CALIBRATION OF DISTRIBUTED SOUND ACQUISITION SYSTEMS USING TOA MEASUREMENTS FROM A MOVING ACOUSTIC SOURCE

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 calibrating a distributed microphone array using time-of-arrival (TOA) measurements. The calibration encompasses localization and gain equalization of the microphones, which are both important in applications such as beamforming. The availability of accurate TOA measurements between the microphones and a set of spatially distributed acoustic events is pivotal to the calibration task. We propose to use a moving acoustic source emitting a calibration signal at known intervals. We then show that the TOAs and the observed signals can be used to estimate the gain differences between microphones in addition to the more established microphone localization. Finally, we provide experimental results with simulated and real measured data to demonstrate that our approach facilitates accurate TOA measurements and hence, accurate localization and gain equalization, even in reverberant and noisy conditions.

Index Terms— microphone localization, array calibration

1. INTRODUCTION

Microphone arrays can be used to capture speech and audio signals in adverse acoustic environments [1–3]. The microphone signals may be processed and combined in order to focus on a signal originating from a particular location with respect to the microphones while suppressing signals from all other locations. This results in reduced noise and reverberation compared to an unprocessed, single-microphone observation. A common approach to microphone array processing is beamforming and while there is a plethora of beamforming techniques [4], there are two common requirements for these to operate correctly: (i) the relative locations of the microphones must be known and (ii) the microphones must be calibrated [3, 5]. Furthermore, it may not always be feasible to measure manually the inter-microphone distances and to carefully select and calibrate the microphones as, for example, in ad-hoc microphone arrays. Therefore, an automatic procedure is preferred.

Several methods for localization of microphones and acoustic sources exist, majority of which rely on time-of-arrival (TOA) measurements from spatially distributed acoustic events [6–12]. Some alternatives such as time-difference-of-arrival (TDOA) [8, 13], signal energy [14, 15] and the diffuse noise field coherence [16] have also been explored. It has also been shown that gain calibration is important for several beamforming algorithms and some algorithms for automatic gain calibration have been proposed [5, 17, 18].

Despite the existence of many TOA-based localization methods, there has been relatively little discussion on how to obtain accurate TOA measurements in practice, and errors in such measurements are typically modelled as additive measurement noise. There are three specific issues to consider for TOA measurements, which were highlighted in [8, 12]. First, the TOA for a specific acoustic event must be identified correctly between all microphones. Second, the onset time of the acoustic event – the time at which the acoustic source begins transmitting the sound, and the internal delay – the time from that the sound reaches the microphone to that it is registered as received by the device must be identified. In [8] it is proposed to use chirps or maximum length sequences (MLSs) to identify the acoustic events; the onset times are estimated under the assumption that each microphone is associated with an acoustic event occurring in its proximity while the internal delays are measured manually. In [12] an algorithm was presented that is able to estimate the onset times and the internal delays from the measured TOAs directly.

In this paper, we show that a complete calibration of a distributed sound acquisition system can be achieved provided that accurate TOAs measurements are available. We propose to use a small device such as a mobile phone to emit a calibration signal from which the TOAs are obtained. The signal is emitted at known intervals and it will be shown in Sect. 3 that this determines the source onset times. We then derive a simplified version of the algorithm in [12] to estimate the internal delays. It will also be shown that such a controlled source facilitates the design of a calibration signal that it is robust to noise and reverberation. We demonstrate the use of the measured TOAs for two different purposes. In Sect. 4 we use these to localize the microphones. More importantly, in Sect. 5 we show that the TOAs can be used in conjunction with the observed calibration signals to estimate the relative gain differences between the microphones. Thus, our approach provides complete calibration for a distributed microphone array and it's performance will be supported by several experimental results in Sect. 6. In the following Sect. 2 we begin by formulating the problem.

2. PROBLEM FORMULATION

Consider a 3-dimensional space where $\mathcal{J}$ calibration signals $s_j(n)$ originating from different and unknown locations $\mathbf{s}_j = [s_{x,j}\, s_{y,j}\, s_{z,j}]^T$ are captured by $\mathcal{I}$ microphones at unknown locations $\mathbf{r}_i = [r_{x,i}\, r_{y,i}\, r_{z,i}]^T$. The signal from source j to microphone i can be written

$$x_{ij}(n) = G_i \left(h_{ij}(n) * s_j(n) + \nu_{ij}(n) \right), \quad (1)$$

where $h_{ij}(n)$ is the acoustic impulse response (AIR) and $\nu_{ij}(n)$ is additive measurement noise and $*$ denotes convolution. Each microphone has an associated unknown gain, G_i.

The objective of this work is threefold: (i) to extract accurately the TOAs between the calibration signal and the microphones from the observed signals $x_{ij}(n)$; (ii) to use the TOAs to calculate the relative microphone locations $\mathbf{r}_i$, and (iii) to use the TOAs and the observed signals $x_{ij}(n)$ to estimate the microphone gains G_i.

This work was supported by Google Inc.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

3. TOA ESTIMATION

Accurately measured TOAs form the foundation in auto-localization of distributed systems [19]. TOA measurements are obtained from the observed signals $x_{ij}(n)$ and the measured TOA of the jth acoustic event at microphone i is

$$t_{ij} = c^{-1}\|\mathbf{r}_i - \mathbf{s}_j\| + \tau_j + \delta_i + \epsilon_{ij}, \quad (2)$$

where c is the speed of sound, $\|\cdot\|$ denotes Euclidean norm and ϵ_{ij} is measurement noise, δ_i is the ith microphone internal delay and τ_j denotes the onset time of the jth acoustic event.

For localization, we are interested in the true TOAs, $\hat{t}_{ij} = c^{-1}\|\mathbf{r}_i - \mathbf{s}_j\|$. Therefore, the source onset times and the internal delays must be compensated for and the measurement error must be kept to a minimum. In the following, we describe a robust approach to achieve this by using a calibration signal from a moving source.

3.1. Calibration signals

We assume that a small portable device capable of audio playback, such as a mobile phone, is available for this procedure and that the microphones have a line of sight of the audio source at all times. A sequence of calibration signals is produced by the audio source at intervals T_{p}, while the device is moved (e.g., by a waving motion with a phone) and the signal is captured by the microphones. If the duration of the individual calibration signals in the sequence is short relative to the velocity of the moving source, each occurrence $s_j(n)$ can be treated as an acoustic event with distinct location $\mathbf{s}_j$ and onset time $\tau_j = \tau_1 + jT_{\mathrm{p}}$. Substituting τ_j with $\tau_1 + jT_{\mathrm{p}}$ in (2), we see that we can correct for the source onset times by subtracting the known jT_{p} from the observed TOAs. If we further set the onset time of the first acoustic event as the time origin, *i.e.*, $\tau_1 = 0$, we get that

$$t_{ij} = c^{-1}\|\mathbf{r}_i - \mathbf{s}_j\| + \delta_i + \epsilon_{ij}. \quad (3)$$

Using a calibration signal has several properties that improve the accuracy of microphone and source localization. Firstly, we have already seen that it eliminates the unknown acoustic source onset time. Secondly, it allows a large number of acoustic events to be generated in a short period of time. Lastly, the excitation signals $s_j(n)$ can be chosen to facilitate accurate TOA measurements.

A question that arises then is what signal to use for $s_j(n)$. The requirements based on the above discussions are that the measurement should be robust to additive noise and reverberation and the signal should be of short duration (compared to the velocity of the sound source). Three possible candidate signals are: (i) a unit impulse; (ii) a time-stretched pulse (TSP) as defined in [20] – this is normally used for measurement of AIRs and has the desirable feature that the pulse can be much shorter than the length of the AIR; (iii) a Gaussian modulated sinusoidal pulse (GMSP) – a pulse that is maximally localized in both time and frequency. One advantage of a frequency localized pulse is that it can be centred on the spectral region where the microphone characteristics can be assumed favourable.

3.2. TOA extraction

The TOAs must be measured from the observed signals $x_{ij}(n)$. One alternative would be to detect the onsets of the sounds [21,22]. The advantage of this is that the line of sight between the source and the microphone is not required, however, finding the onsets is not straightforward and can be very sensitive to additive noise [22]. Instead, we measure the TOAs using peak picking as follows.

The received signals (except for the unit impulse) are post-processed using a matched filter, $h_{\mathrm{mf},j}(n)$, which is the time-reversed version of the excitation signal, $h_{\mathrm{mf},j}(n) = s_j(-n)$, and the TOAs t_{ij} are extracted. First, the input signal $x_{ij}(n)$ is processed in non-overlapping frames of T_{p} s and the peak of highest power is selected in each frame; these form the candidate TOAs. Then the $\mathcal{J}$ peaks with highest energy are chosen from the candidates in the previous step. In the case of the TSP, the matched filter is equivalent to the inverse filter of the sequence and it results in an impulse (leaving the AIR). For the GMSP, it results in a peak at maximum correlation between the two signals. In general matched filtering also efficiently suppresses uncorrelated additive noise; the amount of noise suppression increases with sampling frequency and with filter length [23].

3.3. Internal delay estimation

The TOAs obtained from Sect. 3.2 are accurate up to an internal delay. An algorithm to estimate internal delays and source onset times was proposed in [12]. However, the source onset times are now determined through the use of a calibration signal and only the internal delays need estimating. This leads to a simplified version of the algorithm with faster convergence as described in the following.

Let us assume, without loss of generality, that $c = 1$ and also that the TOAs have been extracted correctly (up to an internal delay) so that $\epsilon_{ij} = 0$. Squaring both sides of (3), subtracting the equation for $i = 1$ then subtracting the equation for $j = 1$ we obtain [11]

$$\begin{aligned} -(\mathbf{r}_i - \mathbf{r}_1)^T(\mathbf{s}_j - \mathbf{s}_1) &= 0.5(t_{ij}^2 - t_{1j}^2 - t_{i1}^2 + t_{11}^2) \\ &\quad - \delta_i(t_{ij} - t_{i1}) + \delta_1(t_{1j} - t_{11}), \end{aligned} \quad (4)$$

$i = 2, \ldots, \mathcal{I}$, $j = 2, \ldots, \mathcal{J}$. We can further write (4) as

$$-\bar{\mathbf{R}}\bar{\mathbf{S}}^T = \mathbf{T} + \mathbf{A}(\boldsymbol{\delta}), \quad (5)$$

where $\bar{\mathbf{R}}$ is the $(\mathcal{I} - 1) \times 3$ location matrix of the microphones relative to $\mathbf{r}_1$, $\bar{\mathbf{S}}$ is the $(\mathcal{J} - 1) \times 3$ location matrix of the acoustic events relative to $\mathbf{s}_1$, $\mathbf{T}_{i-1j-1} = 0.5(t_{ij}^2 - t_{1j}^2 - t_{i1}^2 + t_{11}^2)$, $\mathrm{vec}\{\mathbf{A}(\boldsymbol{\delta})\} = \mathbf{W}\boldsymbol{\delta}$, $\boldsymbol{\delta} = [\delta_1\ \delta_2\ \ldots\ \delta_{\mathcal{I}}]^T$ and $\mathbf{W}$ is a matrix composed of terms $(t_{ij} - t_{i1})$ and $(t_{1j} - t_{11})$; $\mathrm{vec}\{\mathbf{X}\}$ defines an operator that stacks the columns of a matrix into a column vector.

From (5) we can make the important observation that matrix $\bar{\mathbf{R}}\bar{\mathbf{S}}^T$ is at most of rank 3. Consequently, we use a two-stage iterative algorithm as in [12] to estimate the internal delays $\boldsymbol{\delta}$ such that $\widehat{\mathbf{T}} = \mathbf{T} + \mathbf{A}(\boldsymbol{\delta}^{(n)})$ is of rank 3. In the first stage we use the current estimate of $\boldsymbol{\delta}$, $\boldsymbol{\delta}^{(n)}$ and keeping this fixed we find the best estimate of $\bar{\mathbf{R}}\bar{\mathbf{S}}^T$ by solving the following optimization problem

$$\mathbf{T}_3^{(n)} = \arg\min_{\mathbf{T}_3} \|\widehat{\mathbf{T}}^{(n)} - \mathbf{T}_3\|_F \quad \text{s.t. } \mathrm{rank}\{\mathbf{T}_3\} \leq 3, \quad (6)$$

where $\widehat{\mathbf{T}}^{(n)} = \mathbf{T} + \mathbf{A}(\boldsymbol{\delta}^{(n)})$ and $\|\cdot\|_F$ denotes the Frobenius norm; $\mathbf{T}_3^{(n)}$ is the best rank-3 approximation of $\widehat{\mathbf{T}}^{(n)}$ and hence, the best approximation of $\bar{\mathbf{R}}\bar{\mathbf{S}}^T$ (in the Frobenius norm sense). The solution to (6) can be obtained using the Eckart-Young-Mirsky low-rank approximation theorem [24]. Consider the singular value decomposition (SVD) of $\widehat{\mathbf{T}}^{(n)}$,

$$\widehat{\mathbf{T}}^{(n)} = [\mathbf{U}_1\ \mathbf{U}_2] \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 \\ 0 & \boldsymbol{\Sigma}_2 \end{bmatrix} [\mathbf{V}_1\ \mathbf{V}_2]^T, \quad (7)$$

where $\boldsymbol{\Sigma}_1$ contains the 3 largest singular values and $\mathbf{U}_1$ and $\mathbf{V}_1$ are the corresponding left- and right-singular vectors. The optimal rank-3 approximation is obtained as

$$\mathbf{T}_3^{(n)} = \mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{V}_1^T. \quad (8)$$

In the second stage, we update the estimate of the internal delays as

$$
\begin{aligned}
\boldsymbol{\delta}^{(n+1)} &= \arg\min_{\boldsymbol{\delta}} \|\mathbf{E}^{(n)} - \mathbf{A}(\boldsymbol{\delta})\|_F^2 + \lambda \|\widehat{\mathbf{T}}^{(n)}\|_F^2 \\
&= \frac{\mathbf{W}^{+}\mathrm{vec}\{\mathbf{E}^{(n)} - \lambda \mathbf{T}\}}{1+\lambda}
\end{aligned}
\tag{9}
$$

where $\mathbf{E}^{(n)} = \mathbf{T}_3^{(n)} - \mathbf{T}$ and $\mathbf{W}^{+}$ is the pseudo-inverse of $\mathbf{W}$. The additional term $\|\widehat{\mathbf{T}}^{(n)}\|_F$ is introduced to increase the initial convergence rate, but λ needs to be set to zero according to some criterion to allow the algorithm to converge to the correct solution. This is done by monitoring $\|\widehat{\mathbf{T}}^{(n)}\|_F$ and setting $\lambda = 0$ when the change between consecutive iterations is below a threshold. Similarly, $\|\mathbf{E}^{(n)} - \mathbf{A}(\boldsymbol{\delta}^{(n)})\|_F$ is monitored and the algorithm is stopped when the change from one iteration to the next is below a threshold. It was shown in [12] that this algorithm converges in terms of the Frobenius norm when $\lambda \to 0$. An alternative algorithm that uses nuclear norm regularization to solve the same problem was presented in [25]. Note that in [12] there is an additional term in (5) which propagates to (9) and thus, requires non-linear optimization in this step (unless all internal delays are equal). In the above algorithm these terms are not present, which improves the convergence rate of the algorithm compared to [12].

4. MICROPHONE LOCALIZATION

With accurate estimates of the internal delays at hand, we can now estimate the locations of the microphones (and the sources) by the nonlinear least-squares (LS) criterion based on (3)

$$
\widehat{\mathbf{R}}, \widehat{\mathbf{S}}, \widehat{\boldsymbol{\delta}} = \arg\min_{\mathbf{R},\mathbf{S},\boldsymbol{\delta}} \sum_{i=1}^{\mathcal{I}} \sum_{j=1}^{\mathcal{J}} \left((\|\mathbf{r}_i - \mathbf{s}_j\| + c\delta_i) - d_{ij} \right)^2, \tag{10}
$$

where $d_{ij} = ct_{ij}$ and we use $c = 343$ m/s. This optimization is prone to local minima if not initialized carefully. Therefore, we apply the efficient method in [11] to obtain the initial values of the microphone and source locations and we initialize the vector of internal delays with that estimated in Sect. 3.3.

5. MICROPHONE GAIN CALIBRATION

We now show how to use the estimated TOAs together with the received signals to find the unknown gains G_i at each microphone. We consider a free-field AIR and assume that the noise and the excitation signals are uncorrelated. From (1) we can write the energy at the ith microphone due to the jth acoustic event as

$$
E_{x,ij} = \frac{G_i^2 E_{s,j}}{d_{ij}^2} + G_i^2 E_{\nu,ij}, \tag{11}
$$

where $E_{s,j}$ is the energy of the signal at the acoustic source at the jth location, $E_{\nu,ij}$ is the energy of the measurement noise, and $d_{ij} = \|\mathbf{r}_i - \mathbf{s}_j\| = c\hat{t}_{ij}$ is the distance between the jth acoustic source and the ith microphone. Note that we can use the TOA measurements or the estimated source–microphone locations for the distance estimate.

We choose an arbitrary reference microphone, for example, the first, $i = 1$. This leads to the following expression for the microphone gains relative to the reference microphone

$$
\frac{E_{x,ij} d_{ij}^2}{E_{x,1j} d_{1j}^2} = \frac{G_i^2}{G_1^2} \left(\frac{1 + \zeta_{ij}^{-1}}{1 + \zeta_{1j}^{-1}} \right), \tag{12}
$$

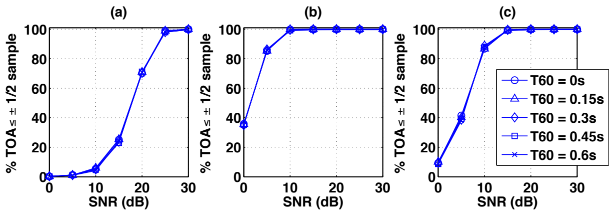

Fig. 1. Estimation error in % correct TOAs to within $\pm 1/2$ sample using (a) unit impulse, (b) GMSP and (c) TSP.

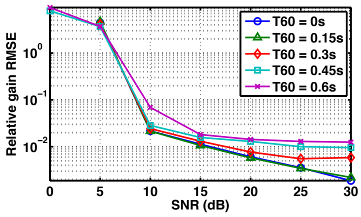

Fig. 2. Gain estimates vs SNR and for different T_{60}.

where ζ_{ij} is the signal-to-noise ratio (SNR) for the jth source at the ith microphone. Consequently, we can estimate the relative gains as

$$
\frac{G_i}{G_1} = \frac{1}{\mathcal{J}} \sum_{j=1}^{\mathcal{J}} \sqrt{\frac{E_{x,ij} d_{ij}^2 (1 + \zeta_{1j}^{-1})}{E_{x,1j} d_{1j}^2 (1 + \zeta_{ij}^{-1})}}, \tag{13}
$$

where we use the measured TOAs and the *a priori* knowledge of the signal durations to evaluate the signal energies. The bias due to noise is corrected by estimating the noise level at the microphones in the intervals between the calibration signals, which are determined by the TOAs. Note that the TOA estimation is independent of gain differences. Interestingly, for spatially diffuse noise and for closely spaced microphones where the SNR can be expected to be the same across all microphones, noise will have little effect. Reverberation is not considered here but its effects can be minimized by making the excitation intervals T_{p} greater than the reverberation time.

6. EXPERIMENTAL RESULTS

We provide experimental results to demonstrate the microphone array calibration method. First, we evaluate the TOA estimation using different calibration signals. Second, we evaluate the internal delay estimation algorithm. Finally, we use the measured TOAs for microphone localization and gain calibration.

The observed signals were generated according to (1). We simulated an acoustic environment using the source-image method [26, 27] for a room with dimensions $6 \times 5 \times 4$ m. The reverberation time T_{60}, was varied between 0 s (free-field) and 0.6 s in steps of 0.15 s. $\mathcal{I} = 8$ microphones were positioned randomly within a rectangular prism of $2 \times 2 \times 1$ m in the centre of the room and $\mathcal{J} = 30$ sources were randomly distributed in a one meter cubic space positioned at random in the room; ten different source-microphone configurations were generated. Each source-point was defined by the AIR $h_{ij}(n)$ and emulates an instantaneous location of the moving acoustic source. The calibration signal $s_j(n)$ was either an impulse, a 1.33 ms TSP or a 1.33 ms GMSP emitted at an

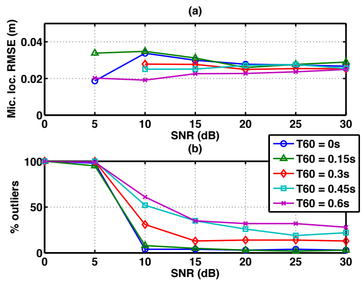

Fig. 3. (a) RMSE of the localized microphones and (b) % outliers defined as $\|\mathbf{R} - \widehat{\mathbf{R}}\| \geq 0.01$ m.

SNR (dB) \ T_{60}(s)	0.00	0.15	0.30	0.45	0.60
0	102.93	119.90	153.28	331.82	202.11
5	58.23	89.15	107.34	151.64	125.88
10	0.77	0.80	1.57	2.13	2.84
15	0.60	0.60	1.06	1.68	1.96
20	0.51	0.59	0.99	1.50	1.99
25	0.54	0.52	0.96	1.16	1.96
30	0.47	0.48	0.94	1.26	1.94

Table 1. RMSE of internal delay estimates in ms.

interval of $T_{\mathrm{p}} = 0.1$ s and a sampling rate of $f_s = 48$ kHz. We assumed white Gaussian additive noise for $\nu_{ij}(n)$. The noise level was adjusted with reference to the free-field unit impulse and with respect to the weakest signal – the calibration signals for the largest source-microphone distance – according to a desired SNR and was then held constant for the other calibration signals; SNRs between 0 dB and 30 dB were considered. In this way, the weakest signal for the impulse was completely obscured by the noise at SNR=0 dB. Ten realizations of the noise signal were generated. Furthermore, a random gain factor G_i varying between 0.5 and 1 and a random internal delay between 0 and 0.1 s was applied to each microphone. The resulting signal was processed with a POTS filter limiting the bandwidth; this was in order to emphasize the benefit of the GMSP, which was designed such that its frequency band coincides with that of the POTS filter. All results are presented as the average of the ten source-microphone locations and the ten noise realizations.

Experiment 1: We evaluated the different calibration signals for extracting the TOAs following the method in Sect. 3.1. The internal delays were assumed known for the purpose of this evaluation. The minimum error depends on the sampling rate and is within one sample. Thus, for each case we calculated the percentage correctly identified TOAs within $\pm 1/2$ sample. The results are shown in Fig. 1 (a)-(c) for the unit impulse, the GMSP and the TSP, respectively. It can be seen that the GMSPs has the greatest tolerance to noise and reverberation and all TOAs are extracted accurately down to 10 dB SNR even with the simple extraction algorithm from Sect. 3.1.

Experiment 2: We used the measured TOAs with the TSP as calibration signal to estimate the internal delays using the algorithm from Sect. 3.3. Consequently, this included realistic measurement errors ϵ_{ij} in (2). The results are shown in Table 1, where we see that acceptable estimates to within 3 ms are obtained for all reverberation times and for SNR$\geq$ 10 dB.

Experiment 3: We used the estimated TOAs to find the gains relative to microphone $i = 1$ following the method from Sect. 5. The results for different noise levels and reverberation are shown in

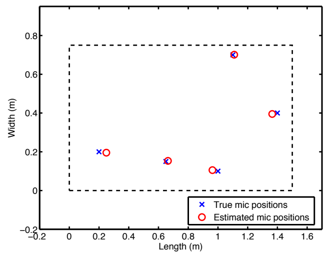

Fig. 4. Localization with measured data using five microphones (crosses) positioned on a table (dashed line). Estimated microphone locations (circles) are accurate to within 2.9 cm.

Fig. 2. It can be seen that in the SNR range where the TOAs are extracted correctly, the gains are estimated to an accuracy of 10^{-2} in terms of root mean squared error (RMSE).

Experiment 4: We used the complete measured TOAs to find the locations of the microphones and the sources with the method described in Sect. 4. The results in Fig. 3 (a) show the average estimation error over the cases where the algorithm converged to an error lower than 0.1 m; the remainder of the results were classified as outliers. Figure 3 (b) shows the percentage outliers as a function of SNR and for different amounts of reverberation. Two important observations can be made: (i) the localization is extremely sensitive to errors in the TOA estimates and even one wrongly detected TOA can result in wrong localization; (ii) when accurate TOAs are obtained the localization is accurate to within $1 - 4$ cm.

Experiment 5: In the final experiment, we performed all of the steps in the experiments above but we used five AKG C417 lapel microphones connected to an RME Fireface 800 through an RME Octamic II microphone preamplifier. The microphones were positioned randomly on a table with dimensions 0.75×1.5 m in a quiet, mildly reverberant room. A mobile phone was used as a sound source and a pulse train with 50 pulses at $T_p = 0.1$ s intervals were produced, while a waving motion was performed with the phone. An additional random delay between 0 and 100 ms was added to each microphone signal to simulate different internal delays. The outcome is shown in Fig. 4. The microphones were localized to an accuracy of 2.9 cm, while using the centralized clock resulted in an RMSE of 1.4 cm.

7. CONCLUSIONS

We demonstrated that accurate estimation of the TOAs is the key to successful calibration of distributed microphone arrays. To address this, we proposed a method for TOA measurements from spatially distributed acoustic events using a calibration signal emitted at known intervals by a moving acoustic source. We showed that this eliminates the unknown source onset times and we derived an algorithm to estimate the internal delays of the microphones. We also showed that the calibration signal can be designed to provide robustness to noise, reverberation and spectral characteristics of the microphones and that the GMSP is a particularly good choice for this purpose. Subsequently, we introduced a new method for gain calibration based on the TOAs and showed that the relative gain differences of the microphones can be estimated accurately even in the presence of noise and reverberation. Finally, using both simulated and real measured data, we demonstrated the use of the measured TOAs for microphone localization resulting in an accuracy of $1 - 4$ cm.

8. REFERENCES

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

[2] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, Springer-Verlag, Berlin, Germany, 2008.

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

[4] B. D. van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Acoustics, Speech and Signal Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[5] I. Tashev, "Gain self-calibration procedure for microphone arrays," in *Proc. Intl. Conf. Multimedia and Expo (ICME)*, Taipei, Taiwan, June 2004.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] N. D. Gaubitch, W. B. Kleijn, and R. Heusdens, "Auto-localization in ad-hoc microphone arrays," in *Proc. IEEE Intl. Conf. on Acoust., Speech, Signal Process. (ICASSP)*, Vancouver, Canada, May 2013, pp. 106–110.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] T. P. Hua, A. Sugiyama, and G. Faucon, "A new self-calibration technique for adaptive microphone arrays," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Eindhoven, Netherlands, Sept. 2005, pp. 237–240.

[18] M. Buck, T. Haulick, and H-J. Pfleiderer, "Self-calibrating microphone arrays for speech signal acquisition: A systematic approach," *Signal Processing*, vol. 86, pp. 1230–1238, 2006.

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

[20] Y. Suzuki, F. Asano, H.-Y. Kim, and Toshio Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *J. Acoust. Soc. Am.*, vol. 97, no. 2, pp. 1119–1123, 1995.

[21] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 1035–1047, Sept. 2005.

[22] G. Defrance, L. Daudet, and J-D. Polack, "Finding the onset of a room impulse response: Straightforward?," *J. Acoust. Soc. Am.*, vol. 124, no. 4, pp. EL248–EL254, Oct. 2008.

[23] R. Heusdens, S. Engelen, P. J. Buist, A. Noroozi, P. Sundaramoorthy, C. Verhoeven, M. Bentum, and E. Gill, "Match filtering approach for signal acquisition in radio-pulsar navigation," in *Proc. Intl. Astronautical Congress*, Napels, Italy, 2012, vol. B2.6.10, pp. 1–5.

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

[25] F. Jiang, Y. Kuang, and K. Åström, "Time delay estimation for TDOA self-calibration using truncated nuclear norm regularization," in *Proc. IEEE Intl. Conf. on Acoust., Speech, Signal Process. (ICASSP)*, Vancouver, Canada, May 2013, pp. 3885 – 3889.

[26] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[27] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. Acoust. Soc. Am.*, vol. 80, no. 5, pp. 1527–1529, Nov. 1986.