IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 September 2023, accepted 13 October 2023, date of publication 24 October 2023, date of current version 3 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3327011

RESEARCH ARTICLE

Minimum-Spanning-Tree-Based Time Delay Estimation Robust to Outliers

KOUEI YAMAOKA[1], (Student Member, IEEE),
TAISHI NAKASHIMA[1], (Student Member, IEEE),
YUKOH WAKABAYASHI[1,2], (Member, IEEE), AND NOBUTAKA ONO[1], (Senior Member, IEEE)
[1]Graduate School of Systems Design, Tokyo Metropolitan University, Tokyo 191-0065, Japan
[2]Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi 441-8580, Japan

Corresponding author: Kouei Yamaoka (yamaoka-kouei@ed.tmu.ac.jp)

This work was supported by both Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant 20H00613 and Japan Science and Technology agency (JST) Core Research for Evolutionary Science and Technology (CREST) under Grant JPMJCR19A3.

ABSTRACT In this paper, we present a novel approach to estimating multiple time delays (TDs) in sensor arrays that is robust to outliers of TD measurements. These measurements are typically obtained from the peak of the cross correlation of two sensor signals but may contain outliers due to noise, significantly degrading the performance of downstream applications. To address this issue, we propose an approach to leverage only the best minimum-necessary TD measurements. First, we describe the general properties of TDs and show that the degree of freedom of TDs is the number of sensors minus one for a signal source, which indicates that the full set of TDs is redundant in this case. We then consider selecting nonredundant TDs from all measurements given the necessary and sufficient condition to reconstruct all TDs uniquely. We represent this condition by utilizing the graph theory, and then, formulate an optimization problem to select the optimal nonredundant TDs while satisfying the condition above. We reduce this problem to the problem of finding a minimum spanning tree and propose an efficient algorithm for TD estimation. Experimental evaluation shows that our method successfully eliminates outliers while ensuring that all TDs can be restored.

INDEX TERMS Time delay estimation, generalized cross correlation, graph theory, microphone array, minimum spanning tree.

I. INTRODUCTION

A time delay (TD) or time difference of arrival (TDOA) observed between two sensors is a fundamental spatial cue for many signal processing techniques regarding the spatial position of a signal source [1]. Source localization and direction of arrival (DOA) estimation are essential techniques in audio [2], [3], [4], [5] and other various engineering fields [6], [7], [8], including sonar [9] and radar [10]. Typically, a TD estimate is obtained by the generalized cross correlation (GCC) method [11], [12]. Any improvement in TD estimation (TDE) directly benefits the subsequent applications, and numerous studies on this subject have been carried out [1], [12], [13], [14], [15], [16].

Recently, devices equipped with more than three sensors have become common. In the field of acoustic signal processing, spatially distributed microphone arrays and wireless acoustic sensor networks (WASNs) have been widely and actively studied [17], [18], [19]. TDs are necessary information in resampling and synchronization [20], [21], [22], which is a general topic in WASNs. For these applications, a TDE method that can estimate multiple TDs with high accuracy is expected.

Let us consider TDE on an M channel sensor array. Fig. 1(a) shows an example of a 5-channel sensor array, where the circles and lines indicate the sensors and TDs, respectively. As shown in this figure, there are ${}_MC_2 = M(M-1)/2$ TDs corresponding to every possible pair of sensors (2-combinations of M sensors). A set of these TDs is referred to as the *full* TD set [23], [24]. In practice, individual TDs can be measured by, e.g., the GCC method [11], [12]. Therefore, the full TD set must be contaminated by errors in the GCC method and may contain outliers (huge estimation errors). This problem becomes more severe in,

The associate editor coordinating the review of this manuscript and approving it for publication was Xuebo Zhang.

e.g., noisy environments and rooms with long reverberations. Since these outliers have a tremendous negative impact on downstream applications, in this paper, we study how to avoid using them but estimate the full TD set.

The key idea to achieve the above is redundancy of the full TD set (overdetermined) [23], [24]. Indeed, $M(M-1)/2$ TDs can be represented by the $M-1$ TDs among them. For example, a TD between sensors i and j, τ_{ij} say, can be obtained indirectly as $\tau_{ik}+\tau_{kj}$ for any k. That is, the degree of freedom of the TDs is $M-1$, whereas $M(M-1)/2$ TDs exist. This redundancy can be utilized to improve the accuracy and robustness of the TDE algorithms.

One approach to leveraging the full TD set is computing $M-1$ nonredundant TDs from every member of the set. A classical solution in this sense is using the average of $\tau_{ik}+\tau_{kj}$ for all k $(k=1,\ldots,M)$ to estimate τ_{ij}. This is the least square estimator under the assumption that TD measurements are contaminated by additive Gaussian noises [25], [26]. Moreover, Velasco et al. have proposed novel TDE methods based on a TD (or TDOA) matrix (TDM) [24]. The TDM itself has been considered in [27], and its properties have been further studied in [24]. The TDM is one representation of a full TD set; thus, its measurement might include errors and outliers. Therefore, the denoising methods have also been proposed to deal with this problem [24], [28], [29].

Another approach is to choose a reference sensor and only use nonredundant $M-1$ TDs between the reference sensor and the others [3], [23], [30]. Fig. 1(b) shows an example where sensor 1 is chosen as the reference sensor. In this case, any TDs can be computed from the TDs between the reference sensor and the remaining ones, e.g., $\tau_{34}=\tau_{31}+\tau_{14}$. Utilizing an appropriate reference sensor may help avoid the use of TD measurements with significant errors. Choosing a reference sensor is equivalent to determining the absolute time origin for TDs, which are relative values. This approach is thus meaningful. Indeed, choosing a reference sensor is also a common practice in downstream applications [3], [31], [32].

However, is this approach optimal for avoiding outliers? There might be a better way to choose nonredundant TDs that can still reconstruct the full TD set; this is the focus of this study. In fact, there is an alternative approach to selecting $M-1$ TDs that can restore the full TD set, even without using the reference sensor, as shown in Fig. 1(c) Note that we cannot always compute the full TD set from the chosen $M-1$ TDs, as will be shown later.

In this paper, we propose a new TDE method based on the selection approach, which is robust against the outliers in the full set of TD measurements. We discuss the methodology and criterion for selecting the best TDs among the TDs in the full set. We introduce fundamental graph theory to the above problem, which is essential in this paper. On the basis of the graph theory, we first show a necessary and sufficient condition to uniquely reconstruct the full TD set from the chosen TDs. We then consider the optimization problem for choosing the best minimum-necessary TDs under that condition. This optimization problem can be reduced to the problem of finding the minimum spanning tree (MST), a common topic in the graph theory. The cost function for the optimization is designed on the basis of the GCC method.

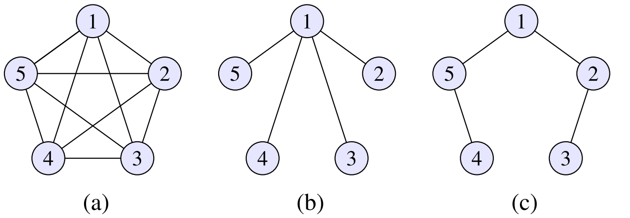


FIGURE 1. Sensor array and TDs for a single source. The circles and lines between them represent the sensors and the TDs to be measured from the sensor pairs, respectively. (a) Set of full redundant $M(M-1)/2$ TDs. (b) Set of nonredundant $(M-1)$ TDs. Only the TDs between the reference sensor and the others are used, where sensor 1 is the reference. (c) Another set of nonredundant TDs. No reference sensor is used in this case.

Note that disambiguation [33], [34] is another topic when considering a graph structure on a sensor array. In contrast to our study, the key constraint in the disambiguation methods is that the sum of cyclic TDs becomes 0, e.g., $\tau_{12}+\tau_{23}+\tau_{31}=0$. Disambiguation aims to estimate the nonredundant TDs under the above constraint. Since multiple sets of TDs that satisfy the constraint may exist, however, a certain type of postprocessing is necessary to choose the best TD set [31]. One advantage of our approach is that the solution is uniquely obtained by finding the MST.

Robust localization [35], [36], [37] is a field that also addresses outliers removal, where sensor positions or distances are necessary to compute DOAs from TDs. In contrast, our approach focuses solely on TD measurements, requiring no prior information, such as microphone positions. Furthermore, the proposed method eliminates outliers and estimates the full set of TDs, which are the fundamental spatial cues for various array signal processing methods. The proposed method can thus serve as preprocessing for a wide range of applications.

The rest of this paper is organized as follows. In Section II, we introduce the background knowledge regarding TDE. In Section III, we explain the motivation and then state the problem addressed in this paper. Section IV is devoted to solving the problem, where we present the necessary and sufficient condition to restore the full set of TD estimates from the chosen TD measurements, and in Section V, we propose the method of estimating optimal $M-1$ TDs. In Section VI, we discuss the efficacy of the proposed method via numerical experiments. Finally, we conclude this paper in Section VII.

A. NOTATION

We write scalars with regular letters (e.g., x), vectors and matrices with bold lowercase and uppercase letters (e.g., $\boldsymbol{x}$ and $\boldsymbol{X}$), respectively, and sets with calligraphic fonts (e.g., $\mathcal{X}$). The superscripts T, H, and $*$ denote transposition, conjugate transposition, and complex conjugate, respectively.

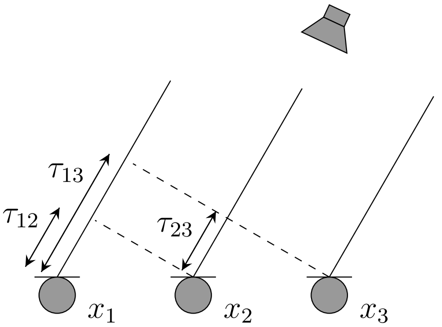


FIGURE 2. TDs defined in a sensor array. TD τ_{13} is equal to the sum of τ_{12} and τ_{23}. This relationship holds for a single source in any sensor array arrangement.

II. PRELIMINARIES

A. SIGNAL MODEL

Let us consider estimating TDs observed by an M channel sensor array. In this paper, we assume that only one target signal propagates to the array. Let x_{mkn} be the short-time Fourier transform (STFT) representation of the signal observed by the mth sensor ($m = 1, \ldots, M$), where $k = -K/2+1, \ldots, K/2$ and $n = 1, \ldots, N$ denote the frequency bin and time frame indices, respectively. We model the discrete observed signals as follows:

$$\begin{aligned} \boldsymbol{x}_{kn} &= s_{kn}\boldsymbol{g}_k(\boldsymbol{\tau}_r) + \boldsymbol{u}_{kn} \\ &= [x_{1kn} \ \cdots \ x_{Mkn}]^{\mathsf{T}} \in \mathbb{C}^M, \quad (1) \\ \boldsymbol{g}_k(\boldsymbol{\tau}_r) &= \left[a_{1k}e^{-\mathbb{i}\omega_k\tau_{r1}} \ \cdots \ a_{Mk}e^{-\mathbb{i}\omega_k\tau_{rM}}\right]^{\mathsf{T}} \in \mathbb{C}^M, \quad (2) \\ \boldsymbol{\tau}_r &= [\tau_{r1} \ \cdots \ \tau_{rM}]^{\mathsf{T}} \in \mathbb{R}^M. \quad (3) \end{aligned}$$

The variable s_{kn} denotes the source image observed at the reference sensor indexed by $r \in \{1, \ldots, M\}$. The vector $\boldsymbol{u}_{kn} = [u_{1kn} \ \cdots \ u_{Mkn}]^{\mathsf{T}}$ denotes noise signals at each sensor, $\omega_k = 2\pi k/K$ denotes the normalized angular frequency, and $\mathbb{i}$ denotes the imaginary unit. The vector $\boldsymbol{g}_k(\boldsymbol{\tau}_r)$ denotes the relative transfer function (RTF) [38], which is defined as the ratio of the transfer function at the reference sensor to those at the other sensors. The relative amplitude is denoted by $a_{mk} \in \mathbb{R}^+$.

Fig. 2 shows an example of a sensor array. The continuous TD parameter between sensors i and j is denoted by τ_{ij}. Given three sensors with the indices i, j, and k, the TDs between two of these three sensors satisfy the following three equations:

$$\tau_{ij} = \tau_{ik} + \tau_{kj}, \quad (4a)$$
$$\tau_{ij} = -\tau_{ji}, \quad (4b)$$
$$\tau_{ii} = 0. \quad (4c)$$

These equations are naturally used to evaluate the so-called consistency of TDs [23], [24]. Clearly, all TDs on an M channel sensor array have the $M-1$ degree of freedom at most, whereas ${}_MC_2 = M(M-1)/2$ TDs can be considered.

Note that the use of a reference sensor is not mandatory in this paper. Nonetheless, we use the reference sensor merely for notation ease and denote the TD vector $\boldsymbol{\tau}_r$ as in (3).

B. TD MEASUREMENT AND TD MATRIX

Hereafter, we will proceed with our discussion by distinguishing between the TD τ_{ij} and the TD measurement θ_{ij}. The TDs are the parameters we seek to estimate. Even though there exist $M(M-1)/2$ TDs, the degree of freedom of TD parameters is $M-1$ and they satisfy (4a), (4b), and (4c) as described in the previous section. On the other hand, the TD measurements obtained from pairs of two sensor signals are a type of observation including errors. We assume that the TD measurements also satisfy $\theta_{ij} = -\theta_{ij}$ and $\theta_{ii} = 0$ as in (4b) and (4c), but $\theta_{ij} = \theta_{ik} + \theta_{kj}$ does not always hold owing to errors.

A popular method of obtaining TD measurements is the GCC method [11], [12]. Given two observations x_{ikn} and x_{jkn}, the GCC function is written as

$$\Phi_{ij}(t) = \frac{1}{K}\sum_{k=-K/2+1}^{K/2} W_{ijk}S_{ijk}e^{\mathbb{i}\omega_k t}, \quad (5)$$
$$S_{ijk} = \frac{1}{N}\sum_n x_{ikn}x^*_{jkn}, \quad (6)$$

where t denotes the discrete lag and $W_k \in \mathbb{R}^+$ denotes an arbitrary weight function. Suitable weights have been proposed, e.g., GCC-phase transform (PHAT) and GCC-smoothed coherence transform (SCOT);

$$W^{\mathrm{PHAT}}_{ijk} = |S_{ijk}|^{-1}, \quad W^{\mathrm{SCOT}}_{ijk} = \left(S_{iik}S_{jjk}\right)^{-\frac{1}{2}}. \quad (7)$$

Then, we can obtain the TD measurement between sensors i and j as the peak of the GCC function:

$$\theta_{ij} = \arg\max_t \ \Phi_{ij}(t). \quad (8)$$

Now, we consider the following TDM [24], [27] to represent the full set of TD measurements:

$$\begin{aligned} \boldsymbol{\Theta} &= [\boldsymbol{\theta}_1 \ \cdots \ \boldsymbol{\theta}_M]^{\mathsf{T}} \\ &= \begin{bmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1M} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{M1} & \theta_{M2} & \cdots & \theta_{MM} \end{bmatrix}, \quad (9) \end{aligned}$$

where $\boldsymbol{\theta}_r = [\theta_{r1} \ \cdots \ \theta_{rM}]^{\mathsf{T}}$ denotes the TD measurements between sensors r and $m = 1, \ldots, M$. The TDM $\boldsymbol{\Theta}$ is a skew-symmetric matrix since $\theta_{ij} = -\theta_{ji}$ and $\theta_{ii} = 0$.

When the measurements contain no errors, i.e., all of them satisfy (4), $\boldsymbol{\Theta}$ can be rewritten as

$$\boldsymbol{\Theta} = \boldsymbol{1}\boldsymbol{\theta}_r^{\mathsf{T}} - \boldsymbol{\theta}_r\boldsymbol{1}^{\mathsf{T}}, \quad (10)$$

where $\boldsymbol{1} = [1 \ \cdots \ 1]^{\mathsf{T}} \in \mathbb{R}^M$. In this case, the degree of freedom of the TDM is essentially $M-1$ because $\boldsymbol{\Theta}$ is represented by only $\boldsymbol{\theta}_r$ as in (10), where $\theta_{rr} = 0$. The other properties, such as the rank of $\boldsymbol{\Theta}$, which is two in theory, have been shown in [24]. However, when the measured TDM contains errors, it has not $M-1$ but $M(M-1)/2$ degrees of freedom, which is redundant.

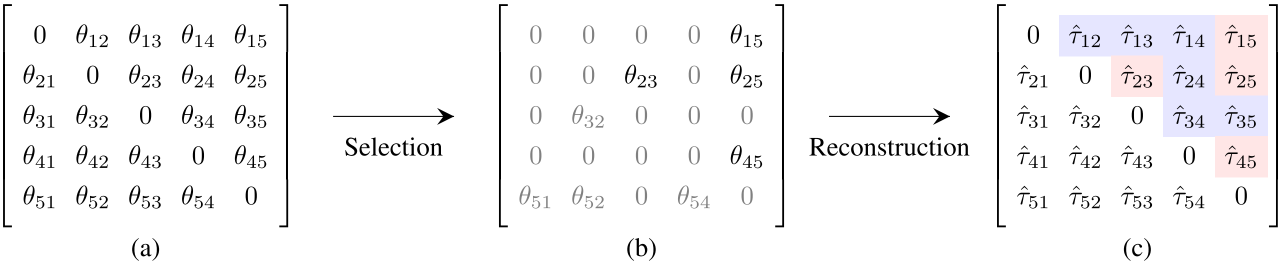


FIGURE 3. Processing flow of the proposed approach with a 5-channel sensor array. (a) TDM measurement Θ, where $\theta_{ij} = -\theta_{ji}$ and $\theta_{ii} = 0$ for all i and j. (b) Perforated TDM with the TD measurements corresponding to the set of selected sensor pairs $\mathcal{E}'$. (c) Estimate of the TDM $\hat{\mathcal{T}}$ computed from (b), where $\hat{\tau}_{ij} = \hat{\tau}_{ik} + \hat{\tau}_{kj}$, $\hat{\tau}_{ij} = -\hat{\tau}_{ji}$, and $\hat{\tau}_{ii} = 0$ for all i, j and k. The TD estimates in red and blue represent direct and indirect estimates, respectively.

III. PROBLEM STATEMENT

A. MOTIVATION

The purpose of this study is to develop a TDE algorithm that is robust against errors in TD measurements. Let us consider errors by dividing them into two types: small ones around the true TD value and outliers far from it. In the context of the GCC method, the former corresponds to the errors around the global maxima. The GCC function, being a nonconvex function, has one main lobe and numerous side lobes. The peak of the main lobe is basically expected to be close to the true TD, whereas the peak and the true TD do not match because they are affected by noise and other factors. Some methods have utilized the redundancy of multichannel observation to further improve the accuracy of GCC-based TDE [24], [39], [40].

The peaks of side lobes are far from that of the main lobe. The value of the GCC function corresponding to these peaks may be reversed in severe environments. The TD measurement obtained in such circumstances may become an outlier that is far from the true TD. Since these outliers significantly degrade the quality of downstream applications, such as DOA estimation and signal enhancement, they should be removed immediately after observation. In this paper, we thus discuss the methodology for avoiding the outliers.

B. APPROACH

To develop a TDE method that is robust to outliers, we adopt a selection approach, as shown in Fig. 3. Specifically, we start with the measurement of TDM, as shown in Fig. 3(a), and identify a minimally necessary set of sensor pairs $\{i, j\}$ denoted by $\mathcal{E}'$, as shown in Fig. 3(b). By using only the TD measurements θ_{ij} such that $\{i, j\} \in \mathcal{E}'$, we restore all TDs, as shown in Fig. 3(c). We aim to mitigate the impact of outliers by selecting the smallest number of reliable measurements for TDE, as the number of outliers is unknown in practical scenarios.

Let $\hat{\tau}_{ij}$ be an estimate of τ_{ij}. We first define the direct and indirect TDEs as follows.

- We refer to the TDE $\hat{\tau}_{ij} \leftarrow \theta_{ij}$ as *direct estimation*, where the sensor pair $\{i, j\}$ is the member of $\mathcal{E}'$;
- We similarly refer to $\hat{\tau}_{ij} \leftarrow \sum_{\ell=1}^{L-1} \theta_{z_\ell z_{\ell+1}}$ as the *indirect estimation*, where z_ℓ denotes the sensor index, especially $z_1 = i$, $z_L = j$. The sensor pair $\{z_\ell, z_{\ell+1}\}$ for all ℓ is the member of $\mathcal{E}'$.

Given $\mathcal{E}'$, we then restore the full set of TD estimates $\hat{\boldsymbol{\mathcal{T}}} = \left[\hat{\boldsymbol{\tau}}_1 \cdots \hat{\boldsymbol{\tau}}_M\right]^\mathsf{T}$ by direct or indirect estimation, where $\hat{\boldsymbol{\tau}}_r$ is the estimate of the TD vector (3).

In this approach, the choice of sensor pairs is very important. Depending on how the sensor pairs are selected, all TD estimates may not be obtained or multiple inconsistent estimates may be obtained. Then, the specific problems addressed in this paper are as follows.

- What is the necessary and sufficient condition of $\mathcal{E}'$ for obtaining all elements of $\hat{\boldsymbol{\mathcal{T}}}$ uniquely?
- How is the optimal one determined among $\mathcal{E}'$ satisfying the condition above?

We devote Section IV to solving the first problem and Section V to the second one.

IV. RECONSTRUCTION OF FULL TD SET

A. SENSOR ARRAY AS GRAPH

In this paper, we consider the problem of sensor pair selection on the basis of graph theory. The graph theory has sometimes been used for representing the pairwise relations in the TDE problem as in [33] and [34].

Let us consider the M channel sensor array. Let $\mathcal{V} = \{i \mid 1 \leq i \leq M\}$ and $\mathcal{E} = \{\{i, j\} \mid i, j \in \mathcal{V} \text{ and } i \neq j\}$ be a set of vertices and all distinct pairs of vertices (edges), respectively. We denote the undirected graph including all edges as $G = (\mathcal{V}, \mathcal{E})$. On the other hand, the set of selected sensor pairs $\mathcal{E}'$, which appeared in the previous section, is a subset of $\mathcal{E}$, that is, $\mathcal{E}' \subseteq \mathcal{E}$, and we use only the TD measurements for sensor pairs contained in $\mathcal{E}'$. Then, let $G' = (\mathcal{V}, \mathcal{E}')$ be a sensor graph. Now, Fig. 1 can be regarded as an example of the sensor graphs with the 5-channel sensor array.

B. RECONSTRUCTION CONDITION

Our objective here is to choose $\mathcal{E}'$ appropriately since a certain choice of $\mathcal{E}'$ would yield a problem such that not all TDs are estimated or some TDs are not uniquely determined. We can visualize this problem with the sensor graph as shown in Fig. 4(a) where the TD measurements associated with the sensor pairs $\{1, 4\}$, $\{1, 5\}$, $\{2, 3\}$, or $\{4, 5\}$ are selected. In this case, we cannot compute the TD estimate $\hat{\tau}_{12}$ since the subgraphs with $\{1, 4, 5\}$ and $\{2, 3\}$ are not connected; then, the relationship between them has been lost. Moreover, the selected edges form a cycle in Fig. 4(a). Then, there exist both direct and indirect estimations, namely, $\hat{\tau}^{(1)}_{14} \leftarrow \theta_{14}$ and $\hat{\tau}^{(2)}_{14} \leftarrow \theta_{15} + \theta_{54}$, where $\hat{\tau}^{(1)}_{14} \neq \hat{\tau}^{(2)}_{14}$ owing to the error in measurements. This indicates that the TDs corresponding to the chosen sensor pairs remain redundant. To avoid this problem, the edges must be chosen appropriately, as shown in Fig. 4(b). We thus consider the condition for reconstructing the full TD set.

First, the following proposition regarding a single TD holds for any sensor graph $G' = (\mathcal{V}, \mathcal{E}')$.

Proposition 1: A TD between arbitrary two sensors i and j can be uniquely determined by either direct or indirect estimation in general[1] *if and only if there exists a unique path between the vertices i and j on the sensor graph* $G' = (\mathcal{V}, \mathcal{E}')$.

Proof: Let us denote the path from a vertex z_1 to another vertex z_L on the graph $\mathcal{E}'$ as $(z_1, \ldots, z_L)$. Suppose there exists a unique path $(z_1, \ldots, z_L)$ from $z_1 = i$ to $z_L = j$ where $L \geq 2$. Since the edge $\{z_\ell, z_{\ell+1}\}$ corresponds to the TD measurement $\theta_{z_\ell z_{\ell+1}}$, the TD estimate is uniquely obtained as $\hat{\tau}_{ij} \leftarrow \sum_{\ell=1}^{L-1} \theta_{z_\ell z_{\ell+1}}$.

Conversely, suppose there is no unique path from the vertices i to j, which is divided into two cases, that is, no path or multiple paths exist. If there is no path from the vertices i to j, $\hat{\tau}_{ij}$ cannot be computed by the direct or indirect estimation. Next, suppose there exist multiple distinct paths, namely, $(z_1, \ldots, z_{L_1})$ and $(z'_1, \ldots, z'_{L_2})$, where $z_1 = z'_1 = i$, $z_{L_1} = z'_{L_2} = j$, and $L_1, L_2 \geq 2$. Then, the TD estimate can be computed in two ways by tracing the two paths: $\hat{\tau}_{ij} \leftarrow \sum_{\ell=1}^{L_1-1} \theta_{z_\ell z_{\ell+1}}$ and $\hat{\tau}'_{ij} \leftarrow \sum_{\ell=1}^{L_2-1} \theta_{z'_\ell z'_{\ell+1}}$. However, these two TD estimates do not match in general since they contain different TD measurements; thus, $\hat{\tau}_{ij}$ cannot be uniquely determined. This holds for any i and j. ∎

Now, we present the following proposition, which we refer to as the reconstruction condition.

Proposition 2: TDs between all distinct sensor pairs can be uniquely determined by either direct or indirect estimation if and only if the sensor graph $G' = (\mathcal{V}, \mathcal{E}')$ *is a spanning tree.*

Proof: From Proposition 1, the following two conditions are identical: TDs between all distinct sensor pairs can be uniquely determined and there exists a unique path between any distinct pair of vertices. Now, a graph with such a property is referred to as a spanning tree [41], [42]. ∎

In accordance with Proposition 2, the full set of TD estimates can be computed from the TD measurements corresponding to an arbitrary spanning tree. The number of minimum-necessary sensor pairs is thus $M - 1$, which corresponds to the same number of nonredundant TD measurements.

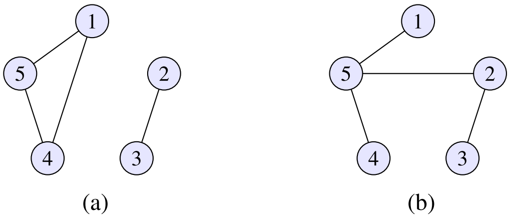


FIGURE 4. Examples of undirected graphs for the 5-channel sensor array with 4 edges. (a) Graph with a cycle where the sensor nodes {1, 4, 5} and {2, 3} are disconnected, indicating that this is not a tree. (b) Graph without any cycles where all sensor nodes are connected, thus indicating that this is a tree.

As an example of the reconstruction process, let us consider computing $\hat{\tau}_{34}$ in the spanning tree shown in Fig. 4(b). The edges are $\{\{1, 5\}, \{2, 3\}, \{2, 5\}, \{4, 5\}\}$ and the corresponding TDs are $\{\theta_{15}, \theta_{23}, \theta_{25}, \theta_{45}\}$ with their sign reversals. With such a spanning tree, there exists a unique path from sensors 3 to 4, and thus, the TD estimate can be computed indirectly as

$$\begin{aligned}\hat{\tau}_{34} &\leftarrow \theta_{32} + \theta_{25} + \theta_{54} \\ &= -\theta_{23} + \theta_{25} - \theta_{45}. \end{aligned} \tag{11}$$

All the elements in $\hat{\boldsymbol{\mathcal{T}}} = [\hat{\tau}_{ij}]_{1 \leq i,j \leq M}$ can be obtained in the same manner, where $\hat{\tau}_{ii} = 0$.

V. ESTIMATION OF NONREDUNDANT TDS

A. OPTIMIZATION PROBLEM

As mentioned in the previous section, estimates of TDs between all distinct sensor pairs can be determined uniquely if and only if the set of selected sensor pairs forms a spanning tree. Let $\mathcal{S}$ be a set of spanning trees on $\mathcal{V}$ and S be a member of $\mathcal{S}$. The problem here is how to find the optimal spanning tree $S^\dagger$ among $\mathcal{S}$.

To evaluate the optimality of a spanning tree, we introduce a cost function for each TD measurement, which corresponds to each edge of the graph, $e \in \mathcal{E}$. The lower the cost, the better the TD measurement. (or: We intend to represent that the lower the cost, the better the estimate.) Let C_{ij} denote the cost for the TD measurement θ_{ij} and $\boldsymbol{C} = [C_{ij}]_{1 \leq i,j \leq M}$ denote the cost matrix. We define that all diagonal elements of $\boldsymbol{C}$ are zero. The matrix $\boldsymbol{C}$ is symmetric as a result, providing the weight for the edge $\{i, j\}$ on G. The cost matrix then represents the adjacency matrix of a weighted graph. With such a cost matrix, we consider the optimal spanning tree to be the one that minimizes the sum of the costs of its edges.

We can then formulate the following optimization problem pertinent to the sensor graph with selected edges, $G' = (\mathcal{V}, \mathcal{E}')$, for choosing the optimal set of sensor pairs.

$$\mathcal{E}^\dagger = \mathop{\arg\min}_{\mathcal{E}'} \sum_{\substack{i,j \in e \\ e \in \mathcal{E}'}} C_{ij} \quad \text{s.t.} \quad G' \in \mathcal{S}, \tag{12}$$

where the abbreviation "arg min" stands for the argument of the minimum. The optimal spanning tree can be obtained as $S^\dagger = (\mathcal{V}, \mathcal{E}^\dagger)$.

[1] Here "in general" means to exclude a particular case such that θ_{ij} and $\theta_{ik} + \theta_{kj}$ take the same value coincidentally. In such a case, the TD estimate can be uniquely determined even if the two paths (i, j) and (i, k, j) exist. However, we here generally discuss excluding such a case.

Algorithm 1 Kruskal's MST Solver

```
Input: $G = (\mathcal{V}, \mathcal{E}; \mathcal{C})$
Output: $\mathcal{S}^\dagger = (\mathcal{V}, \mathcal{E}^\dagger)$
  Initialize:
     $\mathcal{E}' = \emptyset$
     $l = 1, \dots, L$ and $L = M(M-1)/2$
     $\tilde{C}(e) = C_{ij}$, where $i, j \in e$ and $e \in \mathcal{E}$
     Define $e_l$ for all $l$ such that $\tilde{C}(e_1) \leq \cdots \leq \tilde{C}(e_L)$
  $i = 0$
  while $|\mathcal{E}'| \neq |\mathcal{V}| - 1$ do
     if no cycle exists in $\mathcal{E}' \cup \{e_i\}$ then
        $\mathcal{E}' \leftarrow \mathcal{E}' \cup \{e_i\}$
     end if
     $i \leftarrow i + 1$
  end while
  $\mathcal{E}^\dagger \leftarrow \mathcal{E}'$
```

Algorithm 2 Prim's MST Solver

```
Input: $G = (\mathcal{V}, \mathcal{E}; \mathcal{C})$, arbitrary $a \in \mathcal{V}$
Output: $\mathcal{S}^\dagger = (\mathcal{V}, \mathcal{E}^\dagger)$
  Initialize:
     $\mathcal{E}' = \emptyset$
     $\mathcal{A} = \{a\}$                        ▷ Set of selected vertices
     $\mathcal{A}^c = \mathcal{V} \setminus \mathcal{A}$                    ▷ Set of remaining ones
  while $\mathcal{A} \neq \mathcal{V}$ do
     Find the edge $e = \{i, j\}$ such that whose weight $C_{ij}$
     is minimum in $\{C_{ij} \mid \{i, j\} \in \mathcal{E}, i \in \mathcal{A}, j \in \mathcal{A}^c\}$
     $\mathcal{E}' \leftarrow \mathcal{E}' \cup \{e\}$
     $\mathcal{A} \leftarrow \mathcal{A} \cup \{j\}$
     $\mathcal{A}^c \leftarrow \mathcal{V} \setminus \mathcal{A}$
  end while
  $\mathcal{E}^\dagger \leftarrow \mathcal{E}'$
```

B. REFERENCE-BASED SOLUTION

As mentioned in Section I, the reference sensor is commonly used to employ nonredundant TDs [3], [23], [30], [31], [32], with the first sensor often serving as the reference. Then, before solving (12), let us consider the problem of choosing the optimal reference sensor as follows.

$$\mathcal{E}^\dagger = \{\{r^\dagger, j\} \mid 1 \leq j \leq M,\ j \neq r^\dagger\}, \tag{13}$$

$$r^\dagger = \arg\min_r \sum_{j \neq r} C_{rj}. \tag{14}$$

Obviously, this solution satisfies the reconstruction condition [see (10)]. We will show that it is a special case of (12) as follows.

From the graph theory perspective, this solution forms a star [41], as shown in Fig. 1(b). We can also confirm that the reconstruction condition holds since the star is one of the spanning trees. Hereafter, we will refer to this solution as SST (star-spanning tree).

This approach is very intuitive and the computational complexity is low. However, this solution is sub-optimal since the feasible region is limited to only a set of stars among all spanning trees. For instance, let us consider a TDM that has outliers on superdiagonal entries, namely, $(i, i+1)$ entries (and consequently, $(i+1, i)$ entries as well). Then, no matter which reference sensor is chosen, an outlier will inevitably be included. If the selection of sensor pairs is not constrained to the pair of the reference sensor and others, we can still choose $M-1$ TD measurements without outliers even in this case.

C. MST-BASED SOLUTION

Now, we consider solving the optimization problem (12) directly, which is referred to as the MST, one of the well-studied subjects in the graph theory [43], [44], [45], [46], [47], [48], [49]. There exist efficient algorithms that can achieve a unique solution as long as the edge weights are defined appropriately.

Kruskal's algorithm [43], [44] is one of the famous MST solvers. We briefly show the pseudo-code in Algorithm 1, where $\mathcal{C} = \{C_{ij} \mid i, j \in \mathcal{V} \text{ and } i < j\}$ denotes the set of weights corresponding to each edge in $\mathcal{E}$. The inequality for i and j is merely for choosing C_{ij} from the elements above the main diagonal of the cost matrix, which is essentially meaningless. Kruskal's algorithm is a kind of greedy algorithm. The MST is found by iteratively selecting/removing the edge with the lowest weight from a list of sorted edges while ensuring that the selected ones do not form a cycle. The computational complexity is $\mathcal{O}(|\mathcal{E}| \log |\mathcal{V}|)$ or equivalently $\mathcal{O}(|\mathcal{V}|^2 \log |\mathcal{V}|)$ because $|\mathcal{E}| = |\mathcal{V}|(|\mathcal{V}| - 1)/2$ in our problem.

Another famous method is Prim's algorithm [45], [46], [47], which is shown in Algorithm 2. The MST is obtained by iteratively expanding a subset of edges. The subset is grown by adding the edge with the lowest weight between the vertices in the current subset and the remaining ones. The computational complexity[2] is $\mathcal{O}(|\mathcal{V}|^2)$.

To compare these methods, Kruskal's algorithm is preferable when the edges are sparse since it first sorts all the edges. On the other hand, Prim's algorithm evaluates a limited number of edges per loop and is thereby faster for a complete graph. Since the graph associated with the full TD set always has full edges, the latter is a better choice for our problem.

The output of whichever MST solver is the MST with a set of $M-1$ edges. We can then reconstruct the TDM by estimating every TD directly or indirectly, as discussed in Section IV-B.

D. COST MATRIX

The remaining problem is how to define an appropriate cost matrix $\boldsymbol{C} = [C_{ij}]_{1 \leq i,j \leq M}$. In this paper, we propose the GCC-function-based cost matrix for the following reason. Typically, a TD is obtained by measuring the peak of the GCC function. It is expected that the GCC function will take a larger value when the given TD measurement is closer to the unknown true TD. Conversely, outliers are expected to have a relatively small GCC function value. We thus employ the value obtained by substituting the measurement θ_{ij} into the GCC function (5) as the cost for itself. Moreover, we introduce the weighting factor $\alpha \in \{0, 1\}$

[2] The computational complexity depends on the data structure, e.g., it is $\mathcal{O}(|\mathcal{E}| + |\mathcal{V}| \log |\mathcal{V}|)$ with the Fibonacci heap.

for normalization by defining the following cost matrix:

$$C_{ij} = \begin{cases} -\dfrac{1}{K}\displaystyle\sum_{k=-K/2+1}^{K/2} \dfrac{S_{ijk}}{|S_{ijk}|^{\alpha}} e^{-\mathbb{i}\omega_k\theta_{ij}} & i \neq j \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

where $i, j \in e$ and $e \in \mathcal{E}$. The (i, j) element of the cost matrix corresponds to the ordinary cross correlation when $\alpha = 0$, and it corresponds to the GCC-PHAT (7) when $\alpha = 1$. Hence, the properties of the cost matrix with $\alpha = 1$ follow those of the GCC-PHAT.

E. PROPOSED ALGORITHM

Finally, we summarize the algorithm of the MST-based TDE in Algorithm 3. The function $\mathbf{TRIU}(\cdot)$ returns a matrix with the upper triangle elements of the given matrix, where elements below the main diagonal are zero. $\mathbf{MST}(\cdot)$ denotes the MST solver. Although any MST solver can be used, Prim's algorithm is employed in this paper. One implementation of Prim's algorithm is in "NetworkX," a Python package [50].[3]

In Algorithm 3, we assume that the measurements are only the multichannel observation $\boldsymbol{x}_{kn}$. The normalization factor p is the user-defined parameter. The measurement of the TDM is obtained by performing the GCC method, and the cost matrix defined in (15) is employed. The output is the estimate of TDM $\hat{\mathcal{T}}$ or equivalently the full set of TDs.

There are some options regarding the above conditions. For example, given a measurement of the TDM, we can use it directly instead of performing the GCC method. Any existing TDE method can also be utilized as an alternative to the GCC method. It is also possible to define another cost matrix, where the cost matrix must be symmetrical since the TDM is skew-symmetric.

VI. EXPERIMENTS

To investigate the efficacy of the proposed method, we conduct two numerical experiments and compare its performance with that of existing methods. In Section VI-A, we evaluate the robustness of the proposed method against the outliers. In Section VI-B, we evaluate the TDE performance in a simulated reverberant environment.

A. ROBUSTNESS AGAINST OUTLIERS

1) EXPERIMENTAL CONDITIONS

We simulate an $M = 8$ channel microphone array. The target signal s_{kn} is a white noise of 4096 samples. The TD between two adjacent microphones is randomly generated from the uniform distribution whose interval is $(-25, 25]$ samples. TDs are simulated by rotating the phase of the target signal, and the TDM is computed by using the ground truth. We then contaminate the TDM with outliers, which are uniformly generated from $\pm[50, 100]$ samples. The number of outliers is in the range of 0 to ${}_8C_2 = 28$. We perform STFT with

Algorithm 3 MST-Based TDE

```
Input: $\boldsymbol{x}_{kn}, \alpha$
Output: $\hat{\mathcal{T}} = [\hat{\tau}_{ij}]_{1 \leq i,j \leq M}$
  $S_k = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_{kn}\boldsymbol{x}_{kn}^{\mathsf{H}}$
  $C_{ij} = 0$ for all $i = j$
  for $i, j = 1, \ldots M$ and $i \neq j$ do
      $\theta_{ij} = \mathbf{GCC}(S_{ijk})$
      $C_{ij} = -\frac{1}{K}\sum_{k=-K/2+1}^{K/2} \frac{S_{ijk}}{|S_{ijk}|^{\alpha}} e^{-\mathbb{i}\omega_k\theta_{ij}}$
  end for
  $(\mathcal{V}, \mathcal{E}^{\dagger}) = \mathbf{MST}(\mathbf{TRIU}(\boldsymbol{C}))$
  $\hat{\tau}_{ij} = 0$ for all $i = j$
  for $i, j = 1, \ldots M$ and $i \neq j$ do
      Let $(z_1, \ldots, z_L)$ be the path from $z_1 = i$ to $z_L = j$,
      where $\{z_\ell, z_{\ell+1}\} \in \mathcal{E}^{\dagger}$ for all $\ell$
      $\hat{\tau}_{ij} = \sum_{\ell=1}^{L-1} \theta_{z_\ell z_{\ell+1}}$
  end for
  Function $\mathbf{GCC}(S_k)$
      $\Phi(t) = \frac{1}{K}\sum_{k=-K/2+1}^{K/2} W_k S_k e^{\mathbb{i}\omega_k t}$
      return $t = \arg\max_t \Phi(t)$
  end Function
```

4096 samples in a rectangle window, which is equivalent to the one-shot discrete Fourier transform (DFT).

2) COMPARISON METHODS

We evaluate three methods for comparison: RST, SST, and TDM. RST is a crude method that randomly generates a spanning tree and uses it as a solution, implemented merely for comparison. The algorithm employed in this paper to obtain a random spanning tree[4] has a computational complexity of $\mathcal{O}(|\mathcal{V}|)$. SST is the method described in Section V-B. TDM is a robust TDE method based on the TDM, which has been proposed in [24, Sec. VI]. To the best of the authors' knowledge, TDM is one of the state-of-the-art algorithms that does not require prior information, such as microphone positions. Although TDM is a method of computing the TDM by using all TD measurements rather than selection, we employ it as a comparison method because the purpose of removing outliers is the same. Note that TDM requires the maximum number of outliers supposed to be present, and in this experiment, we use the true value.

We also perform the proposed method denoted by MST, where $\alpha = 0$ in this experiment. Since the noisy measurement of the TDM is given, we use it directly instead of performing the GCC method in this experiment.

3) EVALUATION CRITERIA

In this paper, we will evaluate whether all outliers have been removed. We thus define the rejection failure (RF) and RF rate (RFR) as the evaluation criteria as follows:

$$\mathrm{RF}_p = \begin{cases} 0 & \text{if } \left|\left[\hat{\mathcal{T}}_{(p)} - \mathcal{T}^{\star}_{(p)}\right]_{ij}\right|^2 < \mathrm{THR} \quad \forall i, j \\ 1 & \text{otherwise} \end{cases}, \tag{16}$$

[3] Available as networkx.minimum_spanning_tree().

[4] Available as networkx.random_tree().

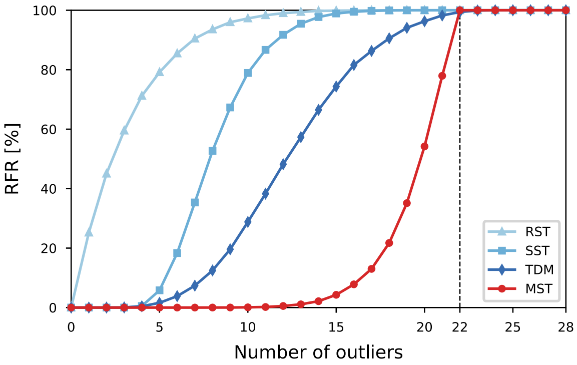


FIGURE 5. RFR as a function of the number of outliers in the TD measurements, where the number of sensors is eight. Each method attempts to eliminate outliers, and its failure rate is shown on the vertical axis. Spanning-tree-based methods require at least $M - 1$ clean TDs; hence, they always fail to remove outliers in the range of 22 or more on the horizontal axis.

$$\mathrm{RFR} = \frac{1}{P}\sum_{p} \mathrm{RF}_p, \tag{17}$$

where $\mathcal{T}^{\star}_{(p)}$ denotes the true TDM and the subscript $p = 1, \ldots, P$ denotes the simulation index. THR denotes the threshold for distinguishing whether the estimate is an outlier or not, and we set THR $= 5$ in this paper. If a TDE method successfully estimates $\hat{\mathcal{T}}$ and eliminates all outliers from $\boldsymbol{\Theta}$, RF_p becomes 0.

4) RESULTS AND DISCUSSION

Fig. 5 shows the results of the experiment, where we perform $P = 10000$ trials. The horizontal axis shows the number of outliers in the noisy TDM and the vertical axis shows RFR. In the spanning-tree-based approach, at least $M - 1 = 7$ TD measurements must be clean. Hence, these methods theoretically fail if the number of outliers equals or exceeds 22.

As shown in Fig. 5, RST shows the worst result. The probability of not choosing one outlier in 28 elements is one-fourth, almost matching the result of RST when the number of outliers is one. The result of SST indicates that estimating the optimal reference sensor helps avoid a few outliers, while it may fail to remove outliers when their count exceeds five. This tendency is almost the same in TDM even if the true number of outliers is given.

In contrast, the proposed method, MST, works well despite the numerous outliers. Even if 10 out of 28 TDs are strongly contaminated by noise, MST successfully removes them almost 100 % of the time and restores the full set of accurate TDs. Additionally, the curvature of the MST result is far from the others. The RFR should be as close to zero as possible, and MST achieves a significantly low RFR. Note that the ideal RFR depends on which elements of the TDM contain outliers and may not always be zero. Let us consider a specific scenario where there are 20 outliers. In some cases, it is possible that, regardless of which seven of the eight clean TDs are selected, a spanning tree cannot be constructed. Then, outlier removal cannot be achieved in principle.

B. PERFORMANCE EVALUATION OF TDE

1) EXPERIMENTAL CONDITIONS

Next, we evaluate the performance of TDE in a reverberant enclosure. We use pyroomacoustics [51], a Python package, in this experiment. We synthesize $M = 4$, 8, and 16 observed signals with simulated room impulse responses (RIRs) with a reverberation time of approximately 400 ms. The sampling frequency is 16 kHz. The target signal is approximately 4 s long and is randomly generated following the normal Gaussian distribution. Noise signals are generated under the same conditions, and the average signal-to-noise ratio (SNR) is set to 20 dB. We perform a one-shot DFT with a rectangle window for the observed signals. The target source and microphones are randomly located in a room of 18 m × 24 m × 6 m size. They are placed at least 1 m away from the surfaces of the room, and the distance between the adjacent microphones is forced to be greater than 0.01 m. Thus, the minimum and maximum distances between the microphones are 0.01 m and approximately 27.50 m, respectively, with the average distance of about 10.77 m. The source-microphone distance is around 10.78 m on average. In general, the larger the microphone spacing and/or the longer the reverberation time, the more difficult it is to estimate the TDs. We adopt this room size as a setting where outliers are likely to occur.

2) COMPARISON METHODS

The comparison methods are generally the same as those in Section VI-A. Since the measurements are the microphone observations in this experiment, we employ the GCC method to compute the TDM, as shown in Algorithm 3. The weight function here is $W_k = 1$ for all k. We then perform SST with the computed TDM, referred to as SST-INT.

The TD measurements obtained by the GCC method alone are limited to an integer multiple of the reciprocal of the sampling frequency. We thus employ parabolic interpolation [52] as a postprocessing to obtain the TDM with subsample TD measurements. Every method except for SST-INT is based on the subsample TDM.

As in the previous section, we perform three comparison methods: RST, SST, and TDM. TDM requires the number of outliers in advance, whereas it is unknown in practical situations. We thus perform TDM several times while changing the hyperparameter, which is denoted by κ.

3) EVALUATION CRITERIA

In addition to the RFR, we use the root mean square error (RMSE) as the evaluation metric in this section defined as[5]

$$\mathrm{RMSE} = \sqrt{\frac{1}{PM(M-1)}\sum_{p} \|\hat{\mathcal{T}}_{(p)} - \mathcal{T}^{\star}_{(p)}\|_{\mathrm{F}}^{2}}, \tag{18}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. The number of simulations, P, is 10000 in this paper. Since we generate

[5] The main diagonal elements of the TDM are always zero. The number of TD estimates in the TDM is thus $M(M-1)$.

TABLE 1. Results of TDE with $M = 4, 8, 16$-channel microphone arrays.

Method	Number of direct estimates used	$M = 4$			$M = 8$			$M = 16$		
		RMSE	RMSE (W/O OLs)	RFR [%]	RMSE	RMSE (W/O OLs)	RFR [%]	RMSE	RMSE (W/O OLs)	RFR [%]
SST-INT	$M - 1$	7.406	0.358	1.90	7.525	0.390	3.93	8.569	0.404	7.63
SST		7.153	0.118	1.82	7.820	0.132	4.22	8.757	0.122	7.91
RST		7.448	0.121	1.93	9.150	0.161	4.54	11.214	0.189	9.50
MST ($\alpha = 0$)		4.398	0.112	1.23	5.201	0.123	2.24	4.808	0.122	3.95
MST ($\alpha = 1$)		2.602	0.108	0.46	1.831	0.120	0.60	1.446	0.122	0.61
TDM ($\kappa = 0$)	$\frac{M(M-1)}{2}$	4.516	0.092	3.29	3.749	0.149	10.51	3.090	0.239	22.07
TDM ($\kappa = 2$)		3.624	0.096	0.80	2.702	0.096	2.75	2.152	0.158	11.38
TDM ($\kappa = 4$)		3.624	0.096	0.80	2.961	0.091	1.16	1.727	0.128	5.96
TDM ($\kappa = 6$)		N/A	N/A	N/A	2.998	0.099	1.17	1.691	0.095	2.67
TDM ($\kappa = 8$)		N/A	N/A	N/A	3.134	0.299	8.16	1.764	0.081	1.47
TDM ($\kappa = 10$)		N/A	N/A	N/A	∞	0.086	22.22	1.774	0.071	1.20

the RIRs regarding the microphone and source positions, the ground truth $\mathcal{T}^{\star}_{(p)}$ is unknown. Therefore, we compute it from the distance between them. The speed of sound is set to 343 m/s, which is equal to that used for RIR generation via pyroomacoustics.

If there is even one outlier in the TD estimates, the RMSE becomes extremely large owing to the effect of the outlier. The criteria excluding such a gross error(s) will help in the interpretation of the results. Therefore, we also evaluate the RMSE using only the results of trials in which outliers were successfully discarded. From now on, we denote the RMSE computed in this manner as "RMSE (W/O OLs)".

4) RESULTS AND DISCUSSION

Table 1 show the results of the experiments. In the case of $M = 4$, the baseline method SST-INT shows the worst RMSEs. SST-INT fails to remove outliers 190 times out of $P = 10000$ trials. SST improves the RMSE (W/O OLs) by applying quadratic interpolation to the TD measurements. The RFR is also slightly improved accordingly. RST is worse than the baseline method, which is similar to the results discussed in the previous section. We show the results of TDM for the different numbers of outliers $\kappa = 0, 2, \ldots, 10$. Since the number of TD measurements is $_4C_2 = 6$, no results are available with $\kappa = 6, 8, 10$. TDM is a method that simultaneously realizes the refinement of TD measurements and the removal of outliers by utilizing all TD measurements. Thus, it shows much better results in RMSE (W/O OLs). However, $\kappa = 0$ assumes that there are no outliers (in other words, all measurements are assumed to be reliable), and the RFR increases as a result. It can be seen that the RFR is less than that of the baseline when considering outliers with $\kappa = 2$ and 4.

The proposed method MST shows the best results when $\alpha = 1$ (with GCC-PHAT-based cost matrix). The number of RFs is the lowest at 46, and the RMSE is accordingly improved. Additionally, we can see that the cost matrix based on the GCC-PHAT significantly improves the performance of MST. Unlike TDM, the spanning-tree-based approach has no refinement mechanism since it uses the minimum-necessary TD measurements. The RMSE (W/O OLs) is thus limited by the accuracy of the given TD measurements. In practice, it is desirable to apply some accurate TDE algorithms, such as those in [40], as the postprocessing of MST.

The trend observed with $M = 8$ microphones is similar to the aforementioned results; however, the benefits of the proposed method become more pronounced. As the number of microphones increases from four to eight, the number of TDs to be estimated also increases, leading to a higher occurrence of outliers. This is the reason for the overall increase in RFR compared with the case of $M = 4$. Additionally, parabolic interpolation is applied regardless of the accuracy of TD measurements. The interpolated TDs are expected to approach the ground truth more closely, whereas interpolated outliers may further stray away from it. Then, the subsample TDM might contain worse outliers, resulting in the decrease in the performance of SST compared with SST-INT. Despite such circumstances, MST ($\alpha = 1$) markedly reduces the number of RFs. SST and MST search for the optimal spanning tree in the same TDM on the basis of their respective criteria. Therefore, we can confirm the effectiveness of the proposed approach that estimates the MST rather than using the best reference sensor. TDM can also reduce the RFR by setting κ appropriately, such as $\kappa = 4, 6$. Basically, TDM is guaranteed to converge when κ is small [24]. However, it sometimes diverges when the number of outliers, κ, is set to 10 (its maximum value here is $_8C_2 = 28$), which leads the RMSE to become infinity. In contrast, the proposed MST ($\alpha = 1$) shows the best results in terms of the RMSE and RFR.

In the case of $M = 16$, similar results as in the case of $M = 8$ are obtained. As shown in the rightmost column of Table 1, the RFR becomes worse than that with $M = 8$ in most cases. On the other hand, MST ($\alpha = 1$) can reduce the RFR to the same level as in the case of $M = 8$ by utilizing the additional redundancy owing to the increase in the number of observations. Since the RMSE is still limited, the application of postprocessing is preferable, as mentioned earlier. Finally, from the above results, we can confirm the efficacy of MST regardless of the number of microphones.

VII. CONCLUSION

In this paper, we proposed a novel method of estimating TDs that is robust against outliers in the measurements.

Specifically, we addressed the problem of selecting non redundant TDs from a full set of TD measurements, while ensuring that all TDs can still be restored. We showed that this problem is reduced to the problem of finding the MST and we developed an efficient algorithm based on graph theory.

In numerical experiments, we demonstrated that the proposed method successfully selects relatively clean TD measurements and reconstructs the full set while discarding outliers. Moreover, we verified the efficacy of the proposed method in a reverberant room environment through computer simulations.

Compared with conventional methods, our approach achieved superior performance in removing outliers, although its estimation accuracy is still limited. The future work thus includes integrating our method with TD refinement methods to realize highly accurate and robust TDE.

REFERENCES

[1] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: An overview," *EURASIP J. Adv. Signal Process.*, vol. 2006, no. 1, pp. 1–19, Dec. 2006.

[2] T. Gustafsson, B. D. Rao, and M. Trivedi, "Source localization in reverberant environments: Modeling and statistical analysis," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 791–803, Nov. 2003.

[3] Y. Huang, J. Benesty, and J. Chen, "Time delay estimation and source localization," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Berlin, Germany: Springer, 2008, pp. 1043–1063.

[4] J. Yang, X. Zhong, W. Chen, and W. Wang, "Multiple acoustic source localization in microphone array networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 334–347, 2021.

[5] C. Evers, H. W. Löllmann, H. Mellmann, A. Schmidt, H. Barfuss, P. A. Naylor, and W. Kellermann, "The LOCATA challenge: Acoustic source localization and tracking," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 1620–1643, 2020.

[6] L. Qu, Q. Sun, T. Yang, L. Zhang, and Y. Sun, "Time-delay estimation for ground penetrating radar using ESPRIT with improved spatial SmoothingTechnique," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 8, pp. 1315–1319, Aug. 2014.

[7] J. Capon, "Applications of detection and estimation theory to large array seismology," *Proc. IEEE*, vol. 58, no. 5, pp. 760–770, May 1970.

[8] J. C. Chen, K. Yao, and R. E. Hudson, "Source localization and beamforming," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 30–39, Mar. 2002.

[9] G. Carter, "Time delay estimation for passive sonar signal processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 3, pp. 463–470, Jun. 1981.

[10] P. Protiva, J. Mrkvica, and J. Machac, "Estimation of wall parameters from time-delay-only through-wall radar measurements," *IEEE Trans. Antennas Propag.*, vol. 59, no. 11, pp. 4268–4278, Nov. 2011.

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

[12] I. J. Tashev, *Sound Capture and Processing* (Practical Approaches). Hoboken, NJ, USA: Wiley, Jul. 2009.

[13] J. Chen, Y. Huang, and J. Benesty, "Time delay estimation," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*. Boston, MA, USA: Kluwer, 2004, pp. 197–227.

[14] B. Qin, H. Zhang, Q. Fu, and Y. Yan, "Subsample time delay estimation via improved GCC PHAT algorithm," in *Proc. 9th Int. Conf. Signal Process.*, Oct. 2008, pp. 2579–2582.

[15] M. Cobos, F. Antonacci, L. Comanducci, and A. Sarti, "Frequency-sliding generalized cross-correlation: A sub-band time delay estimation approach," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 1270–1281, 2020.

[16] X. Wang, G. Huang, J. Benesty, J. Chen, and I. Cohen, "Time difference of arrival estimation based on a Kronecker product decomposition," *IEEE Signal Process. Lett.*, vol. 28, pp. 51–55, 2021.

[17] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," in *Proc. 18th IEEE Symp. Commun. Veh. Technol. Benelux (SCVT)*, Nov. 2011, pp. 1–6.

[18] A. Bertrand, S. Doclo, S. Gannot, N. Ono, and T. van Waterschoot, "Special issue on wireless acoustic sensor networks and ad hoc microphone arrays," *Signal Process.*, vol. 107, pp. 1–3, Feb. 2015.

[19] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee, "A survey of sound source localization methods in wireless acoustic sensor networks," *Wireless Commun. Mobile Comput.*, vol. 2017, pp. 1–24, Aug. 2017.

[20] S. Miyabe, N. Ono, and S. Makino, "Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation," *Signal Process.*, vol. 107, pp. 185–196, Feb. 2015.

[21] L. Wang and S. Doclo, "Correlation maximization-based sampling rate offset estimation for distributed microphone arrays," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 3, pp. 571–582, Mar. 2016.

[22] A. Chinaev, P. Thüne, and G. Enzner, "Double-cross-correlation processing for blind sampling-rate and time-offset estimation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 1881–1896, 2021.

[23] X. Alameda-Pineda and R. Horaud, "A geometric approach to sound source localization from time-delay estimates," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 6, pp. 1082–1095, Jun. 2014.

[24] J. Velasco, D. Pizarro, J. Macias-Guarasa, and A. Asaei, "TDOA matrices: Algebraic properties and their application to robust denoising with missing data," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5242–5254, Oct. 2016.

[25] W. Hahn and S. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 5, pp. 608–614, Sep. 1973.

[26] H. C. So, Y. T. Chan, and F. K. W. Chan, "Closed-form formulae for time-difference-of-arrival estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2614–2620, Jun. 2008.

[27] N. Zhu, "Locating and extracting acoustic and neural signals," Ph.D. dissertation, Dept. Mech. Eng., Wayne State Univ. Detroit, MI, USA, 2011. [Online]. Available: https://digitalcommons.wayne.edu/oa_dissertations/422

[28] M. Compagnoni, A. Pini, A. Canclini, P. Bestagini, F. Antonacci, S. Tubaro, and A. Sarti, "A geometrical–statistical approach to outlier removal for TDOA measurements," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3960–3975, Aug. 2017.

[29] Y. Zhu, B. Deng, A. Jiang, X. Liu, Y. Tang, and X. Yao, "ADMM-based TDOA estimation," *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1406–1409, Jul. 2018.

[30] T.-K. Le, K. C. Ho, and T.-H. Le, "Rank properties for matrices constructed from time differences of arrival," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3491–3503, Jul. 2018.

[31] A. Canclini, P. Bestagini, F. Antonacci, M. Compagnoni, A. Sarti, and S. Tubaro, "A robust and low-complexity source localization algorithm for asynchronous distributed microphone networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 10, pp. 1563–1575, Oct. 2015.

[32] L. Kraljevic, M. Russo, M. Stella, and M. Sikora, "Free-field TDOA-AOA sound source localization using three soundfield microphones," *IEEE Access*, vol. 8, pp. 87749–87761, 2020.

[33] J. Scheuing and B. Yang, "Disambiguation of TDOA estimation for multiple sources in reverberant environments," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 8, pp. 1479–1489, Nov. 2008.

[34] C. M. Zannini, A. Cirillo, R. Parisi, and A. Uncini, "Improved TDOA disambiguation techniques for sound source localization in reverberant environments," in *Proc. IEEE Int. Symp. Circuits Syst.*, Jun. 2010, pp. 2666–2669.

[35] S. S. A. Al-Samahi, Y. Zhang, and K. C. Ho, "Elliptic and hyperbolic localizations using minimum measurement solutions," *Signal Process.*, vol. 167, Feb. 2020, Art. no. 107273.

[36] I. Kang and H. Nam, "Robust localization system using vector combination in wireless sensor networks," *IEEE Access*, vol. 10, pp. 73437–73445, 2022.

[37] Y. Wang, K. C. Ho, and Z. Wang, "Robust localization under NLOS environment in the presence of isolated outliers by full-set TDOA measurements," *Signal Process.*, vol. 212, Jun. 2023, Art. no. 109159.

[38] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 4, pp. 692–730, Apr. 2017.

[39] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 549–557, Nov. 2003.

[40] K. Yamaoka, Y. Wakabayashi, and N. Ono, "Estimation of consistent time delays in subsample via auxiliary-function-based iterative updates," 2022, *arXiv:2203.09723*.

[41] R. Balakrishnan and K. Ranganathan, *A Textbook Graph Theory*, 2nd ed. Universitext. New York, NY, USA: Springer, 2012.

[42] R. Diestel, *Graph Theory* (Graduate Texts in Mathematics), 5th ed. Berlin, Germany: Springer, 2017.

[43] J. B. Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem," *Proc. Amer. Math. Soc.*, vol. 7, no. 1, pp. 48–50, 1956.

[44] H. Loberman and A. Weinberger, "Formal procedures for connecting terminals with a minimum total wire length," *J. ACM*, vol. 4, no. 4, pp. 428–437, Oct. 1957.

[45] V. Jarník, "O jistém problému minimálním," *Práce Moravské Přírodovědecké Společnosti*, vol. 6, no. 4, pp. 57–63, 1930.

[46] R. C. Prim, "Shortest connection networks and some generalizations," *Bell Syst. Tech. J.*, vol. 36, no. 6, pp. 1389–1401, Nov. 1957.

[47] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Math.*, vol. 1, no. 1, pp. 269–271, Dec. 1959.

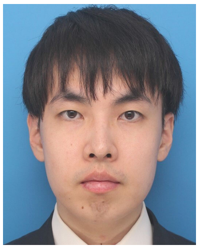
[48] D. Cheriton and R. E. Tarjan, "Finding minimum spanning trees," *SIAM J. Comput.*, vol. 5, no. 4, pp. 724–742, Dec. 1976.

[49] B. Chazelle, "A minimum spanning tree algorithm with inverse-Ackermann type complexity," *J. ACM*, vol. 47, no. 6, pp. 1028–1047, Nov. 2000.

[50] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using NetworkX," in *Proc. 7th Python Sci. Conf.*, G. Varoquaux, T. Vaught, and J. Millman, Eds., 2008, pp. 11–15.

[51] R. Scheibler, E. Bezzam, and I. Dokmanic, "Pyroomacoustics: A Python package for audio room simulation and array processing algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 351–355.

[52] G. Jacovitti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 525–533, Feb. 1993.



KOUEI YAMAOKA (Student Member, IEEE) received the B.Sc. and M.E. degrees in information engineering and engineering from the University of Tsukuba, Tsukuba, Japan, in 2017 and 2019, respectively. He is currently pursuing the Ph.D. degree with Tokyo Metropolitan University, Hino, Japan. His research interests include acoustic signal processing, signal enhancement, source localization, and asynchronous distributed microphone array.

He is a member of the Acoustical Society of Japan.

TAISHI NAKASHIMA (Student Member, IEEE) received the B.E. degree in engineering from Osaka University, Osaka, Japan, in 2019, and the M.S. degree in informatics from Tokyo Metropolitan University, Tokyo, Japan, in 2021, where he is currently pursuing the Ph.D. degree. His research interests include blind source separation and acoustic signal processing. He has received the JSPS Research Fellowship (DC1), in April 2021. He is also an esteemed Student Member of the Acoustical Society of Japan (ASJ) and the IEEE Signal Processing Society (SPS). He received the 24th Best Student Presentation Award of ASJ, the 16th IEEE SPS Japan Student Conference Paper Award, in 2022, and the Top 3% Recognition at ICASSP 2023.

YUKOH WAKABAYASHI (Member, IEEE) received the B.E. and M.E. degrees from Osaka University, Osaka, Japan, in 2008 and 2010, respectively, and the Ph.D. degree from Ritsumeikan University, Kyoto, Shiga, Japan, in 2017. In 2010, he joined Rohm Inc., Kyoto, Japan. From 2012 to 2014, he was an Assistant Researcher with Kyoto University, Kyoto. From 2018 to 2020, he was an affiliate Assistant Professor with Ritsumeikan University. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi, Japan, and the Faculty of Systems Design, Tokyo Metropolitan University, Hino, Japan. His research interests include acoustic signal processing, speech phase processing, array signal processing, and speaker diarization.

He is a member of the Institute of Electronics, Information and Communication Engineers and Acoustical Society of Japan. From 2016 to 2017, he was a recipient of the JSPS Research Fellowship for Young Scientists DC2.

NOBUTAKA ONO (Senior Member, IEEE) received the B.E., M.S., and Ph.D. degrees in mathematical engineering and information physics from The University of Tokyo, Tokyo, Japan, in 1996, 1998, and 2001, respectively.

He was a Research Associate with The University of Tokyo, in 2001, and became a Lecturer, in 2005. He was also an Associate Professor with the National Institute of Informatics, Tokyo, in April 2011, and became a Professor, in 2017. In 2017, he was with Tokyo Metropolitan University, Hino, Japan. He is the author or coauthor of more than 280 articles in international journal articles and peer-reviewed conference proceedings. His research interests include acoustic signal processing, especially microphone array processing, source localization and separation, machine learning, and optimization algorithms.

Dr. Ono is a Senior Member of IEEE Signal Processing Society and a member of the Acoustical Society of Japan (ASJ), Institute of Electronics, Information and Communications Engineers, Information Processing Society of Japan, and Society of Instrument and Control Engineers (SICE), Tokyo. He was a Tutorial Speaker with ISMIR 2010 and ICASSP 2018. He was the Chair of Signal Separation Evaluation Campaign Evaluation Committee, in 2013 and 2015, the Technical Program Chair of IWAENC 2018, the General Chair of DCASE 2020 Workshop, and a member of IEEE Audio and Acoustic Signal Processing Technical Committee, from 2014 to 2019. From 2012 to 2015, he was an Associate Editor of IEEE/ACM Transactions on Audio, Speech, and Language Processing. He is currently the Vice Chair of IEEE Signal Processing Society Tokyo Joint Chapter. He was a recipient of the Awaya Award from ASJ, in 2007, Igarashi Award at the Sensor Symposium from IEEJ, in 2004, Best Paper Award at IEEE ISIE, in 2008, the Measurement Division Best Paper Award from SICE, in 2013, Best Paper Award in IEEE IS3C, in 2014, Excellent Paper Award in IIHMSP, in 2014, Unsupervised Learning ICA Pioneer Award from SPIE, DSS, in 2015, Sato Paper Award from ASJ, in 2000 and 2018, two TAF Telecom System Technology Awards, in 2018, and Best Paper Award in APSIPA ASC, in 2018.