TDOA-BASED SELF-CALIBRATION OF DUAL-MICROPHONE ARRAYS

Mojtaba Farmani[1], *Richard Heusdens*[2], *Michael Syskind Pedersen*[3], *Jesper Jensen*[1,3]

[1]Department of Electronic Systems, Aalborg University, Denmark,{mof, jje}@es.aau.dk
[2]Group of Circuits & Systems, Delft University of Technology, Netherlands, r.heusdens@tudelft.nl
[3]Oticon A/S, Denmark,{micp, jesj}@oticon.com

ABSTRACT

We consider the problem of determining the relative position of dual-microphone sub-arrays. The proposed solution is mainly developed for binaural hearing aid systems (HASs), where each hearing aid (HA) in the HAS has two microphones at a known distance from each other. However, the proposed algorithm can effortlessly be applied to acoustic sensor network applications. In contrast to most state-of-the-art calibration algorithms, which model the calibration problem as a non-linear problem resulting in high computational complexity, we model the calibration problem as a simple linear system of equations by utilizing a far-field assumption. The proposed model is based on target signals time-difference-of-arrivals (TDOAs) between the HAS microphones. Working with TDOAs avoids clock synchronization between sound sources and microphones, and target signals need not be known beforehand. To solve the calibration problem, we propose a least squares estimator which is simple and does not need any probabilistic assumptions about the observed signals.

***Index Terms*—** Microphone array calibration, hearing aid, DOA, TDOA, far-field

1. INTRODUCTION

Performance of many signal processing algorithms using microphone arrays depends on the knowledge of the microphone array geometry. For example, in [1, 2], the microphone array geometry is needed to estimate the direction of arrival (DOA) of the target sound for a binaural hearing aid system (HAS). A binaural HAS consists of two hearing aids (HAs) mounted on the ears of a user. Different heads radii and varying shapes of pinnae of users cause uncertainties about the geometry of the microphone array, e.g. the distance between the HAs, which degrade performance of the DOA estimation algorithms.

The microphone array calibration problem is the problem of determining the relative locations of the microphones in a microphone array. This problem has been studied using different types of measurements such as received signal strength (RSS) [3], time-of-arrival (TOA) [4–6], and time-difference-of-arrival (TDOA) [7]. Among these, TDOA is a suitable choice for HAS applications because it is less vulnerable to reverberation [4], does not require clock-synchronization between sources and microphones, and
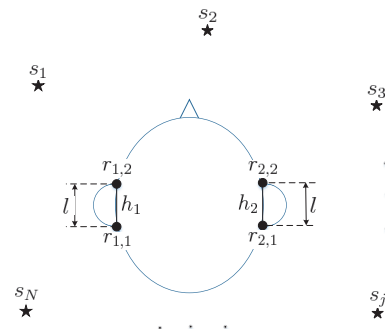

Fig. 1: A typical scenario of microphone array calibration problem for a binaural HAS. We aim to find the relative locations of h_1 and h_2 using signals received from sound sources $s_1, s_2, ..., s_N$ which are distributed randomly around the user.

does not require the time of emission of the target signals.

Different techniques have been proposed to solve the calibration problem. Multi-dimensional scaling (MDS) [8] is one of the earliest methods that implicitly needs each node (HA) to be a compound of a microphone and a sound source, a requirement which in general is not satisfied in HA applications. Another approach has been proposed in [9] based on singular value decomposition (SVD) that finds the coordinates of the microphones up to an invertible matrix by assuming that sources are in the far-field. Finding the appropriate invertible matrix is a non-linear optimization problem [9], which might be trapped in local minima. An SVD-based approach has also been proposed in [10], which avoids the far-field assumption but requires co-location of one of the sources and one of the microphones for a closed-form solution. Recently, an alternative approach was proposed [11] that solves the localization problem for a minimal case, where minimal number of microphones and sound sources are required to solve the problem, without imposing any co-location constraint. However, for overdetermined cases, where more sound sources or microphones than the minimal case are available, an additional non-linear optimization is still required. In [12] a closed-form solution has been proposed for an overdetermined case based on ToA measurements, for which synchronization of sources and microphones is needed. Lately, a new approach has been proposed [6] where pairs of microphones are set on a rigid rack, similar to the problem
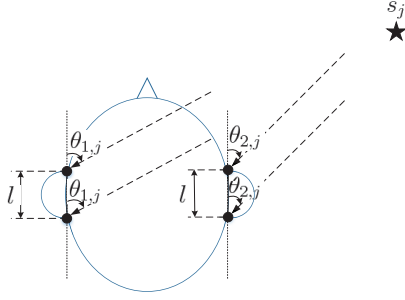
Fig. 2: The special far-field assumption considered in this paper.

considered in this paper. However, the approach in [6] is based on TOA measurements which are not suitable for HAS applications.

Fig. 1 shows an exemplary scenario of the problem considered in this paper. There are two HAs h_k, $k = 1,2$, each with two microphones $r_{k,1}$ and $r_{k,2}$. The distance l between $r_{k,1}$ and $r_{k,2}$ is known, but the relative locations of h_1 and h_2 are unknown (we define the location of h_k as the center of its microphones axis). We aim to find the relative locations of h_1 and h_2 using the signals received by the HAs microphones from N sound sources $s_1, s_2, ..., s_N$. We assume that N is known and, at each time frame, exactly one sound source is active. This assumption is reasonable in HA applications, because when the HAS user moves his/her head, the relative location of a sound source with respect to the microphone array will change, which can be interpreted as a new sound source originating from a different relative location. Therefore, the user's head movements ensure sound signals from several different relative locations as needed.

The main contribution of this paper is in modeling the microphone array calibration problem as a linear system by utilizing a special far-field assumption. The proposed model is based on target signals TDOAs, which do not need clock synchronization between sound sources and microphones, and knowledge of target signals is not necessary. The latter point means that special calibration signals are unnecessary, and we can use signals which are naturally present, e.g. speech signals, for the calibration. To solve the modeled calibration problem, we use a least squares (LS) estimator, which additionally provides estimates of the sound sources locations. The proposed method effectively exploits the extra information about the microphones distance in a HA and needs only two sources when considering the horizontal plane, i.e. two dimensions. For simplicity, we will discuss our estimator in 2D. However, the generalization to three dimensions is straightforward.

2. PROBLEM FORMULATION

Let $t_{k,i,j}$ denote the TOA of the target signal generated by source s_j received at receiver $r_{k,i}$ (microphone $i \in \{1,2\}$ of hearing aid $h_k \in \{h_1, h_2\}$), which is given by

$$t_{k,i,j} = \frac{\|r_{k,i} - s_j\|_2}{c} + t_j + \delta_{k,i}, \tag{1}$$

where $\|.\|_2$ denotes the Euclidean norm, c is the sound speed, t_j is the emission time at source j, and δ_i is the internal delay of microphone $r_{k,i}$. If we assume that the internal delays of the HAS microphones are equal, i.e. $\delta_{k,i} = \delta$ for all i and k, the TDOA of the target signal generated by source j received at $r_{k,i}$ and $r_{u,w}$ (microphone $w \in \{1,2\}$ of hearing aid $h_u \in \{h_1, h_2\}$) is

$$\Delta_{k,i,u,w,j} = t_{k,i,j} - t_{u,w,j} = \frac{\|r_{k,i} - s_j\|_2}{c} - \frac{\|r_{u,w} - s_j\|_2}{c}.$$

Hence, the TDOA depends only on the locations of the sources and the receivers, and it is independent of the δ and t_js. In the following, we will estimate the relative locations of the HAs using TDOAs and a special far-field assumption.

2.1. Far-Field Assumption

Let $d_{k,j}$ denote the distance between s_j and h_k. In HAS applications, the $d_{k,j}$s are usually much larger than the microphones distance within a HA, i.e. $d_{k,j} \gg l$. Therefore, we can assume that the DOAs of the target sounds for the microphones of a HA are almost equal (see Fig. 2). However, we assume the target distances are not much larger than the diameters of the user's head, which means $\theta_{1,j}$ and $\theta_{2,j}$ are not necessarily equal.

The far-field assumption and the given estimated TDOAs allow us to estimate $\theta_{k,j}$, $k = 1,2$ (see Fig. 2), up to a sign as follows:

$$\hat{\Delta}_{k,2,k,1,j} = \frac{l}{c}\cos\left(\hat{\theta}_{k,j}\right)$$
$$\Rightarrow \tilde{\theta}_{k,j} = \pm\hat{\theta}_{k,j} = \pm\arccos\left(\frac{c}{l}\hat{\Delta}_{k,2,k,1,j}\right), \tag{2}$$

where $\hat{\Delta}_{k,2,k,1,j}$ is the estimated TDOA between $r_{k,2}$ and $r_{k,1}$ for the target signal from s_j. Note that the DOAs are expressed clockwise with respect to the microphones axis. Moreover, we define the TDOA of the target signal from s_j between midpoint of h_1 and h_2 as $\Delta_j = \frac{\hat{\Delta}_{2,1,1,1,j} + \hat{\Delta}_{2,2,1,2,j}}{2}$ to estimate $\Delta d_j = d_{2,j} - d_{1,j}$ as

$$\Delta d_j \approx \Delta_j c. \tag{3}$$

Therefore, there are three known parameters for each source s_j: $\tilde{\theta}_{1,j}$, $\tilde{\theta}_{2,j}$ and Δd_j, which leads to $3N$ known parameters in total. On the other hand, the locations of the sound sources, h_1 and h_2 are unknown. Without loss of generality, we will assume $h_1 = [0,0]^T$, and we estimate locations of h_2 and $\{s_1,...,s_N\}$ with respect to h_1. As a consequence, we have $2N+2$ unknown in a two-dimensional scenario, and the calibration problem is solvable when $3N \geq 2N+2$, i.e. $N \geq 2$.

3. LOCALIZATION ALGORITHM

In this section, we propose an algorithm to estimate the relative locations of h_1 and h_2 using the known parameters. The relation between s_j and h_k, $k=1,2$, can be written as

$$s_j = h_k + d_{k,j}\left[\sin(\theta_{k,j}) \quad \cos(\theta_{k,j})\right]^T, \tag{4}$$

which allows us to formulate the relative location of h_2 as

$$h_2 = \begin{bmatrix} X \\ Y \end{bmatrix} = h_1 + d_{1,j} \begin{bmatrix} \sin(\theta_{1,j}) \\ \cos(\theta_{1,j}) \end{bmatrix} - d_{2,j} \begin{bmatrix} \sin(\theta_{2,j}) \\ \cos(\theta_{2,j}) \end{bmatrix}. \quad (5)$$

From Eq. (3), we have $d_{2,j} = d_{1,j} + \Delta d_j$. Therefore,

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} d_{1,j}\sin(\theta_{1,j}) - (d_{1,j} + \Delta d_j)\sin(\theta_{2,j}) \\ d_{1,j}\cos(\theta_{1,j}) - (d_{1,j} + \Delta d_j)\cos(\theta_{2,j}) \end{bmatrix}. \quad (6)$$

Considering the second row of Eq. (6), we can express $d_{1,j}$ as a function of Y and Δd_j:

$$d_{1,j} = \frac{Y + \Delta d_j \cos(\theta_{2,j})}{\cos(\theta_{1,j}) - \cos(\theta_{2,j})}. \quad (7)$$

Substitution of Eq. (7) into the first row of Eq. (6) leads to:

$$\begin{bmatrix} \cos(\theta_{1,j}) - \cos(\theta_{2,j}) \\ -\sin(\theta_{1,j}) + \sin(\theta_{2,j}) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} X \\ Y \end{bmatrix} = \Delta d_j \sin(\theta_{1,j} - \theta_{2,j}), \quad (8)$$

and considering N sound sources together leads to a linear system of equations

$$\mathbf{A}h_2 = \boldsymbol{b}, \quad (9)$$

where $A \in \mathcal{R}^{N \times 2}$ and $\boldsymbol{b} \in \mathcal{R}^N$. The first and second columns of row j of $\mathbf{A}$ are $A_{j1} = \cos(\theta_{1,j}) - \cos(\theta_{2,j})$, $A_{j2} = -\sin(\theta_{1,j}) + \sin(\theta_{2,j})$ respectively, and row j of $\boldsymbol{b}$ is $b_j = \Delta d_j \sin(\theta_{1,j} - \theta_{2,j})$. Because in practice observations are always noisy, to obtain the location of h_2 based on Eq. (9), we will compute a LS estimate of h_2 which is given by

$$\hat{h}_2 = \mathbf{A}^+ \boldsymbol{b}, \quad (10)$$

where $\mathbf{A}^+$ denotes the pseudo-inverse of $\mathbf{A}$. and straightforwardly, the LS estimators of $s_j \in \{s_1, s_2, ..., s_N\}$ can be obtained by replacing $\hat{h}_2$ in Eqs. (7) and (4), respectively.

One remaining issue is that, as showed in Sec. 2.1, we can estimate $\theta_{k,j}$ only up to a sign (see Eq. (2)). Therefore, for each s_j, three different cases are conceivable (see Fig. 3):

- Case 1: s_j is on the right sides of h_1 and h_2 (Fig. 3a), i.e. $\theta_{1,j} = +\hat{\theta}_{1,j}$ and $\theta_{2,j} = +\hat{\theta}_{2,j}$.
- Case 2: s_j is between h_1 and h_2 (Fig. 3b), i.e. $\theta_{1,j} = +\hat{\theta}_{1,j}$ and $\theta_{2,j} = -\hat{\theta}_{2,j}$.
- Case 3: s_j is on the left sides of h_1 and h_2 (Fig. 3c), i.e. $\theta_{1,j} = -\hat{\theta}_{1,j}$ and $\theta_{2,j} = -\hat{\theta}_{2,j}$.

We can distinguish Case 1 and Case 3 by Δ_j:

- If $\Delta_j > 0$, the target signal reached h_1 before h_2, i.e. case 3 cannot be the case.
- If $\Delta_j < 0$, the target signal reached h_2 before h_1, i.e. case 1 cannot be the case.

However, cases 1 and 2, and cases 2 and 3 are not distinguishable from each other based on Δd_j. In other words:

$$[\theta_{1,j}, \theta_{2,j}] = \begin{cases} [\pm\hat{\theta}_{1,j}, -\hat{\theta}_{2,j}], & \text{if } \Delta d_j > 0 \\ [+\hat{\theta}_{1,j}, \pm\hat{\theta}_{2,j}], & \text{otherwise} \end{cases}. \quad (11)$$

(a) The source is on the right side of the both HAs.

(b) The source is between the HAs.

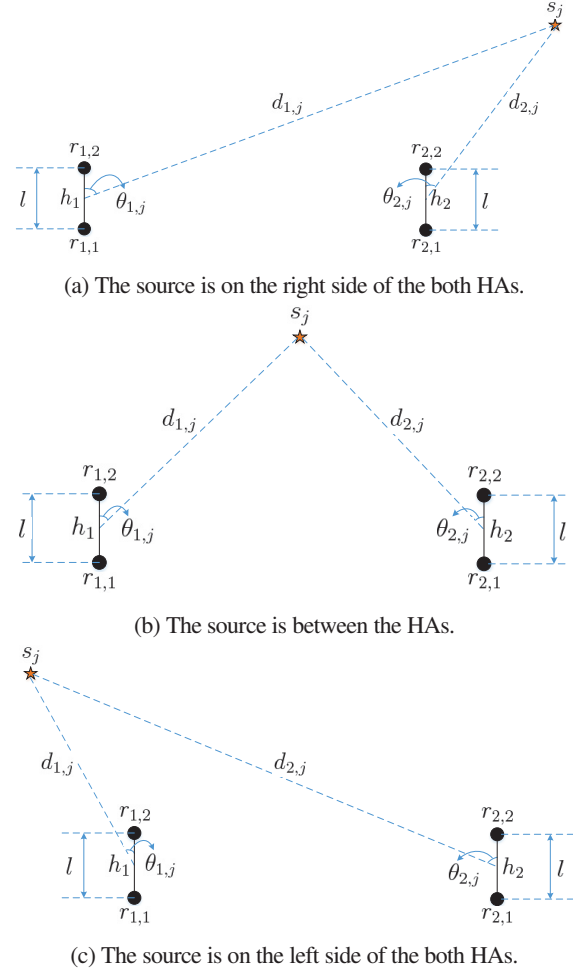(c) The source is on the left side of the both HAs.

Fig. 3: Different relative locations of a sound source with respect to a binaural HAS.

Therefore, for each source, we have two different cases which cannot be distinguished based on Δd_j. To resolve this ambiguity, we solve the calibration problem for all possible combinations of different cases of the $\theta_{k,j}$s, and the combination of the cases that can justify all the estimated parameters best is the solution to the problem. Two different cases for each source result in 2^N different combinations of cases considering all sources. Therefore, the problem must be solved for 2^N different combinations of the cases, and the best combination b^* is given by:

$$b^* = \arg\min_{b \in \mathcal{B}} \sum_{j=1}^{N} \|\Delta d_j - \hat{\Delta} d_{j,b}\|_2, \quad (12)$$

where $\mathcal{B}$ is the set of all possible combinations of the cases, and $\hat{\Delta} d_{j,b} = \hat{d}_{2,j,b} - \hat{d}_{1,j,b}$, where $\hat{d}_{1,j,b}$ is obtained by Eq. 7 for combination b and $\hat{d}_{2,j,b} = \|\hat{h}_{2,b} - \hat{s}_{j,b}\|_2$, ($\hat{h}_{2,b}$ and $\hat{s}_{j,b}$ denote the estimated locations of h_2 and s_j for combination b, respectively). The outputs of the localization algorithm are $\hat{h}_{2,b^*}$ and $\{\hat{s}_{1,b^*}, ..., \hat{s}_{N,b^*}\}$.

3.1. TDOA estimation

The last issue is how to estimate the TDOAs upon which the above algorithm relies. The most well-known approach for time delay estimation (TDE) is based on the Generalized Cross Correlation (GCC) method [13]: the GCC of two correlated signals has a maximum at a lag τ corresponding to the delay.

Let $r_{k,i,j}(n)$ and $r_{u,w,j}(n)$ denote the signals received from source j by microphone i of hearing aid k, and microphone w of hearing aid u, respectively. Furthermore, let $R_{k,i,j}(f)$ and $R_{u,w,j}(f)$ denote their discrete Fourier transforms (DFTs), respectively. The GCC is then given by [13]:

$$\mathcal{R}_{k,i,u,w,j}(\tau) = \sum_{f=1}^{M} \psi(f) R_{k,i,j}^*(f) R_{u,w,j}(f) e^{j2\pi f\tau}, \quad (13)$$

where M is the DFT order, $*$ represents complex conjugation and $\psi(.)$ is a weighting function. Then, the estimated $\Delta_{k,i,u,w,j}$ is given by:

$$\hat{\Delta}_{k,i,u,w,j} = \underset{\tau}{\arg\max}\ \mathcal{R}_{k,i,u,w,j}(\tau). \quad (14)$$

Because microphone array calibrations are usually performed in high SNR situations, we simply use the conventional cross-correlation method for TDOA estimation, i.e. $\psi(f)=1$ for all f in Eq. (13). However, to improve the TDE performance in noisy situations, there are more complex weighting functions which take into account the noise characteristics [13].

Because TDOAs are estimated based on sampled signals, the estimation accuracy is limited by the sampling interval. Moreover, the small distance between the microphones of a HA limits the possible discrete TDOA values. Therefore, subsample TDE is necessary, and we need interpolation methods to tackle this problem [14, 15]. In this paper, we use the cubic spline method [16] to interpolate the microphone signals before computing the GGC.
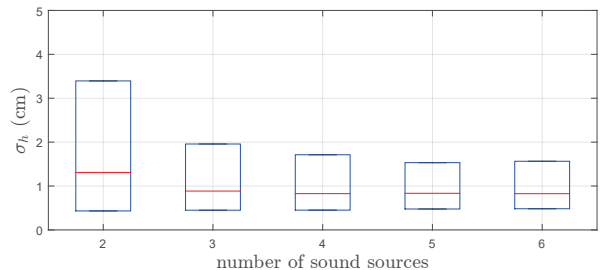
4. SIMULATION RESULTS

4.1. Setup

To evaluate the performance of the proposed algorithm, we consider a free-field situation, i.e. head presence is ignored in the simulations. Moreover, we set $l=1$cm and consider the head diameter, or more precisely, the distance between h_1 and h_2 to be 16 cm. We distribute the sound sources randomly according to a uniform distribution on a disc or a circle (depending on the experiment) around the user. We use the TSP database [17] for generating speech sound sources. The sampling frequency is 48 kHz, the estimation window length is 1024 samples, and we run the simulations for 200 different realizations. The number of query points for interpolation between each two consecutive sample points of the signal is 100.
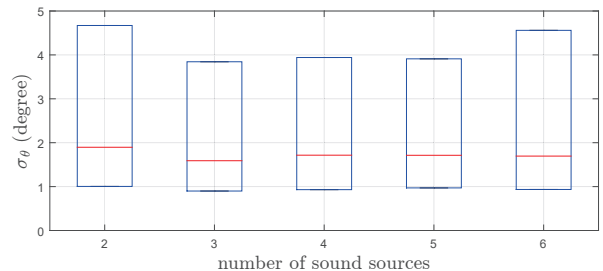
4.2. Performance Measures

To evaluate the estimated location of h_2, we use

$$\sigma_h = \|h_2 - \hat{h}_2\|_2, \quad (15)$$

(a) Performance of $\hat{h}_2$.

(b) Performance of the estimated DOAs.

Fig. 4: The box plot of the performance of the proposed algorithm as a function of number of sound sources. The bottom and top of the boxes are the first and third quartiles, and the bands inside the boxes are the median.

where $\|.\|$ denotes the 2-norm. As another performance metric, we use the mean absolute error of the obtained DOAs:

$$\sigma_\theta = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{|\theta_{1,j}-\tilde{\theta}_{1,j}|+|\theta_{2,j}-\tilde{\theta}_{2,j}|}{2}\right), \quad (16)$$
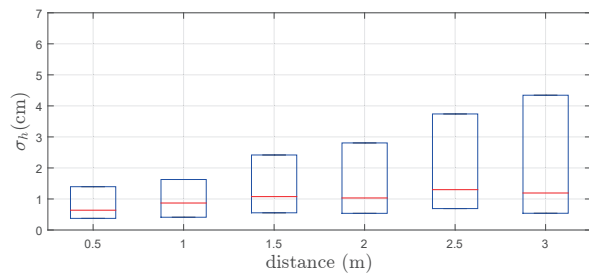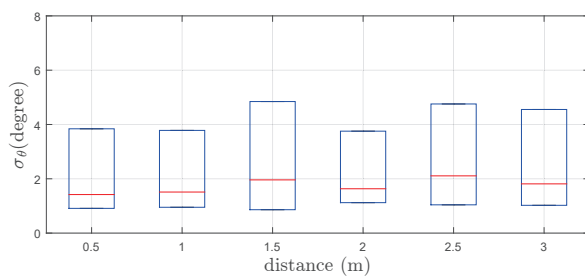
where $\tilde{\theta}_{1,j}$ and $\tilde{\theta}_{2,j}$ obtained from $h_1 = [0,0]^{\mathrm{T}}$, $\hat{h}_{2,b^*}$ and $\hat{s}_{j,b^*}$, and $\theta_{1,j}$ and $\theta_{2,j}$ are the true DOAs of the target signal from s_j to h_1 and h_2, respectively.

To demonstrate the results, we use box plots (Figs. 4 and 5), where the bottom and top of the box are the first and third quartiles, and the band inside the box is the median.

4.3. Results and Discussion

The effect of the number of sound sources on the proposed algorithm has been shown in a box plot in Fig. 4. As can be seen, increasing the number of sound sources from two to three would improve the estimation performance. However, increasing the number of the sound sources to more than three does not offer any advantages because the fundamental subsample error of the TDOA estimation cannot be overcome by increasing the number of the sound sources. Overall, the estimated medians of σ_h and σ_θ are around 1 cm and 2 degree, respectively. It should be mentioned that $d_j \in [0.5, 1.5]$ in these simulations.

Fig. 5 shows the box plot of the proposed algorithm as a function of d_j. We distribute three sound sources randomly on a circle centered at the user's head for different distances. Generally,

(a) Performance of $\hat{h}_2$.

(b) Performance of the estimated DOAs.

Fig. 5: The box plot of the performance of the proposed algorithm as a function of the distance of the sound sources from the user.

increasing the distance degrades the performance because the distance increment would put the sound sources in a far-field situation regarding both HAs—we modeled the problem in a way that the sound sources are in far-field with respect to each HA individually, not both HAs. Overall, as before, the estimated medians of σ_h and σ_θ are around 1 cm and 2 degree, respectively.

5. CONCLUSION AND FUTURE WORK

In this paper, we studied the microphone array calibration problem for binaural hearing aid systems. The proposed algorithm is based on the estimated TDOAs of the target signals received by hearing aid microphones. We used a far-field assumption to model the problem as a linear system, and we proposed a least squares estimator to estimate the locations. As future work, we plan to study the proposed algorithm under more realistic situations by considering presence of the head, microphone noise and reverberation.

REFERENCES

[1] M. Farmani et al., "Informed TDoA-based direction of arrival estimation for hearing aid applications," in *Proc. of IEEE GlobalSIP*, December 2015.

[2] M. Farmani et al., "Informed direction of arrival estimation using a spherical-head model for hearing aid applications," in *Proc. of IEEE ICASSP*, March 2016.

[3] M. Chen and others., "Energy-based position estimation of microphones and speakers for ad hoc microphone arrays," in *Proc. of IEEE WASPAA*, 2007, pp. 22–25.

[4] R. Heusdens and N. D. Gaubitch, "Time-delay estimation for TOA-based localization of multiple sensors," in *Proc. of IEEE ICASSP*, 2014, pp. 609–613.

[5] N. D. Gaubitch et al., "Auto-localization in ad-hoc microphone arrays," in *Proc. of IEEE ICASSP*, May 2013, pp. 106–110.

[6] S. Zhayida et al., "TOA-based self-calibration of dual-microphone array," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 791–801, Aug 2015.

[7] R. Kaune, "Accuracy studies for TDOA and TOA localization," in *Proc. of FUSION*, July 2012, pp. 408–415.

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

[9] S. Thrun, "Affine structure from sound," in *Adv. Neural Inf. Proc. Syst*, 2005, pp. 1353–1360.

[10] M. Crocco et al., "A bilinear approach to the position self-calibration of multiple sensors," *IEEE Trans. Signal Processing*, vol. 60, no. 2, pp. 660–673, Feb 2012.

[11] Y. Kuang et al., "A complete characterization and solution to the microphone position self-calibration problem," in *Proc. of IEEE ICASSP*, May 2013, pp. 3875–3879.

[12] M. Pollefeys and D. Nister, "Direct computation of sound and microphone locations from time-difference-of-arrival data," in *Proc. of IEEE ICASSP*, March 2008, pp. 2445–2448.

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

[14] F. Viola and W. F. Walker, "A spline-based algorithm for continuous time-delay estimation using sampled data," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 52, no. 1, pp. 80–93, Jan 2005.

[15] X. Lai and H. Torp, "Interpolation methods for time-delay estimation using cross-correlation method for blood velocity measurement," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 46, no. 2, pp. 277–290, March 1999.

[16] C. B. Moler, *Numerical Computing with MATLAB: Revised Reprint*, SIAM e-books. Society for Industrial and Applied Mathematics, 2008.

[17] P. Kabal, "TSP speech database," Tech. Rep., Department of Electrical and Computer Engineering, McGill University, 2002.