Structured Total Least Squares Based Internal Delay Estimation For Distributed Microphone Auto-Localization

Zhang, Jie; Hendriks, Richard C.; Heusdens, Richard

DOI
10.1109/iwaenc.2016.7602958

Publication date
2016

Document Version
Accepted author manuscript

Published in
2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)

STRUCTURED TOTAL LEAST SQUARES BASED INTERNAL DELAY ESTIMATION FOR DISTRIBUTED MICROPHONE AUTO-LOCALIZATION

Jie Zhang, Richard C. Hendriks, Richard Heusdens

Signal and Information Processing Lab., Delft University of Technology, The Netherlands
E-mail: {j.zhang-7, r.c.hendriks, r.heusdens}@tudelft.nl

ABSTRACT

Auto-localization in wireless acoustic sensor networks (WASNs) can be achieved by time-of-arrival (TOA) measurements between sensors and sources. Most existing approaches are centralized, and they require a fusion center to communicate with other nodes. In practice, WASN topologies are time-varying with nodes joining or leaving the network, which poses scalability issues for such algorithms. In particular, for an increasing number of nodes, the total transmission power required to reach the fusion center increases. Therefore, in order to facilitate scalability, we present a structured total least squares (STLS) based internal delay estimation for distributed microphone localization where the internal delay refers to the time taken for a source signal reaching a sensor to that it is registered as received by the capture device. Each node only needs to communicate with its neighbors instead of with a remote host, and they run an STLS algorithm locally to estimate local internal delays and positions (i.e., its own and those of its neighbors), such that the original centralized computation is divided into many subproblems. Experiments demonstrate that the decentralized internal delay estimation converges to the centralized results with increasing signal-to-noise ratio (SNR). More importantly, less computational complexity and transmission power are required to obtain comparable localization accuracy.

***Index Terms*—** Time-of-arrival, structured total least squares, internal delay estimation, auto-localization

1. INTRODUCTION

Wireless acoustic sensor networks (WASNs) have attracted increasing attention in the area of speech processing, due to their flexibility in sensor placement, e.g., [1–4]. However, in many applications like beamforming and source localization, the locations of microphones are assumed to be known as $a\ priori$. This is not always true, especially for dynamic network topologies with some nodes joining or disappearing.

Recently, many methods have been derived for microphone auto-localization, which can be generally categorised into methods based on received signal strength (RSS) [5], time-of-arrival (TOA) [6], angle-of-arrival (AOA) [7], time difference of arrival (TDOA) [8] and Euclidean distance matrices (EDM) [9, 10]. TOA and TDOA based techniques are popular in many applications because they are less vulnerable to multipath reflections, and they only require one receiver per sensor. Actually, TDOA based localization can be viewed as a special case of TOA based ones as the TDOA matrix can be obtained from TOA matrix [8]. Given the inter-sensor distances matrix, which can be obtained by multiplying the TOAs by the speed of sound, many techniques exist to estimate the coordinates of the sensors. One of the most common methods is multi-dimensional scaling (MDS) [11, 12]. MDS is designed to find the sensor locations given the inter sensor distances. However, MDS is based on an implicit assumption that the sensors and sources are co-located, thus limiting its use in practical applications. Alternatively, auto-localization can also be solved by using non-linear least-squares (LS), e.g., [9, 12–14].

The TOAs are usually incomplete with unknown source onset times or device capture times, such that before localization, it is required to estimate the unknown parameters. The source onset time denotes the time when a source signal is transmitted. The device capture time, also known as internal delay, denotes the time taken from a source signal reaching a sensor until it is registered as received by the capturing device. A number of algorithms are available to solve this problem, see e.g., [14–21]. Most make use of the low-rank information of TOA matrices, which is determined by the dimension of the space the sensors are located in. Although the unknowns are accurately computed, to some extent, these methods assume that the sources are located in the far field and often have slow rates of convergence. To remove these limitations, [6] presented a data fitting method based on structured total least squares (STLS), which is guaranteed to converge to the optimal solution. In [6], the STLS based internal delay estimation is realized by the Gauss-Newton iteration, with ultimately achieves a quadratic rate of convergence.

Most of the existing approaches for microphone localization are based on a centralized computation strategy, which requires a fusion centre (or host) for facilitating computations, communication and time-synchronization in the WASNs. This is a serious bottleneck for the reliability, scalability, communication and hardware costs. For instance, if the fusion center were to fail, the operation of the WASN will be compromised. In other words, the centralized algorithms are not scalable, while localization methods are required not to be influenced by changing network topologies.

Motivated by [6] and supposing each node is a simple microphone sensor (having a small CPU, e.g., smartphone, laptop) with some limited computational ability, this paper proposes a fully distributed microphone localization method based on STLS. We only employ the local TOA measurements to estimate the internal delays, which can be formulated as a low-rank approximation problem. After that, the sensor locations can be computed through a singular value decomposition of the matrix containing the relative arrival times, up to a $d \times d$ invertible matrix where d denotes the dimension of localization space. To this end, each node has a copy of the neighbors' internal delays and coordinates. Finally, we can calculate the positions of sensors by averaging information between neighbors. Experimental results show that the decentralized internal delay estimation converges to the centralized results with increasing sam-

Supported by the China Scholarship Council (NO. 201506010331).

978-1-5090-2007-2/16/$31.00

pling frequency or signal-to-noise ratio (SNR). More importantly, the computational complexity and transmission power are less than those of the centralized approach to obtain comparable accuracy.

This paper is organised as follow. In Sec. 2 we formalize the TOA-based localization problem. In Sec. 3 we describe the STLS for distributed internal delay estimation. Localization is discussed in Sec. 4 and the experimental results are shown in Sec. 5. Finally, the conclusions are drawn in Sec. 6.

2. PROBLEM FORMULATION

Consider the situation where we have to localize M receivers (e.g., wireless microphones, cellphones) (either near field or far field) using N sources (e.g., loudspeakers). The source locations $\boldsymbol{s}_j \in \mathbb{R}^d$ ($j \in \{1, ..., N\}$) and receiver locations $\boldsymbol{r}_i \in \mathbb{R}^d$ ($i \in \{1, ..., M\}$) are placed in space (d denotes the dimension of the linear manifold the sensors are located in). Let τ_j and δ_i denote the onset time of source $\boldsymbol{s}_j$ and the internal delay of receiver $\boldsymbol{r}_i$, respectively. Thus, the TOA measurement of the event generated by source $\boldsymbol{s}_j$ at receiver $\boldsymbol{r}_i$ is given by

$$t_{ij} = \frac{\|\boldsymbol{r}_i - \boldsymbol{s}_j\|}{c} + \tau_j + \delta_i, \tag{1}$$

where $\|\cdot\|$ denotes the Euclidean norm, c is the velocity of the calibration signal and we assume the measurement is noise free. The source onset times can be regarded as being known $a\ priori$, because we can generate the sources at known time instances, e.g., by using periodically generated wavelets [20]. This means that we can assume $\tau_j = 0$ for all j without loss of generality. Furthermore, setting $c = 1$ for notational convenience, the inter sensor distances satisfy

$$\|\boldsymbol{r}_i - \boldsymbol{s}_j\|^2 = (t_{ij} - \delta_i)^2, \text{ for all } i, j. \tag{2}$$

To simplify notations, we index the neighboring nodes of node k as $k_1, k_2, ..., k_{M_k}$, where M_k denotes the number of node k's neighbors, i.e., $M_k = |\mathcal{N}_k|$. Considering node k and its neighbors k_i ($k_i \in \mathcal{N}_k$), we have four equations similar to Eq. (2), given by

$$\|\boldsymbol{r}_{k_i} - \boldsymbol{s}_j\|^2 = (t_{k_i j} - \delta_{k_i})^2, \tag{3}$$

$$\|\boldsymbol{r}_{k_i} - \boldsymbol{s}_1\|^2 = (t_{k_i 1} - \delta_{k_i})^2, \tag{4}$$

$$\|\boldsymbol{r}_k - \boldsymbol{s}_j\|^2 = (t_{kj} - \delta_k)^2, \tag{5}$$

$$\|\boldsymbol{r}_k - \boldsymbol{s}_1\|^2 = (t_{k1} - \delta_k)^2. \tag{6}$$

With the operation of (4)+(5)-(3)-(6), we obtain

$$\begin{aligned}(\boldsymbol{r}_{k_i} - \boldsymbol{r}_k)^T(\boldsymbol{s}_j - \boldsymbol{s}_1) = \delta_{k_i}(t_{k_i j} - t_{k_i 1}) - \delta_k(t_{kj} - t_{k1}) \\ - (t_{k_i j}^2 - t_{k_i 1}^2 - t_{kj}^2 + t_{k1}^2)/2, \ \ k_i \in \mathcal{N}_k,\end{aligned} \tag{7}$$

which is bilinear with respect to the sensor and source locations. As a consequence, we can define the following matrices for the kth node as

$$\boldsymbol{R}_k = (\boldsymbol{r}_{k_1} - \boldsymbol{r}_k, ..., \boldsymbol{r}_{k_{M_k}} - \boldsymbol{r}_k) \in \mathbb{R}^{d \times M_k},$$

$$\boldsymbol{S} = (\boldsymbol{s}_2 - \boldsymbol{s}_1, ..., \boldsymbol{s}_N - \boldsymbol{s}_1) \in \mathbb{R}^{d \times (N-1)},$$

$$\boldsymbol{T}_k(i-1, j-1) = -(t_{ij}^2 - t_{i1}^2 - t_{kj}^2 + t_{k1}^2)/2 \in \mathbb{R}^{M_k \times (N-1)},$$
$$\text{for } i \in \{k_1, k_2, ..., k_{M_k}\}, j \in \{2, ..., N\},$$

$$\boldsymbol{W}_k(i, j-1) = t_{ij} - t_{i1} \in \mathbb{R}^{(M_k+1) \times (N-1)},$$
$$\text{for } i \in \{k, k_1, k_2, ..., k_{M_k}\}, j \in \{2, ..., N\}$$

$$\boldsymbol{E}_k = \begin{pmatrix} -\delta_k & \delta_{k_1} & 0 & \cdots & 0 \\ -\delta_k & 0 & \delta_{k_2} & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ -\delta_k & 0 & \cdots & 0 & \delta_{k_{M_k}} \end{pmatrix} \in \mathbb{R}^{M_k \times (M_k+1)}.$$

The internal delays that node k can access directly compose a vector $\boldsymbol{\delta}_k = (\delta_k, \delta_{k_1}, \delta_{k_2}, ..., \delta_{k_{M_k}})^T \in \mathbb{R}^{M_k+1}$. With this we can express Eq. (7) in a compact form given by

$$\boldsymbol{R}_k^T \boldsymbol{S} = \boldsymbol{T}_k + \boldsymbol{E}_k \boldsymbol{W}_k. \tag{8}$$

Assumption 1. *In a random geographic graph (RGG), we assume that $M_k \geq d$, which is a sufficient condition for rank$(\boldsymbol{R}_k) = r \leq d$.*

Remark 1. *With $M_k \geq d$, the matrices $\boldsymbol{R}_k$ and $\boldsymbol{S}$ have at least d rows, such that rank$(\boldsymbol{R}_k)$ = rank$(\boldsymbol{S}) = r \leq d$. In the absence of measurement errors the matrix $\boldsymbol{R}_k^T \boldsymbol{S}$ also has rank r as well as the right side of Eq. (8) $\boldsymbol{T}_k + \boldsymbol{E}_k \boldsymbol{W}_k$.*

To this end, the goal of this paper is to estimate the internal delays $\boldsymbol{\delta}_k$ (contained in the matrix $\boldsymbol{E}_k$) as well as the source/sensor positions (contained in the matrices $\boldsymbol{R}_k$ and $\boldsymbol{S}$) using the local TOA-based matrices $\boldsymbol{T}_k, \boldsymbol{W}_k$ and the low-rank information in a distributed way.

3. DISTRIBUTED STLS

In this section, we will estimate the internal delays based on the fact that the matrix $\boldsymbol{R}_k^T \boldsymbol{S}$ has rank r. We formulate this low-rank approximation problem as a structured total least squares (STLS) problem [22] similar to what has been done in [6]. In order to find a rank-r approximation matrix for $\boldsymbol{T}_k + \boldsymbol{E}_k \boldsymbol{W}_k$, we firstly write $\boldsymbol{T}_k = [\boldsymbol{A}_k \ \boldsymbol{B}_k]$, where $\boldsymbol{A}_k \in \mathbb{R}^{M_k \times r}, \boldsymbol{B}_k \in \mathbb{R}^{M_k \times (N-1-r)}$ and $\boldsymbol{W}_k = [\boldsymbol{F}_k \ \boldsymbol{G}_k]$, where $\boldsymbol{F}_k \in \mathbb{R}^{(M_k+1) \times r}, \boldsymbol{G}_k \in \mathbb{R}^{(M_k+1) \times (N-1-r)}$. Note that the matrices $\boldsymbol{T}_k$ and $\boldsymbol{W}_k$ can be calculated from the measured TOAs, such that the perturbation matrix of $\boldsymbol{T}_k$ is given by $\boldsymbol{E}_k \boldsymbol{W}_k$. We assume that $\boldsymbol{T}_k$ and $\boldsymbol{W}_k$ have full rank, and rank$(\boldsymbol{A}_k)$ = rank$(\boldsymbol{F}_k) = r$. Therefore, the rank-r approximation matrix for $\boldsymbol{T}_k + \boldsymbol{E}_k \boldsymbol{W}_k$ can be expressed as the following optimization problem,

$$\begin{aligned}&\min_{\boldsymbol{X}, \boldsymbol{\delta}_k} \|\boldsymbol{E}_k \boldsymbol{W}_k\|_F \\ &\text{s.t. } (\boldsymbol{A}_k + \boldsymbol{E}_k \boldsymbol{F}_k)\boldsymbol{X} = \boldsymbol{B}_k + \boldsymbol{E}_k \boldsymbol{G}_k.\end{aligned} \tag{9}$$

In practice, each node can be viewed as a micro-processor, such that it is capable of computation. Then, the optimization problem in Eq. (9) can be solved by the kth node separately. Given the rank information, the internal delay estimation based on the above optimization can be solved by rank approximation. And best rank-r approximation of a matrix has an analytic solution in terms of its singular value decomposition (SVD), which is given by the Eckart-Young-Mirsky theorem [23]. Actually, the STLS is an extension to TLS problem in the sense that it permits a known structure (e.g., rank-r) in $[\boldsymbol{A}_k \ \boldsymbol{B}_k]$ to be preserved in $[\boldsymbol{A}_k + \boldsymbol{E}_k \boldsymbol{F}_k, \ \boldsymbol{B}_k + \boldsymbol{E}_k \boldsymbol{G}_k]$.

In order to solve Eq. (9), we need to formulate the relationship between $\boldsymbol{E}_k$ and $\boldsymbol{\delta}_k$. Through observing the structure of $\boldsymbol{E}_k$, for each node k we can induce the sparse matrices $\boldsymbol{P}_{k,i} \in \mathbb{R}^{(M_k+1) \times (M_k+1)}, i = 1, ..., M_k$ with $\boldsymbol{P}_{k,i}(1,1) = -1, \boldsymbol{P}_{k,i}(i+1, i+1) = 1$, such that $\boldsymbol{E}_k = (\boldsymbol{P}_{k,1}\boldsymbol{\delta}_k, ..., \boldsymbol{P}_{k,M_k}\boldsymbol{\delta}_k)^T$. Note that $\boldsymbol{P}_{k,i}$ represents the ith sparse matrix of the kth node. As a result, we have

$$\|\boldsymbol{E}_k \boldsymbol{W}_k\|_F = \boldsymbol{\delta}_k^T \boldsymbol{Z}_k \boldsymbol{\delta}_k, \tag{10}$$

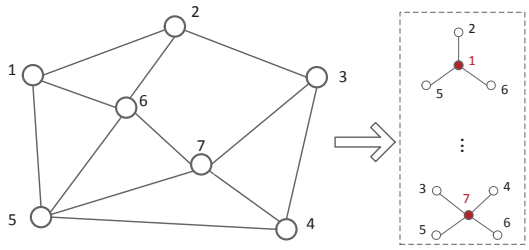


Fig. 1. An example of WASN with 7 nodes (each node has 3 neighbors at least). The original network topology is divided into many local star-like networks.

where $\boldsymbol{Z}_k = \sum_{i=1}^{M_k} \boldsymbol{P}_{k,i}\boldsymbol{W}_k\boldsymbol{W}_k^T\boldsymbol{P}_{k,i}^T$. With the fact that $\boldsymbol{Z}_k$ is positive definite and symmetric, it has an eigenvalue decomposition as $\boldsymbol{Z}_k = \boldsymbol{Q}_k\boldsymbol{\Lambda}_k\boldsymbol{Q}_k^T$ with $\boldsymbol{Q}_k$ unitary and $\boldsymbol{\Lambda}_k \succ 0$. Hence, $\boldsymbol{Z}_k$ can be decomposed as $\boldsymbol{Z}_k = \boldsymbol{D}_k\boldsymbol{D}_k^T$ with $\boldsymbol{D}_k = \boldsymbol{Q}_k\boldsymbol{\Lambda}_k^{\frac{1}{2}}\boldsymbol{Q}_k^T \succ 0$, i.e., the symmetric matrix $\boldsymbol{D}_k$ is the square root of $\boldsymbol{Z}_k$. Therefore, the optimization problem in Eq. (9) is equivalent to

$$\begin{aligned} &\min_{\boldsymbol{X},\boldsymbol{\delta}_k} \|\boldsymbol{D}_k\boldsymbol{\delta}_k\|_2 \\ &\text{s.t. } (\boldsymbol{A}_k + \boldsymbol{E}_k\boldsymbol{F}_k)\boldsymbol{X} = \boldsymbol{B}_k + \boldsymbol{E}_k\boldsymbol{G}_k, \end{aligned} \tag{11}$$

which is non-convex, because the constraint in terms of $\boldsymbol{\delta}_k$ is non-convex. Alternatively, we change Eq. (11) into an unconstrained minimization problem as

$$\min_{\boldsymbol{X},\boldsymbol{\delta}_k} \frac{1}{2}\|\boldsymbol{D}_k\boldsymbol{\delta}_k\|_2^2 + \frac{\omega^2}{2}\|\text{vec}(\boldsymbol{\rho}_k(\boldsymbol{X},\boldsymbol{\delta}_k))\|_2^2, \tag{12}$$

where $\boldsymbol{\rho}_k(\boldsymbol{X},\boldsymbol{\delta}_k) = \boldsymbol{B}_k + \boldsymbol{E}_k\boldsymbol{G}_k - (\boldsymbol{A}_k + \boldsymbol{E}_k\boldsymbol{F}_k)\boldsymbol{X}$, $\text{vec}(\cdot)$ is the vectorization operator and ω is a sufficiently large penalty value. As a consequence, we can solve $\boldsymbol{X}$ and $\boldsymbol{\delta}_k$ using a Gauss-Newton method in a decentralized manner. For the sake of brevity, we refer to [6] for an overview of this approach.

4. MICROPHONE AUTO-LOCALIZATION

After the internal delays are estimated by the Gauss-Newton algorithm, each node has an estimate of its own and its neighbors' internal delays. For example, in Fig. 1 node 1 has internal delay estimates of nodes $\{1, 2, 5, 6\}$. Then, we can compute the final internal delays by collecting data from neighbors and averaging over a local star network, like,

$$\delta_k = \frac{1}{1+M_k}\left(\delta_k + \sum_{i\in\mathcal{N}_k}\delta_i\right), \;\; k \in \{1, ..., M\}, \tag{13}$$

to reduce the estimation error. As a result, the right side of Eq. (8) is known (E_k is known), such that $\boldsymbol{R}_k^T\boldsymbol{S}$ has an SVD given by

$$\boldsymbol{R}_k^T\boldsymbol{S} = \boldsymbol{U}_k\boldsymbol{\Sigma}_k\boldsymbol{V}_k^T, \tag{14}$$

where $\boldsymbol{U}_k \in \mathbb{R}^{M_k\times r}$, $\boldsymbol{V}_k \in \mathbb{R}^{(N-1)\times r}$ and $\boldsymbol{\Sigma}_k \in \mathbb{R}^{r\times r}$, which determines $\boldsymbol{R}_k$ up to an $r \times r$ invertible matrix. The locations of the receivers and sources can be formulated as $\boldsymbol{R}_k = (\boldsymbol{U}_k\boldsymbol{C})^T$ and $\boldsymbol{S} = \boldsymbol{C}^{-1}\boldsymbol{\Sigma}_k\boldsymbol{V}_k^T$, where the matrix $\boldsymbol{C}$ can be obtained by non-linear optimization or LS approximation (if one source-receiver pair is co-located, a closed-form solution is known) [18].

To this end, each node has access to its own estimated position, those of its neighbors, as well as the positions of all sources. Let's consider again the example of a WASN with 7 nodes as depicted in Fig. 1. Node 1 has the estimated positions of $\{1, 2, 5, 6\}$, and it also holds the estimates of all source locations. This is true for all other sensors. Hence, for microphone auto-localization, node k only needs to collect data from its neighbors, and then do averaging as

$$\boldsymbol{r}_k = \frac{1}{1+M_k}\left(\boldsymbol{r}_k + \sum_{i\in\mathcal{N}_k}\boldsymbol{r}_i\right), \;\; k \in \{1, ..., M\}, \tag{15}$$

to reduce the estimation error. Of course, the procedure of collecting data can be also viewed as averaging consensus [24] over a local star network. Every node also has an estimate of the source positions in matrix $\boldsymbol{S}$. They will be different in general due to measurement noise. To reduce these variations, it is necessary to calculate the averaged source positions using averaging consensus. In this work, we only focus on the task of microphone self-localization with source localization left as future work.

5. EXPERIMENTAL RESULTS

In this section, we present experimental results and analysis for the internal delay estimation and microphone localization, respectively, and compare the decentralized STLS algorithm with the centralized STLS algorithm. In the following experiments, there are 15 sources placed uniformly at random in a room of dimensions $4 \times 4 \times 2.5$ m. The receivers are wirelessly connected as a random geographic graph (RGG), where the transmission range is determined by $\sqrt{\log M/M}$. Note that in practice, for a fixed enclosure as the number of receivers increases, the distribution of nodes becomes denser, and each node will have more neighbours, because the number of receivers increases (in linear sense) faster than the increase of transmission range (in logarithmic sense). For the receivers, the internal delays are generated according to an uniform distribution over the time interval $[0, 100]$ ms. The sound velocity is set to $c = 343$ m/s and the penalty value ω in Eq. (12) is chosen to be 10^9, which is kept the same for both the centralized and decentralized STLS algorithms. Furthermore, the programming platform is MATLAB 2014b, and the processor is i5-4690 CPU@3.50GHz.

Most literatures, like [9, 25], use signal-to-noise ratio (SNR) to represent the measurement noise level. With measurement noise present, the TOAs can be expressend as

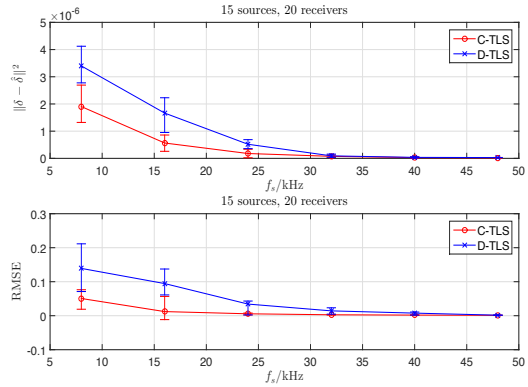
$$\hat{t}_{ij} = \frac{\|\boldsymbol{r}_i - \boldsymbol{s}_j\|}{c} + \delta_i + \nu_{ij}, \tag{16}$$

where ν_{ij} denotes the measurement errors, which are randomly draw from an uniform distribution over the interval $[-T_s/2, T_s/2]$ (similar to [6]) with T_s representing the sampling period (in seconds) of the calibration signals. Using the matrix formulations $\hat{\boldsymbol{t}} = [\hat{t}_{ij}], \boldsymbol{\nu} = [\nu_{ij}]$, the SNR is then defined as

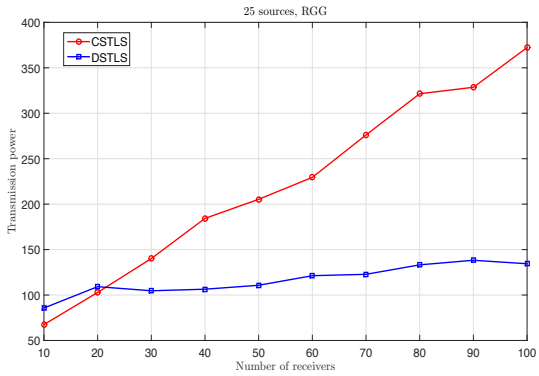
$$\text{SNR} = 20\log_{10}\frac{\|\hat{\boldsymbol{t}} - \boldsymbol{\nu}\|_F}{\|\boldsymbol{\nu}\|_F}. \tag{17}$$

Hereby, we clarify the relationship between the SNR and sampling frequency f_s with $f_s = 1/T_s$ in Fig. 3. We conclude that for a fixed sampling frequency, $\|\boldsymbol{\nu}\|_F$ becomes larger with an increasing number of receivers, such that the SNR gets slightly smaller.

Firstly, we show the results of internal delay estimation in terms of the sampling frequency f_s in the upper subplot of Fig. 2(a). The bars denote the statistical qualities (i.e., mean and variance) of internal delay estimation, which are averaged over 100 trails. It can be



(a) Results of internal delay estimation and localization



(b) Transmission power

Fig. 2. Comparisons between decentralized and centralized STLS based internal delay estimation for localization over RGG.

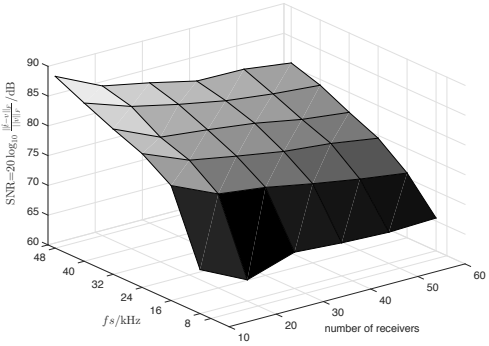


Fig. 3. Relationship between the SNR and f_s.

concluded that with noisy TOA measurements, the centralized STLS (CSTLS) always outperforms the decentralized STLS (DSTLS), but with increasing sampling frequency (i.e., SNR) the performance of DSTLS converges to that of CSTLS gradually. However, the computational complexity of the two methods are of order $O(MN)$ and $O(M_k N)$ with $M_k \ll M$, respectively. Hence, with increasing receivers the runtime of CSTLS increases linearly, while that of DSTLS stays approximately constant.

Secondly, the root-mean-square errors (RMSEs) of the microphone localization (i.e., $\sqrt{\frac{1}{M}\sum_{i=1}^{M}\|\boldsymbol{r}_i - \hat{\boldsymbol{r}}_i\|^2}$) are shown in the bottom subplot of Fig. 2(a) as a function of the sampling frequency. These results are quite consistent to the results of internal delay estimation. The localization error of DSTLS decreases with increasing sampling frequency, and it converges to that of CSTLS.

Finally, the transmission powers (TPs) of the two approaches are compared in Fig. 2(b). Since the TP is proportional to the squared Euclidean distance between sensors, we will measure the TP in terms of inter-sensor distances. For CSTLS, each sensor must transmit its TOA to the fusion center (assumed to be placed at the center, say $\boldsymbol{r}_c$, of the room), and the fusion center needs to broadcast the localization results back to the sensors after computation. Thus, the minimum TP of CSTLS will be

$$P_{T,CSTLS} = \sum_{i=1}^{M}\|\boldsymbol{r}_i - \boldsymbol{r}_c\|_2^2 + \max_{i\in\{1,\ldots,M\}}\|\boldsymbol{r}_i - \boldsymbol{r}_c\|_2^2. \quad (18)$$

For DSTLS, each sensor only requires to communicate with its neighbors, its minimum TP can be formulated as

$$\begin{aligned} P_{T,DSTLS} = &\sum_{i=1}^{M}\|\boldsymbol{r}_i - \arg\max_{\boldsymbol{r}_j,\, j\in\mathcal{N}_i}\{\boldsymbol{r}_i - \boldsymbol{r}_j\}\|_2^2 \\ &+ \sum_{i=1}^{M}\sum_{j=1}^{M_i}\|\boldsymbol{r}_i - \boldsymbol{r}_j\|_2^2, \end{aligned} \quad (19)$$

where the two terms on the right side of Eq. (19) represent the power consumption used to broadcast the measured TOAs and to collect data, respectively. Note that we ignore the influence of the size of transmitted data on the TP here. From Fig. 2(b), we can conclude that DSTLS requires significantly less TPs, such that it can save resources to prolong the lifetime of the WASNs.

6. CONCLUSIONS

In this paper, we extended the centralized STLS based internal delay estimation for microphone localization presented in [6] to a fully distributed framework. With the assumption that the minimum number of neighbors of each sensor node is the dimension of a localization space, the results of the proposed decentralized method converge to those of the centralized method with increasing SNR (i.e., sampling frequency). When the TOA measurements are noise free, the localization errors of the two approaches are identical. Futhermore, for an increasing number of sensors, the proposed algorithm achieves a significant reduction in transmission power and computational complexity as compared to the centralized case. Hence, the proposed method can improve the scalability, flexibility, and lifetime of WASNs.

7. REFERENCES

[1] A. Bertrand and M. Moonen, "Distributed node-specific LCMV beamforming in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 233 –246, Jan. 2012.

[2] S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed GSC beamforming using the relative transfer function," in *EURASIP Europ. Signal Process. Conf. (EUSIPCO)*, Aug. 2012, pp. 1274 – 1278.

[3] R. Heusdens, G. Zhang, R. C. Hendriks, Y. Zeng, and W. B. Kleijn, "Distributed MVDR beamforming for (wireless) microphone networks using message passing," in *Int. Workshop Acoustic Echo, Noise Control (IWAENC)*, 2012.

[4] Y. Zeng and R. C. Hendriks, "Distributed delay and sum beamformer for speech enhancement via randomized gossip," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 1, pp. 260–273, 2014.

[5] X. Sheng and Y. Hu, "Energy based acoustic source localization," in *ACM/IEEE Int. Conf. Inform. Process. in Sensor Networks*. Springer, 2003, pp. 285–300.

[6] R. Heusdens and N. Gaubitch, "Time-delay estimation for toa-based localization of multiple sensors," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2014, pp. 609–613.

[7] R. Peng and M. L. Sichitiu, "Angle of arrival localization for wireless sensor networks," in *IEEE Comm. Soc. Conf. Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, 2006, vol. 1, pp. 374–382.

[8] L. Wang, T. Hon, J. Reiss, and A. Cavallaro, "Self-localization of ad-hoc arrays using time difference of arrivals," *IEEE Trans. Signal Process.*, vol. 64, no. 4.

[9] M. J. Taghizadeh, A. Asaei, S. Haghighatshoar, P. N. Garner, and H. Bourlard, "Spatial sound localization via multipath euclidean distance matrix recovery," *IEEE J. of Selected Topics in Signal Process.*, vol. 9, no. 5, pp. 802–814, 2015.

[10] I. McCowan, M. Lincoln, and I. Himawan, "Microphone array shape calibration in diffuse noise fields," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 3, pp. 666–670, 2008.

[11] W. S. Torgerson, "Multidimensional scaling: I. theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, 1952.

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

[13] V. C. Raykar, I. V. Kozintsev, and R. Lienhart, "Position calibration of microphones and loudspeakers in distributed computing platforms," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 70–83, 2005.

[14] N. Ono, H. Kohno, N. Ito, and S. Sagayama, "Blind alignment of asynchronously recorded signals for distributed microphone array," in *IEEE Workshop Appl. Signal Process. Audio, Acoust. (WASPAA)*, 2009, pp. 161–164.

[15] S. Thrun, "Affine structure from sound," in *Advances in Neural Information Processing Systems*, 2005, pp. 1353–1360.

[16] M. Pollefeys and D. Nister, "Direct computation of sound and microphone locations from time-difference-of-arrival data.," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2008, pp. 2445–2448.

[17] M. H. Hennecke and G. A. Fink, "Towards acoustic self-localization of ad hoc smartphone arrays," in *Int. Workshop Hands-Free Speech Commun.*, 2011, pp. 127–132.

[18] M. Crocco, A. Del Bue, and V. Murino, "A bilinear approach to the position self-calibration of multiple sensors," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 660–673, 2012.

[19] M. Crocco, A. Del Bue, M. Bustreo, and V. Murino, "A closed form solution to the microphone position self-calibration problem," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2012, pp. 2597–2600.

[20] N. Gaubitch, W. B. Kleijn, and R. Heusdens, "Auto-localization in ad-hoc microphone arrays," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2013, pp. 106–110.

[21] Y. Kuang, S. Burgess, A. Torstensson, and K. Astrom, "A complete characterization and solution to the microphone position self-calibration problem," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2013, pp. 3875–3879.

[22] J. B. Rosen, H. Park, and J. Glick, "Total least norm formulation and solution for structured problems," *SIAM Journal on Matrix Analysis and Applications*, vol. 17, no. 1, pp. 110–126, 1996.

[23] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[24] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[25] A. Asaei, N. Mohammadiha, M. J. Taghizadeh, S. Doclo, and H. Bourlard, "On application of non-negative matrix factorization for ad hoc microphone array calibration from incomplete noisy distances," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2015, pp. 2694–2698.