Cross-Correlation Based Under-Modelled Multichannel Blind Acoustic System Identification with Sparsity Regularization

Wei Xue, Mike Brookes, Patrick A. Naylor
Dept. of Electrical and Electronic Engineering, Imperial College London, UK
Email: {w.xue, mike.brookes, p.naylor}@imperial.ac.uk

***Abstract*—In room acoustics, the under-modelled blind system identification (BSI) problem arises when the identified room impulse response (RIR) is shorter than the real one. Conventional BSI methods can perform poorly under these circumstances. In this paper, we propose an algorithm for multichannel BSI in under-modelled situations. Instead of minimizing the cross-relation error, a new optimization criterion is formulated, which is based on maximizing a cross-correlation criterion. We show that under the statistical model of reverberant signals, the cross-correlation based criterion helps to reduce the adverse effects of system under-modelling on BSI. Moreover, the optimization problem is regularized by including a sparsity term in the cost function. The optimization problem is finally solved based on the split Bregman method in the least-mean-square (LMS) framework. Experimental results show that the proposed method can perform effectively in the under-modelled situations in which conventional methods fail.**

I. Introduction

By using multiple microphones, the room impulse responses (RIRs) from a sound source to the microphones can be estimated using multichannel blind system identification (BSI) methods. Traditional methods aim to estimate the entire RIR, but many applications require only the early part of the RIR, which consists of the direct propagation path and a few early reflections [1]. These applications include the acoustic RAKE receiver for speech dereverberation [2], sound source localization [3, 4], and room geometry inference [5, 6]. If only the early RIR is identified using under-modelled BSI, the resulting estimation is inaccurate in general.

Conventional multichannel BSI methods are formulated non-adaptively [7, 8] or adaptively [4, 9–13], and adaptive methods are more popular due to the computational efficiency and better performance. Widely used adaptive algorithms include multichannel LMS (MCLMS) algorithm [9, 10], multichannel Newton (MCN) algorithm [9], multichannel frequency-domain LMS (MCFLMS) algorithm [11] and normalized MCFLMS (NMCFLMS) algorithm [11]. In addition to the general purpose BSI algorithms, others have been proposed which utilize the a priori characteristics of RIRs. For instance, the sparsity of (early) RIRs has been used to improve the convergence speed [12] and the robustness to noise [4, 13].

The fundamental rule for conventional adaptive multichannel BSI methods is the cross-relation (CR) property [8, 9], which will be introduced in Section III in this paper, and RIRs of different channels can be estimated based on minimizing the CR error between microphone pairs. However, when the system is under-modelled, there is always a mismatch between the RIR used for computing the CR error and the true RIR that would make the CR property valid. In such cases, conventional BSI methods may fail to work reliably. One method of early RIR estimation is to identify the entire RIR and then extract the early part by truncation. However, as this strategy introduces more irrelevant parameters to estimate, it increases the computational complexity and noise sensitivity, and decreases the convergence rate. Moreover, setting the RIR length requires knowledge of the true channel length which is normally unknown and difficult to estimate in practice.

In this paper, we propose an algorithm for under-modelled multichannel BSI. A new optimization problem is formulated, which is based on maximizing a cross-correlation criterion. With the statistical reverberant signal model, we show that the negative effect caused by the system under-modelling can be alleviated using our approach. Furthermore, as the early RIR only contains the direct path and sparse early reflections, a sparsity term is further integrated into the optimization objective. We finally derive an adaptive LMS algorithm, which is based on the split Bregman method [14], to solve the optimization problem. The experiments conducted in system under-modelling cases show that the proposed algorithm can perform reliably in cases in which conventional methods fail..

II. Signal Model

In a reverberant environment with a single sound source and an M-element microphone array, we have the following time-domain equation in matrix form for the i-th microphone at time index n:

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

where $\mathbf{x}_i^n = [x_i^n\ x_i^{n-1}\ \ldots\ x_i^{n-K+1}]^{\mathrm{T}}$ is the $K \times 1$ signal vector, with K as the length of the final identified RIR. The additive noise vector $\mathbf{v}_i^n$ is defined in a similar way to $\mathbf{x}_i^n$. $\mathbf{H}_i$ is a $K \times (K + L - 1)$ Toeplitz matrix constructed from the true L-tap RIR, $\mathbf{h}_i = [h_{i,0}\ h_{i,1}\ \ldots\ h_{i,L-1}]^{\mathrm{T}}$, as

$$\mathbf{H}_i = \begin{bmatrix} h_{i,0} & h_{i,1} & \cdots & h_{i,L-1} & 0 & \cdots & 0 \\ 0 & h_{i,0} & \cdots & h_{i,L-2} & h_{i,L-1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{i,0} & h_{i,1} & \cdots & h_{i,L-1} \end{bmatrix}.$$

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

In addition, $\mathbf{s}^n = [s^n \; s^{n-1} \; \dots \; s^{n-K-L+1}]^{\mathrm{T}}$ is a $(K+L-1)\times 1$ vector of the source signal.

In this paper, we are interested in only identifying the early RIR, which mainly contains the direct path and early reflections, thus $K < L$. Let us rewrite $\mathbf{h}_i$ as $\mathbf{h}_i = [\mathbf{h}_{i,e}^{\mathrm{T}} \; \mathbf{h}_{i,l}^{\mathrm{T}}]^{\mathrm{T}}$, where $\mathbf{h}_{i,e} = [h_{i,0} \; \cdots \; h_{i,K-1}]^{\mathrm{T}}$, and $\mathbf{h}_{i,l} = [h_{i,K} \; \cdots \; h_{i,L-1}]^{\mathrm{T}}$, which correspond to the early RIR and late RIR components respectively. Then we have $\mathbf{H}_i = \mathbf{H}_{i,e} + \mathbf{H}_{i,l}$, where $\mathbf{H}_{i,e}$ and $\mathbf{H}_{i,l}$ are two $K \times (K+L-1)$ matrices constructed by using $[\mathbf{h}_{i,e}^{\mathrm{T}} \; \mathbf{0}_{(L-K)\times 1}^{\mathrm{T}}]^{\mathrm{T}}$ and $[\mathbf{0}_{K\times 1}^{\mathrm{T}} \; \mathbf{h}_{i,l}^{\mathrm{T}}]^{\mathrm{T}}$, having the same structure as $\mathbf{H}_i$. In the absence of $\mathbf{v}_i^n$, the clean reverberant signal in (1) can be further expressed as:

$$\mathbf{x}_i^n = \mathbf{H}_{i,e}\mathbf{s}^n + \mathbf{H}_{i,l}\mathbf{s}^n = \mathbf{x}_{i,e}^n + \mathbf{x}_{i,l}^n, \tag{2}$$

where the reverberant signal is decomposed into the early reverberation $\mathbf{x}_{i,e}^n$ and the late reverberation $\mathbf{x}_{i,l}^n$. It can be assumed that $\mathbf{x}_{i,e}^n$ and $\mathbf{x}_{j,l}^n$ for $i,j = 1,...,M$ are uncorrelated [15]. Spatially, the late reverberation can be modelled as the diffuse sound field [15], and the pairwise cross-correlation is usually much smaller than the autocorrelations unless the microphones are very close in space (here we ignore this case as it does not provide enough spatial diversity for BSI).

III. Performance Analysis of CR Error Based BSI

In this section, we first briefly introduce the concept of the traditional CR error based methods in the time domain, and then discuss the problem encountered when the system is under-modelled. We note that the conclusions drawn in this section also apply to the frequency-domain methods.

In the absence of noise $\mathbf{v}_i^n$, the following CR property may be deduced from (1) [8, 9]:

$$\mathbf{h}_i^{\mathrm{T}}\mathbf{x}_j^n = \mathbf{h}_j^{\mathrm{T}}\mathbf{x}_i^n, \; i,j = 1,2,...,M, i \neq j. \tag{3}$$

Given the $K \times 1$ vector $\mathbf{g}_i = [g_{i,0} \; g_{i,1} \; \dots \; g_{i,K-1}]^{\mathrm{T}}$ as one estimate of the i-th channel impulse response, and letting $\mathbf{g} = [\mathbf{g}_1^{\mathrm{T}} \; \mathbf{g}_2^{\mathrm{T}} \; \dots \mathbf{g}_M^{\mathrm{T}}]$, then:

1) Fully-modelled BSI: For BSI of the entire RIR, i.e., $K = L$, define the CR error as

$$\varepsilon_{ij}^n = \tilde{x}_{ij}^n - \tilde{x}_{ji}^n, \tag{4}$$

where

$$\tilde{x}_{ij}^n = \mathbf{g}_i^{\mathrm{T}}\mathbf{x}_j^n, \; i,j = 1,2,...,M, \tag{5}$$

is called the cross-filtered signal.

Next, define $\chi(\mathbf{g})$ as the summation of mean squared error (MSE) values $\mathbb{E}\{[\varepsilon_{ij}^n]^2\}$ for all microphone pairs [9]:

$$\chi(\mathbf{g}) = \sum_{i=1}^{M-1}\sum_{j=i+1}^{M} \mathbb{E}\{[\varepsilon_{ij}^n]^2\}. \tag{6}$$

where $\mathbb{E}\{\cdot\}$ denotes expectation.

According to (3), if $\mathbf{g}_i = \alpha\mathbf{h}_i$ for $i = 1,2,...,M$, where α is an arbitrary non-zero scale factor, then $\chi(\mathbf{g}) = 0$. Thus the RIRs can be estimated by finding the RIRs which minimize the $\chi(\mathbf{g})$ based criterion, under the unit-norm constraint $||\mathbf{g}||_2^2 = 1$. This constraint is used to avoid the trivial solution with all RIRs equal to zero.

2) Under-modelled BSI: Now we analyse the case when the RIR is under-modelled, i.e., $K < L$. Given the RIR estimates, the cross-filtered signal in (5) becomes:

$$\tilde{x}_{ij}^n = \mathbf{g}_i^{\mathrm{T}}\mathbf{x}_{j,e}^n + \mathbf{g}_i^{\mathrm{T}}\mathbf{x}_{j,l}^n = \tilde{x}_{ij,e}^n + \tilde{x}_{ij,l}^n, \tag{7}$$

thus $\tilde{x}_{ij}^n$ is decomposed into $\tilde{x}_{ij,e}^n$ and $\tilde{x}_{ij,l}^n$ which we call respectively the early and late cross-filtered signals.

It can be easily verified that if $\mathbf{g}_i = \alpha\mathbf{h}_{i,e}$ for $i = 1,2,...,M$, where α is a non-zero scale factor, then $\tilde{x}_{ij,e}^n = \tilde{x}_{ji,e}^n$, while $\tilde{x}_{ij,l}^n \neq \tilde{x}_{ji,l}^n$. Thus in the under-modelled case, the CR property can hold *only* for the early cross-filtered signals, and the late cross-filtered signals act as unwanted noise.

Under the assumption in Section II, we have:

$$\begin{aligned}\mathbb{E}\{\tilde{x}_{ij,e}^n \cdot \tilde{x}_{ij,l}^n\} &= \mathbf{g}_i^{\mathrm{T}}\mathbb{E}\{\mathbf{x}_{j,e}^n[\mathbf{x}_{j,l}^n]^{\mathrm{T}}\}\mathbf{g}_i = 0,\\ \mathbb{E}\{\tilde{x}_{ij,e}^n \cdot \tilde{x}_{ji,l}^n\} &= \mathbf{g}_i^{\mathrm{T}}\mathbb{E}\{\mathbf{x}_{j,e}^n[\mathbf{x}_{i,l}^n]^{\mathrm{T}}\}\mathbf{g}_j = 0.\end{aligned} \tag{8}$$

Substituting (7) into (4), according to (8), $\mathbb{E}\{[\varepsilon_{ij}^n]^2\}$ in (6) is given by:

$$\mathbb{E}\{[\varepsilon_{ij}^n]^2\} = \mathbb{E}\{[\tilde{x}_{ij,e}^n - \tilde{x}_{ji,e}^n]^2\} + \mathbb{E}\{[\tilde{x}_{ij,l}^n - \tilde{x}_{ji,l}^n]^2\}. \tag{9}$$

In (9), the MSE is biased by the term $\mathbb{E}\{[\tilde{x}_{ij,l}^n - \tilde{x}_{ji,l}^n]^2\}$, which is also related $\mathbf{g}_i$ and $\mathbf{g}_j$ according to (7). With the diffuse late reverberation model [15], it is reasonable to assume that the cross-correlation of late cross-filtered signals is much smaller than the autocorrelations. Thus the bias term is dominated by the summation $\mathbb{E}\{[\tilde{x}_{ij,l}^n]^2\}+\mathbb{E}\{[\tilde{x}_{ji,l}^n]^2\}$. If the energy of the un-modelled part is strong, the large autocorrelation of late cross-filtered signals will result in a large MSE bias, which will mean that the optimal solution cannot approximate the true early RIR. On the other hand, when the estimated RIRs approach the true values, as $\mathbb{E}\{[\tilde{x}_{ij,e}^n - \tilde{x}_{ji,e}^n]^2\} \to 0$, the bias term will become more dominant in the optimization criterion so that the CR error based methods will not converge to the truncated true RIRs.

IV. Proposed Method

A. Cross-Correlation of Cross-Filtered Signals

The MSE values, $\mathbb{E}\{[\varepsilon_{ij}^n]^2\}$, in (6) can be regarded as a measure of the similarity of the cross-filtered signals, $\tilde{x}_{ij}^n$ and $\tilde{x}_{ji}^n$. Thus the conventional CR error based BSI algorithms can be regarded as finding the $\mathbf{g}$ that maximizes the similarity between all the pairs of cross-filtered signals. However when the system is under-modelled, the MSE does not reflect the similarity of the early components of the cross-filtered signals. One way to improve the performance is to find a similarity measure which is robust to the system under-modelling. In this subsection, similar to Section III, the cross-correlation between a pair of cross-filtered signals will be analysed.

The cross-correlation between a pair of cross-filtered signals is defined as:

$$\gamma_{ij} = \mathbb{E}\{\tilde{x}_{ij}^n \cdot \tilde{x}_{ji}^n\}. \tag{10}$$

Obviously, when the RIR is fully-modeled, as long as $\mathbf{g}_i = \alpha\mathbf{h}_i^{\mathrm{T}}$ for $i = 1,2,...,M$, then $\tilde{x}_{ij}^n = \tilde{x}_{ji}^n$, and the similarity between the signals reaches the maximum. If we take the cross-correlation as the similarity measure, the RIRs can be estimated based on maximizing γ_{ij} under the unit-norm constraint $||\mathbf{g}||_2^2 = 1$. The unit-norm constraint is needed to avoid the estimated RIR coefficients tending towards infinity.

In the system under-modelling case, since the CR property is only valid on the early cross-filtered signals, we can only rely on the cross-correlation between these signals for BSI. Substituting (7) into (10), according to (8), we have

$$\begin{aligned}\gamma_{ij} &= \mathbb{E}\{[\tilde{x}_{ij,e}^n + \tilde{x}_{ij,l}^n][\tilde{x}_{ji,e}^n + \tilde{x}_{ji,l}^n]\} \\ &= \mathbb{E}\{[\tilde{x}_{ij,e}^n \cdot \tilde{x}_{ji,e}^n]\} + \mathbb{E}\{[\tilde{x}_{ij,l}^n \cdot \tilde{x}_{ji,l}^n]\}. \end{aligned} \tag{11}$$

We can see that the cross-correlation based criterion is still biased by $\mathbb{E}\{[\tilde{x}_{ij,l}^n \cdot \tilde{x}_{ji,l}^n]\}$. However, as the autocorrelations of late cross-filtered signals have been totally removed in (11), and the remaining cross-correlation bias term is much smaller than autocorrelations, compared with the bias in (9), the bias in (11) is greatly reduced. Moreover, in contrast with the CR error based criterion, the more the estimated RIRs approach the true values, the higher the cross-correlation between early cross-filtered signals will be. This makes the bias term less dominant in the overall optimization criterion. Therefore, in the cross-correlation between the cross-filtered signals, the adverse effect of system under-modelling has been alleviated greatly.

B. Optimization Problem Formulation

Inspired by the analysis in the above, in this subsection, a new optimization problem is formulated, which is based on maximizing the cross-correlation of the cross-filtered signals. Moreover, as the early RIR generally consists of the direct path and sparse early reflections, an l_1-norm term is used to promote sparsity in the estimated system.

With (5), the cross-correlation in (10) becomes

$$\gamma_{ij} = \mathbf{g}_i^{\mathrm{T}} \mathbb{E}\{\mathbf{x}_j^n [\mathbf{x}_i^n]^{\mathrm{T}}\} \mathbf{g}_j. \tag{12}$$

Similarly to the CR error based BSI in [9], with (12), by using all microphone pairs, we first define

$$\Upsilon^n(\mathbf{g}) = \sum_{i=1}^{M} \sum_{j=1, j\neq i}^{M} \mathbf{g}_i^{\mathrm{T}} \mathbf{x}_j^n [\mathbf{x}_i^n]^{\mathrm{T}} \mathbf{g}_j = \mathbf{g}^{\mathrm{T}} \mathbf{R}^n \mathbf{g}, \tag{13}$$

where $\mathbf{R}^n$ is a $MK \times MK$ matrix having the form

$$\mathbf{R}^n = \begin{bmatrix} \mathbf{0}_{K\times K} & \mathbf{R}_{21}^n & \cdots & \mathbf{R}_{M1}^n \\ \mathbf{R}_{12}^n & \mathbf{0}_{K\times K} & \cdots & \mathbf{R}_{M2}^n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{1M}^n & \mathbf{R}_{2M}^n & \cdots & \mathbf{0}_{K\times K} \end{bmatrix} \tag{14}$$

with $\mathbf{R}_{ij}^n = \mathbf{x}_i^n [\mathbf{x}_j^n]^{\mathrm{T}}$. By enforcing the unit-norm constraint on $\mathbf{g}$ at all times, we then define the cross-correlation based cost function as

$$J^n(\mathbf{g}) = -\frac{\Upsilon^n(\mathbf{g})}{||\mathbf{g}||_2^2}. \tag{15}$$

We expect to find a unit-norm solution $\hat{\mathbf{g}}$ which has low sparsity and maximizes the cross-correlation of the cross-filtered signals. Maximizing the cross-correlation of the cross-filtered signals under the unit-norm constraint is equivalent to minimizing the expectation of $J^n(\mathbf{g})$. Let $\bar{J}(\mathbf{g}) = \mathbb{E}\{J^n(\mathbf{g})\}$, and taking l_1-norm of $\mathbf{g}$ as the sparsity measure, finally the optimization problem is formulated as:

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} \{\bar{J}(\mathbf{g}) + \rho |\mathbf{g}|_1\}, \text{ s.t. } ||\mathbf{g}||_2^2 = 1. \tag{16}$$

where ρ is a regularization parameter.

C. Adaptive LMS Updating

An LMS-type algorithm is derived here to solve the optimization problem (16) in an iterative manner.

Since $|\mathbf{g}|_1$ is included in the minimization,, the problem in (16) cannot be directly solved in the LMS framework. The cost function in (16) is a combination of the convex differential term and the l_1-norm term, which is similar to the optimization problem in [4]. According to [4], if first omitting the unit-norm constraint, (16) can be reformulated into an unconstrained optimization problem using a quadratic penalty function as:

$$(\hat{\mathbf{g}}, \hat{\mathbf{d}}) = \arg\min_{\mathbf{g},\mathbf{d}} \{\bar{J}(\mathbf{g}) + \rho |\mathbf{d}|_1 + \lambda ||\mathbf{d} - \mathbf{g}||_2^2\}, \tag{17}$$

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

Then the split Bregman iteration method [14] can be applied to (17), which iteratively solves the problem as

$$(\hat{\mathbf{g}}, \hat{\mathbf{d}})^{k+1} = \arg\min_{\mathbf{g},\mathbf{d}} \{\bar{J}(\mathbf{g}) + \rho |\mathbf{d}|_1 + \lambda ||\mathbf{d} - \mathbf{g} - \mathbf{b}^k||_2^2\}, \tag{18a}$$

$$\mathbf{b}^{k+1} = \mathbf{b}^k + \hat{\mathbf{g}}^{k+1} - \hat{\mathbf{d}}^{k+1}. \tag{18b}$$

where $\mathbf{b}$ is a $(KM) \times 1$ Bregman variable vector, and k denotes the iteration index. The problem of (18a) can be finally transformed into two sub-problems which can be solved with respect to $\mathbf{g}$ and $\mathbf{d}$, respectively [4]:

$$\hat{\mathbf{g}}^{k+1} = \arg\min_{\mathbf{g}} \{\bar{J}(\mathbf{g}) + \lambda ||\hat{\mathbf{d}}^k - \mathbf{g} - \mathbf{b}^k||_2^2\}, \text{ s.t. } ||\mathbf{g}||_2^2 = 1 \tag{19a}$$

$$\hat{\mathbf{d}}^{k+1} = \arg\min_{\mathbf{d}} \{\rho |\mathbf{d}|_1 + \lambda ||\mathbf{d} - \hat{\mathbf{g}}^{k+1} - \mathbf{b}^k||_2^2\}. \tag{19b}$$

1) Solving (19a)*:* $\hat{\mathbf{g}}$ is updated at each new sample, therefore, the iteration index k equals the time index n.

In the LMS framework, the expectation $\bar{J}(\mathbf{g})$ is replaced by the instantaneous value, $J^n(\mathbf{g})$. First ignoring the unit-norm constraint, by using the gradient descent method, we have

$$\begin{aligned}\hat{\mathbf{g}}^{n+1} &= \hat{\mathbf{g}}^n - \mu \frac{\partial J^n(\mathbf{g})}{\partial \mathbf{g}} - \mu\lambda \frac{\partial ||\hat{\mathbf{d}}^n - \mathbf{g} - \mathbf{b}^n||_2^2}{\partial \mathbf{g}} \\ &= \hat{\mathbf{g}}^n + 2\mu \frac{\mathbf{R}^n \hat{\mathbf{g}}^n + J^n(\hat{\mathbf{g}}^n) \hat{\mathbf{g}}^n}{||\hat{\mathbf{g}}^n||_2^2} - 2\mu\lambda(\hat{\mathbf{g}}^n + \mathbf{b}^n - \hat{\mathbf{d}}^n), \end{aligned} \tag{20}$$

where μ is the step-size. Then enforcing the unit-norm constraint by normalization after each update, we have the final form:

$$\begin{aligned}&\hat{\mathbf{g}}^{n+1} \\ &= \frac{\hat{\mathbf{g}}^n + 2\mu[\mathbf{R}^n \hat{\mathbf{g}}^n - \Upsilon^n(\hat{\mathbf{g}}^n) \cdot \hat{\mathbf{g}}^n - \lambda(\hat{\mathbf{g}}^n + \mathbf{b}^n - \hat{\mathbf{d}}^n)]}{||\hat{\mathbf{g}}^n + 2\mu[\mathbf{R}^n \hat{\mathbf{g}}^n - \Upsilon^n(\hat{\mathbf{g}}^n) \cdot \hat{\mathbf{g}}^n - \lambda(\hat{\mathbf{g}}^n + \mathbf{b}^n - \hat{\mathbf{d}}^n)||_2^2}. \end{aligned} \tag{21}$$

2) Solving (19b)*:* Similar to [4], in order to reduce the computational complexity, $\hat{\mathbf{d}}$ is updated only once for every P samples. If $(n+1) \mod P = 0$, according to [14], the optimal solution of $\hat{\mathbf{d}}$ can be obtained by updating each element as

$$\hat{\mathbf{d}}_i^{n+1} = \text{sign}(\hat{\mathbf{g}}_i^{n+1} + \mathbf{b}_i^n) \cdot \max\{|\hat{\mathbf{g}}_i^{n+1} + \mathbf{b}_i^n| - \frac{\rho}{2\lambda}, 0\}, \tag{22}$$

where $\mathbf{u}_i$ denotes the i-th element of the vector $\mathbf{u}$. If $(n+1) \mod P \neq 0$, $\hat{\mathbf{d}}$ keeps unchanged, then $\hat{\mathbf{d}}_i^{n+1} = \hat{\mathbf{d}}_i^n$.

Once $\hat{\mathbf{d}}$ is updated, the Bregman variable vector $\mathbf{b}$ is also updated in accordance with $\hat{\mathbf{d}}$ according to (18b).

V. Experiment

In this section, the performance of the proposed algorithm is evaluated in the simulated room environment. Four CR error based algorithms, including the widely used time-domain MCLMS algorithm [9], frequency-domain NMCFLMS algorithm [11], as well as the recently proposed sparse MCLMS (SMCLMS) algorithm and sparse NMCFLMS (SNMCFLMS) algorithm [4], are used for comparison.

A rectangular room with size 5 m × 6 m × 3 m is simulated using the image-source method [16]. We use $M = 2$ microphones, with positions (2.4, 2.0, 1.6) m and (2.6, 2.0, 1.6) m respectively to capture the signal from the source. The source is located at (2.05, 3.95, 1.67) m, a white Gaussian noise of duration 10 s is used as the source signal, and the sampling rate is 8 kHz. We set the reverberation time to be 300 ms, then simulate two RIRs from source to microphones, and truncate the RIRs to be 128-taps long. The signals captured by microphones are generated by convolving the source signal with the truncated RIRs. Therefore, in the experiment, the length of the target RIR in the signal model (1) is $L = 128$.

We evaluate the BSI performances with different values of the identified RIR length K, which decreases from $L = 128$ to $L/4 = 32$ with step size $L/4 = 32$. Thus totally one fully modelled case and three under-modelled cases are tested. In all experiments, we empirically choose the parameters of the proposed method as: $\mu = 0.02$, $\rho = 4 \cdot 10^{-4}$, $\lambda = 0.5$, $P = 30$.

The comparison results for channel 2 are illustrated in Fig. 1 to Fig. 4 (the results for channel 1 are similar and not shown for brevity). For each algorithm, the normalized projection misalignment (NPM) [17] is computed by comparing the estimated RIR with the first K taps of the target RIR, and a lower NPM indicates a better estimation result. We can observe from Fig. 1 that, when the RIR is fully-modelled, all the conventional methods can accurately estimate the target RIR. However, the RIR estimated by proposed method has larger errors in the late part. This can be mainly explained as follows. According to Section IV-A, the effect of late RIR is suppressed in the cross-correlation criterion. As a consequence, compared with the conventional methods, the proposed method becomes less sensitive to the error in the late RIR during the filter update, and the fluctuations in the late RIR cannot be well tracked. On the other hand, the sparsity regularization in (16) also makes the algorithm unable to estimate the late RIR which is actually not sparse. From Fig. 2 to Fig. 4, we can see that the proposed algorithm's under-modelled BSI is of comparable accuracy to the fully-modelled case, whereas for the existing algorithms, the BSI has failed in the under-modelled case. Moreover, it is shown for the proposed algorithm that key features of the early part of the estimated RIR are estimated almost independent of K.

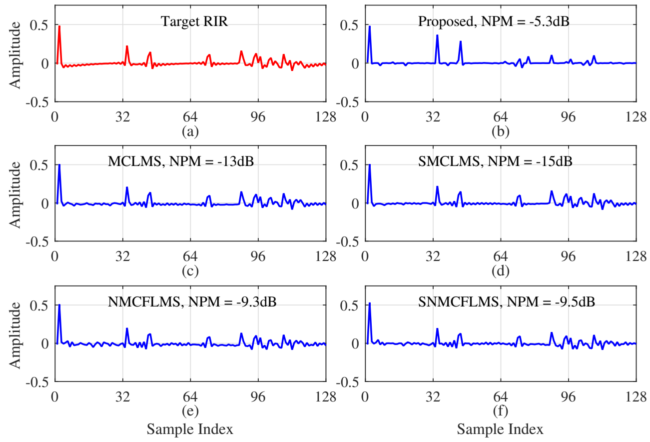

Fig. 1. Comparison results for $K = 128$: (a) The target RIR, and estimated RIRs by using (b) the proposed method, (c) MCLMS, (d) SMCLMS, (e) NMCFLMS and (f) SNMCFLMS, respectively.

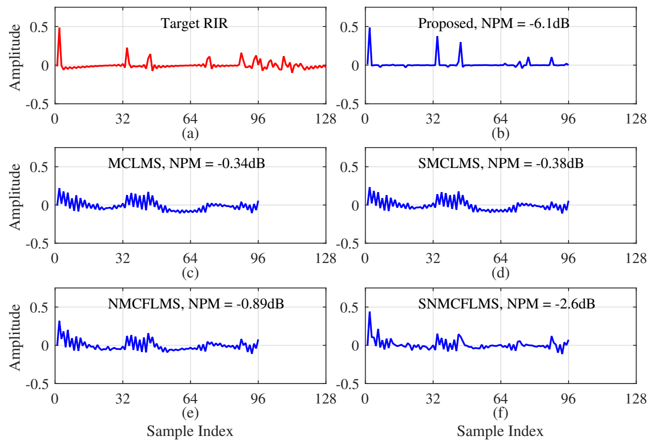

Fig. 2. Comparison results for $K = 96$: (a) The target RIR, and estimated RIRs by using (b) the proposed method, (c) MCLMS, (d) SMCLMS, (e) NMCFLMS and (f) SNMCFLMS, respectively.

VI. Conclusion

In this paper, we have proposed a novel algorithm for under-modelled multichannel BSI. A new optimization problem has been formulated, based on maximizing the cross-correlation of cross-filtered signals. By exploiting the sparse nature of the early RIRs in real acoustic environments, sparsity is additionally promoted in the optimization scheme. We solve the optimization problem by using a LMS algorithm based on the split Bregman method. By conducting experiments in different system under-modelling cases, we have demonstrated the effectiveness of the proposed method.

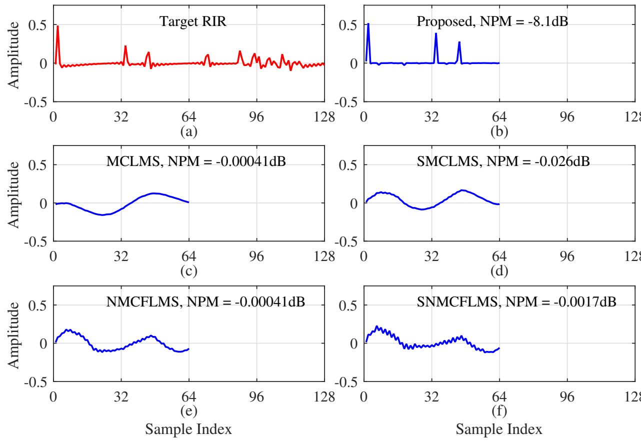

Fig. 3. Comparison results for $K = 64$: (a) The target RIR, and estimated RIRs by using (b) the proposed method, (c) MCLMS, (d) SMCLMS, (e) NMCFLMS and (f) SNMCFLMS, respectively.

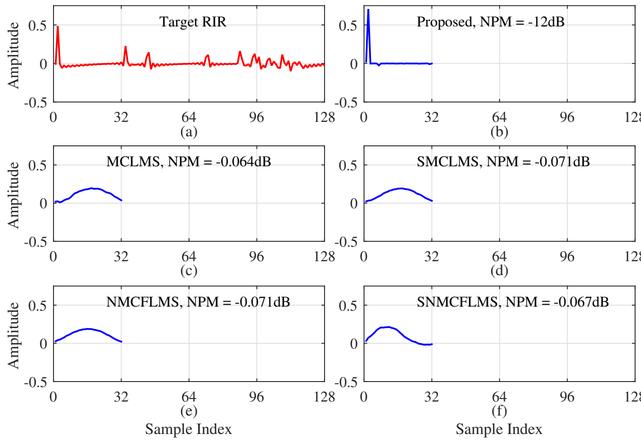

Fig. 4. Comparison results for $K = 32$: (a) The target RIR, and estimated RIRs by using (b) the proposed method, (c) MCLMS, (d) SMCLMS, (e) NMCFLMS and (f) SNMCFLMS, respectively.

Acknowledgment

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

References

[1] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*. Springer, 2010.

[2] 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.

[3] Y. Huang and J. Benesty, "Adaptive multichannel time delay estimation based on blind system identification for acoustic source localization," in *Adaptive Signal Processing*. Springer Berlin Heidelberg, 2003, pp. 227–247.

[4] 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.

[5] J. Filos, A. Canclini, M. R. P. Thomas, F. Antonacci, A. Sarti, and P. A. Naylor, "Robust inference of room getometry from acoustic impulse responses," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Barcelona, Spain, Aug. 2011.

[6] 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.

[7] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, Feb. 1995.

[8] 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.

[9] 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.

[10] J. Huang, J. Benesty, and J. Chen, "Optimal step size of the adaptive multichannel LMS algorithm for blind SIMO identification," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 173–176, March 2005.

[11] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[12] R. Ahmad, A. W. H. Khong, and P. A. Naylor, "Proportionate frequency domain adaptive algorithms for blind channel identification," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 5, Toulouse, France, 2006, pp. 2932–2935.

[13] Y. Lin, J. Chen, Y. Kim, and D. Lee, "Blind Sparse-Nonnegative (BSN) channel identification for acoustic Time-Difference-of-Arrival estimation," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2007, pp. 106–109.

[14] 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.

[15] A. Schwarz and W. Kellermann, "Coherent-to-diffuse power ratio estimation for dereverberation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 6, pp. 1006–1018, Jun. 2015.

[16] 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.

[17] D. R. Morgan, J. Benesty, and M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 174–176, Jul. 1998.