AUTO-LOCALIZATION IN AD-HOC MICROPHONE ARRAYS

Nikolay D. Gaubitch[1], *W. Bastiaan Kleijn*[1,2], *Richard Heusdens*[1]

[1] Signal and Information Processing Lab., Delft University of Technology, Netherlands
[2] School of Engineering and Computer Science, Victoria University of Wellington, New Zealand

ABSTRACT

We present a method for automatic microphone localization in ad-hoc microphone arrays. The localization is based on time-of-arrival (TOA) measurements obtained from spatially distributed acoustic events. In practice, measured TOAs are an incomplete representation of the true TOAs due to unknown onset times of the acoustic events and internal delays in the capturing devices and make the localization problem insoluble if not addressed appropriately. The main contribution of the proposed method is an algorithm that identifies and corrects for such internal delays and acoustic event onset times in the measured TOAs. Experimental results using both simulated and real-world data demonstrate the performance of the method and highlight the significance of correct estimation of the internal delays and onset times.

Index Terms— microphone array, auto-localization

1. INTRODUCTION

Microphone arrays facilitate the use of both spatial and temporal information and can provide superior results in speech enhancement compared to a single microphone [1]. Many established methods for multi-microphone audio processing such as beamforming, require knowledge of the relative positions of the microphones. This requirement is easily satisfied in conventional microphone arrays where the microphones are positioned in a fixed configuration. On the other hand, ad-hoc microphone arrays have an unknown configuration that changes every time the microphones are deployed and a method to automatically localize the microphones is necessary.

Auto-localization methods [2, 3] typically use acoustic events emitted from varying spatial locations and measure the time-of-arrivals (TOAs), which are used to localize the microphones. However, practical TOA measurements are incomplete because they include an unknown source onset time – the time the acoustic event was generated, and an unknown internal delay – the time taken from the sound reaching the microphone to that it is registered as received by the capturing device. Some alternatives to TOA have been explored such as, the time-difference-of-arrival (TDOA) [4, 5, 6], the diffuse noise field coherence [7] and received signal power [8, 9].

One of the earliest methods for auto-localization is the multi-dimensional scaling (MDS) [10, 3, 11], which provides the relative configuration of sensors given the distances between all sensors. This implicitly assumes that sensors and sources are co-located, which is of limited applicability for microphone arrays. Non-linear least-squares (LS) formulations for simultaneous estimation of the sources and the sensors have been proposed by several authors [4, 5, 12, 13, 14]. The solution is typically found through gradient descent optimization and suffers from the common problem of local minima, if not carefully initialized. The initialization is addressed in [4, 5] where it is assumed that sources and microphones are closely spaced – approximately co-located. Then, MDS is used to find approximate locations of the sources and the microphones, which are used to initialize the non-linear LS optimization. This yields improved results but requires at least seven devices with closely spaced microphones and loudspeakers. An alternative approach was formulated in [13]. It assumes sources in the far-field leading to a least squares solution that is accurate up to an arbitrary matrix, which is estimated through non-linear optimization. While some of the above methods do address the onset times [4, 5, 12, 13, 15] and/or the internal delays [4, 5, 15], they all impose geometrical constraints on the placement of the microphones and the sound sources.

Crocco *et al.* [16, 17] generalized the method in [13] such that it does not rely on the far-field assumption and showed that if one source and one sensor are co-located, it can lead to an exact closed-form solution. This method is attractive for localization in ad-hoc microphone arrays but it does not consider acoustic event onset times and internal delays – these are assumed to be known. In practice it may be difficult or even impossible to measure these and they must be estimated from the observed data.

In this paper, we begin by formulating the problem of microphone localization in Section 2. Then, building on the work of [16, 17], we propose an algorithm in Section 3 that estimates and corrects for internal delays and acoustic event onset times, and localizes the microphones. In Section 4 we present results using both simulated and real-world experiments and conclusions are drawn in Section 5.

2. PROBLEM FORMULATION

Consider $\mathcal{I}$ microphones and $\mathcal{J}$ acoustic events distributed in a 3-dimensional Euclidean space. We specify the ith microphone and the jth source locations by the Cartesian coordinates $\mathbf{r}_i = [r_{x,i}\, r_{y,i}\, r_{z,i}]^T$ and $\mathbf{s}_j = [s_{x,j}\, s_{y,j}\, s_{z,j}]^T$, respectively; the coordinates of all microphones and sound sources are represented by the $\mathcal{I} \times 3$ matrix $\mathbf{R} = [\mathbf{r}_1\ \mathbf{r}_2\ \dots\ \mathbf{r}_\mathcal{I}]^T$ and the $\mathcal{J} \times 3$ matrix $\mathbf{S} = [\mathbf{s}_1\ \mathbf{s}_2\ \dots\ \mathbf{s}_\mathcal{J}]^T$.

The localization is performed using measured TOAs. The measurements arise from acoustic events such as hand claps, from unknown locations $\mathbf{s}_j$ and onset times τ_j captured by microphones with unknown locations $\mathbf{r}_i$ and internal delays δ_i. The measured TOA of acoustic event $\mathbf{s}_j$ at microphone $\mathbf{r}_i$ is given by

$$t_{ij} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|}{c} + \tau_j + \delta_i + \nu_{ij}, \tag{1}$$

where c is the speed of sound, $\|\cdot\|$ denotes Euclidean norm and ν_{ij} is measurement noise.

This work was supported by Google Inc.

The objective of auto-localization is to identify the internal delays δ_i and the onset times τ_j, and to find an estimate, $\widehat{\mathbf{R}}$, of the microphone locations $\mathbf{R}$, using only the measured TOAs, t_{ij}.

3. MICROPHONE LOCALIZATION

A straightforward solution to the localization can be obtained by finding the minimum of the non-linear LS problem [12, 4, 5, 13, 14]

$$\widehat{\mathbf{R}}, \widehat{\mathbf{S}} = \arg\min_{\mathbf{R},\mathbf{S}} \sum_{i=1}^{\mathcal{I}} \sum_{j=1}^{\mathcal{J}} \left(\frac{\|\mathbf{r}_i - \mathbf{s}_j\|}{c} - t_{ij} \right)^2, \tag{2}$$

which is the maximum likelihood solution if the measurement noise, ν_{ij}, is assumed Gaussian. The solution to (2) is generally obtained using gradient descent optimization. This approach does not guarantee a unique solution; it generally finds a local rather than the global minimum if not initialized with care. It also assumes that the internal delays and onset times are known. In the remainder of this section we describe a two-stage method. In the first stage, the internal delays and acoustic event onset times are identified. These are used to estimate the correct TOAs, which are applied for the microphone localization in the second stage.

3.1. Internal delay and onset time estimation

We assume, without loss of generality, that $c = 1$ and we set the time reference for the acoustic events to $\tau_1 = 0$; we also assume absence of observation noise, $\nu_{ij} = 0$. Expanding the equation of observed TOAs in (1) we obtain

$$\mathbf{r}_i^T\mathbf{r}_i + \mathbf{s}_j^T\mathbf{s}_j - 2\mathbf{r}_i^T\mathbf{s}_j = t_{ij}^2 + \tau_j^2 + \delta_i^2 - 2(t_{ij}\tau_j + t_{ij}\delta_i - \delta_i\tau_j) \tag{3}$$

Next, we subtract the corresponding equation for $i = 1$ from the general form of (3), which results in

$$\begin{aligned}\mathbf{r}_i^T\mathbf{r}_i - \mathbf{r}_1^T\mathbf{r}_1 - 2(\mathbf{r}_i - \mathbf{r}_1)^T\mathbf{s}_j = {} & t_{ij}^2 - t_{1j}^2 + \delta_i^2 - \delta_1^2 \\ & - 2t_{ij}(\delta_i + \tau_j) + 2t_{1j}(\delta_1 + \tau_j) \\ & + 2\tau_j(\delta_i - \delta_1),\end{aligned} \tag{4}$$

and then we subtract the equation for $j = 1$ from (4)

$$\begin{aligned}-2(\mathbf{r}_i - \mathbf{r}_1)^T(\mathbf{s}_i - \mathbf{s}_1) = {} & t_{ij}^2 - t_{1j}^2 - t_{i1}^2 + t_{11}^2 \\ & - 2\delta_i(t_{ij} - t_{i1}) + 2\delta_1(t_{1j} - t_{11}) \\ & - 2\tau_j(t_{ij} - t_{1j}) + 2\tau_j(\delta_i - \delta_1).\end{aligned} \tag{5}$$

Let $\mathcal{C}\{\mathbf{X}\}$ be an operator that transforms a matrix into a column vector and $\mathcal{C}_{M\times N}^{-1}\{\mathbf{x}\}$ the corresponding inverse operator that transforms a vector $\mathbf{x}$ into a $M \times N$ matrix $\mathbf{X}$. Then we can write (5) in a matrix form as

$$-2\bar{\mathbf{R}}\bar{\mathbf{S}}^T = \mathbf{T} + \mathbf{A}(\mathbf{p}) + \mathbf{\Gamma}, \tag{6}$$

where $\mathbf{T}_{ij} = t_{ij}^2 - t_{1j}^2 - t_{i1}^2 + t_{11}^2$ for $i = 2, \ldots, \mathcal{I}$, $j = 2, \ldots, \mathcal{J}$, $\mathbf{A}(\mathbf{p}) = \mathcal{C}_{\mathcal{I}-1\times\mathcal{J}-1}^{-1}\{\mathbf{W}\mathbf{p}\}$, $\mathbf{\Gamma}_{ij} = 2(\delta_i - \delta_1)\tau_j$, $\mathbf{p} = [\delta_1\ \delta_2\ \ldots\ \delta_{\mathcal{I}}\ \tau_2\ \tau_3\ \ldots\ \tau_{\mathcal{J}}]^T$ and $\mathbf{W}$ is a matrix composed of terms $(t_{ij} - t_{i1})$, $(t_{1j} - t_{11})$ and $(t_{ij} - t_{1j})$; $\bar{\mathbf{R}}$ is the $(\mathcal{I}-1)\times 3$ location matrix of the microphones relative to the first microphone and $\bar{\mathbf{S}}$ is the $(\mathcal{J}-1)\times 3$ location matrix of the acoustic events relative to the first event. It can be seen that $\mathbf{A}(\mathbf{p})$ and $\mathbf{\Gamma}$ act as correction matrices to $\mathbf{T}$ that compensate for the internal delays and onset times; if these are assumed known, then the formulation in (6) is equivalent to that presented in [16] and a solution can be found accordingly. However, in most practical scenarios this will not be the case, and the unknown onset times and internal delays must be estimated.

An important observation can be made in (6): $\bar{\mathbf{R}}\bar{\mathbf{S}}^T$ is a matrix of rank 3 provided that we have three or more microphones and acoustic events, which must be the case as will be shown in Section 3.3. If $\mathbf{A}(\mathbf{p})$ and $\mathbf{\Gamma}$ are not considered, as in [16], the relationship in (6) does not hold and it is not possible to localize the acoustic events or the microphones. However, this insight can be used to devise an algorithm, which finds an estimate $\widehat{\mathbf{p}}$ of the unknown onset times and internal delays such that $\widehat{\mathbf{T}} = \mathbf{T} + \mathbf{A}(\widehat{\mathbf{p}}) + \widehat{\mathbf{\Gamma}}$ is rank 3.

Evidently, this is a rank-reduction problem and the estimation of $\mathbf{p}$ can be based on the Eckart-Young-Mirsky low-rank approximation theorem [18]: the best rank-r approximation, $\widetilde{\mathbf{X}}$, of a matrix $\mathbf{X}$ such that the Frobenius norm $\|\mathbf{X} - \widetilde{\mathbf{X}}\|_F$ is minimised is given by

$$\widetilde{\mathbf{X}} = \mathbf{U}\widetilde{\mathbf{\Sigma}}_r\mathbf{V}^T \tag{7}$$

where $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is the singular value decomposition (SVD) of $\mathbf{X}$ and $\widetilde{\mathbf{\Sigma}}_r = \mathrm{diag}(\sigma_1\ \ldots\ \sigma_r\ 0\ \ldots\ 0)$ is a diagonal matrix with the r largest singular values of $\mathbf{\Sigma}$; the rest are set to zero. Low-rank approximation has been applied to the solution of several problems in signal processing [19, 20]. We can use this result to estimate $\mathbf{p}$ iteratively by minimizing the following cost function at each iteration

$$\widehat{\mathbf{p}}^{(n+1)} = \arg\min_{\widehat{\mathbf{p}}^{(n)}} \|\mathbf{E}^{(n)} - (\mathbf{A}(\widehat{\mathbf{p}}^{(n)}) + \widehat{\mathbf{\Gamma}})\|_F^2 + \lambda\|\widehat{\mathbf{T}}^{(n)}\|_F^2, \tag{8}$$

with

$$\mathbf{E}^{(n)} = \widetilde{\mathbf{T}}_3^{(n)} - \mathbf{T}, \tag{9}$$

where $\widehat{\mathbf{T}}^{(n)} = \mathbf{T} + \mathbf{A}(\widehat{\mathbf{p}}^{(n)}) + \widehat{\mathbf{\Gamma}}^{(n)}$ and $\widetilde{\mathbf{T}}_3^{(n)}$ is the best rank-3 approximation of $\widehat{\mathbf{T}}^{(n)}$ obtained from (7).

Consider first the case when $\lambda = 0$ and the internal delays are all equal; for equal internal delays $\mathbf{\Gamma} = \mathbf{0}$. This leads to a least squares solution of (8)

$$\widehat{\mathbf{p}}^{(n+1)} = \mathbf{W}^+\mathbf{e}^{(n)}, \tag{10}$$

where $\mathbf{W}^+$ is the pseudo-inverse of $\mathbf{W}$ and $\mathbf{e}^{(n)} = \mathcal{C}\{\mathbf{E}^{(n)}\}$. It follows from the optimization procedure that $\|\mathbf{E}^{(n)} - \mathbf{A}(\widehat{\mathbf{p}}^{(n+1)})\|_F \le \|\mathbf{E}^{(n)} - \mathbf{A}(\widehat{\mathbf{p}}^{(n)})\|_F$ and therefore, $\|\widehat{\mathbf{T}}_3^{(n)} - \widehat{\mathbf{T}}^{(n+1)}\|_F \le \|\widehat{\mathbf{T}}_3^{(n)} - \widehat{\mathbf{T}}^{(n)}\|_F$. By the Eckart-Young-Mirsky theorem we must have that $\|\widehat{\mathbf{T}}_3^{(n+1)} - \widehat{\mathbf{T}}^{(n+1)}\|_F \le \|\widehat{\mathbf{T}}_3^{(n)} - \widehat{\mathbf{T}}^{(n+1)}\|_F$ and therefore, the algorithm converges. The algorithm is summarized in Algorithm 1.

Although, the algorithm is shown here to converge, the convergence rate can be very slow in practice. Therefore, the additional constraint to minimize $\|\widehat{\mathbf{T}}^{(n)}\|_F$ is introduced in order to force the solution to be within reasonable dimensions. This increases the initial convergence rate but λ needs to be set to zero according to some criterion to allow the algorithm to converge fully. We monitor $\|\widehat{\mathbf{T}}^{(n)}\|_F$ and set $\lambda = 0$ when the change from one iteration to the next is below a threshold.

In the general case when the internal delays are different, the solution to (8) has to be found through non-linear LS optimization.

3.2. Localizing the microphones

Having identified the internal delays and the acoustic event onset times, we can now use the corrected matrix of relative TOAs, $\widehat{\mathbf{T}}$, to estimate the locations of the sources and the microphones as in [16]. First, we have to convert $\widehat{\mathbf{T}}$ to a squared distance matrix multiplying it by c^2 such that $\mathbf{D} = c^2\widehat{\mathbf{T}}$. Here the correct speed of sound in air

Algorithm 1 Internal delay and onset time estimation

Initialize: $\widehat{\mathbf{p}}^{(0)}$

Computation: $n = 0, 1, 2 \ldots$

1. Compute: $\widehat{\mathbf{T}}^{(n)} = \mathbf{T} + \mathbf{A}(\widehat{\mathbf{p}}^{(n)})$
2. Rank-3 approximation of $\widehat{\mathbf{T}}^{(n)}$: $\widetilde{\mathbf{T}}_3^{(n)} = \mathbf{U}\widetilde{\mathbf{\Sigma}}_3\mathbf{V}^T$
3. Compute the error: $\mathbf{E}^{(n)} = \widetilde{\mathbf{T}}_3^{(n)} - \mathbf{T}$
4. Update estimate of $\mathbf{p}$ from (8): $\widehat{\mathbf{p}}^{(n+1)} = \arg\min_{\widehat{\mathbf{p}}^{(n)}} \|\mathbf{E}^{(n)} - \mathbf{A}(\widehat{\mathbf{p}}^{(n)})\|_F^2$
5. Update estimate of $\mathbf{A}(\mathbf{p})$: $\mathbf{A}(\widehat{\mathbf{p}}^{(n+1)}) = \mathcal{C}^{-1}_{\mathcal{I}-1\times\mathcal{J}-1}\left\{\mathbf{W}\widehat{\mathbf{p}}^{(n+1)}\right\}$
6. Compute Frobenius norm $F^{(n)} = \|\mathbf{E}^{(n)} - \mathbf{A}(\widehat{\mathbf{p}}^{(n+1)})\|_F$
7. Exit if converged

is required and we assume $c = 343$ m/s in room temperature. We can now write (6) as

$$-2\bar{\mathbf{R}}\bar{\mathbf{S}}^T = \mathbf{D}, \tag{11}$$

and by taking the SVD of $\mathbf{D}$, we have

$$\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \tag{12}$$

Since $\bar{\mathbf{R}}\bar{\mathbf{S}}^T$ is a matrix of rank 3, as discussed in Section 3.1, only the three largest singular values in $\mathbf{\Sigma}$ need to be considered and the remaining should ideally be zero or close to zero. Consequently, $\mathbf{U}$, $\mathbf{\Sigma}$ and $\mathbf{V}$ can be truncated and using (11) and (12) we can write

$$\bar{\mathbf{R}} = \mathbf{U}'\mathbf{C} \tag{13}$$

$$-2\bar{\mathbf{S}}^T = \mathbf{C}^{-1}\mathbf{\Sigma}'\mathbf{V}'^T, \tag{14}$$

where $\mathbf{C}$ is an arbitrary invertible 3×3 matrix and $\mathbf{X}'$ denotes a truncated version of $\mathbf{X}$. The minimum solution for the localization is invariant to rotations or translations of the source-microphone configuration. Therefore, we can constrain the coordinates of the first microphone to be at the origin, $\mathbf{r}_1 = [0\ 0\ 0]^T$, and the coordinates of the first source to be on the x-axis, $\mathbf{s}_1 = [s_{x,1}\ 0\ 0]^T$. An estimate of the matrix $\mathbf{C}$ and of $s_{x,1}$ can then be found based on (4) with the corrected TOAs, $\hat{t}_{ij}$, according to [16]

$$\mathbf{r}_i^T\mathbf{r}_i - \mathbf{r}_1^T\mathbf{r}_1 - 2(\mathbf{r}_i - \mathbf{r}_1)^T\mathbf{s}_j = d_{ij}^2 - d_{1j}^2, \tag{15}$$

where $d_{ij} = c\hat{t}_{ij}$. The non-linear LS optimization criterion is formed by substituting the known values from (11) and (12) into (15)

$$\widehat{\mathbf{C}} = \arg\max_{\mathbf{C}} \sum_{i=1}^{\mathcal{I}-1}\sum_{j=1}^{\mathcal{J}-1} \left\{([\mathbf{U}'\mathbf{C}]_{i1}^2 + [\mathbf{U}'\mathbf{C}]_{i2}^2 + [\mathbf{U}'\mathbf{C}]_{i3}^2)^2 + [\mathbf{U}'\mathbf{\Sigma}'\mathbf{V}'^T]_{ij} - 2[\mathbf{U}'\mathbf{C}]_{i1}s_{x,1} - d_{i+1j+1}^2 + d_{1j+1}^2\right\}^2. \tag{16}$$

Finally, we find the estimate of the the microphone positions by substituting $\widehat{\mathbf{C}}$ into (13). This method results in the estimation of only ten parameters, independent of the number of acoustic events and microphones.

3.3. Identifiability conditions

From (1) we see that every microphone and every acoustic event introduces four unknowns (three for the spatial coordinates and one for the acoustic event onset or internal delay) and there are $\mathcal{I}\mathcal{J}$ equations. The localization is invariant to rotations or translations of the source-microphone configuration. Therefore, we can constrain to the origin the coordinates of the first item (source or microphone), two of the coordinates of a second item and one coordinate of a third item. We have also set the time origin of the acoustic events to $\tau_1 = 0$. Thus, we have that $\mathcal{I}\mathcal{J} \geq 4\mathcal{I} + 4\mathcal{J} - 7$, which leads to

$$\mathcal{J} \geq \frac{4\mathcal{I} - 7}{\mathcal{I} - 4}. \tag{17}$$

In other words, we require a minimum of five microphones and 13 sound source events. Note that different acoustic events means that they have to be in different spatial locations, the actual signal can be the same.

4. EXPERIMENTS AND RESULTS

In this section, we present experimental results with the objective to highlight two particular features of the localization method described in Section 3. First, we demonstrate the properties of the algorithm for identification of the internal delays and the acoustic event onset times described in Section 3.1. Then, we show results for the localization algorithm in Section 3.2 with and without the compensation for delays and onset times in order to emphasize its importance. Finally, we will show an example using real-world data.

The accuracy of the estimated delays and onset times was measured in terms of the root mean squared error (RMSE) between the true and the estimated values and the accuracy of the estimated microphone locations was evaluated with the RMSE between the true and the estimated locations. Since the estimated microphone locations are a rotated and translated version of the true locations, they were first aligned to the actual locations using the Procrustes optimal alignment [3] – a common practice in the evaluation of auto-localization methods.

We simulated an open cubic space with a side of 5 m. Inside this space $\mathcal{J} = 20$ acoustic events and a varying number of microphones, $\mathcal{I} = \{5, 10, 15, 20, 25, 30\}$, were distributed at randomly chosen locations. For each $\mathcal{I}$ between five and 30, we simulated source and microphone location coordinates using 100 different realizations drawn from a uniform distribution. The internal delays were assumed to be the same for all microphones and were generated randomly for each scenario according to $\delta_i = \alpha, \forall i$, where α is a uniformly distributed random variable; this results in an internal delay between zero and one second. The acoustic event onsets were generated to occur approximately every three seconds according to $\tau_j = 3j + \beta, \quad j = 1, 2, \ldots, \mathcal{J}$, where $\beta \sim \mathcal{N}(0, 1)$ is a zero-mean unit-variance normally distributed random variable. At each iteration, we identified the internal delays and the acoustic event onset times following the algorithm in Section 3.1 and the microphones following the algorithm in Section 3.2. The algorithm parameters $\widehat{\mathbf{p}}$ and $\widehat{\mathbf{C}}$ were initialized to non-informative random values and $\lambda = 0.7$. Algorithm 1 was considered to have converged when the change in Frobenius norm from one iteration to the next was less than $\epsilon = 10^{-12}$, $|F^{(n-1)} - F^{(n)}| < \epsilon$.

Figure 1a shows the resulting RMSE for the identification of the internal delays and onset times and Fig. 1b shows the Frobenius norm $F^{(n)}$ at convergence. It can be seen that the algorithm converges in terms of the Frobenius norm and that the estimated

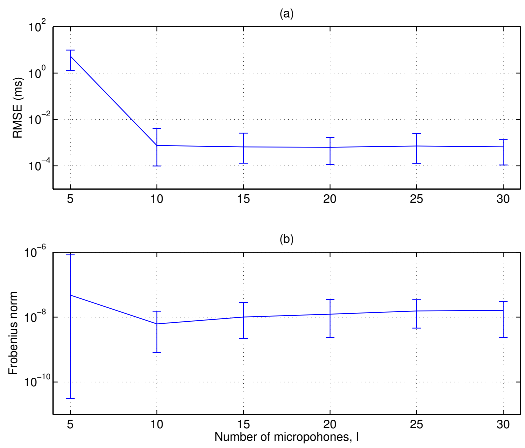

Fig. 1. Resulting (a) RMSE for $\mathbf{p}$ and (b) the Frobenius norm minimized in (8) vs. number of microphones. The results are averaged over 100 different source-microphone locations; error bars indicate maximum and minimum values.

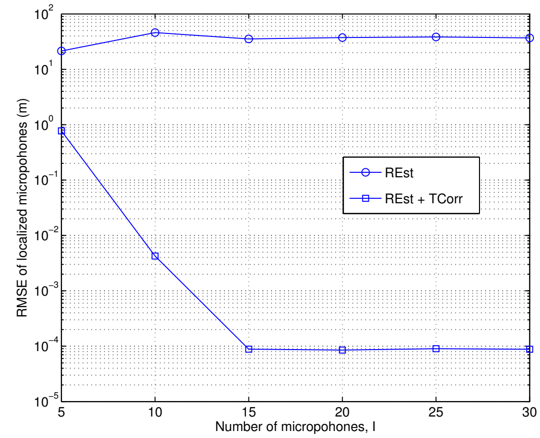

Fig. 2. RMSE of the localized microphones with correction for the internal delays and acoustic event onset times (REst+TCorr) and using the measured TOAs (REst) directly. The results are an average over 100 different source-microphone configurations.

times are within 5 ms for five microphones and down to 10^{-3} ms as the number of microphones increases to 30. The corresponding results for the microphone localization with corrected TOA matrix (squares) are given in Fig. 2. The figure also shows a result for localization without correction (circles); in this case the measured TOAs are used directly with the onset time taken as the beginning of the recording. It is not surprising that this results in wrong estimates but it emphasizes the importance of correct estimation of the onset times and internal delays. Without these, the localization algorithm is incomplete. Based on the current experiments, the two-stage method is able to localize the microphones to an accuracy of down to 10^{-4} m with increasing number of microphones. Note that keeping the number of microphones fixed but increasing the number of acoustic events results in similar gradual improvement in estimation accuracy.

Finally, we present an example of using this algorithm with real measured data. We used eight AKG C417 lapel microphones positioned randomly on a table with dimensions 0.75×1.5 m. The microphones were connected to an RME Fireface 800 through an RME Octamic II microphone preamplifier. A recording was made of 20 handclaps produced at different locations in the room.This was used in the localization procedure; the TOAs of the hand-claps were labelled manually in the recorded data as the largest peaks in each hand-clap. Figure. 3 shows the table (dashed line), the true microphone locations (crosses) and the estimated microphone locations (circles). The RMSE of this result is 0.02 m, which is in good agreement with the simulated results.

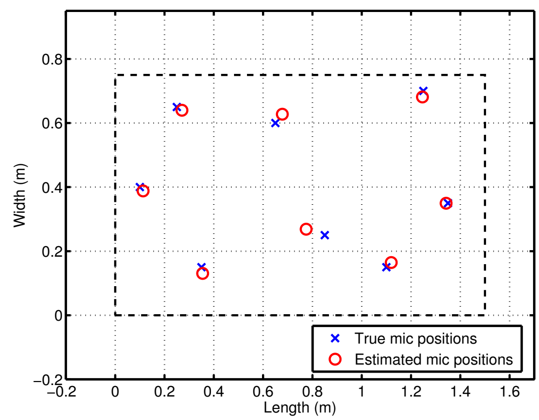

Fig. 3. Example with measured data; TOAs extracted from recorded handclaps with microphones (crosses) randomly positioned on a table (dashed line). Estimated microphone locations (circles) are accurate to within 0.02 m.

5. CONCLUSIONS

We proposed a method for auto-localization in ad-hoc microphone arrays that operates on time-of-arrival measurements from spatially distributed acoustic events. The proposed method consists of a localization algorithm and an algorithm to estimate the internal delays of the devices and the onset times of the acoustic events. We showed through experiments with simulated and real measured data that auto-localization can be achieved in practice. However, it is important to account for the internal delays of the capturing devices and the onset times of the acoustic events; ignoring these parameters results in an incomplete localization procedure.

6. REFERENCES

[1] M. S. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, Springer-Verlag, Berlin, Germany, 2001.

[2] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero III, R. L. Moses, and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54 – 69, July 2005.

[3] J. N. Ash and R. L. Moses, *Self-localization of sensor networks*, chapter 13, pp. 409 – 437, In Haykin and Liu [21], 2009.

[4] V. C. Raykar, I. Kozintsev, and R. Lienhart, "Self localization of acoustic sensors and actuators on distributed platforms," in *Proc. Intl. Workshop on Multimedia Technologies in ELearning and Collaboration*, 2003.

[5] V. C. Raykar, I. V. Kozintsev, and R. Lienhart, "Position calibration of microphones and loudspeakers in distributed computing platforms," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 70 – 83, Jan. 2005.

[6] N. Ono, H. Kohno, N. Ito, and S. Sagayama, "Blind alignment of asynchronously recorded signals for distributed microphone array," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 2009, pp. 161–164.

[7] I. McCowan, M. Lincoln, and I. Himawan, "Microphone array shape calibration in diffuse noise fields," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 3, pp. 666 – 670, Mar. 2008.

[8] Z. Liu, Z. Zhang, L.-W. He, and P. Chou, "Energy-based sound source localization and gain normalization for ad hoc microphone arrays," in *Proc. IEEE Intl. Conf. on Acoust., Speech, Signal Process. (ICASSP)*, 2007, pp. II–761–II–764.

[9] M. Chen, Z. Liu, L.-W. He, P. Chou, and Z. Zhang, "Energy-based position estimation of microphones and speakers for ad hoc microphone arrays," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2007.

[10] W. S. Torgerson, "Multidimensional scaling: I. theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, Dec. 1952.

[11] S. T. Birchfield and A. Subramanaya, "Microphone array position calibration by basis-point classical multidimensional scaling," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 1025 – 1034, Sept. 2005.

[12] R. Biswas and S. Thrun, "A passive approach to sensor network localization," in *Proc. IEEE/RSJ Intl. Conf. Intelligent Robots and Systems*, Sept. 2004, vol. 2, pp. 1544 – 1549.

[13] S. Thrun, "Affine structure from sound," in *Advances in Neural Information Processing Systems 18*, Y. Weiss, B. Schölkopf, and J. Platt, Eds., pp. 1353–1360. MIT Press, Cambridge, MA, 2006.

[14] P. D. Jager and A. Hashemi-Sakhatsari, "Automatic microphone array position calibration using an acoustic sounding source," in *Proc. IEEE Conf. Industrial Electronics and Applications (ICIEA)*, May 2009, pp. 2110 – 2113.

[15] M. H. Hennecke and G. A. Fink, "Towards acoustic self-localization of ad hoc smartphone arrays," in *Proc. Hands-free Speech Comm. and Microphone Arrays*, May 2011, pp. 127 –131.

[16] M. Crocco, A. Del Bue, and V. Murino, "A blinear approach to the position self-calibration of multiple sensors," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 660 – 673, Feb. 2012.

[17] M. Crocco, A. Del Bue, M. Bustreo, and V. Murino, "A closed form solution to the microphone position self-calibration problem," in *Proc. IEEE Intl. Conf. on Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2012, pp. 2597 – 2600.

[18] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, pp. 211–218, 1936.

[19] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica*, vol. 44, pp. 891–909, 2008.

[20] I. Markovsky, *Low Rank Approximation: Algorithms, Implementation, Applications*, Springer, 2012.

[21] S. Haykin and K. J. Ray Liu, Eds., *Handbook on array processing and sensor networks*, Wiley, 2009.