



**University of  
Sunderland**

Cao, Faxian, Cheng, Yongqiang, Khan, Adil Mehmood, Yang, Zhijing and Guo, Zhongwai (2024) Are Microphone Signals Alone Sufficient for TOA Self-Positioning? In: 2024 29th International Conference on Automation and Computing (ICAC). IEEE, pp. 1-6. ISBN 979-8-3503-6088-2

Downloaded from: <http://sure.sunderland.ac.uk/id/eprint/18435/>

#### **Usage guidelines**

Please refer to the usage guidelines at <http://sure.sunderland.ac.uk/policies.html> or alternatively contact [sure@sunderland.ac.uk](mailto:sure@sunderland.ac.uk).



# Are Microphone Signals Alone Sufficient for TOA Self-Positioning?

1<sup>st</sup> Faxian Cao  
School of Computer Science  
University of Hull  
Hull, U.K.  
faxian.cao-2022@hull.ac.uk

2<sup>nd</sup> Yongqiang Cheng  
School of Computer Science  
University of Sunderland  
Sunderland, U.K.  
yongqiang.cheng@sunderland.ac.uk

3<sup>rd</sup> Adil Mehmood Khan  
School of Computer Science  
University of Hull  
Hull, U.K.  
a.m.khan@hull.ac.uk

4<sup>th</sup> Zhijing Yang  
School of Information Engineering  
Guangdong University of Technology  
Guangzhou, China  
yzhj@gdut.cn

5<sup>th</sup> Zhongwai Guo  
Shenfa Bearings Ltd.  
Zhuji, Zhejiang, China  
gzwai@163.com

**Abstract**—In an era where asynchronous environments pose challenges to traditional self-positioning methods, we propose a new transformation to the existing paradigm. Traditionally, time of arrival (TOA) measurements require both microphone and source signals, limiting their applicability in environments with unknown emission time of human voices or sources and unknown recording start time of independent microphones. To address this issue, our research pioneers a mapping function capable of transforming both TOA and time difference of arrival (TDOA) formulas, demonstrating, for the first time, that they can be identical to one another. This implies that microphone signals alone are sufficient for self-positioning without the need for source signals waveform, a groundbreaking advancement in the field that carries the potential to revolutionize self-positioning techniques, expanding their applicability in challenging environments. Supported by a robust mathematical proof and compelling experimental results, this research represents a timely and significant contribution to the current discourse in signal, and audio processing.

**Index Terms**—Time of arrival, time difference of arrival, self-positioning, mapping function

## I. INTRODUCTION

**T**HE ability to accurately localize distributed microphones and sound sources is a fundamental requirement in various acoustic tasks, including noise reduction, source signal enhancement, and separation [1]–[3]. This is conventionally achieved through the utilization of time of arrival (TOA) or time difference of arrival (TDOA) measurements [4]. However, these techniques have significant limitations, particularly in asynchronous environments where the unknown timing information (UTIm) of signal emission and recording are unknown in advance [5], resulting unknown microphones start time and sources emission time contained in TOA while the unknown time offset of a pair of microphones contained in TDOA.

In scenario where the waveform of the source signals is available, including information on frequency, amplitude, and

duration, TOA measurements can be estimated through cross-correlation methods [6]. This has led to the development of various self-positioning methodologies such as probabilistic generative models [7], maximal likelihood estimation [8], Gram matrix and semi-definite relaxation [5], and techniques utilizing the low-rank property (LRP) [9] with alternating minimization method [10], [11], and structure total least square [12], [13].

Alternatively, when source signals waveform is hard to obtain, self-positioning techniques pivot towards TDOA measurements, which can be estimated with audio signals from a pair of microphones [6]. This shift has led to a plethora of methodologies, such as maximal likelihood estimation [8], [17], auxiliary function method [14], LRP with nuclear truncation minimization [15], [16], and distributed damped Newton optimization [4], [18].

Yet, amidst these developments, a significant and pressing question has lingered: Can microphone signals alone be sufficient for self-positioning, thereby negating the need for source signals? In details, on one side, after the UTIm in TOA/TDOA is estimated, we can obtain the range measurements between microphones and sources for self-localization with TOA measurements while the range difference of a pair of microphones with respect to corresponding sources can be used for self-localization with TDOA measurements. However, range measurements between microphones and sources contain richer information than range difference. For example, with both range measurements and AOA measurements [19], the location of both microphones and sources can be obtained directly [20], alternatively, with range measurements only, closed form solutions [21] can also be derived directly for localization of microphones and sources. On the other side, TDOA measurements need microphones signals only while both microphones and sources are necessary to be used for obtaining TOA measurements. Therefore, with microphones signals only, how to prove TOA can be identical to TDOA

This work was supported by China Scholarship Council (*Corresponding author: Yongqiang Cheng*)

is the key problem. The answer to this question carries profound implications for the field, as it can streamline and make the self-positioning process more efficient and adaptable. Moreover, it presents a timely advancement, given the growing complexity of audio environments and the increasing need for flexible and efficient localization methods.

This research takes a groundbreaking step towards answering this crucial question. We introduce an innovative mapping function that transforms both TOA and TDOA formulas to an identical representation/form. In other words, our findings illustrate that their transformations can mirror one another perfectly, confirming that relying solely on microphone signals is sufficient for self-positioning tasks. Therefore, our novel approach unveils, for the first time, the exact relationship between TOA and TDOA measurements, challenging the long-standing assumption that TOA necessitates both microphone signals and the source signal waveform.

This revolutionary insight doesn't merely simplify the self-positioning process by eliminating the need for additional information from source signals. It also broadens its applicability, as properties initially designed for TOA-based localization, such as rank 3 [3] and rank 5 [22], can now be applied to TDOA-based localization. In essence, our work represents a significant, novel, and timely contribution, with the potential to dramatically reshape self-positioning techniques in asynchronous environments and catalyze further advancements in signal, and audio processing.

## II. PROBLEM FORMULATION

Consider a setup where we have  $M$  asynchronous microphones and  $N$  asynchronous sound sources, located at  $R = [r_1, \dots, r_M]_{3 \times M}$  and  $S = [s_1, \dots, s_N]_{3 \times N}$ , respectively, with 3 denoting three dimensions. After sources have emitted the audio signals and microphones have received the corresponding signals, we can encounter two possible scenarios.

In the scenario where the waveform from the source signals can be acquired, by defining the unknown recording start time of  $i^{th}$  microphone as  $\delta_i$  and unknown emission time of  $j^{th}$  source as  $\eta_j$  as well as the speed of sound as  $c$ , the known TOA ( $t_{i,j}$ ) between  $i^{th}$  microphone and  $j^{th}$  source can be calculated as [5]

$$t_{i,j} = \frac{\|r_i - s_j\|}{c} + \eta_j - \delta_i, \quad (1)$$

where  $i = 1, \dots, M$  and  $j = 1, \dots, N$ , and  $\|\bullet\|$  is the  $l_2$  norm. In addition, without loss of generality, the location of the first source can be set as  $s_1 = [0, 0, 0]^T$  because of the invariance of translation and rotation as well as reflection regarding the geometry of microphones and sources [3].

The second scenario arises when it is challenging to obtain the waveform from the source signals. Here, we define the known TDOA ( $\tau_{i,j}$ ) of  $j^{th}$  source between  $i^{th}$  microphone and  $1^{st}$  microphone as [5]

$$\tau_{i,j} = t_{i,j} - t_{1,j} = \frac{\|r_i - s_j\|}{c} - \frac{\|r_1 - s_j\|}{c} + \delta_1 - \delta_i. \quad (2)$$

Upon inspection of Eq. (2), it can be observed that after the source  $j$  emits the audio signal and both  $i^{th}$  and  $1^{st}$  microphones receive the corresponding signal, the  $i^{th}$  microphone signal contains information about the start time of  $i^{th}$  microphone  $\delta_i$ , the emitted time of  $j^{th}$  source  $\eta_j$ , as well as the time difference in signal propagation from the  $j^{th}$  source to the  $i^{th}$  microphone. Similarly, the signal at the  $1^{st}$  microphone contains information about the start time of the  $1^{st}$  microphone  $\delta_1$ , the emitted time of  $j^{th}$  source  $\eta_j$ , and the time difference in signal propagation from the  $j^{th}$  source to the  $1^{st}$  microphone. Thus, employing the generalized cross-correlation with phase transform [23] method, TDOA ( $\tau_{i,j}$ ) can be estimated using only the audio signals from the  $j^{th}$  and  $1^{st}$  microphones, demonstrating the independence of TDOA from the source signal. Besides, according to the definition of TDOA, it measures the time difference between a pair of microphones when they receive the corresponding source signal, therefore, the TDOA ( $\tau_{i,j}$ ) of  $j^{th}$  source in Eq. (2) can also be measured by the  $i^{th}$  microphone signal and any other one of remaining microphone signals. Fig. (II) illustrates these two cases for TOA formula in Eq. (1) and TDOA formula in Eq. (2).

Let's denote  $\check{\delta}_i = \delta_i - \delta_1$  and  $\check{\eta}_j = -\frac{\|r_1 - s_j\|}{c}$ , then TDOA formula in Eq. (2) can be re-written as [5]

$$\tau_{i,j} = \frac{\|r_i - s_j\|}{c} + \check{\eta}_j - \check{\delta}_i. \quad (3)$$

Interestingly, this equation shares the same structural form as the TOA formula in Eq. (1). However, the exact relationships between TOA formula in Eq. (1) and TDOA formula in Eq. (2) remain elusive. No existing works have demonstrated this relationship so far, and as a result, the sufficiency of utilizing only microphone signals for self-positioning is still unknown. Our research objective, therefore, is to investigate the feasibility of utilizing the microphone signals alone for self-positioning when the waveform of source signals is unavailable. The results of our study have the potential to challenge the long-standing assumption that the acquisition of source signal waveform is a necessity for TOA-based self-positioning. This can lead to an expansion of self-positioning techniques, enhancing their utility in challenging environments.

## III. MAPPING FUNCTION FOR TOA AND TDOA FORMULAS

In this section, a novel mapping function is derived for TOA formula in Eq. (1) and TDOA formula in Eq. (2). We first present the novel mapping function in Subsection A, then the proof of the proposed mapping function is shown in Subsection B.

### A. Mapping Function $f(\bullet)$

TOA measurements are unavailable when the waveform of source signals is missing, and only TDOA measurements can be used for localization once this situation happens. Since there are no existing works in the state-of-the-art investigate

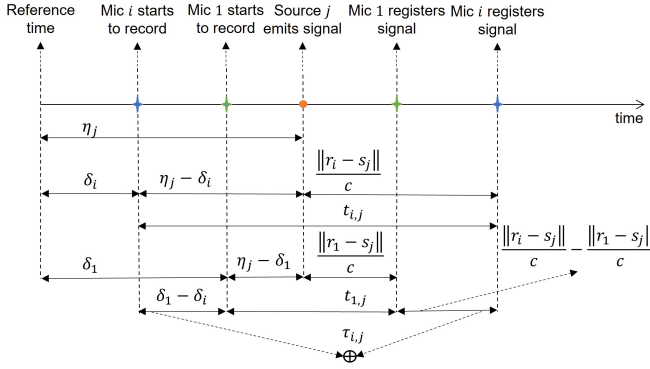


Fig. 1. The illustrations for TOA and TDOA formulas with asynchronous microphones and sources ( $r_1$  and  $r_i$ : locations of 1<sup>st</sup> and  $i^{th}$  microphones, respectively;  $s_j$ : location of  $j^{th}$  source;  $t_{i,j}$ : TOA measurement between  $i^{th}$  microphone and  $j^{th}$  source;  $\tau_{i,j}$ : TDOA measurement of  $j^{th}$  source between 1<sup>st</sup> microphone and  $i^{th}$  microphone;  $\delta_1$  and  $\delta_i$ : start time of 1<sup>st</sup> and  $i^{th}$  microphones, respectively;  $\eta_j$ : emission time of  $j^{th}$  source;  $c$ : speed of sound;  $\|\bullet\|$ :  $l_2$  norm).

the relationships between TOA and TDOA measurements, here, we present a novel general form of mapping function to show the sufficiency of using microphone signals alone for both TOA and TDOA-based self-positioning. The proposed general form of mapping function,  $f(\bullet)$ , for TOA formula in Eq. (1) and TDOA formula in Eq. (2), is defined as

$$f(t_{i,j}) = t_{i,j} - t_{i,1} - \frac{\sum_{i=1}^M (t_{i,j} - t_{i,1})}{M}, \quad (4)$$

and

$$f(\tau_{i,j}) = \tau_{i,j} - \tau_{i,1} - \frac{\sum_{i=1}^M (\tau_{i,j} - \tau_{i,1})}{M}, \quad (5)$$

respectively, then by applying the mapping function,  $f(\bullet)$ , to TOA formula in Eq. (1) and TDOA formula in Eq. (2) and defining two variables

$$\begin{cases} \hat{\delta}_i = \frac{\|r_i - s_1\|}{c} \\ \hat{\eta}_j = \frac{\sum_{i=1}^M (\|r_i - s_1\| - \|r_i - s_j\|)}{cM} \end{cases}, \quad (6)$$

we state that

$$f(t_{i,j}) = f(\tau_{i,j}) = \frac{\|r_i - s_j\|}{c} - \hat{\delta}_i + \hat{\eta}_j, \quad (7)$$

where  $i = 1, \dots, M$  and  $j = 1, \dots, N$ . From the statement in Eq. (7), we can see that once this relationship is proved, this mapping function indicates the same structure as TOA formula in Eq. (1), showing the location of both microphones and sources can be obtained with  $f(\bullet)$  by utilizing the same methods that are designed for TOA-based self-positioning. More importantly, the sufficiency of utilizing microphone signals alone can be revealed for self-positioning, providing the potential to challenge the long-standing assumption that TOA necessitates both microphone signals and the waveform of source signals for self-positioning. Thus, the process of self-positioning can be more adaptable and efficient, and the abilities of self-positioning techniques can be expanded in challenging environments.

## B. Proof for mapping function $f(\bullet)$

We first derive the transformation of TOA formula in Eq. (4), then the derivation of transformation of TDOA formula in Eq. (5) is displayed. Finally, we validate the statement in Eq. (7) by comparing the transformation of TOA formula in Eq. (4) with the transformation of TDOA formula in Eq. (5).

1) *Transformation of TOA formula:* From TOA formula in Eq. (1), we can have

$$t_{i,1} = \frac{\|r_i - s_1\|}{c} + \eta_1 - \delta_i, \quad (8)$$

then with Eqs. (1) and (8), the difference between  $t_{i,j}$  and  $t_{i,1}$  can be written as

$$t_{i,j} - t_{i,1} = \frac{\|r_i - s_j\|}{c} - \frac{\|r_i - s_1\|}{c} + \eta_j - \eta_1. \quad (9)$$

From Eq. (9), we can see the mean value for  $t_{i,j} - t_{i,1}$  with respect to the index  $i$  is

$$\frac{\sum_{i=1}^M (t_{i,j} - t_{i,1})}{M} = \frac{\sum_{i=1}^M (\|r_i - s_j\| - \|r_i - s_1\|)}{cM} + \eta_j - \eta_1. \quad (10)$$

Finally, with Eqs. (9) and (10), the mapping function for TOA formula in Eq. (4),  $f(t_{i,j})$ , can be formulated as

$$\begin{aligned} f(t_{i,j}) &= t_{i,j} - t_{i,1} - \frac{\sum_{i=1}^M (t_{i,j} - t_{i,1})}{M} \\ &= \frac{\|r_i - s_j\|}{c} - \frac{\|r_i - s_1\|}{c} + \frac{\sum_{i=1}^M (\|r_i - s_1\| - \|r_i - s_j\|)}{cM}, \end{aligned} \quad (11)$$

where  $i = 1, \dots, M$  and  $j = 1, \dots, N$ .

2) *Transformation of TDOA formula:* From TDOA formula in Eq. (2), we can display the difference between  $\tau_{i,j}$  and  $\tau_{i,1}$  as

$$\begin{aligned} \tau_{i,j} - \tau_{i,1} &= \frac{\|r_i - s_j\|}{c} - \frac{\|r_1 - s_j\|}{c} \\ &\quad - \frac{\|r_i - s_1\|}{c} + \frac{\|r_1 - s_1\|}{c}, \end{aligned} \quad (12)$$

then from Eq. (12), the mean value for  $\tau_{i,j} - \tau_{i,1}$  with respect to the index  $i$  can be written as

$$\begin{aligned} \frac{\sum_{i=1}^M (\tau_{i,j} - \tau_{i,1})}{M} &= \frac{\sum_{i=1}^M (\|r_i - s_j\| - \|r_i - s_1\|)}{cM} \\ &\quad - \frac{\|r_1 - s_j\|}{c} + \frac{\|r_1 - s_1\|}{c}. \end{aligned} \quad (13)$$

Finally, with Eqs. (12) and (13), the mapping function for TDOA formula in Eq. (5),  $f(\tau_{i,j})$ , can be formulated as

$$\begin{aligned} f(\tau_{i,j}) &= \tau_{i,j} - \tau_{i,1} - \frac{\sum_{i=1}^M (\tau_{i,j} - \tau_{i,1})}{M} \\ &= \frac{\|r_i - s_j\|}{c} - \frac{\|r_i - s_1\|}{c} + \frac{\sum_{i=1}^M (\|r_i - s_1\| - \|r_i - s_j\|)}{cM} \end{aligned} \quad (14)$$

where  $i = 1, \dots, M$  and  $j = 1, \dots, N$ .

3) *Validation of statement:* Based on the definitions of the two variables  $\hat{d}_i$  and  $\hat{r}_j$  in Eq. (6), then with the transformations of both TOA formula in Eq. (11) and TDOA formula in Eq. (14), we can see that Eq. (11) and Eq. (14) are identical to one another, this completes the proof of mapping function  $f(\bullet)$  in Eq. (7).

With the proof of the statement in Eq. (7), we can see that the transformations of TOA and TDOA formulas are identical to one another, revealing the sufficiency of utilizing microphone signals for both TOA and TDOA-based self-positioning, providing the potentials to challenge the long-standing assumption that TOA necessitates both microphone signals and the waveform of source signals for self-positioning. In addition, the statement in Eq. (7) indicates that many properties, such as rank 3 [3] and rank 5 [22], that are used for TOA-based localization can also be used for TDOA-based localization, this makes the tasks of self-positioning more efficient and adaptable. Besides, by eliminating the need for additional information from source signals, a wide range of other applications, such as noise reduction, sources signals enhancement and separation [1]–[3] can also be facilitated since the importance of self-positioning for those applications above.

#### IV. EXPERIMENTAL VALIDATIONS

In this section, experimental results are shown to validate the proposed mapping function. The experimental setups are illustrated in subsection A first, then the evaluation metric is defined and the validations of both the proposed mapping function and the property of the proposed mapping function are shown in subsection B.

##### A. Setups

1) *Simulation data:* All the simulation data is randomly generated by MATLAB with uniform distribution, both the start time of microphones and emission time of sources are in the range of  $[-1, 1]s$ , the locations of microphone and source are distributed in the room with size of  $10 \times 10 \times 3 m^3$  [5] and the speed of sound is set to be  $340 m/s$ . In addition, both the number of microphones  $M$  and the number of sources  $N$  are set to be 20, and the number of configurations is set to be 1000. Thus, there are 400000 data points for simulated data.

2) *Real-Life data:* The real data [24] was collected in an office of size of  $5 \times 3 m^2$ , where most of the furniture inside the office was removed. There are 12 microphones which were fixed, and a chirp was played by a loudspeaker from 65 positions. This real-life data for  $12 \times 65$  TOA matrix can be downloaded at Github<sup>1</sup> [5], [24] and the TDOA matrix is calculated by Eq. (2). For more details of this real-life data, readers can refer to references [5], [24]. Also, both the start time of microphones and the emission time of sources are in the range of  $[-1, 1]s$ . In addition, the number of data points for real-life data is 780 since there are 12 microphones and 65 sources.

<sup>1</sup>This real-life data is available at <https://github.com/swing-research/xtdoa/tree/master/matlab>

##### B. Evaluations and Results

We first show the values of proposed mapping function for the transformations of both TOA and TDOA measurements in both simulation data and real-life data sets, then the statement for the proposed mapping function in Eq. (7) is evaluated by measuring the difference of transformations of TOA and TDOA formulas

$$\Delta f_{i,j} = f(t_{i,j}) - f(\tau_{i,j}), \quad (15)$$

where  $i = 1, \dots, M$  and  $j = 1, \dots, N$ . As can be seen from Eq. (15), once  $\Delta f_{i,j}$  is equal to zero, the values of transformation of both TOA and TDOA formulas are the same as each other, it indicates the validation of the proposed mapping function  $f(\bullet)$ .

Fig. 2 shows the experimental results with both simulated data and real data. From Fig. 2(a), it can be observed that the values of  $f(\bullet)$  for both TOA and TDOA measurements in simulated data are in the range of  $[-0.05, 0.05]s$  while the corresponding values in real data are in the range of  $[-0.02, 0.02]s$ , this is because of the different sizes of the rooms are used for simulation data and real life data, respectively. In addition, from Fig. 2(b), we can also see that the value of  $\Delta f_{i,j}$  is also with a magnitude of  $10^{-16}$  due to the machine calculation accuracy of MATLAB, therefore,  $\Delta f_{i,j} = 0$  in Eq. (15) is validated. This implies that the transformations of TOA formula in Eq. (1) and TDOA formula in Eq. (2) are identical to one another with our proposed mapping function  $f(\bullet)$ , so that the statement for proposed mapping function in Eq. (7) is validated. TOA measurements are obtained with both microphone and source signals while TDOA measurements are obtained with microphone signals only, and from Fig. 2(b), it is obvious that the transformation of TOA and TDOA measurements are the same as each other, therefore, our novel mapping function shows the sufficiency of utilizing microphone signals alone for self-positioning, negating the need of source signals for self-positioning, presenting a timely advancement for tasks of self-positioning.

#### V. CONCLUSION

This paper investigated the sufficiency of using microphone signals alone for self-positioning that has never been investigated in the state-of-the-art. When both the emission time of the source signal and the recording start times of the microphones are unknown, by presenting a novel mapping function that has never been shown in the literature to transform both TOA and TDOA formulas, we demonstrated that the transformations of TOA and TDOA formulas are identical to one another, showing the sufficiency that uses microphone signals alone for self-positioning, enabling the tasks of self-positioning more flexible and adaptable. Therefore, the proposed mapping function can be regarded as a timely advancement for the tasks of self-positioning.

For future works, based on the existing TOA and TDOA-based methods, it would be interesting to apply our novel mapping function to estimate the unknown source emission time and microphone start time as well as the locations of

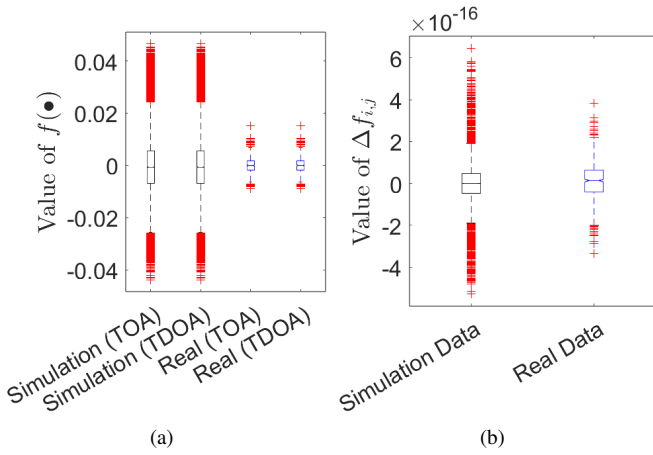


Fig. 2. Validation of proposed mapping function with both simulation and real-life data sets: (a) the value of proposed mapping function  $f(\bullet)$  for transformations of both TOA measurements in Eq. (4) and TDOA measurements in Eq. (5); (b) the value of  $\Delta f_{i,j}$  in Eq. (15).

microphones and sources. Besides, it might also be interesting to apply this mapping function to other applications, such as noise reduction, sources signals enhancement and separation.

## REFERENCES

- [1] X. Dang, Q. Cheng, and H. Zhu, "Indoor multiple sound source localization via multi-dimensional assignment data association," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 12, pp. 1944–1956, 2019.
- [2] Q. Zhang, Z. Chen, and F. Yin, "Distributed marginalized auxiliary particle filter for speaker tracking in distributed microphone networks," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 11, pp. 1921–1934, 2016.
- [3] T. K. Le and N. Ono, "Closed-form and near closed-form solutions for TOA-based joint source and sensor localization," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4751–4766, 2016.
- [4] D. Hu, Z. Chen, F. Yin, "Geometry calibration for acoustic transceiver networks based on network newton distributed optimization," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 1023–1032, 2021.
- [5] D. E. Badawy, V. Larsson, M. Pollefeys, and I. Dokmanic, "Localizing unsynchronized sensors with unknown sources," *IEEE Trans. Signal Process.*, vol. 71, pp. 641–654, 2023.
- [6] X. Wang, and D. Hu, "Distributed self-localization for acoustic transceiver networks," *IEEE Signal Process. Lett.*, 2023.
- [7] R. Biswas and S. Thrun, "A passive approach to microphone network localization," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, pp. 1544–1549, 2004.
- [8] V. C. Raykar, I. V. Kozintsev, and R. Lienhart, "Position calibration of microphones and loudsources in distributed computing platforms," *IEEE Trans. Speech, Audio Process.*, vol. 13, no. 1, pp. 70–83, 2004.
- [9] P. H. Schönemann, "On metric multidimensional unfolding," *Psychometrika*, vol. 35, no. 3, pp. 349–366, 1970.
- [10] N. D. Gaubitch, W. B. Kleijn, and R. Heusdens, "Calibration of distributed sound acquisition systems using toa measurements from a moving acoustic source," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, pp. 7455–7459, 2014.
- [11] N. D. Gaubitch, W. B. Kleijn, and R. Heusdens, "Auto-localization in ad-hoc microphone arrays," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, pp. 106–110, 2013.
- [12] J. Zhang, R. C. Hendriks, and R. Heusdens, "Structured total least squares based internal delay estimation for distributed microphone auto-localization," in *Proc. Int. Workshop Acoustic Signal Enhancement*, pp. 1–5, 2016.
- [13] R. Heusdens and N. Gaubitch, "Time-delay estimation for toa-based localization of multiple microphones," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, pp. 609–613, 2014.
- [14] N. Ono, H. Kohno, N. Ito and S. Sagayama, "Blind alignment of asynchronously recorded signals for distributed microphone array," in *Proc. WASPAA*, pp. 161–164, 2009.
- [15] F. Jiang and Y. Kuang, "Time delay estimation for tdoa self-calibration using truncated nuclear norm regularization," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, pp. 3885–3889, 2013.
- [16] Y. Kuang, and K. Åström, "Stratified microphone network self-calibration from tdoa measurements," in *Proc. EUSIPCO*, pp. 1–5, 2013.
- [17] S. Woźniak, and K. Kowalczyk, "Passive joint localization and synchronization of distributed microphone arrays," *IEEE Signal Process. Lett.*, vol. 26, no. 2, pp. 292–296, 2018.
- [18] D. Hu, Z. Chen, and F. Yin, "Passive Geometry Calibration for Microphone Arrays Based on Distributed Damped Newton Optimization," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 118–131, 2020.
- [19] T. K. Le and K. C. Ho, "Joint source and sensor localization by angles of arrival," *IEEE Trans. Signal Process.*, vol. 68, pp. 6521–6534, 2020.
- [20] T. K. Le and K. C. Ho, "Uncovering source ranges from range differences observed by sensors at unknown positions: Fundamental theory," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2665–2678, 2019.
- [21] T. K. Le and N. Ono, "Closed-form and near closed-form solutions for toa-based joint source and sensor localization," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4751–4766, 2016.
- [22] M. Pollefeys and D. Nister, "Direct computation of sound and microphone locations from time-difference-of-arrival data," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, pp. 2445–2448, 2008.
- [23] M. S. Brandstein, and H. F. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, vol. 1, pp. 375–378, 1997.
- [24] K. Batstone, G. Flood, T. Beleyur, V. Larsson, H. R. Goerlitz, M. Oskarsson, and K. Åström, "Robust self-calibration of constant offset time-difference-of-arrival," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, pp. 4410–4414, 2019.