FREQUENCY-DOMAIN UNDER-MODELLED BLIND SYSTEM IDENTIFICATION BASED ON CROSS POWER SPECTRUM AND SPARSITY REGULARIZATION

Wei Xue, Mike Brookes, Patrick A. Naylor

Department of Electrical and Electronic Engineering, Imperial College London, UK

ABSTRACT

In room acoustics, under-modelled multichannel blind system identification (BSI) aims to estimate the early part of the room impulse responses (RIRs), and it can be widely used in applications such as speaker localization, room geometry identification and beamforming based speech dereverberation. In this paper we extend our recent study on under-modelled BSI from the time domain to the frequency domain, such that the RIRs can be updated frame-wise and the efficiency of Fast Fourier Transform (FFT) is exploited to reduce the computational complexity. Analogous to the cross-correlation based criterion in the time domain, a frequency-domain cross power spectrum based criterion is proposed. As the early RIRs are usually sparse, the RIRs are estimated by jointly maximizing the cross power spectrum based criterion in the frequency domain and minimizing the l_1-norm sparsity measure in the time domain. A two-stage LMS updating algorithm is derived to achieve joint optimization of these two targets. The experimental results in different under-modelled scenarios demonstrate the effectiveness of the proposed method.

***Index Terms*—** Blind system identification, microphone arrays, system under-modelling.

1. INTRODUCTION AND PRIOR WORK

In an enclosure, the room impulse responses (RIRs) from a source to a microphone array can be estimated by multichannel blind system identification (BSI) using the multichannel outputs. In many applications such as speaker localization [1, 2, 3], room geometry identification [4, 5], and acoustic RAKE receivers [6, 7], only the early part of the RIR is of interest. This motivates the under-modelled BSI problem which allows the identified system to be shorter than the real one.

Over the past few decades, different algorithms for multichannel BSI have been proposed. Widely used multichannel BSI algorithms are formulated adaptively in the least mean squares (LMS) framework, including the multichannel LMS (MCLMS) algorithm [8], multichannel Newton (MCN) algorithm [8], multichannel frequency-domain LMS (MCFLMS) algorithm [9] and normalized MCFLMS (NMCFLMS) algorithm [9]. These algorithms are generally based on the cross-relation (CR) property [10], and the RIRs are estimated by minimizing the CR error. However, it has been shown that when the system is under-modelled, the CR property is no longer valid [11]. Therefore, conventional adaptive algorithms cannot yield accurate RIR estimates with system under-modelling.

In our recent work, a time-domain cross-correlation and sparsity regularization (CSR) based algorithm was proposed for under-modelled multichannel BSI [11]. Instead of minimizing the CR error, the RIRs were estimated by maximizing a cross-correlation based criterion. It was shown that, in this way, the adverse effect of system under-modelling is greatly reduced. Furthermore, the sparsity of the early RIR was exploited to improve the BSI performance.

For adaptive multichannel BSI, frequency-domain methods have received more attention since the filter updating can be performed frame-wise, and the convolution and cross-correlation operation which are computationally intensive in the time domain can be efficiently implemented by fast Fourier transform (FFT) [9]. Inspired by this, a frequency-domain under-modelled multichannel BSI algorithm is proposed in this paper. Analogous to maximizing a cross-correlation based criterion in the time-domain method, in the proposed method, the RIRs are estimated by maximizing a cross power spectrum based criterion. In addition, the l_1-norm sparsity measure is also integrated into the optimization problem to promote the sparsity of the estimated RIRs. As the sparsity measure is computed in the time domain, a two-stage LMS updating algorithm is derived to achieve joint optimization of both frequency-domain and time-domain targets. By conducting experiments in different under-modelled scenarios, we demonstrate the effectiveness of the proposed method.

2. TIME-DOMAIN UNDER-MODELLED BSI

We begin with formulating the BSI problem in the time domain and briefly introduce our previous time-domain algorithm [11].

Consider a reverberant room with a single source and an M-element microphone array. The time-domain signal vector received by the i-th microphone is expressed:

$$\mathbf{x}_i(n) = \mathbf{H}_i \cdot \mathbf{s}(n) + \mathbf{v}_i(n),\ i = 1, 2, ..., M, \tag{1}$$

This research has received funding from the EU 7th Framework Programme (FP7/2007-2013) project DREAMS under grant agreement ITN-GA-2012-316969.

where $\mathbf{x}_i(n) = [x_i(n)\ x_i(n-1)\ \ldots\ x_i(n-K+1)]^T$ is the $K \times 1$ signal vector, K is the length of the final identified RIR. $\mathbf{v}_i(n)$ is the additive noise vector. $\mathbf{H}_i$ is a $K \times (K + L - 1)$ upper-rectangular Toeplitz matrix whose first row is $[\mathbf{h}_i^T\ \mathbf{0}_{1\times K-1}]$, where $\mathbf{h}_i = [h_{i,0}\ h_{i,1}\ \ldots\ h_{i,L-1}]^{\mathrm{T}}$ is the true L-tap RIR of the i-th channel. $\mathbf{s}(n) = [s(n)\ s(n-1)\ \ldots\ s(n-K-L+2)]^T$ is the $(K+L-1)\times 1$ source signal vector.

Multichannel BSI aims to estimate the RIRs with $\mathbf{x}_i(n)$ for $i = 1, 2, ..., M$. When $K < L$, the BSI problem becomes under-modelled. Given the estimated RIR $\hat{\mathbf{h}}_i$ for $i = 1, 2, ..., M$, the cross-filtered signals can be computed pairwise by convolving the output of one channel with the RIR of the other channel, as

$$\tilde{x}_{ij}(n) = \hat{\mathbf{h}}_i^T \mathbf{x}_j(n),\ i, j = 1, 2, ..., M. \tag{2}$$

In [11], we propose to estimate RIRs based on maximizing the cross-correlation of the cross-filtered signals. It has been shown that compared with minimizing the CR error, in the system under-modelling case, maximizing the cross-correlation of the cross-filtered signals produces much smaller bias term to the objective function, thus the adverse effect of system under-modelling is reduced. To avoid a scale ambiguity, the vector $\hat{\mathbf{h}} = [\hat{\mathbf{h}}_1^T\ \hat{\mathbf{h}}_2^T\ \cdots\ \hat{\mathbf{h}}_M^T]^T$ is constrained to be unit norm. By using all microphone pairs, and imposing the unit norm constraint on $\hat{\mathbf{h}}$ at all times, a cross-correlation based cost function is constructed as

$$J(\hat{\mathbf{h}}) = \frac{\sum_{i=1}^{M-1}\sum_{j=i+1}^{M} \mathbb{E}\{\tilde{x}_{ij}(n)\cdot\tilde{x}_{ji}(n)\}}{||\hat{\mathbf{h}}||_2^2}. \tag{3}$$

By promoting sparsity, an optimization problem is formulated as

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} \{-J(\mathbf{h}) + \rho|\mathbf{h}|_1\},\ \text{s.t.}\ ||\mathbf{h}||_2 = 1, \tag{4}$$

where ρ is a regularization parameter. In [11], the optimization problem in (4) is solved adaptively based on the split Bregman method [12]. The details of the adaptive updating are omitted here for brevity.

3. PROPOSED METHOD

In this section, we extend the time-domain algorithm to the frequency domain such that frame-level updating can be performed, and the efficiency of FFT is exploited to reduce the computational complexity.

3.1. Cross Power Spectrum of Cross-Filtered Signals

The cross-correlation of cross-filtered signals in the time domain corresponds to the cross power spectrum in the frequency domain.

Based on the overlap-save technique [13], with $2K$ frame length and 50% overlap, the $K \times 1$ time-domain cross-filtered signal vector $\tilde{\mathbf{x}}_{ij}(m)$ for the m-th frame can be written as the second half of the circular convolution between $\hat{\mathbf{h}}_i$ and $\mathbf{x}_j$ [9]:

$$\tilde{\mathbf{x}}_{ij}(m) = \mathbf{W}^{01}_{K\times 2K}\mathbf{C}_{x_j}(m)\mathbf{W}^{10}_{2K\times K}\hat{\mathbf{h}}_i(m), \tag{5}$$

where $\tilde{\mathbf{x}}_{ij}(m) = [\tilde{x}_{ij}(m)\ \tilde{x}_{ij}(m+1)\ \ldots\ \tilde{x}_{ij}(m+K-1)]^T$, $\mathbf{W}^{01}_{K\times 2K} = [\mathbf{0}_{K\times K}\ \mathbf{I}_{K\times K}]$, $\mathbf{W}^{10}_{2K\times K} = [\mathbf{I}_{K\times K}\ \mathbf{0}_{K\times K}]^T$, and $\mathbf{C}_{x_j}(m)$ is a circulant matrix with the first column as $[x_j(mK-K)\ \ldots\ x_j(mK)\ \ldots\ x_j(mK+K-1)]^T$.

By using FFT to efficiently perform circular convolution, the frequency-domain expression of the cross-filtered signal is obtained as:

$$\begin{aligned}\underline{\tilde{\mathbf{x}}}_{ij}(m) &= \mathbf{F}_K\tilde{\mathbf{x}}_{ij}(m)\\ &= \mathbf{F}_K\mathbf{W}^{01}_{K\times 2K}\mathbf{C}_{x_j}(m)\mathbf{W}^{10}_{2K\times K}\hat{\mathbf{h}}_i(m)\\ &= \boldsymbol{\mathcal{W}}^{01}_{K\times 2K}\boldsymbol{\mathcal{D}}_{x_j}(m)\boldsymbol{\mathcal{W}}^{10}_{2K\times K}\underline{\hat{\boldsymbol{h}}}_i(m),\end{aligned} \tag{6}$$

where $\mathbf{F}_K$ is the $K \times K$ discrete Fourier transform (DFT) matrix, $\boldsymbol{\mathcal{D}}_{x_j}(m) = \mathbf{F}_{2K}\mathbf{C}_{x_j}\mathbf{F}_{2K}^{-1}$ is a diagonal matrix whose diagonal elements are the DFT of the first column of $\mathbf{C}_{x_j}(m)$, $\underline{\hat{\boldsymbol{h}}}_i(m) = \mathbf{F}_K\hat{\mathbf{h}}_i(m)$ consists of the K-point DFTs of $\hat{\mathbf{h}}_i(m)$,

$$\begin{aligned}\boldsymbol{\mathcal{W}}^{01}_{K\times 2K} &= \mathbf{F}_K\mathbf{W}^{01}_{K\times 2K}\mathbf{F}_{2K}^{-1},\\ \boldsymbol{\mathcal{W}}^{10}_{2K\times K} &= \mathbf{F}_{2K}\mathbf{W}^{10}_{2K\times K}\mathbf{F}_K^{-1}.\end{aligned} \tag{7}$$

Accordingly, the cross power spectrum of a pair of cross-filtered signals is computed as

$$\begin{aligned}\underline{\gamma}_{ij} &= \mathbb{E}\{\underline{\tilde{\mathbf{x}}}_{ij}^H(m)\underline{\tilde{\mathbf{x}}}_{ji}(m)\}\\ &= \mathbb{E}\{\underline{\hat{\boldsymbol{h}}}_i^H(m)(\boldsymbol{\mathcal{W}}^{10}_{2K\times K})^H\boldsymbol{\mathcal{P}}_{ij}(m)\boldsymbol{\mathcal{W}}^{10}_{2K\times K}\underline{\hat{\boldsymbol{h}}}_j(m)\}\\ &= \mathbb{E}\{[\underline{\hat{\boldsymbol{h}}}_i^{10}(m)]^H\boldsymbol{\mathcal{P}}_{ij}(m)\underline{\hat{\boldsymbol{h}}}_j^{10}(m)\},\end{aligned} \tag{8}$$

where $\underline{\hat{\boldsymbol{h}}}_i^{10}(m) = \boldsymbol{\mathcal{W}}^{10}_{2K\times K}\underline{\hat{\boldsymbol{h}}}_i(m)$, and $\boldsymbol{\mathcal{P}}_{ij}(m) = \boldsymbol{\mathcal{D}}_{x_j}^H(m) \times (\boldsymbol{\mathcal{W}}^{01}_{K\times 2K})^H\boldsymbol{\mathcal{W}}^{01}_{K\times 2K}\boldsymbol{\mathcal{D}}_{x_i}(m)$.

3.2. Optimization Problem Formulation

Analogous to the cross-correlation based criterion in the time-domain, a new frequency-domain criterion based on the cross power spectrum of the cross-filtered signals is constructed. As the early RIR is generally sparse in the time-domain, an l_1-norm term computed using the time-domain RIRs is integrated into the optimization problem.

It should be noted that the cross power spectrum in (8) is a complex number, therefore it cannot be optimized directly. By utilizing the fact that $\underline{\tilde{\mathbf{x}}}_{ij}^H(m)\underline{\tilde{\mathbf{x}}}_{ji}(m) = (\underline{\tilde{\mathbf{x}}}_{ji}^H(m)\underline{\tilde{\mathbf{x}}}_{ij}(m))^*$, and using all microphone pairs, a real-valued cost function can be first formed as:

$$\begin{aligned}J_f(\underline{\hat{\boldsymbol{h}}}^{10}(m)) &= \sum_{i=1}^{M}\sum_{j=1,j\neq i}^{M}\underline{\tilde{\mathbf{x}}}_{ij}^H(m)\underline{\tilde{\mathbf{x}}}_{ji}(m)\\ &= \sum_{i=1}^{M}\sum_{j=1,j\neq i}^{M}[\underline{\hat{\boldsymbol{h}}}_i^{10}(m)]^H\boldsymbol{\mathcal{P}}_{ij}(m)\underline{\hat{\boldsymbol{h}}}_j^{10}(m),\end{aligned} \tag{9}$$

where $\underline{\hat{\boldsymbol{h}}}^{10}(m) = [[\underline{\hat{\boldsymbol{h}}}_1^{10}(m)]^H \; [\underline{\hat{\boldsymbol{h}}}_2^{10}(m)]^H \; \ldots \; [\underline{\hat{\boldsymbol{h}}}_M^{10}(m)]^H]^H$.

Similar to the time-domain approach, the RIRs can be estimated by maximizing the cross power spectrum based cost function. However, simply maximizing $J_f(\underline{\hat{\boldsymbol{h}}}^{10}(m))$ will lead to the trivial infinity solution. To avoid this, we restrict $\|\hat{\mathbf{h}}\|_2^2 = \hat{\mathbf{h}}^T\hat{\mathbf{h}} = 1$ and equivalently $[\underline{\hat{\boldsymbol{h}}}^{10}(m)]^H\underline{\hat{\boldsymbol{h}}}^{10}(m) = 2K$ from the Parseval's theorem. Ignoring the scaling factor $2K$, a new cost function is defined as

$$\begin{aligned}&\mathcal{J}_f(\underline{\hat{\boldsymbol{h}}}^{10}(m))\\&= \frac{J_f(\underline{\hat{\boldsymbol{h}}}^{10}(m))}{[\underline{\hat{\boldsymbol{h}}}^{10}(m)]^H\underline{\hat{\boldsymbol{h}}}^{10}(m)} = \frac{J_f(\underline{\hat{\boldsymbol{h}}}^{10}(m))}{\sum_{i=1}^{M}[\underline{\hat{\boldsymbol{h}}}_i^{10}(m)]^H\underline{\hat{\boldsymbol{h}}}_i^{10}(m)}.\end{aligned} \tag{10}$$

We aim to find the RIRs which maximize $\mathcal{J}_f(\underline{\hat{\boldsymbol{h}}}^{10}(m))$ in the frequency domain and also being sparse in the time domain. This sparsity is typically measured by the l_1-norm of time-domain RIR vector $\hat{\mathbf{h}}$. To combine the two targets into one optimization problem, note that $\underline{\hat{\boldsymbol{h}}}_i^{10}(m) = F_{2K}\mathbf{W}_{2K\times K}^{10}\hat{\mathbf{h}}_i(m)$, then $\mathcal{J}_f(\underline{\hat{\boldsymbol{h}}}^{10}(m))$ in (10) is also a function of $\hat{\mathbf{h}}(m)$ and can be written as $\mathcal{J}_f(\hat{\mathbf{h}}(m))$. By imposing the unit norm constraint, an optimization problem is formulated as:

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}}\{-\bar{\mathcal{J}}_f(\mathbf{h}) + \rho|\mathbf{h}|_1\}, \text{ s.t. } \|\mathbf{h}\|_2 = 1, \tag{11}$$

where $\bar{\mathcal{J}}_f(\mathbf{h}) = \mathbb{E}\{\mathcal{J}_f(\hat{\mathbf{h}}(m))\}$, and ρ is a regularization parameter for sparsity.

3.3. LMS Updating

We notice that although $\bar{\mathcal{J}}_f(\mathbf{h})$ in (11) is calculated in the frequency domain, we can, by expressing it as a function of the time-domain vector $\mathbf{h}$, obtain a optimization problem which has a similar form to (4). Following [11], an LMS-type algorithm is proposed in this subsection, which is based on the split Bregman algorithm [12]. A two-stage updating scheme is derived, which first updates the RIRs in the frequency domain and then promotes sparsity in the time domain.

According to [2], temporarily omitting the unit norm constraint for clarity, (11) can be reformulated as

$$(\hat{\mathbf{h}}, \hat{\mathbf{d}}) = \arg\min_{\mathbf{h},\mathbf{d}}\{-\bar{\mathcal{J}}_f(\mathbf{h}) + \rho|\mathbf{d}|_1 + \lambda\|\mathbf{d} - \mathbf{h}\|_2^2\}, \tag{12}$$

where $\mathbf{d}$ is a $(MK)\times 1$ auxiliary variable vector, with $\hat{\mathbf{d}}$ as the estimate, and λ is a Lagrange multiplier.

Based on the split Bregman iteration method [12], (12) can be iteratively solved by

$$(\hat{\mathbf{h}}, \hat{\mathbf{d}})^{k+1} = \arg\min_{\mathbf{h},\mathbf{d}}\{-\bar{\mathcal{J}}_f(\mathbf{h}) + \rho|\mathbf{d}|_1 + \lambda\|\mathbf{b}^k + \mathbf{h} - \mathbf{d}\|_2^2\}, \tag{13a}$$

$$\mathbf{b}^{k+1} = \mathbf{b}^k + \hat{\mathbf{h}}^{k+1} - \hat{\mathbf{d}}^{k+1}, \tag{13b}$$

where $\mathbf{b}$ is a $(KM)\times 1$ Bregman variable vector, and k denotes the iteration index. Reinserting the unit norm constraint, (13a) can be transformed into two sub-problems [2]:

$$\hat{\mathbf{h}}^{k+1} = \arg\min_{\mathbf{h}}\{-\bar{\mathcal{J}}_f(\mathbf{h}) + \lambda\|\mathbf{b}^k + \mathbf{h} - \hat{\mathbf{d}}^k\|_2^2\}, \quad \text{s.t. } \|\mathbf{h}\|_2 = 1. \tag{14a}$$

$$\hat{\mathbf{d}}^{k+1} = \arg\min_{\mathbf{d}}\{\rho|\mathbf{d}|_1 + \lambda\|\mathbf{b}^k + \hat{\mathbf{h}}^{k+1} - \mathbf{d}\|_2^2\}. \tag{14b}$$

Solving (14a): In the frequency-domain method, the RIRs are updated in each frame. The problem can be solved using gradient descent, and the gradient consists of components from both the frequency-domain and time-domain objective functions. A two-stage updating scheme is derived here.

In the first stage, we update the RIRs in the frequency domain. Rewriting $\mathcal{J}_f(\mathbf{h})$ as $\mathcal{J}_f(\underline{\hat{\boldsymbol{h}}}^{10}(m))$, the RIR of the i-th channel is updated as

$$\underline{\hat{\boldsymbol{h}}}_i^{10}(m+1) = \underline{\hat{\boldsymbol{h}}}_i^{10}(m) + \mu\frac{\partial\mathcal{J}_f(\underline{\hat{\boldsymbol{h}}}^{10}(m))}{\partial(\underline{\hat{\boldsymbol{h}}}_i^{10}(m))^*}, \tag{15}$$

where μ is the step size. According to (10), we further have

$$\begin{aligned}&\frac{\partial\mathcal{J}_f(\underline{\hat{\boldsymbol{h}}}^{10}(m))}{\partial(\underline{\hat{\boldsymbol{h}}}_i^{10}(m))^*}\\
&= \frac{\frac{\partial J_f(\underline{\hat{\boldsymbol{h}}}^{10}(m))}{\partial(\underline{\hat{\boldsymbol{h}}}_i^{10}(m))^*}}{[\underline{\hat{\boldsymbol{h}}}^{10}(m)]^H\underline{\hat{\boldsymbol{h}}}^{10}(m)} - \frac{J_f(\underline{\hat{\boldsymbol{h}}}^{10}(m))\underline{\hat{\boldsymbol{h}}}_i^{10}(m)}{\{[\underline{\hat{\boldsymbol{h}}}^{10}(m)]^H\underline{\hat{\boldsymbol{h}}}^{10}(m)\}^2}\\
&= \frac{\sum_{j=1,j\neq i}^{M}\boldsymbol{\mathcal{P}}_{ij}(m)\underline{\hat{\boldsymbol{h}}}_j^{10}(m)}{[\underline{\hat{\boldsymbol{h}}}^{10}(m)]^H\underline{\hat{\boldsymbol{h}}}^{10}(m)} - \frac{J_f(\underline{\hat{\boldsymbol{h}}}^{10}(m))\underline{\hat{\boldsymbol{h}}}_i^{10}(m)}{\{[\underline{\hat{\boldsymbol{h}}}^{10}(m)]^H\underline{\hat{\boldsymbol{h}}}^{10}(m)\}^2}\\
&= \frac{\sum_{j=1,j\neq i}^{M}\boldsymbol{\mathcal{P}}_{ij}(m)\underline{\hat{\boldsymbol{h}}}_j^{10}(m)}{2K} - \frac{J_f(\underline{\hat{\boldsymbol{h}}}^{10}(m))\underline{\hat{\boldsymbol{h}}}_i^{10}(m)}{4K^2}.\end{aligned} \tag{16}$$

The last step of (16) is because of the unit norm constraint.

In the second stage, the RIRs are transformed into the time domain, and updated to promote sparsity:

$$\begin{aligned}&\hat{\mathbf{h}}_i(m+1)\\
&= W_{K\times 2K}^{10}F_{2K}^{-1}\underline{\hat{\boldsymbol{h}}}_i^{10}(m+1) - \mu\lambda\frac{\partial\|\mathbf{b}^m + \mathbf{h}(m) - \mathbf{d}^m\|_2^2}{\partial\hat{\mathbf{h}}_i(m)}\\
&= W_{K\times 2K}^{10}F_{2K}^{-1}\underline{\hat{\boldsymbol{h}}}_i^{10}(m+1) - 2\mu\lambda(\mathbf{b}_i^m + \hat{\mathbf{h}}_i(m) - \mathbf{d}_i^m).\end{aligned} \tag{17}$$

After updating, by reimposing the unit norm constraint, the RIRs are normalized as

$$\begin{aligned}&\hat{\mathbf{h}}_i(m+1)\\
&= \frac{W_{K\times 2K}^{10}F_{2K}^{-1}\underline{\hat{\boldsymbol{h}}}_i^{10}(m+1) - 2\mu\lambda(\mathbf{b}_i^m + \hat{\mathbf{h}}_i(m) - \mathbf{d}_i^m)}{\|W_{K\times 2K}^{10}F_{2K}^{-1}\underline{\hat{\boldsymbol{h}}}_i^{10}(m+1) - 2\mu\lambda(\mathbf{b}_i^m + \hat{\mathbf{h}}_i(m) - \mathbf{d}_i^m)\|_2}.\end{aligned} \tag{18}$$

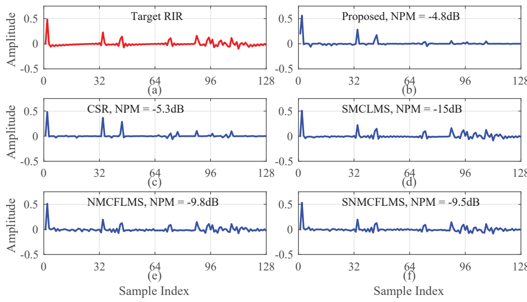

Fig. 1. Comparison results for $K = 128$: (a) The target RIR, and RIRs estimated by (b) the proposed method, (c) CSR, (d) SMCLMS, (e) NMCFLMS and (f) SNMCFLMS.

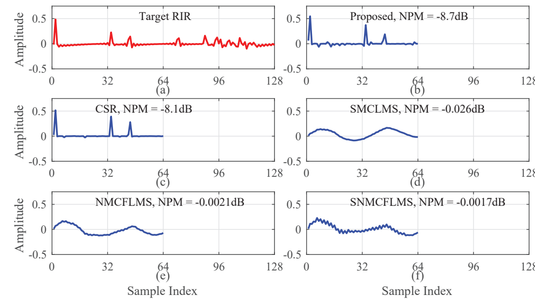

Fig. 2. Comparison results for $K = 64$: (a) The target RIR, and RIRs estimated by (b) the proposed method, (c) CSR, (d) SMCLMS, (e) NMCFLMS and (f) SNMCFLMS.

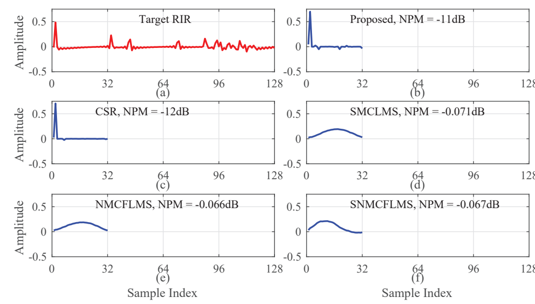

Fig. 3. Comparison results for $K = 32$: (a) The target RIR, and RIRs estimated by (b) the proposed method, (c) CSR, (d) SMCLMS, (e) NMCFLMS and (f) SNMCFLMS.

Solving (14b)*:* Following [12], we update $\hat{\mathbf{d}}$ in frame m as

$$\begin{aligned}\hat{\mathbf{d}}_i^{m+1} =& \text{sign}(\hat{\mathbf{h}}_i(m+1) + \hat{\mathbf{b}}_i^m) \\ & \times \max\{|\hat{\mathbf{h}}_i(m+1) + \hat{\mathbf{b}}_i^m| - \frac{\rho}{2\lambda}, 0\}. \end{aligned} \tag{19}$$

$\mathbf{b}$ is updated according to (13b).

4. EVALUATION

In this section, we conduct experiments on simulated data. The proposed method is compared with the NMCFLMS algorithm in [9], the sparse MCLMS (SMCLMS) algorithm and sparse NMCFLMS (SNMCFLMS) algorithm in [2], and our previous time-domain CSR algorithm in [11].

We model a 5 m $\times$ 6 m $\times$ 3 m room using the image-source method [14]. A two-element microphone array with microphones at (2.4, 2.0, 1.6) m and (2.6, 2.0, 1.6) m is used. The source signal is 10 s white Gaussian noise with 8 kHz sampling rate. The source position is (2.05, 3.95, 1.67) m. RIRs are simulated by setting the reverberation time as 300 ms and filter length L as 128.

The BSI is evaluated with fully-modelled $K = L = 128$ and under-modelling $K = \{64, 32\}$. The parameters of the proposed method are empirically chosen as $\mu = 3$, $\rho = 6 \cdot 10^{-5}$, $\lambda = 0.08$ and kept fixed in the experiments.

The RIRs estimated are shown from Fig. 1 to Fig. 3 for 5 different algorithms. Only the second channel is displayed for brevity. It can be seen that in the fully-modelled case (Fig. 1), conventional algorithms achieve better performance than the proposed method and the CSR for reasons analyzed in [11]. However, when the system is under-modelled, conventional methods fail to work reliably whereas the proposed method and the CSR algorithm can accurately estimate the target RIR (Fig. 2, 3). We note that the performance of the proposed method is similar to that of the CSR algorithm.

We next compare the computational complexity of different algorithms. For each K, the BSI is performed by using each algorithm for 100 times, and only the execution time of LMS updating is counted. The average execution time is normalized with respect to that of the NMCFLMS algorithm for $K = 32$. From Fig. 4 we see that while achieving similar performance with CSR, the average execution time of the proposed method is extremely lower.

5. CONCLUSION

In this paper, we propose a frequency-domain under-modelled multichannel BSI algorithm based on cross power spectrum and sparsity regularization. By exploiting the efficiency of FFT, we achieve similar BSI performance with our previous time-domain under-modelled BSI method while the computational complexity is greatly reduced.

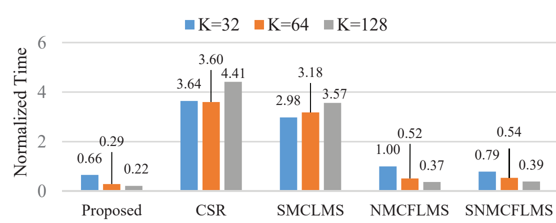

Fig. 4. Average execution time of different algorithms.

6. REFERENCES

[1] W. Xue, M. Brookes, and P. A. Naylor, “Under-modelled blind system identification for time delay estimation in reverberant environments,” in *Proc. Intl. Workshop on Acoustic Signal Enhancement (IWAENC)*, Xi’an, China, Sep. 2016.

[2] K. Kowalczyk, E. A. P. 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, pp. 653–656, 2013.

[3] S. Doclo and M. Moonen, “Robust adaptive time delay estimation for speaker localisation in noisy and reverberant acoustic environments,” *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 11, pp. 1110–1124, Oct. 2003.

[4] A. H. Moore, M. Brookes, and P. A. Naylor, “Room geometry estimation from a single channel acoustic impulse response,” in *Proc. European Signal Processing Conf. (EUSIPCO)*. Marrakech, Morocco: IEEE, Sep. 2013, p. TBC. [Online]. Available: http://www.commsp.ee.ic.ac.uk/~sap/uploads/publications/Moore2013.pdf

[5] F. Antonacci, J. Filos, M. Thomas, E. Habets, A. Sarti, P. Naylor, and S. Tubaro, “Inference of room geometry from acoustic impulse responses,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 10, pp. 2683 –2695, dec. 2012.

[6] H. A. Javed, A. H. Moore, and P. A. Naylor, “Spherical microphone array acoustic rake receivers,” in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016.

[7] ——, “Spherical harmonic rake receivers for dereverberation,” in *Proc. Intl. Workshop on Acoustic Signal Enhancement (IWAENC)*, 2016.

[8] Y. Huang and J. Benesty, “Adaptive multi-channel least mean square and Newton algorithms for blind channel identification,” *Signal Processing*, vol. 82, pp. 1127–1138, Aug. 2002.

[9] ——, “A class of frequency-domain adaptive approaches to blind multichannel identification,” *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[10] G. Xu, H. Liu, L. Tong, and T. Kailath, “A least-squares approach to blind channel identification,” *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.

[11] W. Xue, M. Brookes, and P. A. Naylor, “Cross-correlation based under-modelled multichannel blind acoustic system identification with sparsity regularization,” in *Proc. European Signal Processing Conf. (EUSIPCO)*, Budapest, Hungary, Aug. 2016.

[12] T. Goldstein and S. Osher, “The split Bregman method for L1-regularized problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, Jan. 2009.

[13] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ, USA: Prentice Hall, 1989.

[14] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.