Self-rotation angle estimation of circular microphone array based on sound field interpolation

Guansan Lian, Yukoh Wakabayashi, Taishi Nakashima, Nobutaka Ono
*Tokyo Metropolitan University, Japan
E-mail: {lian-guansan@ed., wakayuko@, nakashima-taishi@ed., onono@}tmu.ac.jp

***Abstract*—In this paper, we present a novel method for estimating the rotation angle of a circular microphone array (CMA) when multiple sound sources do not move but the CMA rotates instantly. We assume that the multichannel observation of the CMA at the reference position follows the multidimensional complex-valued Gaussian and model it. We utilize our previous research results on the interpolation of the sound field on the circle's circumference. Then, using the spatial modeling and sound field interpolation, we estimate the rotation angle of the CMA by the maximum likelihood method. We conducted numerical simulations that indicated the estimation accuracy of our proposed method to be within approximately five degrees and the algorithm to be feasible online.**

I. Introduction

The position estimation of sensors such as microphones and sound sources such as speakers in a room is an important technique for perceiving the spatial information of the room. An approach of estimating the target's spatial location, direction, or speaker's head orientation using only acoustic signals is inexpensive and practical. For example, if the self-position information of a smartphone can be known from acoustic signals observed by the microphones attached to it, it will be useful for improving the accuracy of surrounding source localization. Also, the use of direction information enables the users to be provided with user-specialized signal processing. Specifically, the user's head orientation information can be used to estimate the user's attention on sound source in a room [1] and provide the user with an acoustic signal suitable for the user, such as one that emphasizes the sound source of the user's their preference. This can also be applied to hearing aids and situations in which a microphone array is attached to a robot or human head. Moreover, some research, such as [2], requires the angle of a microphone array for array signal processing. In that study, the use of the sound field interpolation method for following the change in the acoustic transfer system owing to the microphone array's rotation was presented.

There are various research groups actively studying the self-position estimation of sensors or microphone arrays. Research on self-position estimation using a single microphone array has been conducted. In the early work of self-localization [3], a maximum likelihood estimator for the correct position and orientation of the array was derived using the proposed specialized function based on the microphone array integration technique. In [4], to estimate the self-position of the mobile device, an algorithm using acoustic signals coming from the surroundings was proposed. Also, in some research on self-position estimation, distributed multiple microphones were used as an ad hoc array. Hennecke and Fink [5] proposed an method of estimating the spatial position of multiple distributed smartphones by maximum likelihood estimation using the arrival time measurement of short pulses, considering an application to a telephone conferencing system. Using common off-the-shelf devices such as mobile phones, Parviainen *et al.* [6] presented a passive acoustic self-localization and synchronization system. Wang *et al.* [7] proposed a two-stage algorithm using the Gauss–Newton low-rank approximation method for estimating the sensor and source joint localization without prior knowledge, such as the sensor and source positions. Furthermore, in research on source localization, Felsheim *et al.* [8] estimated the head orientation of a single speaker using a small-scale neural network and the speech observed by multiple microphone arrays distributed in a room.

In this study, we focus on the rotation movement of an equally spaced circular microphone array (CMA) as a movement of sensors. We suppose a situation where a robot or user wears a CMA on its head. Then, the CMA rotates when it turns its head. We consider a problem to estimate the rotation angle using acoustic signals. In this method, we utilize the previous work by Wakabayashi *et al.* [2], in which the sound field on a circumference that changes by the rotational movement of the CMA was interpolated. We can calculate the rotation angle of the CMA by estimating the observation signal at the reference position by the sound field interpolation method using the acoustic signal recorded in the array after rotation. Estimating the rotation angle allows the head direction estimation of a robot or a human attached with a CMA, which in turn, enables user-specialized signal processing. In most of the previous studies on position estimation, as described above, source location information and the time differences between the microphone pair were used, whereas our method does not explicitly handle such information but expresses them by probabilistic spatial modeling and utilizes the change in sound field to estimate the rotation angle.

II. Previous work: Sound field interpolation using CMA

In this section, we explain the previous research [2]. Although in the previous work the authors proposed a sound field interpolation method and its application to beamforming, we especially focus on the former topic in this paper. To

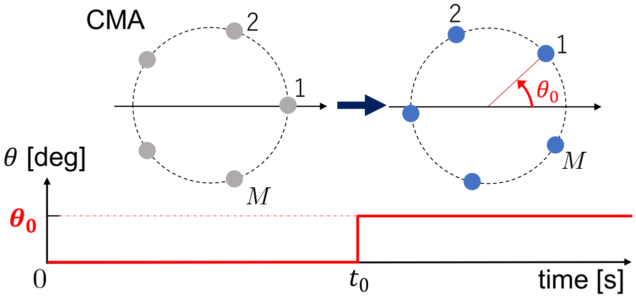

Fig. 1: Problem setting in this study: fixed sound sources are observed by a CMA, and the CMA rotates θ_0 at time t_0 instantly. Estimate θ_0 from observations in this situation.

interpolate the sound field, the authors use the spatial periodicity of the sound field observed using an equally spaced CMA. The authors considered a continuous sound field on the circumference of the circle, $z(\phi)$ at an angle of $\phi \in [0, 2\pi)$. In this case, they used the fact that when sensing the sound field with the M channel CMA, the mth observed signal z_m is written as

$$z_m = z\left(2\pi\frac{m}{M}\right), \quad m = 0, \dots, M-1. \tag{1}$$

They also utilized the fact that by shifting z_m by δ samples, $z_{m+\delta}$ is equivalent to the sound field observed using an equally spaced CMA rotated by θ $(= 2\pi\delta/M)$ rad. By the shift theorem of the discrete Fourier transform, $z_{m+\delta}$ is defined as

$$z_{m+\delta} = z\left(2\pi\frac{m}{M} + \theta\right). \tag{2}$$

$z_{m+\delta}$ can also be formulated with a linear operation as

$$z_{m+\delta} \equiv \sum_{n=0}^{M-1} z_n u_{m,n}(\theta). \tag{3}$$

At this time, the coefficient $u_{m,n}(\theta)$ is defined as

$$u_{m,n}(\theta) = \begin{cases} \frac{1+(-1)^{n-m}}{M} + \frac{\mathrm{sinc}\left(\frac{L}{2}\right)\cos\left(\frac{M+2}{2M}L\pi\right)}{\mathrm{sinc}\left(\frac{L}{M}\right)}, & (M\,\text{:even}), \\ \frac{1}{M} + \frac{M-1}{M}\frac{\mathrm{sinc}\left(\frac{L(M-1)}{2M}\right)\cos\left(\frac{M+1}{2M}L\pi\right)}{\mathrm{sinc}\left(\frac{L}{M}\right)}, & (M\,\text{:odd}), \end{cases} \tag{4}$$

where $L = n - m - \theta M/2\pi$ and $j = \sqrt{-1}$. Here, let $\mathbf{U}(\theta)$ be the $M \times M$ matrix that has $u_{m,n}(\theta)$ in the (m,n) component, i.e., $\mathbf{U}(\theta) = \left(u_{m,n}(\theta)\right)$, and the observation vector sensed using the equally spaced CMA rotated by θ rad from the reference position be

$$\boldsymbol{z} = \begin{bmatrix} z_\delta & \cdots & z_{M-1+\delta} \end{bmatrix}^\mathsf{T}. \tag{5}$$

Then, the sound field sensed using the CMA rotated by angle θ is formulated with matrix operation as

$$\boldsymbol{z} = \mathbf{U}(\theta)\begin{bmatrix} z_0 & \cdots & z_{M-1} \end{bmatrix}^\mathsf{T}. \tag{6}$$

(6) shows that the observation using the CMA at the reference position can be estimated by the observation using the CMA after rotation and the rotation angle.

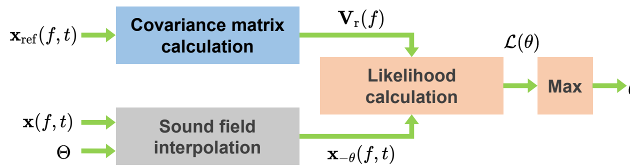

Fig. 2: Block diagram of rotation angle estimation, where Θ is a set of θ, defined as $\Theta = \{\theta : -\pi \le \theta < \pi\}$.

III. Rotation Angle Estimation Based on Maximum Likelihood Estimation of Complex Multidimensional Normal Distribution

A. Rotation and problem setting

We assume that an equally spaced CMA rotates by angle θ_0 from a reference position at time t_0 instantly, as shown in Fig. 1, and the sources are fixed. In this case, we consider estimating the rotation angle by observation using the CMA before and after rotation. We perform the short-time Fourier transform (STFT) to obtain the time–frequency representation of the time domain signal. Here, let $\mathbf{x}(f,t)$ and $\mathbf{x}_{\mathrm{ref}}(f,t)$ be the time–frequency domain observation vectors obtained using the CMA with and without rotation, respectively, where f is the frequency bin index, and t is the time frame index. We define the observation vectors as

$$\mathbf{x}_{\mathrm{ref}}(f,t) = \begin{bmatrix} x_{\mathrm{r}1}(f,t) & \cdots & x_{\mathrm{r}M}(f,t) \end{bmatrix}^\mathsf{T}, \tag{7}$$

$$\mathbf{x}(f,t) = \begin{bmatrix} x_1(f,t) & \cdots & x_M(f,t) \end{bmatrix}^\mathsf{T}, \tag{8}$$

where M is the number of microphones, and $x_{\mathrm{r}m}(f,t)$ and $x_m(f,t)$ are the mth signal observed before and after CMA rotation, respectively.

In summary, we estimate θ_0 using the observation vector obtained using the CMA before rotation, $\mathbf{x}_{\mathrm{ref}}(f,t)$, and that after rotation, $\mathbf{x}(f,t)$.

B. Methodology

In this subsection, we explain this research approach. To estimate the rotation angle, we consider taking advantage of sound field interpolation by the existing method explained in Section II and probabilistic modeling. We propose a framework for rotation angle estimation as shown in Fig. 2. First, we assume that the observation vector follows the complex multidimensional normal distribution and model the observation vector at the reference position, $\mathbf{x}_{\mathrm{ref}}(f,t)$, as a covariance matrix. Next, we estimate $\mathbf{x}_{\mathrm{ref}}(f,t)$ by rotating the current observation vector $\mathbf{x}(f,t)$ at every angle with interpolation. Finally, we calculate the likelihood from the estimated $\mathbf{x}_{\mathrm{ref}}(f,t)$ and the modelled covariance. We obtain the rotation angle by maximizing the likelihood as the angle estimate. We describe the detailed explanation below.

1) Probabilistic modeling of observed signal at reference position: A complex multidimensional normal distribution of zero mean is used as a model of an Mch observation vector $\mathbf{x}$. At this time, the probability density function can be expressed as

$$P(\mathbf{x}, \mathbf{V}) = \frac{1}{\pi^M \det \mathbf{V}} \exp\left(-\mathbf{x}^{\mathrm{H}} \mathbf{V}^{-1} \mathbf{x}\right), \tag{9}$$

where $\mathbf{V}$ is the covariance matrix and $*^{\mathrm{H}}$ is the Hermitian transpose operation. In this study, the covariance matrix of the observed signal at the reference position at frequency f is defined as

$$\mathbf{V}_{\mathrm{r}}(f) = \frac{1}{N_T} \sum_{t=1}^{N_T} \mathbf{x}_{\mathrm{ref}}(f,t) \mathbf{x}_{\mathrm{ref}}^{\mathrm{H}}(f,t), \tag{10}$$

where N_T is the number of frames for calculating covariance, during which we suppose that the CMA is located at the reference position. In this way, the probabilistic modeling of the observation vector at the reference position can be performed.

2) Rotation angle estimation by maximum likelihood estimation: With the method described in Section II, we can express the rotation of an equally spaced CMA as matrix operations. In other words, let $\mathbf{o}(f,t)$ be an observation vector before CMA rotation; then, the observation vector after the CMA rotates by angle θ, $\mathbf{o}_\theta(f,t)$, can be expressed as

$$\mathbf{o}_\theta(f,t) = \mathbf{U}(\theta)\mathbf{o}(f,t). \tag{11}$$

We can estimate the rotation angle from the interpolated observation vector and the probability model of the observed signal at the reference position designed in advance, i.e., the covariance $\mathbf{V}_{\mathrm{r}}(f)$, using the maximum likelihood estimation.

When the CMA rotates by angle θ_0, the observation vector at the reference position, i.e., before CMA rotation, can be estimated by sound field interpolation along an opposite rotation as

$$\hat{\mathbf{x}}_{\mathrm{ref}}(f,t) = \mathbf{x}_{-\theta_0}(f,t) = \mathbf{U}(-\theta_0)\mathbf{x}(f,t). \tag{12}$$

On the basis of this idea, we can design the likelihood with respect to angle. In this study, we employ the sum of Log-likelihood of (9) in the time and frequency directions, using the pre-estimated covariance $\mathbf{V}_{\mathrm{r}}(f)$ and $\mathbf{x}_{-\theta}(f,t)$ $(-\pi \le \theta < \pi)$.

$$\begin{aligned}\mathcal{L}(\theta) &= \sum_{f,t} \ln P\Big(\mathbf{x}_{-\theta}(f,t), \mathbf{V}_{\mathrm{r}}(f)\Big) \\ &= \sum_{f,t} \Big[-\mathbf{x}^{\mathrm{H}}(f,t)\mathbf{U}^{\mathrm{H}}(-\theta)\mathbf{V}_{\mathrm{r}}^{-1}(f)\mathbf{U}(-\theta)\mathbf{x}(f,t) \\ &\qquad - \ln\det \mathbf{V}_{\mathrm{r}}(f) - M \ln \pi\Big]. \end{aligned} \tag{13}$$

It is expected that the angle that maximizes $\mathcal{L}(\theta)$, θ_{max}, will bring the interpolated observation vector $\mathbf{x}_{-\theta}(f,t)$ closest to $\mathbf{x}_{\mathrm{ref}}(f,t)$, resulting in it being equal to the rotation angle θ_0. In this study, we calculate the Log-likelihood value for each angle by an exhaustive search, that is, scanning all angles θ from $-\pi$ to π. Then, we regard angle θ_{max} as the estimated rotation angle $\hat{\theta}_0$.

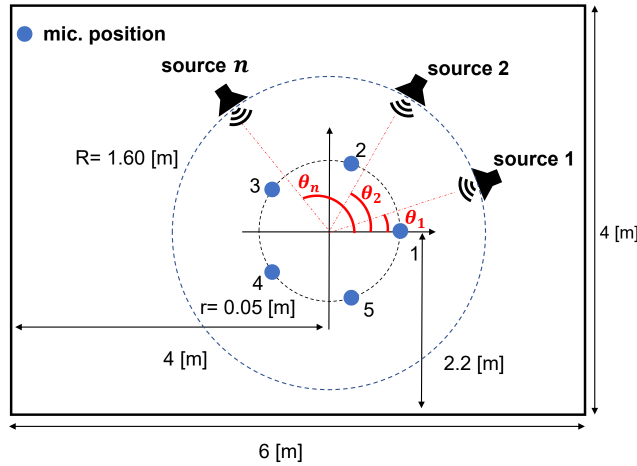

Fig. 3: Numerical simulation setup.

IV. Experimental Evaluation

We conducted three simulation experiments to evaluate the performance of rotation angle estimation by the proposed method. First, we confirmed the Log-likelihood with changing angles and frequencies. Second, we evaluated the estimation accuracy of the rotation angle using a batch estimation scheme. Finally, we evaluated the online estimation performance, that is, the rotation angle was estimated frame by frame.

A. Experimental conditions

Fig. 3 shows the arrangement of the equally spaced CMA and sources. We placed an equally spaced five-channel CMA with a radius of $0.05\,\mathrm{m}$ horizontally in a $6 \times 4 \times 3.5\,\mathrm{m}$ room. We placed the sources along the circumference of a circle with a radius of $1.6\,\mathrm{m}$. We set the reference position of the CMA such that the first channel microphone was in a positive direction on the horizontal axis. We used ten samples (five female and five male voices) with the sampling rate of $16\,\mathrm{kHz}$ from the CMU ARCTIC database [9] as sources. We positioned the sources as shown in Fig. 3 and expressed these patterns as $[\theta_1, \cdots, \theta_n]$, which means that the nth source is located at angle θ_n. We performed the simulation using the Python package *pyroomacoustics* [10]. The reverberation time RT60 was approximately $100\,\mathrm{ms}$. For analysis, we conducted the STFT using a 1/2-shifted Hann window with a length of $64\,\mathrm{ms}$. The length of the source samples was about $20\,\mathrm{s}$. We rotated the CMA instantly by different degrees at time $10\,\mathrm{s}$. In this experiment, to simulate such a situation, we concatenated the two different multichannel signals observed by the CMA with and without rotation. We used the signals before rotation to design the probability model and the signals after rotation to estimate the rotation angle, where we used the first 200 frames (about $6\,\mathrm{s}$) of signals before rotation to calculate the covariance matrix, that is, N_T in (10) was set to 200.

B. Log-likelihood of the rotation angle estimation

To confirm the Log-likelihood values, we show the Log-likelihood values at different angles and frequencies. Here,

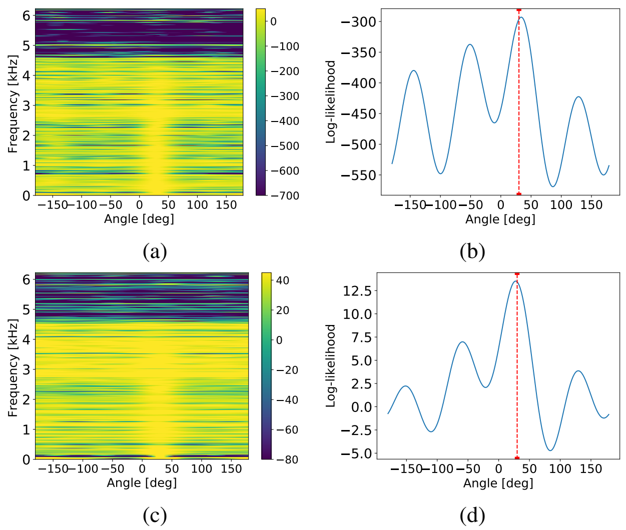

Fig. 4: Examples of Log-likelihood: (a), (c) frequency-wise Log-likelihood and (b), (d) frequency-averaged Log-likelihood, where the vertical red line indicates true rotation angle. The top two are for one source located at [0°], and the bottom two are for two sources located at [0°, 100°].

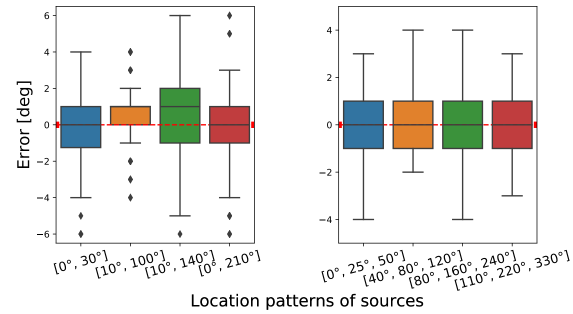

Fig. 5: Boxplot of angle estimation error when the number of sources is two (left) or three (right) sources, where the number of data for the plot is 432 (= 4 source locations $\times$ 4 angles $\times$ 27 source combinations) or 176 (= 4 source locations $\times$ 4 angles $\times$ 11 source combinations), respectively.

the rotation angle is 30°. Figs.4 (a) and (c) show the Log-likelihood in the angle–frequency domain when one source and two sources are respectively placed at 0° and [0°, 100°]. The yellow area means that the Log-likelihood is high. The graph shows that the yellow area is concentrated at the area of the true rotation angle 30° in the frequency domain below about 4 kHz, but not in the high-frequency domain. We consider that the proposed method can estimate the rotation angle accurately in the low-frequency domain but not in the high-frequency domain. This is because the sound field interpolation in the high-frequency domain is inaccurate, as described in [2]. Figs. 4(b) and (d) show the frequency-averaged Log-likelihood values. We can see that the highest peaks of the Log-likelihood graphs point to the true rotation angles, as expected. From these figures, we know that the Log-likelihood values were distributed in the frequency–angle domain correctly. Also, from the findings in Figs. 4(a) and (c), to improve the estimation performance, we used the Log-likelihood values up to 3.2 kHz to estimate the rotation angle in the following experiment as described in Sections IV-C and IV-D.

C. Evaluation of batch rotation angle estimation

In this evaluation, we estimated one rotation angle using all frames of the signal after rotation; this is called batch estimation. We calculated the Log-likelihoods averaged over all frames (defined as t in (13)) in the observed signal after rotation (from 10 s to 20 s). We set four patterns of the source location and four rotation angles of $-30°$, $60°$, $90°$ and $120°$. We randomly selected two and three sources from the ten source samples. To evaluate the estimation performance, we used boxplots. Fig. 5 illustrates that the estimation performance varies with changing source location pattern and the number of sources. We suppose that increasing the number of sources will reduce the estimation performance by our proposed method because the covariance matrix in the probabilistic model involves the time differences between every microphone pair. When the CMA receives source signals from various directions, the information about time differences will be lost, making it difficult to model the observation. However, Fig. 5 shows that the performance for three sources is almost the same as that for two sources, and the error is within 5°. We consider that the proposed method worked well in batch estimation when the number of sources was two or three. In future work, we will confirm the performance when using many more sound sources.

D. Evaluation of online rotation angle estimation

In the previous experiment described in Section IV-C, we used the observed signals after rotation (from 10 s to 20 s) to estimate one rotation angle. In comparison with the batch estimation, we used the Log-likelihood averaged for five frames comprising the current frame and four past frames, and estimated the rotation angle frame by frame in this evaluation. We estimated the rotation angle from 6 s to 20 s, that is, after modeling. We selected three patterns of the source location for this experiment. Fig. 6 shows the results of online performance. The estimation performance worsened for about 1 s just after the rotation time and then improved. We can see that the estimation performance is not as good as the results shown in Fig. 5, but is sufficient to apply our proposed method to online processing.

V. Conclusion

In this study, we proposed a rotation angle estimation method by observation using a CMA before and after rotation in a situation where the CMA rotates by some angle from a specific reference position. We modeled the observation vector at the reference position using the multidimensional complex-valued normal distribution. On the basis of the model, we

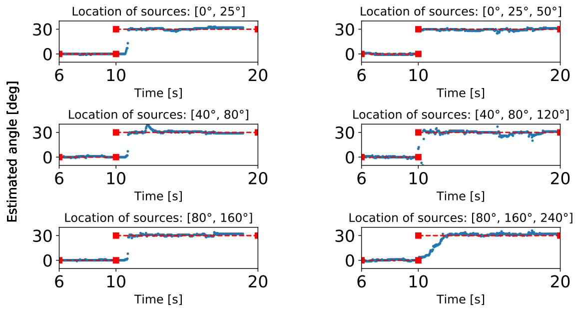

Fig. 6: Online estimation result when the number of sources is two or three, where the CMA rotates from the reference position (0°) to 30° at time 10 s. Red lines show true angles.

designed a maximum likelihood problem of the rotation angle using sound field interpolation and estimated the rotation angle by an exhaustive search. Experimental results showed that the proposed method could be used to estimate the rotation angle within a small error range by both the batch and the online processing. Future work includes improving the estimation accuracy of the rotation angle, introducing a faster algorithm of angle search without an exhaustive search, and confirming the performance in a highly reverberant real environment.

Acknowledgment

This work was supported by JST CREST Grant Number JPMJCR19A3, Japan.

References

[1] C. Segura, C. Canton-Ferrer, A. Abad, J. Casas, and J. Hernando, "Multimodal head orientation towards attention tracking in smartrooms," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 2, 2007, pp. II–681–II–684.

[2] Y. Wakabayashi, K. Yamaoka, and N. Ono, "Rotation-robust beamforming based on sound field interpolation with regularly circular microphone array," in *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 301–305.

[3] P. Aarabi, "Self-localizing dynamic microphone arrays," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 32, no. 4, pp. 474–484, 2002.

[4] D. B. Haddad, W. A. Martins, M. d. V. M. da Costa, L. W. P. Biscainho, L. O. Nunes, and B. Lee, "Robust acoustic self-localization of mobile devices," *IEEE Transactions on Mobile Computing*, vol. 15, no. 4, pp. 982–995, 2016.

[5] M. H. Hennecke and G. A. Fink, "Towards acoustic self-localization of ad hoc smartphone arrays," in *2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays*, 2011, pp. 127–132.

[6] M. Parviainen, P. Pertilä, and M. S. Hämäläinen, "Self-localization of wireless acoustic sensors in meeting rooms," in *2014 4th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, 2014, pp. 152–156.

[7] L. Wang, T. Hon, J. D. Reiss, and A. Cavallaro, "Self-localization of ad-hoc arrays using time difference of arrivals," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 1018–1033, 2016.

[8] R. C. Felsheim, A. Brendel, P. A. Naylor, and W. Kellermann, "Head orientation estimation from multiple microphone arrays," in *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 491–495.

[9] J. Kominek and A. Black, "The CMU Arctic speech databases," *SSW5-2004*, 01 2004.

[10] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 351–355.