Time Difference of Arrival Estimation Based on a Kronecker Product Decomposition

Xianrui Wang , Gongping Huang , *Graduate Student Member, IEEE*, Jacob Benesty , Jingdong Chen , *Fellow, IEEE*, and Israel Cohen , *Fellow, IEEE*

Abstract—**Time difference of arrival (TDOA) estimation, which often serves as the fundamental step for a source localization or a beamforming system, has a significant practical importance in a wide spectrum of applications. To deal with reverberation, the TDOA estimation problem is often transformed into one of identifying the relative acoustic impulse responses. This letter presents a method to efficiently identify the relative acoustic impulse response between two microphones for TDOA estimation based on the so-called Kronecker product decomposition. By decomposing the relative impulse response into a series of Kronecker products of shorter filters, the original channel identification problem with a long impulse response is converted into one of identifying a number of short filters. Since the TDOA information is embedded only in the direct path of the relative impulse response, the dimension of the Kronecker product decomposition can be very small and, as a result, the developed algorithm is expected to work well in real environments with a small number of data snapshots.**

Index Terms—**Kronecker product, time delay estimation.**

I. INTRODUCTION

TIME difference of arrival (TDOA) estimation, which aims at determining the difference of the arrival time of a signal between two sensors, has a significant importance in many applications [1]–[4]. It serves as an essential part of many voice communication systems such as acoustic source localization and beamforming [5]–[8]. Various TDOA algorithms have been developed in the literature, such as the generalized cross-correlation (GCC) and the steered response power (SRP) methods [15], [16], etc. An alternative way to estimate TDOA is based on the blind channel identification technique, which converts the TDOA estimation problem into one of identifying

the direct paths of the two channel impulse responses [13], or the direct path of the relative impulse response [14]. This way of TDOA estimation is found more robust to reverberation and multipath effects than the correlation based methods. However, how to accurately estimate long acoustic impulse responses in a blind manner in reverberant environments is still a challenging problem, especially when the length of the observation signal is short [17], [18].

The Kronecker product decomposition, which can decompose a long impulse response into the sum of several short ones, is very appealing in many applications [19], [20], such as system identification [21] and beamforming [22]–[25]. In [21], the authors proposed a system identification method for echo cancellation by approximating long acoustic impulse responses with Kronecker product decompositions, which demonstrated great advantages in terms of complexity and accuracy.

In this letter, we show that the principle of Kronecker product decomposition can also be applied to time delay estimation. By decomposing the relative impulse response as a series of Kronecker products of two shorter filters, the original channel identification problem with a long impulse response is converted into one of identifying a number of short filters. The advantage of this decomposition is that those short filters are much easier to estimate. For time delay estimation, the TDOA information is only embedded in the direct path so as long as the direct path is correctly identified, the problem would be solved and the overall accuracy of the relative impulse response identification should not matter much. As a result, the dimension of the Kronecker product can be very small, which makes TDOA estimation feasible even with a small number of data snapshots.

II. SIGNAL MODEL AND PROBLEM FORMULATION

To simplify the problem formulation and algorithm derivation, we consider only a dual-microphone system in this letter, but the presented method can be extended to the more general case of multiple microphones. The signals observed with the two microphones at the discrete-time index k can be expressed as

$$y(k) = g_1(k) * s(k) + v_1(k), \qquad (1a)$$

$$x(k) = g_2(k) * s(k) + v_2(k), \qquad (1b)$$

where $*$ denotes convolution, $g_1(k)$ and $g_2(k)$ are the acoustic impulse responses from the unknown source $s(k)$ to the two sensors, and $v_1(k)$ and $v_2(k)$ are the additive noise signals. All signals are assumed to be real, zero-mean, and reasonably broadband random processes.

If we choose the first microphone as the reference sensor, the observation signal at the second sensor can be written as

$$x(k) = h(k) * y(k) + v(k), \tag{2}$$

where $h(k)$ represents the relative acoustic impulse response between the first and second sensors with respect to the desired source and $v(k)$ is the noise, which consists of contributions from both $v_1(k)$ and $v_2(k)$.

Suppose that the relative acoustic impulse response can be modeled by a finite-impulse-response (FIR) filter of length L, i.e., $\mathbf{h} = [h_0 \ h_1 \ \cdots \ h_{L-1}]^T$. Then, the TDOA estimate can be determined as follows:

$$\hat{\tau} = \arg\max_l |h_l|. \tag{3}$$

Now, the TDOA estimation problem is converted to an identification one. Generally, $\mathbf{h}$ can be estimated by minimizing the mean-squared error (MSE) criterion defined as

$$\mathcal{J}(\mathbf{h}) \triangleq \mathbb{E}\left[\left| x(k - D) - \mathbf{h}^T \mathbf{y}(k) \right|^2 \right], \tag{4}$$

where $\mathbb{E}[\cdot]$ denotes mathematical expectation, D is a delay parameter configured to ensure that the relative acoustic impulse response can be modeled as a causal FIR filter [1], and

$$\mathbf{y}(k) = [y(k) \ y(k-1) \ \cdots \ y(k-L+1)]^T \tag{5}$$

is the observation vector of length L. Minimizing the MSE in (4) yields the Wiener solution:

$$\mathbf{h} = \mathbf{R}^{-1}\mathbf{r}, \tag{6}$$

where $\mathbf{R} \triangleq \mathbb{E}[\mathbf{y}(k)\mathbf{y}^T(k)]$ and $\mathbf{r} \triangleq \mathbb{E}[\mathbf{y}(k)x(k-D)]$.

Since the relative acoustic impulse response is in general time varying due to reasons such as source movement, the estimates obtained from (6) are generally not satisfactory in practice. One way to deal with this is to estimate $\mathbf{h}$ in an adaptive manner using, for example, the least-mean-square (LMS) algorithm [1], [14]. However, how to achieve good steady-state and tracking performances with adaptive algorithms is challenging in room acoustic environments, where the relative impulse response is generally long and time varying [17], [18].

III. TDOA ESTIMATION WITH KRONECKER PRODUCT DECOMPOSITION

The relative impulse response can be approximated by a Kronecker product decomposition as [21]

$$\mathbf{h} \approx \sum_{p=1}^{P} \mathbf{h}_{2,p} \otimes \mathbf{h}_{1,p}, \tag{7}$$

where $\otimes$ is the Kronecker product, $\mathbf{h}_{2,p}$ and $\mathbf{h}_{1,p}$, $p = 1, 2, \ldots, P$, are short impulse responses of lengths L_2 and L_1, respectively, with $L = L_1 L_2$, $L_1 \geq L_2$, and $P \leq L_2$ [if $P = L_2$, then (7) becomes a strict equality]. Since our objective is to estimate TDOA, which is reflected only in the direct path of the relative impulse response, we can choose a small value of P.

To illustrate this decomposition, we provide an example in which room impulse responses are simulated with the image model method [26] and the corresponding reverberation time, T_{60}, is set to be approximately 300 ms (see Sec. IV for details). The lengths of the two simulated impulse responses are both
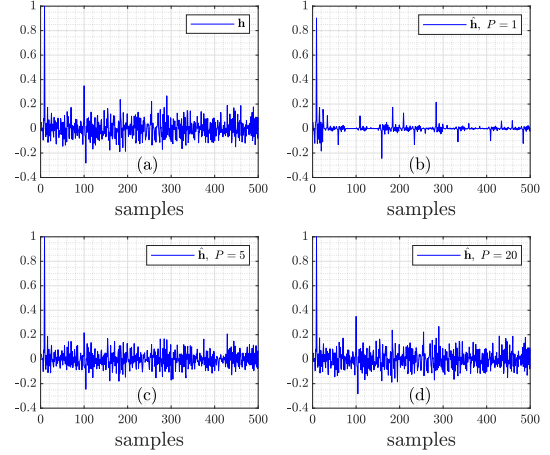

Fig. 1. Relative impulse responses: (a) ground truth, (b) Kronecker product decomposition with $P = 1$, (c) Kronecker product decomposition with $P = 5$, and (d) Kronecker product decomposition with $P = 20$. Conditions: $L = 500$, $L_1 = 25$, and $L_2 = 20$.

1024. The relative impulse response is then computed using the spectral division method and its length is truncated to $L = 500$. This computed relative impulse response is then treated as the ground truth. In the Kronecker decomposition, the parameters are set to $L_1 = 25$ and $L_2 = 20$, respectively. Figure 1 plots the relative impulse response and the approximated ones with $P = 1$, $P = 5$, and $P = 20$. As seen, a good approximation of the relative impulse response, particularly the direct path, is achieved even with a small value of P.

Now, using the following relationships:

$$\mathbf{h}_{2,p} \otimes \mathbf{h}_{1,p} = \mathbf{H}_{2,p}\mathbf{h}_{1,p} = \mathbf{H}_{1,p}\mathbf{h}_{2,p}, \tag{8}$$

where

$$\begin{aligned} \mathbf{H}_{2,p} &= \mathbf{h}_{2,p} \otimes \mathbf{I}_{L_1}, \\ \mathbf{H}_{1,p} &= \mathbf{I}_{L_2} \otimes \mathbf{h}_{1,p}, \end{aligned} \tag{9}$$

and $\mathbf{I}_{L_1}$ and $\mathbf{I}_{L_2}$ are the identity matrices of sizes $L_1 \times L_1$ and $L_2 \times L_2$, respectively, we have

$$
\begin{aligned}
\mathbf{h}^T \mathbf{y}(k) &= \left(\sum_{p=1}^{P} \mathbf{h}_{2,p} \otimes \mathbf{h}_{1,p} \right)^T \mathbf{y}(k) \\
&= \sum_{p=1}^{P} \mathbf{h}_{1,p}^T \mathbf{H}_{2,p}^T \mathbf{y}(k) \\
&= \sum_{p=1}^{P} \mathbf{h}_{1,p}^T \mathbf{y}_{2,p}(k) = \underline{\mathbf{h}}_1^T \underline{\mathbf{y}}_2(k),
\end{aligned}
\tag{10}
$$

where

$$\mathbf{y}_{2,p}(k) = \mathbf{H}_{2,p}^T \mathbf{y}(k),$$

$$\underline{\mathbf{h}}_1 = \left[\mathbf{h}_{1,1}^T \ \mathbf{h}_{1,2}^T \ \cdots \ \mathbf{h}_{1,P}^T \right]^T,$$

$$\underline{\mathbf{y}}_2(k) = \left[\mathbf{y}_{2,1}^T(k) \ \mathbf{y}_{2,2}^T(k) \ \cdots \ \mathbf{y}_{2,P}^T(k) \right]^T.$$

Similarly, $\mathbf{h}^T\mathbf{y}(k)$ can also be written as

$$\mathbf{h}^T\mathbf{y}(k) = \sum_{p=1}^{P} \mathbf{h}_{2,p}^T \mathbf{H}_{1,p}^T \mathbf{y}(k)$$

$$= \sum_{p=1}^{P} \mathbf{h}_{2,p}^T \mathbf{y}_{1,p}(k) = \underline{\mathbf{h}}_2^T \underline{\mathbf{y}}_1(k), \qquad (11)$$

where

$$\mathbf{y}_{1,p}(k) = \mathbf{H}_{1,p}^T \mathbf{y}(k),$$

$$\underline{\mathbf{h}}_2 = \left[\mathbf{h}_{2,1}^T \; \mathbf{h}_{2,2}^T \; \cdots \; \mathbf{h}_{2,P}^T \right]^T,$$

$$\underline{\mathbf{y}}_1(k) = \left[\mathbf{y}_{1,1}^T(k) \; \mathbf{y}_{1,2}^T(k) \; \cdots \; \mathbf{y}_{1,P}^T(k) \right]^T.$$

Substituting (10) into (4), one can write the MSE as

$$\mathcal{J}(\mathbf{h}) = \mathcal{J}(\underline{\mathbf{h}}_1, \underline{\mathbf{h}}_2) = E\left[\left| x(k-D) - \underline{\mathbf{h}}_1^T \underline{\mathbf{y}}_2(k) \right|^2 \right]$$

$$= \sigma^2 - 2\underline{\mathbf{h}}_1^T \underline{\mathbf{r}}_2 + \underline{\mathbf{h}}_1^T \underline{\mathbf{R}}_2 \underline{\mathbf{h}}_1$$

$$= E\left[\left| x(k-D) - \underline{\mathbf{h}}_2^T \underline{\mathbf{y}}_1(k) \right|^2 \right]$$

$$= \sigma^2 - 2\underline{\mathbf{h}}_2^T \underline{\mathbf{r}}_1 + \underline{\mathbf{h}}_2^T \underline{\mathbf{R}}_1 \underline{\mathbf{h}}_2, \qquad (12)$$

where

$$\sigma^2 = \mathbb{E}\left[|x(k-D)|^2 \right],$$

$$\underline{\mathbf{r}}_1 = \mathbb{E}\left[\underline{\mathbf{y}}_1(k) x(k-D) \right],$$

$$\underline{\mathbf{r}}_2 = \mathbb{E}\left[\underline{\mathbf{y}}_2(k) x(k-D) \right],$$

$$\underline{\mathbf{R}}_1 = \mathbb{E}\left[\underline{\mathbf{y}}_1(k) \underline{\mathbf{y}}_1^T(k) \right],$$

$$\underline{\mathbf{R}}_2 = \mathbb{E}\left[\underline{\mathbf{y}}_2(k) \underline{\mathbf{y}}_2^T(k) \right].$$

Since $\underline{\mathbf{h}}_1$ and $\underline{\mathbf{h}}_2$ are embedded in each other, it is difficult to derive a closed-form solution. Hence we resort to an iterative approach [21]. We first initialize $\underline{\mathbf{h}}_2$ and compute the corresponding $\underline{\mathbf{r}}_2$ and $\underline{\mathbf{R}}_2$. When $\underline{\mathbf{h}}_2$ is fixed, the MSE in (4) can be written as

$$\mathcal{J}(\underline{\mathbf{h}}_1|\underline{\mathbf{h}}_2) = \sigma^2 - 2\underline{\mathbf{h}}_1^T \underline{\mathbf{r}}_2 + \underline{\mathbf{h}}_1^T \underline{\mathbf{R}}_2 \underline{\mathbf{h}}_1 \qquad (13)$$

and the estimate of $\underline{\mathbf{h}}_1$ is computed as

$$\underline{\mathbf{h}}_1 = \underline{\mathbf{R}}_2^{-1} \underline{\mathbf{r}}_2. \qquad (14)$$

We then compute the corresponding $\underline{\mathbf{r}}_1$ and $\underline{\mathbf{R}}_1$. With $\underline{\mathbf{h}}_1$ being fixed, the MSE in (4) can be written as

$$\mathcal{J}(\underline{\mathbf{h}}_2|\underline{\mathbf{h}}_1) = \sigma^2 - 2\underline{\mathbf{h}}_2^T \underline{\mathbf{r}}_1 + \underline{\mathbf{h}}_2^T \underline{\mathbf{R}}_1 \underline{\mathbf{h}}_2. \qquad (15)$$

Minimizing this MSE yields

$$\underline{\mathbf{h}}_2 = \underline{\mathbf{R}}_1^{-1} \underline{\mathbf{r}}_1. \qquad (16)$$

Continuing the iterations up to n times, we obtain the final solution of $\underline{\mathbf{h}}_1$ and $\underline{\mathbf{h}}_2$, and the optimal solution of $\mathbf{h}$ is calculated using (7). After an estimate of $\mathbf{h}$ is achieved, the TDOA estimate is determined by identifying the direct path as

$$\hat{\tau} = \arg\max_l |h_l| - D. \qquad (17)$$

The proposed algorithm is summarized in Table I.

TABLE I
PROPOSED ALGORITHM

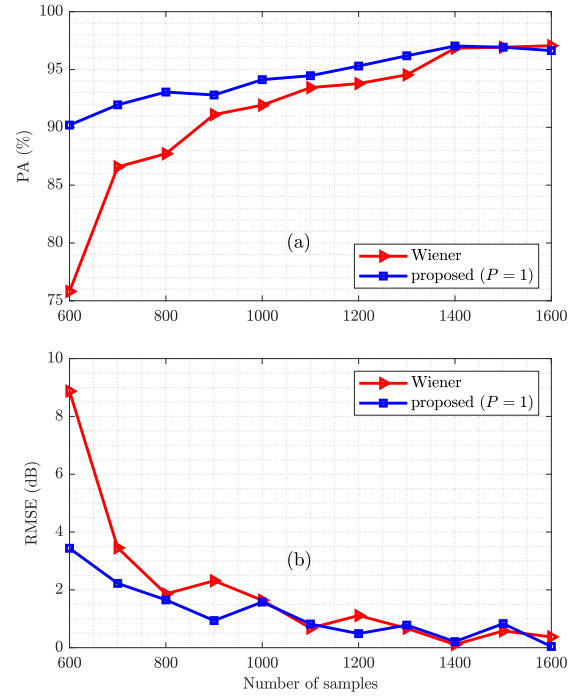|	
Inputs:	$\mathbf{x}(k)$ and $\mathbf{y}(k)$ (signal vectors)
Initialization:	$\underline{\mathbf{h}}_2^{(0)}$
	Compute $\underline{\mathbf{y}}_2^{(0)}(k)$
	Compute $\underline{\mathbf{R}}_2^{(0)}$ and $\underline{\mathbf{r}}_2^{(0)}$
Iteration:	for $i = 1, 2, \ldots, n$
	$\quad \underline{\mathbf{h}}_1^{(i)} = \left(\underline{\mathbf{R}}_2^{(i-1)} + \delta\mathbf{I}_{PL_1} \right)^{-1} \underline{\mathbf{r}}_2^{(i-1)}$
	$\quad$ Compute $\underline{\mathbf{y}}_1^{(i)}(k)$
	$\quad$ Compute the matrix $\underline{\mathbf{R}}_1^{(i)}$ and the vector $\underline{\mathbf{r}}_1^{(i)}$
	$\quad \underline{\mathbf{h}}_2^{(i)} = \left(\underline{\mathbf{R}}_1^{(i)} + \delta\mathbf{I}_{PL_2} \right)^{-1} \underline{\mathbf{r}}_1^{(i)}$
	$\quad$ Compute $\underline{\mathbf{y}}_2^{(i)}(k)$
	$\quad$ Compute the matrix $\underline{\mathbf{R}}_2^{(i)}$ and the vector $\underline{\mathbf{r}}_2^{(i)}$
	end

Fig. 2. Performance of the Wiener and proposed (with $P = 1$) methods as a function of the number of snapshot samples: (a) PA and (b) RMSE. Condition: $T_{60} = 300$ ms, input SNR $= 15$ dB, and the background noise is white Gaussian noise.

IV. SIMULATIONS

In this section, we study the performance of the proposed method in reverberant and noisy acoustic environments. We consider a room of size $7 \times 6 \times 4$ m, where 2 omnidirectional microphones are located at $(3.4, 3.0, 1.8)$ and $(3.6, 3.0, 1.8)$, respectively, and a sound source is placed at $(1.95, 3.96, 1.55)$. The acoustic channel impulse responses from the source to the microphones are generated with the image method [26]. The microphone signals are generated by convolving the clean source signal with the generated impulse responses and white Gaussian noise is then added to control the input SNR. The sampling rate is 16 kHz.

The observation signals are partitioned into non-overlapping time frames and non-speech frames are detected by a voice
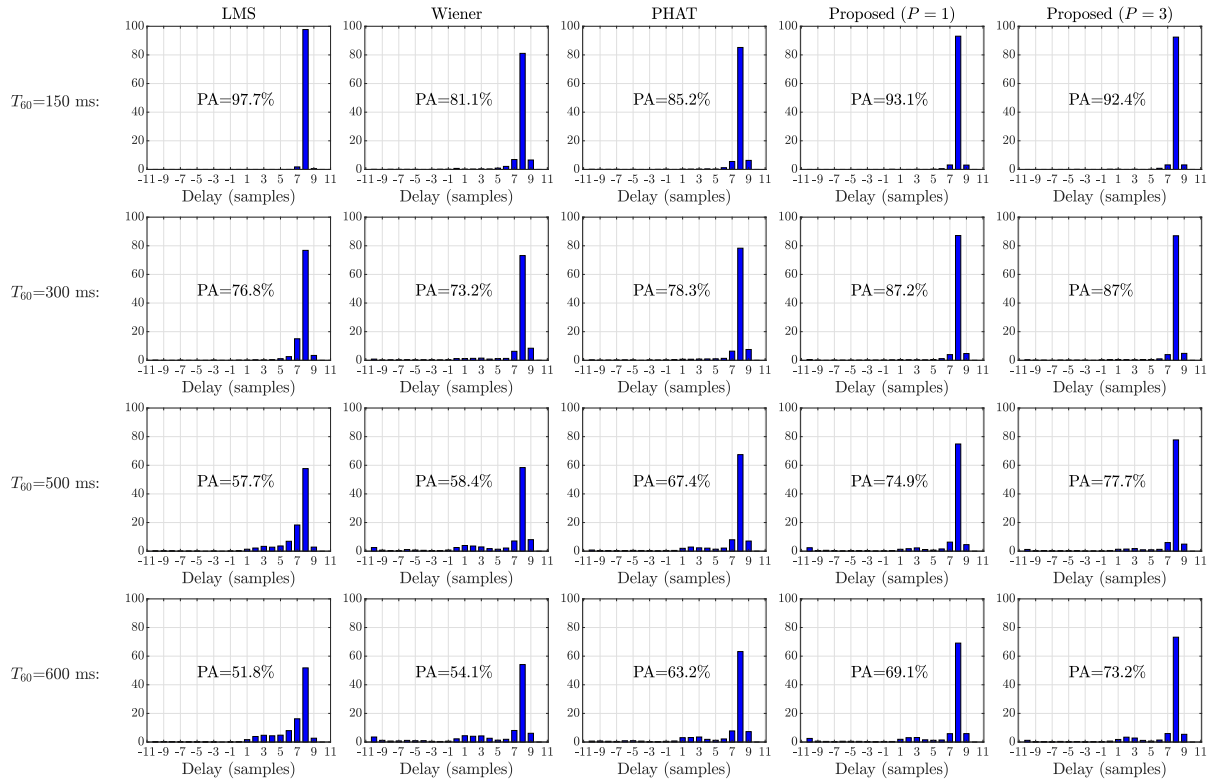
Fig. 3.   Histograms of the TDOA estimates of the LMS, Wiener, PHAT, and proposed methods. Conditions: $T_{60} = 150$ ms, 300 ms, 500 ms, and 600 ms, input SNR = 15 dB, and the background noise is white Gaussian noise.

activity detector (VAD) and are then removed. For frames with the presence of speech, TDOA estimation is performed on a frame-by-frame basis. The proposed method is implemented with the following parameter configuration: $L = 500$, $L_1 = 25$, $L_2 = 20$, $D = 10$ (slightly larger than the maximum possible delay given the spacing), and the number of iterations $n = 5$. The filter $\underline{\mathbf{h}}_2$ is initialized as $\mathbf{h}_{2,p} = [0.1 \ 0 \ \cdots \ 0]^T$, $p = 1, 2, \ldots, P$.

We first show the performance of the Wiener and proposed (with $P = 1$) methods as a function of the number of snapshots (length of frames) in noisy and reverberant environments, where the reverberation time, T_{60}, is approximately 300 ms and the input SNR is 15 dB. The number of snapshots varies from 600 to 1600 with an increment of 100 samples. A total of 1000 Monte Carlo simulations are carried out. We use the percentage of accurate (PA) estimates (where the term accurate estimate means that the TDOA estimate is the same as the ground truth) and the root mean-squared error (RMSE) of all the estimates as the evaluation metrics. The results are plotted in Fig. 2. As seen, when the number of snapshots is small, the proposed method demonstrates better performance, which can be explained as follows. The sizes of the matrices $\underline{\mathbf{R}}_1$ and $\underline{\mathbf{R}}_2$ are, respectively, $PL_2 \times PL_2$ and $PL_1 \times PL_1$. For $P \ll L_2$, the dimensions of the two matrices are much smaller than the dimension of the matrix $\mathbf{R}$, which is used in the Wiener method. As a result, less data is needed to achieve reliable estimates of these two matrices, leading to better performance of the proposed method.

We also compare the proposed method with the LMS, Wiener, and phase transform (PHAT) methods [12]. We consider four

different reverberation conditions, i.e., $T_{60} = 150$ ms, 300 ms, 500 ms, and 600 ms, all with an input SNR of 15 dB. The number of snapshots is set to 600. The histograms of the TDOA estimates with 2500 Monte Carlo simulations are plotted in Fig. 3. As seen, the LMS method achieves the best performance for the condition with $T_{60} = 150$ ms, but it suffers from significant performance degradation as reverberation becomes stronger, indicating that this method is sensitive to reverberation. In comparison, the proposed method outperforms the other compared methods except for $T_{60} = 150$ ms (note that the proposed approach is computationally more demanding than the PHAT and LMS methods). It is also observed that using a reasonably larger value of P can improve the performance of the proposed method in highly reverberant environments.

V. CONCLUSION

This letter studied the problem of TDOA estimation and presented a method based on the Kronecker product decomposition. We decomposed the relative impulse response between two sensors as a series of Kronecker products of two short filters, and the optimal filters are solved with an iterative method. Since the core issue is to identify the direct path of the relative impulse response, a small dimension of the Kronecker product is sufficient for the problem. When the number of data snapshots is small, the proposed approach yields better TDOA estimation than traditional channel identification based methods. Simulation results demonstrated that the proposed method outperforms several widely-used conventional methods in reverberant environments.

REFERENCES

[1] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: An overview," *EURASIP J. Appl. Signal Process.*, vol. 2006, no. 1, pp. 170–170, 2006.

[2] A. Lombard, Y. Zheng, H. Buchner, and W. Kellermann, "TDOA estimation for multiple sound sources in noisy and reverberant environments using broadband independent component analysis," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 6, pp. 1490–1503, Aug. 2011.

[3] F. Nesta and M. Omologo, "Generalized state coherence transform for multidimensional TDOA estimation of multiple sources," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 1, pp. 246–260, Jan. 2012.

[4] 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.

[5] T. Long, J. Chen, G. Huang, J. Benesty, and I. Cohen, "Acoustic source localization based on geometric projection in reverberant and noisy environments," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 1, pp. 143–155, Mar. 2019.

[6] F. Borra, F. Antonacci, A. Sarti, and S. Tubaro, "Localization of acoustic sources in the ray space for distributed microphone sensors," in *Proc. IEEE WASPAA*, 2017, pp. 170–174.

[7] G. Huang, J. Benesty, and J. Chen, "On the design of frequency-invariant beampatterns with uniform circular microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 5, pp. 1140–1153, May 2017.

[8] G. Huang, J. Chen, and J. Benesty, "Insights into frequency-invariant beamforming with concentric circular microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 12, pp. 2305–2318, Dec. 2018.

[9] M. Compagnoni *et al.*, "A geometrical–statistical approach to outlier removal for TDOA measurements," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3960–3975, Aug. 2017.

[10] D. Salvati, C. Drioli, and G. L. Foresti, "Incoherent frequency fusion for broadband steered response power algorithms in noisy environments," *IEEE Signal Process. Lett.*, vol. 21, no. 5, pp. 581–585, May 2014.

[11] K. Kowalczyk, E. A. Habets, W. Kellermann, and P. A. Naylor, "Blind system identification using sparse learning for TDOA estimation of room reflections," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 653–656, Jul. 2013.

[12] 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.

[13] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *J. Acoust. Soc. Amer.*, vol. 107, no. 1, pp. 384–391, 2000.

[14] F. A. Reed, P. L. Feintuch, and N. J. Bershad, "Time delay estimation using the LMS adaptive filter–static behavior," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-29, no. 3, pp. 561–571, Jun. 1981.

[15] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays*, New York, NY, USA: Springer, 2001, pp. 157–180.

[16] G. Huang, J. Chen, and J. Benesty, "Direction-of-arrival estimation of passive acoustic sources in reverberant environments based on the Householder transformation," *J. Acoust. Soc. Amer.*, vol. 138, no. 5, pp. 3053–3060, 2015.

[17] R. Talmon, I. Cohen, and S. Gannot, "Relative transfer function identification using convolutive transfer function approximation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 4, pp. 546–555, May 2009.

[18] I. Cohen, "Relative transfer function identification using speech signals," *IEEE Trans. Speech, Audio Process.*, vol. 12, no. 5, pp. 451–459, Sep. 2004.

[19] Y. I. Abramovich, G. J. Frazer, and B. A. Johnson, "Iterative adaptive Kronecker MIMO radar beamformer: Description and convergence analysis," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3681–3691, Jul. 2010.

[20] M. Bousse, O. Debals, and L. De Lathauwer, "A tensor-based method for large-scale blind source separation using segmentation," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 346–358, Jan. 2017.

[21] C. Paleologu, J. Benesty, and S. Ciochina, "Linear system identification based on a Kronecker product decomposition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 10, pp. 1793–1809, Oct. 2018.

[22] W. Yang, G. Huang, J. Benesty, I. Cohen, and J. Chen, "On the design of flexible Kronecker product beamformers with linear microphone arrays," in *Proc. IEEE ICASSP*, 2019, pp. 441–445.

[23] J. Benesty, I. Cohen, and J. Chen, *Array Processing–Kronecker Product Beamforming*. Berlin, Germany: Springer-Verlag, 2019.

[24] G. Huang, J. Benesty, J. Chen, and I. Cohen, "Robust and steerable Kronecker product differential beamforming with rectangular microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 211–215.

[25] G. Huang, I. Cohen, J. Benesty, and J. Chen, "Kronecker product beamforming with multiple differential microphone arrays," in *Proc. IEEE SAM*, 2020, pp. 1–5.

[26] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.