 = 56$. The reason for our intentional overestimation is the conservative detection in (15) in order to avoid miss detections.

As one example, we consider the cross-correlation between microphone 1 and 2. The microphones have a distance 1.87 m, corresponding to an upper bound of $n_{\max} = 523$ samples for the TDOA. From the source and microphone positions, we calculated the true direct path TDOAs to 25.8 and 326.7 samples. The cross-correlation $r_{12}(n)$ is shown in Fig. 17 with

$$\begin{aligned}\mathbb{P}_{12} = \{&-81, -31, -4, 21, 48, 109, 162, 188,\\ &267, 327, 337, 347, 358, 438, 448\}.\end{aligned} \tag{27}$$

The autocorrelation extrema are located at

$$\mathbb{P}_{11} = \{10, 21, 28, 142\}, \quad \mathbb{P}_{22} = \{35, 52, 79, 224\}. \tag{28}$$

By applying the raster matching, we found a number of TDOA pairs which match certain autocorrelation extremum

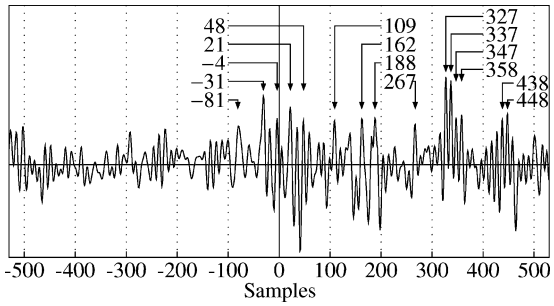

Fig. 17. Cross-correlation between microphone 1 and 2 and its 15 maxima.

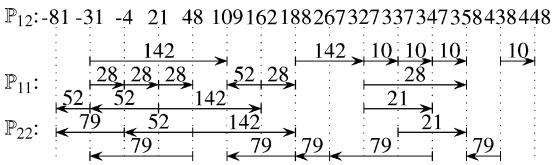

Fig. 18. Pairs of TDOAs whose distances match autocorrelation extremum positions.

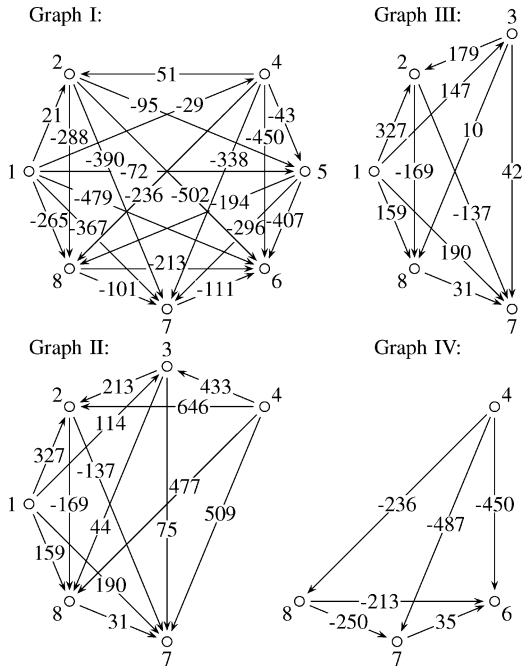

Fig. 19. Synthesized approximately consistent TDOA graphs.

positions from $\mathbb{P}_{11} \cup \mathbb{P}_{22}$. They are depicted in Fig. 18. If we only choose the TDOA without arrowheads, there would be just one valid TDOA at the position 438 which does not correspond to any of the true source positions. By using the quality value $q(n_\mu)$ in (13) with $\Gamma_{\mathrm{TWRM}} = 7$ and the detection rule in (15), the set $\mathbb{P}_{12}$ is reduced to

$$\mathbb{P}'_{12} = \{-31, 21, 48, 267, 327, 337, 438\}. \tag{29}$$

Obviously, eight echo path TDOAs have been rejected, while the true direct path TDOAs 21 and 327 are still contained in $\mathbb{P}'_{12}$.

Now we combine the seven selected TDOA candidates from $\mathbb{P}'_{12}$ with those of the other microphone pairs by synthesizing consistent graphs. We used $\Gamma_{\mathrm{TWTM}} = 9$ while searching for consistent triples. Our synthesis algorithm as proposed in Section VI returned 17 approximately consistent TDOA graphs with the following connectivity values: 18.2, 12.2, 8.9, 3.9, 3.8, 3.7, 3.3, 3.0, 3.0, 3.0, 3.0, 3.0, 3.0, 3.0, 2.9, 2.7, 2.5 as defined in (24). The first four graphs are shown in Fig. 19. We only study graph I to III further. The other graphs contain only four connected nodes like graph IV and will be discarded due to their low connectivity.

Graph I connects seven of the eight microphones. A total number of 21 TDOA estimates fit together to one big approximately consistent graph corresponding to one source position. Graphs II and III are quite similar. They connect six and five microphones, respectively. Interestingly, both graphs share the same TDOA values between microphone 1, 2, 7, and 8. They seem to originate from the same second source. Only the TDOAs involving sensor 3 are different in both graphs. The explanation is that sensor 3 is in one graph the true microphone and in the other graph the mirrored microphone with respect to a wall; see Fig. 10. Which graph is the correct one cannot be answered here. In the next section, we will resolve this ambiguity by using the residual TDOA error after source position estimation.

A similar explanation applies to graph IV. It is a modification of graph I caused by a mirrored microphone. The triple (4,6,8) is identical in both graphs. The sensor 7 in graph IV represents a mirror of the true microphone 7 in graph I.

C. Source Position Estimation and Accuracy

All TDOA estimates in graphs I–III are converted to distances by using the sound speed of 343 m/s. We used the simple spherical interpolation (SI) method [17] to estimate the position of the two sources. Microphone 7 served as the reference sensor for the SI method. Correspondingly, only 6, 5, and 4 TDOA estimates with respect to microphone 7 from graph I to III are used in source localization. The estimated source positions for the three TDOA graphs are

$$\hat{\mathbf{p}}_a = \hat{\mathbf{p}}_{\mathrm{III}} = \begin{bmatrix} 1.671 \\ 1.693 \\ 0.713 \end{bmatrix}, \hat{\mathbf{p}}_b = \hat{\mathbf{p}}_{\mathrm{I}} = \begin{bmatrix} 2.725 \\ 0.667 \\ 1.278 \end{bmatrix}, \hat{\mathbf{p}}_{\mathrm{II}} = \begin{bmatrix} 1.712 \\ -0.688 \\ 1.291 \end{bmatrix}. \tag{30}$$

Clearly, $\hat{\mathbf{p}}_{\mathrm{I}}$ and $\hat{\mathbf{p}}_{\mathrm{III}}$ are good estimates for $\mathbf{p}_b$ and $\mathbf{p}_a$ in (25), although further improvement could be achieved by using the complete TDOA graph. $\hat{\mathbf{p}}_{\mathrm{II}}$ does not correspond to any source position since graph II contains echo path TDOAs caused by a mirrored microphone.

We introduce two accuracy measures for the source localization. The *residual position error*

$$\Delta_{\hat{\mathbf{p}}} = ||\hat{\mathbf{p}} - \mathbf{p}|| \ [\mathrm{m}] \tag{31}$$

TABLE II
COMPARISON OF LOCALIZATION ACCURACIES

	$\Delta_{\hat{\mathbf{p}}}$ [cm]		$\Delta_{\hat{\mathbf{n}}}$ [sample]		
	$\hat{\mathbf{p}}_a$	$\hat{\mathbf{p}}_b$	$\hat{\mathbf{n}}_a$	$\hat{\mathbf{n}}_b$	$\hat{\mathbf{n}}_{\mathrm{II}}$
Our method	4	3	5	28	282
GCC-PHAT	31	48	317	239	—
SRP-PHAT	14	12	—	—	—

is simply the norm of the difference between the true source position vector $\mathbf{p}$ and its estimate $\hat{\mathbf{p}}$. It is, however, only computable in simulations since it requires the knowledge of the true source position. The *residual TDOA error* is defined as

$$\Delta_{\hat{\mathbf{n}}} = \frac{1}{\sqrt{L_{\hat{\mathbf{n}}}}} \|\hat{\mathbf{n}} - \tilde{\mathbf{n}}\| \,[\text{sample}]. \tag{32}$$

$\hat{\mathbf{n}}$ is the TDOA vector estimate from a synthesized TDOA graph. It is the input for computing the source position estimate $\hat{\mathbf{p}}$. Then we calculate the expected TDOA vector $\tilde{\mathbf{n}}$ from the source position $\hat{\mathbf{p}}$ and the known microphone positions. Since, in general, $\hat{\mathbf{n}}$ and $\tilde{\mathbf{n}}$ have a varying (but equal) vector length $L_{\hat{\mathbf{n}}}$ depending on the synthesized TDOA graph, we normalize $\|\hat{\mathbf{n}} - \tilde{\mathbf{n}}\|$ against $\sqrt{L_{\hat{\mathbf{n}}}}$.

Notice that, in contrast to $\Delta_{\hat{\mathbf{p}}}$, $\Delta_{\hat{\mathbf{n}}}$ can be computed for any TDOA vector $\hat{\mathbf{n}}$ in practice since it does not need the true source position. Taking the fact into account that $\tilde{\mathbf{n}}$ is a function of the true microphone positions while $\hat{\mathbf{n}}$ depends on the implicitly available microphone positions in a TDOA graph, the residual TDOA error $\Delta_{\hat{\mathbf{n}}}$ actually compares the *implicit microphone positions* in the TDOA graph with the true ones. If $\hat{\mathbf{n}}$ contains direct path TDOA estimates only, the value of $\Delta_{\hat{\mathbf{n}}}$ is small. If, however, $\hat{\mathbf{n}}$ also contains echo path TDOAs caused by mirrored microphones or TDOAs of different sources, $\Delta_{\hat{\mathbf{n}}}$ will become large.

Table II shows both accuracies of our algorithm for the signal block in Fig. 16. Indeed, $\Delta_{\hat{\mathbf{n}}}$ of graph II is much larger than that of graph I and III as expected because the implicit sensor 3 in graph II is a mirrored microphone. We see that the residual TDOA error $\Delta_{\hat{\mathbf{n}}}$ after source position estimation provides an additional mean to resolve the ambiguity of TDOA graphs. It can also be used to estimate the number of sources if it is unknown.

In order to study the robustness of our method with respect to noise, we added additional white noise to the loudspeaker signals. For the same signal block as before, the estimated source positions vary up to 1 cm for an SNR of up to 20 dB. Up to 15 dB, both sources were still detected. If we further reduce the SNR, the sets of GCC-PHAT maxima often do not contain the true TDOA values, and thus no consistent TDOA graphs can be constructed anymore.

D. Comparison to Other Localization Methods

We also compared our algorithm to other localization techniques. We applied two different approaches to the same signal block. First, we simply selected the two largest maxima of GCC-PHAT for each microphone pair and assigned them *manually* to the two sources, assuming that we know *a priori* the true TDOAs. We applied the same SI method to all seven TDOA estimates with microphone 7 being the reference sensor. The source position estimates are

$$\hat{\mathbf{p}}_{\mathrm{a}}^{\mathrm{GCC}} = [1.88, 1.73, 0.91]^T, \;\; \hat{\mathbf{p}}_{\mathrm{b}}^{\mathrm{GCC}} = [2.29, 0.68, 1.05]^T. \tag{33}$$

We also implemented the SRP-PHAT method in [11]. After a time-consuming full search of the SRP spectrum with a spatial resolution of 2 cm, the two maxima for the same signal block was found at

$$\hat{\mathbf{p}}_{\mathrm{a}}^{\mathrm{SRP}} = [1.66, 1.52, 0.72]^T, \;\; \hat{\mathbf{p}}_{\mathrm{b}}^{\mathrm{SRP}} = [2.68, 0.74, 1.32]^T. \tag{34}$$

Its computation time is about several thousand times of that of our algorithm (assuming efficient C implementations in both cases) even for our small lab. Table II summarizes the residual position and residual TDOA error for these two methods. Clearly, our algorithm outperforms both of them.

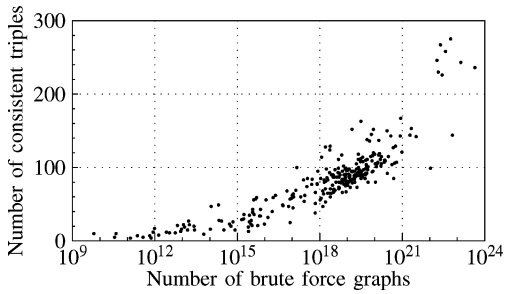

Fig. 20. Number of brute force graphs versus number of consistent triples.

E. Evaluation of Continuous Localization

In a continuous operation of our localization system, the average CPU load is about 40%. For each signal block of length 43 ms, the average computation time of our complete localization algorithm consisting of

- preprocessing like voice activity detection,
- cross- and autocorrelations,
- raster matching,
- synthesis of consistent graphs,
- source position estimation,

is roughly 17 ms. The main computational effort is the calculation of the cross- and autocorrelations. This low complexity is mainly due to the significant reduction of ambiguous TDOAs by raster matching and the synthesis of consistent graphs.

Fig. 20 illustrates the efficiency of the synthesis of consistent graphs. Each point represents one signal block. The abscissa denotes the total number of possible graphs $G_{\mathrm{brute-force}}$ for this block according to (17) if we perform a brute force search. The ordinate shows the number of consistent triples for this block which are used in the synthesis of consistent graphs. For the particular signal block we studied in this section, the total number of 148 TDOA estimates after the raster matching would lead to $G_{\mathrm{brute-force}} \approx 2 \cdot 10^{21}$ different brute force graphs. Our approach finds only 153 consistent TDOA triples.

In order to simplify the evaluation of the accuracy of a continuous localization, we replaced the speech source signals by two white noise signals. This ensures a constant number of simultaneously active sources for each signal block and there is no

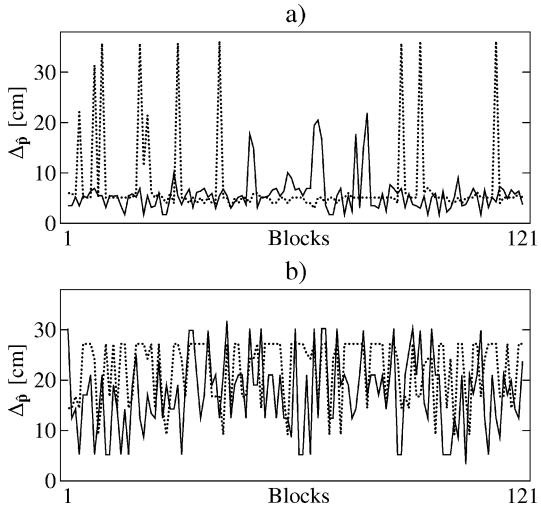

Fig. 21. Residual position errors Δ_ρ in cm for source a (solid) and source b (dotted) in a continuous localization. (a) Our method. (b) SRP-PHAT.

need to estimate the number of sources. The experimental setup and the algorithms are exactly the same as before. Fig. 21 shows the residual position error of both sources for a total number of 121 blocks. Apart from a few outliers, the localization accuracy of our method in (a) is around 5 cm. This is pretty good since the loudspeakers have a membrane diameter of roughly 5 cm. For the SRP-PHAT method, we only performed a local search for each source within a cube of edge length 60 cm around the true source position to reduce the computation time. The spatial resolution of the SRP-PHAT search is again 2 cm. The localization accuracy is shown in (b). GCC-PHAT is not evaluated in this continuous localization due to its difficulty of assigning the cross-correlation maxima to different sources.

VIII. Conclusion

In this paper, we have presented a novel approach for TDOA disambiguation. By using additional extremum positions of autocorrelations of the microphone signals, we have developed a raster matching algorithm to identify and reject wrong TDOA estimates caused by the echo paths. Based on the zero cyclic sum condition of TDOAs originating from the same source, we have formulated the TDOA disambiguation problem in the framework of consistent graphs. We have developed an efficient synthesis algorithm of TDOA graphs based on consistent triples. We also introduced different levels of quality for TDOA estimate, TDOA triple, TDOA graph, and residual TDOA error. Finally, the efficiency and the real-time capability of our algorithms are demonstrated in a real experiment. We believe that our algorithms can also be combined with other localization techniques like the impulse response estimation and SRP search.

References

[1] M. S. Brandstein and H. F. Silverman, "A practical methodology for speech source localization with microphone arrays," *Comput. Speech, Lang.*, vol. 11, pp. 91–126, 1997.

[2] W. L. Kellermann, "Acoustic echo cancellation for beamforming microphone arrays," in *Microphone Arrays*, M. Brandstein and D. Ward, Eds. New York: Springer-Verlag, 2001, ch. 13.

[3] W. R. Hahn and S. A. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 5, pp. 608–614, Sep. 1973.

[4] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal, Process.*, vol. ASSP-24, no. 4, pp. 320–327, Aug. 1976.

[5] J. Benesty, "Adaptive eigenvalue decomposition algorithms for passive acoustic source localization," *J. Acoust. Soc. Amer.*, vol. 107, pp. 384–391, 2000.

[6] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to seperation and dereverberation of speech signals in a reverberant environment," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 882–895, Sep. 2005.

[7] Y. Huang, J. Benesty, and J. Chen, "Speech acquisition and enhancement in a reverberant, cocktail-party-like environment," in *Proc. IEEE ICASSP*, 2006, pp. V-25–V-28.

[8] R. Nickel and A. Iyer, "A novel approach to automated source separation in multispeaker environments," in *Proc. IEEE ICASSP*, 2006, pp. V-629–V-632.

[9] H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation for convolutive mixtures: A unified treatment," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Y. Huang and J. Benesty, Eds. Norwell, MA: Kluwer, 2004.

[10] R. Aichner, H. Buchner, S. Wehr, and W. Kellermann, "Robustness of acoustic multiple-source localization in adverse environments," in *Proc. 7. ITG-Fachtagung Sprach-Kommunikation*, 2006, CD-ROM.

[11] J. H. DiBiase, H. F. Silverman, and M. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays*, M. Brandstein and D. Ward, Eds. New York: Springer-Verlag, 2001, ch. 8.

[12] J. Scheuing and B. Yang, "Disambiguation of TDOA estimates in multi-path multi-source environments (DATEMM)," in *Proc. IEEE ICASSP*, 2006, vol. 4, pp. 837–840.

[13] J. Scheuing and B. Yang, "Efficient synthesis of approximately consistent graphs for acoustic multi-source localization," in *Proc. IEEE ICASSP*, 2007, vol. 4, pp. 501–504.

[14] R. O. Schmidt, "A new approach to geometry of range difference location," *IEEE Trans. Aerospace Electron. Syst.*, vol. AES-8, no. 6, pp. 821–835, Nov. 1972.

[15] Y. Lin, D. D. Lee, and L. K. Saul, "Nonnegative deconvolution for time of arrival estimation," in *Proc. IEEE ICASSP*, 2004, vol. II, pp. 377–380.

[16] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–949, 1979.

[17] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 12, pp. 1661–1669, Dec. 1987.

Jan Scheuing received the Dipl.-Ing. and Ph.D. degrees in electrical engineering from the Universität Stuttgart, Stuttgart, Germany, in 2001 and 2007, respectively.

In 2002, he joined the research team at Chair of System Theory and Signal Processing, Universität Stuttgart. Since 2007, he has been a System Engineer at Bosch Engineering GmbH, Abstatt, Germany. His research interests include time delay estimation and speaker localization.

Bin Yang (SM'06) received the Dipl.-Ing. and Ph.D. degrees in electrical engineering from the Ruhr University Bochum, Bochum, Germany, in 1986 and 1991, respectively.

From 1996 to 2001, he was a Senior Researcher on mobile communications at Infineon Technologies, Germany. Since 2001, he has been a Professor and head of the Chair of System Theory and Signal Processing of Universität Stuttgart, Stuttgart, Germany. His research interests include signal processing for localization, emotion recognition, array processing, automotive safety systems, blind methods, etc.

Dr. Yang is a member of EURASIP and VDE.