ACOUSTIC SYSTEM IDENTIFICATION WITH PARTIALLY TIME-VARYING MODELS BASED ON TENSOR DECOMPOSITIONS

Gongping Huang[1], *Jacob Benesty*[2], *Jingdong Chen*[3], *Constantin Paleologu*[4], *Silviu Ciochină*[4], *Walter Kellermann*[1], *and Israel Cohen*[5]

[1]LMS, University Erlangen-Nuremberg, 91058 Erlangen, Germany
[2]INRS-EMT, University of Quebec, Montreal, QC H5A 1K6, Canada
[3]CIAIC, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China
[4] University Politehnica of Bucharest, 061071, Bucharest, Romania
[5]Technion–Israel Institute of Technology, Haifa 3200003, Israel

ABSTRACT

Acoustic system identification, which aims at estimating the channel impulse response from a source of interest to the microphone position, plays an important role in many applications, e.g., echo cancellation for full-duplex speech communication. Generally, an acoustic channel impulse response is modeled as a linear finite-impulse-response (FIR) filter, so the objective of system identification is to identify it. While much effort has been devoted to this topic over the last five decades, identifying the room FIR filters accurately with only a small number of observation data snapshots remains a significant challenge. This paper studies this problem and proposes to model the acoustic impulse response, i.e., the FIR filter, with a tensor decomposition, which can be expressed as a multidimensional Kronecker product of a series of shorter filters. Then, a partially time-varying model is applied to acoustic system identification, where the global filter is decomposed into two parts: a time-invariant part, which captures the common properties of acoustic channels, and a time-varying part, which, as its name indicates, represents the components of acoustic channels that change with time. During the identification process, the time-invariant filters can be identified or learned in advance, while the time-varying filters are optimized through an iterative procedure. Simulation results demonstrate that the proposed technique can achieve better acoustic system identification performance with a small number of data snapshots.

***Index Terms*—**Acoustic system identification, Kronecker product decomposition, tensor decomposition, Wiener filter, iterative algorithm.

1. INTRODUCTION

Acoustic system identification, studied intensively over the last few decades, is of great importance in many applications [1–4]. In most scenarios, the output signal of an acoustic system is a linear convolution between the excitation (input) signal and the room impulse response, which is generally modeled as a linear finite-impulse-response (FIR) filter. So, the problem is identifying the acoustic FIR filter, which is typically achieved by optimizing the mean-squared error (MSE) criterion [5,6]. While this approach has been around for decades, achieving accurate and robust estimation with a small number of observation data snapshots remains a significant challenge.

The Kronecker product approach, which decomposes a long filter into the sum of several short ones, is very appealing in many applications [7–15]. In [16, 17], the authors propose a system identification method by exploiting the so-called nearest Kronecker product decomposition, which can obtain more accurate identification results with fewer observations as compared to the conventional Wiener solution. Another benefit of this method is that it is computationally more efficient as the matrices to be inverted are smaller than those with the conventional approaches [18–20].

Most existing methods model the acoustic impulse responses from different source positions as separate FIR filters though they are excited in the same acoustic environment. We find through our investigation that the impulse responses from adjacent source positions usually share some common properties even though the misalignments between those impulse responses may be large. Those common properties would be helpful for acoustic system identification, which will be investigated in this work. We first model the acoustic impulse response with a tensor decomposition [21], where the existing Kronecker product decomposition can be considered as a special case, i.e., two-way tensor [17]. Then, we propose a partially time-varying model for acoustic impulse response, where the global filter is divided into two parts: a time-invariant one and a time-varying part. The former can be identified or learned in advance, which we do not discuss in detail here, and can then be used as *a priori* information for the time-varying part. The focus of this work is on the second part, and a method is presented to identify the time-varying part of the acoustic FIR filter.

The rest of this paper is organized as follows. Section 2 describes the signal model and system identification problem. Section 3 introduces the tensor decomposition of acoustic impulse responses. Section 4 presents the proposed acoustic system identification method with partially time-varying model. Sections 5 and 6 present, respectively, some simulation results and conclusions.

2. SIGNAL MODEL AND PROBLEM FORMULATION

Consider an acoustic system with an excitation signal $x(k)$, where k denotes the time index; the output signal is expressed as

$$y(k) = \sum_{l=0}^{L-1} h_l x(k-l) + v(k) \\ = \mathbf{h}^T \mathbf{x}(k) + v(k), \quad (1)$$

where $h_l,\ l = 0, 1, \ldots, L-1$ are the real-valued coefficients of the acoustic impulse response of length L, $v(k)$ is the corrupting additive noise at the output signal, $\mathbf{h} = [\ h_0 \quad h_1 \quad \cdots \quad h_{L-1}\ ]^T$, the superscript ^{T} denotes the transpose of a vector or a matrix, and $\mathbf{x}(k) = [\ x(k) \quad x(k-1) \quad \cdots \quad x(k-L+1)\ ]^T$ is the observation signal of length L.

This research was supported in part by the Alexander von Humboldt Foundation, in part by the Pazy Research Foundation, in part by the National Key Research and Development Program of China under Grant No. 2018AAA0102200 and in part by the Key Program of National Science Foundation of China (NSFC) under Grants 61831019 and 62192713, in part by CNCS-UEFISCDI Romania under Grant PN-III-P4-PCE-2021-0438.

978-1-6654-6867-1/22/$31.00 ©2022 IEEE

Linear system identification aims at estimating the impulse response, $\mathbf{h}$. Let us denote the modeling filter as $\widehat{\mathbf{h}}$ and assume that its length is also L. Passing the reference signal, i.e., $x(k)$, through the modeling filter gives an estimate of the system output, and the resulting estimation error is written as

$$e(k) = y(k) - \widehat{\mathbf{h}}^T \mathbf{x}(k). \quad (2)$$

The MSE criterion can then be defined as

$$\mathcal{J}\left(\widehat{\mathbf{h}}\right) = E\left[e^2(k)\right] = E\left[\left|y(k) - \widehat{\mathbf{h}}^T \mathbf{x}(k)\right|^2\right], \quad (3)$$

where $E[\cdot]$ denotes mathematical expectation. The minimization of $\mathcal{J}\left(\widehat{\mathbf{h}}\right)$ leads to the well-known Wiener filter:

$$\widehat{\mathbf{h}} = \mathbf{R}^{-1}\mathbf{r}, \quad (4)$$

where $\mathbf{R} = E\left[\mathbf{x}(k)\mathbf{x}^T(k)\right]$ (of size $L \times L$) is the covariance matrix of the input signal and $\mathbf{r} = E\left[\mathbf{x}(k)y(k)\right]$ (of length L) is the cross-correlation vector between the input and output signals. Generally, to avoid ill-conditioned problems, the covariance matrix is regularized as $\mathbf{R} + \epsilon\mathbf{I}_L$, where $\epsilon \geq 0$ is the regularization parameter and $\mathbf{I}_L$ is the $L \times L$ identity matrix.

To evaluate the identification performance, the normalized misalignment (NM) between the true channel impulse response, i.e., $\mathbf{h}$, and its estimate, i.e., $\widehat{\mathbf{h}}$, is often used, which is defined as [22]

$$\mathcal{M}\left(\mathbf{h}, \widehat{\mathbf{h}}\right) = \frac{\|\mathbf{h} - \widehat{\mathbf{h}}\|^2}{\|\mathbf{h}\|^2}, \quad (5)$$

where $\|\cdot\|$ denotes the Euclidean norm.

3. TENSOR DECOMPOSITION OF ACOUSTIC IMPULSE RESPONSES

We shape the components of $\mathbf{h}$ into a tensor $\boldsymbol{\mathcal{H}}$ of size $L_1 \times L_2 \times \cdots \times L_N$ (where $L = L_1 L_2 \cdots L_N$ and, without loss of generality, we assume that $L_1 \geq L_2 \geq \cdots \geq L_N$). The tensor $\boldsymbol{\mathcal{H}}$ can be decomposed as the sum of a finite number of rank-1 tensors, given by [23–25]

$$\boldsymbol{\mathcal{H}} = \sum_{i=1}^{R} \mathbf{h}_{1,i} \circ \mathbf{h}_{2,i} \circ \cdots \circ \mathbf{h}_{N,i}, \quad (6)$$

where the symbol $\circ$ denotes the outer product, $\mathbf{h}_{n,i}$, $n = 1, 2, \ldots, N$, $i = 1, 2, \ldots, R$ are filters of length L_n, and R is the rank of the tensor $\boldsymbol{\mathcal{H}}$. This means that each element of the tensor is the product of the corresponding vector elements:

$$(\boldsymbol{\mathcal{H}})_{l_1, l_2, \ldots, l_N} = \sum_{i=1}^{R} h_{1,i,l_1} h_{2,i,l_2} \cdots h_{N,i,l_N}, \quad (7)$$

where $l_n = 1, 2, \ldots, L_n$ for $n = 1, 2, \ldots, N$ and h_{n,p,l_n} is the l_nth element of $\mathbf{h}_{n,p}$.

Because of the structure of acoustic impulse responses, we may assume that the tensor $\boldsymbol{\mathcal{H}}$ can be approximated by a rank-P ($P \leq R$) tensor:

$$\widehat{\boldsymbol{\mathcal{H}}} = \sum_{p=1}^{P} \mathbf{h}_{1,p} \circ \mathbf{h}_{2,p} \circ \cdots \circ \mathbf{h}_{N,p}. \quad (8)$$

Then, an approximation of $\mathbf{h}$ can be obtained from the vectorization of $\widehat{\boldsymbol{\mathcal{H}}}$ as [26]

$$\widehat{\mathbf{h}} = \text{vect}\left(\widehat{\boldsymbol{\mathcal{H}}}\right) = \sum_{p=1}^{P} \mathbf{h}_{N,p} \otimes \mathbf{h}_{N-1,p} \otimes \cdots \otimes \mathbf{h}_{1,p}, \quad (9)$$

where the symbol $\otimes$ denotes the Kronecker product and $\text{vect}(\cdot)$ is the vectorization operation, which consists of converting a tensor into a long vector. Note that the notion of best rank approximation of a tensor is not well defined as in a matrix [27]; however, this does not affect the method developed in this paper.

4. ACOUSTIC SYSTEM IDENTIFICATION BASED ON PARTIALLY TIME-VARYING MODELS

Generally, room acoustic impulse responses, which may consist of thousands of reflections from surfaces and boundaries, have some interesting properties. Let us illustrate this with the conventional image-model method [28]. The simulation setup is as follows. We consider a room of size $6\,\text{m} \times 5\,\text{m} \times 3\,\text{m}$, as shown in Fig. 1. For ease of exposition, the 3-dimensional Cartesian coordinate system is used to specify the position of a point in the room. An omnidirectional microphone is placed at (2.5, 2.0, 1.5). To simulate a moving source, we first place it at $(3.0, 4.0, 1.5)$, and then move it to $(3.1, 4.0, 1.5)$, $(3.5, 4.0, 1.5)$, and $(4.0, 4.0, 1.5)$ successively. The acoustic channel impulse responses from the source to the microphones are gener-

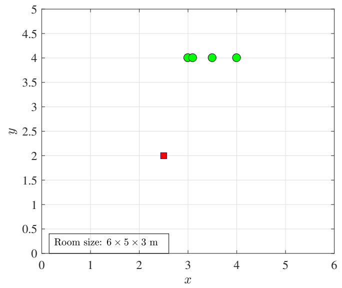

Fig. 1. Experimental setup with the image model, where we consider a room of size 6 m × 5 m × 3 m, an omnidirectional microphone is placed at (2.5, 2.0, 2.5), and a source is first located at $(3.0, 4.0, 1.5)$, and then moved to $(3.1, 4.0, 1.5)$, $(3.5, 4.0, 1.5)$, and $(4.0, 4.0, 1.5)$, subsequently.

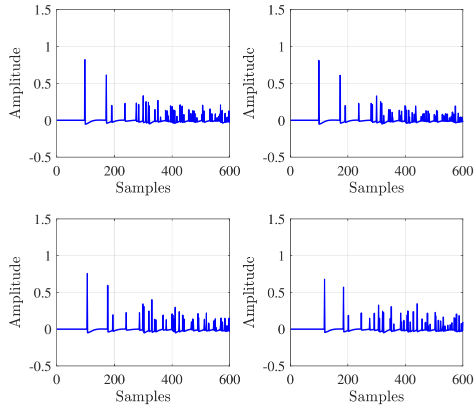

Fig. 2. Acoustic impulse responses: (a) $\mathbf{h}_{(a)}$ from $(3.0, 4.0, 1.5)$, (b) $\mathbf{h}_{(b)}$ from $(3.1, 4.0, 1.5)$, (c) $\mathbf{h}_{(c)}$ from $(3.5, 4.0, 1.5)$, and (d) $\mathbf{h}_{(d)}$ from $(4.0, 4.0, 1.5)$. The reverberation time T_{60} is approximately 300 ms.

Table 1. Normalized misalignment between the impulse responses from the sound source at different positions to the microphone.

	$\mathcal{M}\left(\mathbf{h}_{(a)}, \mathbf{h}_{(b)}\right)$	$\mathcal{M}\left(\mathbf{h}_{(a)}, \mathbf{h}_{(c)}\right)$	$\mathcal{M}\left(\mathbf{h}_{(a)}, \mathbf{h}_{(d)}\right)$
direct	4.16 dB	6.96 dB	4.05 dB
aligned	−2.14 dB	3.28 dB	2.77 dB

ated with the image model [28], where the reverberation time, T_{60}, is approximately 300 ms. Figure 2 plots these impulse responses from the sound source to the microphone at $(3.0, 4.0, 1.5)$, $(3.1, 4.0, 1.5)$, $(3.5, 4.0, 1.5)$, and $(4.0, 4.0, 1.5)$, which are represented as $\mathbf{h}_{(a)}$, $\mathbf{h}_{(b)}$, $\mathbf{h}_{(c)}$, and $\mathbf{h}_{(d)}$. As seen, the impulse responses from different source positions look similar. We compute the normalized misalignment between these impulse responses (both in direct and aligned ways, where "aligned" means different impulse responses are aligned based on the direct path). The results are shown in Table 1. The normalized misalignments between them are large, which explains the fact that there is a large difference between these impulses from a qualitative analysis perspective though they seem similar.

It could be helpful to consider this "similarity" in acoustic system identification. Based on the tensor factorization, we propose a partially time-varying model for acoustic impulse responses. Let us express the modeling filter as

$$\begin{aligned}\widehat{\mathbf{h}} &= \sum_{p=1}^{P} \underbrace{\mathbf{h}_{N,p} \otimes \cdots \otimes \mathbf{h}_{n+1,p}}_{\text{time-invariant}} \otimes \underbrace{\mathbf{h}_{n,p} \otimes \cdots \otimes \mathbf{h}_{1,p}}_{\text{time-varying}} \\ &= \sum_{p=1}^{P} \mathbf{h}_{\mathrm{TI},p} \otimes \mathbf{h}_{n,p} \otimes \cdots \otimes \mathbf{h}_{1,p}, \quad (10)\end{aligned}$$

where $\mathbf{h}_{\mathrm{TI},p} = \mathbf{h}_{N,p} \otimes \cdots \otimes \mathbf{h}_{n+1,p},\ p = 1, 2, \ldots, P$ are the time-invariant filters, which attempt to capture the "similarity" between different room impulse responses (or common properties of acoustic channels). In acoustic system identification, the time-invariant filters can be optimized or learned in advance. So, only time-varying filters need to be identified and updated in real-time, a method to achieve this is discussed as follows.

We have the following relationships for the Kronecker product [17, 29]:

$$\begin{aligned}&\mathbf{h}_{\mathrm{TI},p} \otimes \mathbf{h}_{n,p} \otimes \cdots \otimes \mathbf{h}_{1,p} \\ &\quad= \left(\mathbf{h}_{\mathrm{TI},p} \otimes \mathbf{h}_{n,p} \otimes \cdots \otimes \mathbf{h}_{2,p} \otimes \mathbf{I}_{L_1}\right) \mathbf{h}_{1,p} \quad (11a)\\ &\quad= \left(\mathbf{h}_{\mathrm{TI},p} \otimes \mathbf{h}_{n,p} \otimes \cdots \otimes \mathbf{I}_{L_2} \otimes \mathbf{h}_{1,p}\right) \mathbf{h}_{2,p} \quad (11b)\\ &\quad\ \ \vdots \\ &\quad= \left(\mathbf{h}_{\mathrm{TI},p} \otimes \mathbf{I}_{L_n} \otimes \cdots \otimes \mathbf{h}_{2,p} \otimes \mathbf{h}_{1,p}\right) \mathbf{h}_{n,p}, \quad (11c)\end{aligned}$$

where $\mathbf{I}_{L_{n'}},\ n' = 1, 2, \ldots, n$ are identity matrices of sizes $L_{n'} \times L_{n'}$. Substituting (11a)–(11c) into (10), we have

$$\begin{aligned}\widehat{\mathbf{h}} &= \sum_{p=1}^{P} \mathbf{h}_{\mathrm{TI},p} \otimes \mathbf{h}_{n,p} \otimes \cdots \otimes \mathbf{h}_{1,p} \\ &= \sum_{p=1}^{P} \left(\mathbf{h}_{\mathrm{TI},p} \otimes \mathbf{h}_{n,p} \otimes \cdots \otimes \mathbf{h}_{2,p} \otimes \mathbf{I}_{L_1}\right) \mathbf{h}_{1,p} \\ &= \sum_{p=1}^{P} \mathbf{H}_{1,p} \mathbf{h}_{1,p} \quad (12a)\\ &\ \ \vdots \\ &= \sum_{p=1}^{P} \mathbf{H}_{n,p} \mathbf{h}_{n,p}, \quad (12b)\end{aligned}$$

where

$$\begin{aligned}\mathbf{H}_{1,p} &= \mathbf{h}_{\mathrm{TI},p} \otimes \mathbf{h}_{n,p} \otimes \cdots \otimes \mathbf{h}_{2,p} \otimes \mathbf{I}_{L_1}, \quad (13a)\\ &\ \ \vdots \\ \mathbf{H}_{n,p} &= \mathbf{h}_{\mathrm{TI},p} \otimes \mathbf{I}_{L_n} \otimes \cdots \otimes \mathbf{h}_{2,p} \otimes \mathbf{h}_{1,p} \quad (13b)\end{aligned}$$

are matrices of sizes $L \times L_1$, ..., $L \times L_n$, respectively.

Substituting (12a) into (2), the error signal can be rewritten as

$$\begin{aligned}e(k) &= y(k) - \widehat{\mathbf{h}}^T \mathbf{x}(k) \\ &= y(k) - \sum_{p=1}^{P} \left(\mathbf{h}_{\mathrm{TI},p} \otimes \mathbf{h}_{n,p} \otimes \cdots \otimes \mathbf{h}_{1,p}\right)^T \mathbf{x}(k) \\ &= y(k) - \sum_{p=1}^{P} \mathbf{h}_{1,p}^T \mathbf{H}_{1,p}^T \mathbf{x}(k) \\ &= y(k) - \sum_{p=1}^{P} \mathbf{h}_{1,p}^T \mathbf{x}_{1,p}(k) \\ &= y(k) - \underline{\mathbf{h}}_1^T \underline{\mathbf{x}}_1(k), \quad (14)\end{aligned}$$

where

$$\begin{aligned}\mathbf{x}_{1,p}(k) &= \mathbf{H}_{1,p}^T \mathbf{x}(k),\ p = 1, 2, \ldots, P, \\ \underline{\mathbf{h}}_1 &= \begin{bmatrix} \mathbf{h}_{1,1}^T & \mathbf{h}_{1,2}^T & \cdots & \mathbf{h}_{1,P}^T \end{bmatrix}^T, \\ \underline{\mathbf{x}}_1(k) &= \begin{bmatrix} \mathbf{x}_{1,1}^T(k) & \mathbf{x}_{1,2}^T(k) & \cdots & \mathbf{x}_{1,P}^T(k) \end{bmatrix}^T.\end{aligned}$$

Similarly, we have

$$\begin{aligned}e(k) &= y(k) - \sum_{p=1}^{P} \mathbf{h}_{2,p}^T \mathbf{H}_{2,p}^T \mathbf{x}(k) \\ &= y(k) - \sum_{p=1}^{P} \mathbf{h}_{2,p}^T \mathbf{x}_{2,p}(k) \\ &= y(k) - \underline{\mathbf{h}}_2^T \underline{\mathbf{x}}_2(k) \quad (15a)\\ &\ \ \vdots \\ &= y(k) - \underline{\mathbf{h}}_n^T \underline{\mathbf{x}}_n(k), \quad (15b)\end{aligned}$$

where

$$\begin{aligned}\mathbf{x}_{n',p}(k) &= \mathbf{H}_{n',p}^T \mathbf{x}(k),\ p = 1, 2, \ldots, P, \\ \underline{\mathbf{h}}_{n'} &= \begin{bmatrix} \mathbf{h}_{n',1}^T & \mathbf{h}_{n',2}^T & \cdots & \mathbf{h}_{n',P}^T \end{bmatrix}^T, \\ \underline{\mathbf{x}}_{n'}(k) &= \begin{bmatrix} \mathbf{x}_{n',1}^T(k) & \mathbf{x}_{n',2}^T(k) & \cdots & \mathbf{x}_{n',P}^T(k) \end{bmatrix}^T,\end{aligned}$$

for $n' = 2, 3, \ldots, n$.

Substituting (14) and (15b) into (3), the MSE criterion can be written as

$$\begin{aligned}\mathcal{J}(\mathbf{h}) &= E\left[\left|y(k) - \underline{\mathbf{h}}_1^T \underline{\mathbf{x}}_1(k)\right|^2\right] \\ &= \sigma_y^2 - 2\underline{\mathbf{h}}_1^T \underline{\mathbf{r}}_1 + \underline{\mathbf{h}}_1^T \underline{\mathbf{R}}_1 \underline{\mathbf{h}}_1 \quad (16a)\\ &\ \ \vdots \\ &= \sigma_y^2 - 2\underline{\mathbf{h}}_n^T \underline{\mathbf{r}}_n + \underline{\mathbf{h}}_n^T \underline{\mathbf{R}}_n \underline{\mathbf{h}}_n, \quad (16b)\end{aligned}$$

where $\sigma_y^2 = E\left[y^2(k)\right]$ and

$$\underline{\mathbf{r}}_{n'} = \left[\begin{array}{cccc} \mathbf{r}^T\mathbf{H}_{n',1} & \mathbf{r}^T\mathbf{H}_{n',2} & \cdots & \mathbf{r}^T\mathbf{H}_{n',P} \end{array}\right]^T,$$

$$\underline{\mathbf{R}}_{n'} = \left[\begin{array}{cccc} \mathbf{R}_{n',11} & \mathbf{R}_{n',12} & \cdots & \mathbf{R}_{n',1P} \\ \mathbf{R}_{n',21} & \mathbf{R}_{n',22} & \cdots & \mathbf{R}_{n',2P} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{n',P1} & \mathbf{R}_{n',P2} & \cdots & \mathbf{R}_{n',PP} \end{array}\right]^T, \quad (17)$$

for $n' = 1, 2, \ldots, n$, with

$$\mathbf{R}_{n',ij} = \mathbf{H}_{n',i}^T \mathbf{R} \mathbf{H}_{n',j}, \; i, j = 1, 2 \ldots, P. \quad (18)$$

Since the n filters are coupled together, it is challenging to optimize them simultaneously; but a pragmatic solution can be found iteratively [17]. Suppose that we first initialize $\underline{\mathbf{h}}_2, \ldots, \underline{\mathbf{h}}_n$ and then compute $\underline{\mathbf{h}}_1$. Subsequently, the updated $\underline{\mathbf{h}}_1$ and other initialized filters are used to update $\underline{\mathbf{h}}_2$. Continuing this process, one can obtain the estimates of all the filters. Every time, when all filters are fixed except for $\underline{\mathbf{h}}_{n'}$, $n' = 2, \ldots, n$, the MSE criterion is written as

$$J\left(\underline{\mathbf{h}}_{n'} | \underline{\mathbf{h}}_1, \ldots, \underline{\mathbf{h}}_{n'-1}, \underline{\mathbf{h}}_{n'+1}, \ldots, \underline{\mathbf{h}}_n\right) = \sigma_y^2 - 2\underline{\mathbf{h}}_{n'}^T \underline{\mathbf{r}}_{n'} + \underline{\mathbf{h}}_1^T \underline{\mathbf{R}}_{n'} \underline{\mathbf{h}}_{n'}, \quad (19)$$

from which the optimal solution is computed as

$$\underline{\mathbf{h}}_{n'} = \underline{\mathbf{R}}_{n'}^{-1} \underline{\mathbf{r}}_{n'}. \quad (20)$$

The iteration continues until the pre-specified conditions of convergence are met, thereby giving the final solution of $\underline{\mathbf{h}}_{n'}$, $n' = 1, 2, \ldots, n$, and the optimal solution of $\widehat{\mathbf{h}}$ is obtained according to (10).

5. SIMULATIONS

Now, we study the performance of the proposed method with measured acoustic impulse responses. The impulse responses used are from the multichannel room impulse responses database [30], which are measured in a room of size 6 m × 6 m × 2.4 m, and the corresponding reverberation time is $T_{60} \approx 360$ ms. The impulse responses are measured with a uniform linear array consisting of 8 microphones with an inter-distance of 8 cm. The source positions are 1m away from the microphone array, two impulse responses are measured at two points on a spatial angle of 0° and 15°, where the distance between the two points is about 8.33 cm. In our experiment, we use the measurement on the first microphone and assume the sound source is first at 0° and then moves to 15°. Based on our previous discussion, we expect that the two impulse responses share some "similarity" (the normalized misalignment between the impulse responses at the two points is 3.54 dB). The excitation signal is an autoregressive (AR) process, generated by filtering a white Gaussian process with a first-order system $1/(1 - 0.9z^{-1})$. The observation signal is generated by convolving the excitation signal with the acoustic impulse response, and white Gaussian noise is then added to control the SNR to be 20 dB. All the signals are sampled at 8 kHz. We are interested in estimating the truncated impulse response with a length of 512 samples since their tails after 512 have a very small magnitude and can be neglected, so the length of the modeling filter is set to 512. With the proposed method, the global filter is factorized as a 3-way tensor of size 8 × 8 × 8, i.e., $N = 3$ and $L_1 = L_2 = L_3 = 8$. The covariance matrix, $\mathbf{R}$, and the cross-correlation vector, $\mathbf{r}$, are estimated using a short-time average method.

To simulate the process, we applied tensor decomposition of the impulse response measured at 0° to find its rank-P approximation [31, 32]. Then, $\mathbf{h}_{1,p}$, $p = 1, 2, \ldots, P$ are used as the time-invariant filters, $\mathbf{h}_{3,p}$, $p = 1, 2, \ldots, P$ are initialized as $[1, 1, \cdots, 1]/L_3$, based on which we identify the time-varying part of the impulse response measured at 15°. The iteration continues 10 times to obtain the estimates of the final filters. Note that in implementing matrix inversion, a small regularization parameter is added to the diagonal elements of the matrices. (As pointed out previously, this paper focuses on how to decompose an acoustic channel into time-invariant and time-varying parts and identify the time-varying part, so improving the estimation of the time-invariant part is worth further study but beyond the scope of this work.) We evaluate the performance of the proposed method using NM as the performance metric. Figure 3 show plots of the NM (averaged over 10 Monte-Carlo runs) as a function of the order P, where Fig. 3 (a), (b), (c) and (d) with 600, 650, 700, and 750 samples observation, respectively. For comparison, we also show the results of the conventional Wiener filter (a small regularization parameter is also added to the diagonal elements of the matrix $\mathbf{R}$ to implement matrix inversion). The proposed method achieves better performance with proper values of P (i.e., not too small). The results demonstrate that the proposed method can achieve better acoustic system identification performance with a small number of observation data.

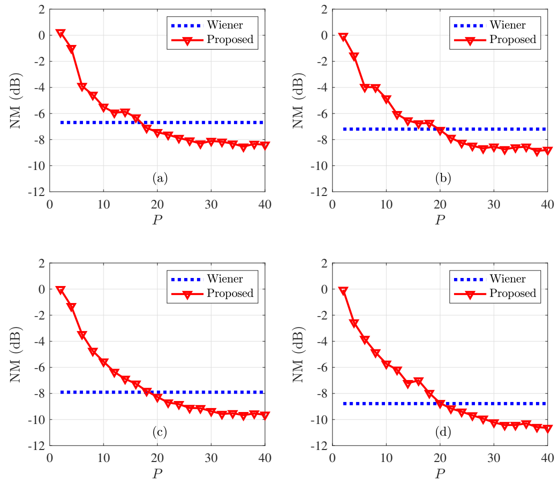

Fig. 3. NM of the conventional Wiener filter and proposed iterative tensor Wiener filter for acoustic system identification with: (a) 600 samples observation, (b) 650 samples observation, (c) 700 samples observation, and (d) 750 samples observation. Conditions: $L = 512$, $L_1 = L_2 = L_3 = 8$, and SNR = 20 dB.

For the conventional Wiener filter, the computations of the matrix $\mathbf{R}^{-1}$ have a computational complexity proportional to $O(L^3)$. For the proposed method, the computations of the matrices $\underline{\mathbf{R}}_{n'}$, $n' = 1, 2$, has a computational complexity proportional to $O(P^3 L_{n'} L^2 + P^4 L_{n'}^2 L)$, and the computations of the matrices $\underline{\mathbf{R}}_{n'}^{-1}$ has a computational complexity proportional to $O(P^3 L_{n'}^3)$. If the value of P is small, the proposed method has lower computational complexity. If the value of P is large, the proposed method has higher computational complexity than the conventional Wiener filter.

6. CONCLUSIONS

This paper studied the problem of acoustic channel identification with a partially time-varying model. Based on the observation that the acoustic channel impulse responses in the same acoustic environment share some common and time-invariant properties, we proposed decomposing the acoustic channel to be identified into a time-invariant part and a time-varying part under the framework of tensor decompositions. Assuming that the time-invariant part is either given as the *a priori* information or can be estimated using, e.g., the batch method, we developed an iterative technique based on the tensor model to identify the time-varying part of the acoustic system.

7. REFERENCES

[1] L. Ljung, "System identification," in *Signal Analysis and Prediction*, pp. 163–173, Springer, 1998.

[2] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*. SciTech Publishing, Inc, Raleigh, NC, 2004.

[3] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Signal Process.*, vol. 8, no. 5, pp. 508–518, 2000.

[4] P. Loganathan, A. W. Khong, and P. A. Naylor, "A class of sparseness-controlled algorithms for echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 8, pp. 1591–1601, 2009.

[5] C. Paleologu, J. Benesty, and S. Ciochina, "Sparse adaptive filters for echo cancellation," *Synthesis Lectures on Speech and Audio Processing*, vol. 6, no. 1, pp. 1–124, 2010.

[6] J. Liu and S. L. Grant, "Proportionate adaptive filtering for block-sparse system identification," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 4, pp. 623–630, 2015.

[7] 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, 2010.

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

[9] G. Huang, J. Benesty, J. Chen, and I. Cohen, "Robust and steerable Kronecker product differential beamforming with rectangular microphone arrays," in *Proc. IEEE ICASSP*, 2020.

[10] X. Wang, J. Benesty, J. Chen, G. Huang, and I. Cohen, "Beamforming with cube microphone arrays via Kronecker product decompositions," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 1774–1784, 2021.

[11] 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.

[12] G. Huang, J. Benesty, I. Cohen, and J. Chen, "Kronecker product multichannel linear filtering for adaptive weighted prediction error-based speech dereverberation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 1277–1289, 2022.

[13] 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.

[14] W. Yang, G. Huang, J. Chen, J. Benesty, I. Cohen, and W. Kellermann, "Robust dereverberation with Kronecker product based multichannel linear prediction," *IEEE Signal Process. Lett.*, vol. 28, pp. 101–105, 2021.

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

[16] J. Benesty, C. Paleologu, and S. Ciochină, "On the identification of bilinear forms with the Wiener filter," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 653–657, 2017.

[17] C. Paleologu, J. Benesty, and S. Ciochina, "Linear system identification based on a Kronecker product decomposition," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 10, pp. 1793–1808, 2018.

[18] C. Paleologu, J. Benesty, and S. Ciochină, "Adaptive filtering for the identification of bilinear forms," *Digital Signal Process.*, vol. 75, pp. 153–167, 2018.

[19] C. Elisei-Iliescu, C. Paleologu, J. Benesty, C. Stanciu, C. Anghel, and S. Ciochină, "Recursive least-squares algorithms for the identification of low-rank systems," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 5, pp. 903–918, 2019.

[20] S. Bhattacharjee and N. George, "Nearest Kronecker product decomposition based linear-in-the-parameters nonlinear filters," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, no. 5, pp. 2111–2121, 2021.

[21] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, 2015.

[22] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*. Berlin, Germany: Springer-Verlag, 2001.

[23] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[24] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[25] R. A. Harshman, *Foundations of the PARAFAC Procedure: Models and Conditions for an "Explanatory" Multimodal Factor Analysis*, UCLA Working Papers in Phonetics, 1970.

[26] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*, 1998.

[27] L. Grasedyck, D. Kressner, and C. Tobler, "A literature survey of low-rank tensor approximation techniques," *GAMM-Mitteilungen*, vol. 36, no. 1, pp. 53–78, 2013.

[28] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[29] C. F. Van Loan, "The ubiquitous Kronecker product," *J. Comput. Appl.Math.*, vol. 123, pp. 85–100, Nov. 2000.

[30] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *Proc. IWAENC*, pp. 313–317, 2014.

[31] B. W. Bader, T. G. Kolda, *et al.*, "Matlab tensor toolbox version 3.0-dev." Available online, Aug. 2017.

[32] B. W. Bader and T. G. Kolda, "Algorithm 862: MATLAB tensor classes for fast algorithm prototyping," *ACM Transactions on Mathematical Software*, vol. 32, pp. 635–653, Dec. 2006.