A MULTICHANNEL TIME-DOMAIN SUBSPACE APPROACH EXPLOITING MULTIPLE TIME-DELAYS FOR ACOUSTIC CHANNEL EQUALIZATION

Rajan S. Rashobh

BeingThere Centre, Institute for Media Innovation
Research Techno Plaza, Singapore, 637553
E-mail: rajansr@ntu.edu.sg

Andy W. H. Khong

School of EEE
Nanyang Technological University, Singapore, 639798
E-mail: andykhong@ntu.edu.sg

ABSTRACT

It is well-known that the performance of acoustic multichannel equalization (MCEQ) algorithms depend on the modeling delay that has to be pre-defined. In this work, we propose a MCEQ algorithm which achieves full equalization of acoustic room impulse responses in the presence of blind-system identification error. We achieve the above by modeling the inverse filters using multiple delays via columns of the identity matrix which serve as basis vectors for this time-domain sub-space approach. We further show that the proposed algorithm allows one to determine an optimal set of inverse filters corresponding to an appropriate modeling delay.

***Index Terms*—** Multichannel equalization, dereverberation, acoustic impulse responses

1. INTRODUCTION

Speech signals acquired by distant microphones in an enclosed space are often degraded by reverberation which adversely affects speech quality and intelligibility. One approach to achieve dereverberation is to equalize received signals using inverse filters that are computed from the acoustic impulse response (AIR) estimates provided by blind system identification (BSI) algorithms [1]–[4].

Since direct inversion results in an unstable inverse filter, single- or multi-channel equalization techniques have been employed for the equalization of non-minimum phase AIRs. Unlike single-channel techniques which result in approximate equalization only [5]–[7], the multiple-input/output inverse theorem (MINT) algorithm, which assumes co-prime channels [8], computes exact inverse filters [6]. In practice however, BSI algorithms do not provide error-free AIR estimates. Since MINT is not robust to such system mismatches, equalization using MINT in the presence of AIR estimation errors will introduce further distortion in the equalized output signal. A regularized least-squares approach has been proposed in [9] to improve the robustness of MINT. Channel shortening (CS) techniques [10]–[13] which achieve equalization of the late reverberation only (as opposed to the whole AIR) have also become increasingly popular in recent years. Although CS algorithms only perform partial equalization, they achieve higher robustness to BSI errors and have shown to enhance speech intelligibility. Recently, a fast frequency-domain algorithm showing higher robustness to BSI errors has also been proposed in [14].

In this paper, we propose a subspace-based approach to achieve full equalization of AIRs. As opposed to existing MINT [6] and CS [11]–[13] algorithms which estimate a single set of inverse filters across multiple channels, we employ multiple sets of inverse filters in a time-domain subspace for estimation and equalization. Results presented in [9], [15], [16] have shown that the equalization performance of existing algorithms is highly dependent on the choice of the modeling delay that has to be pre-defined. By employing columns of the identity matrix as basis vectors, we show that our proposed time-domain subspace approach is equivalent to modeling the inverse filters using multiple delays. Compared with existing algorithms, our formulation further allows one to determine an optimal set of inverse filters (with the appropriate delay) to achieve minimum equalization error. In addition, we show that the proposed approach not only is robust to BSI error, it can also achieve full equalization resulting in an improvement in PESQ scores compared to existing techniques.

2. MULTICHANNEL EQUALIZATION

Under noiseless condition, a signal $x_m(n)$ received by the mth microphone of an M-channel acoustic system is given by $x_m(n) = h_m * s(n), m = 1, \dots, M$, where $s(n)$ is the source signal, h_m is the AIR between the source and the mth microphone and $*$ is the linear convolution operator. In multichannel equalization (MCEQ), a set of filters $\mathbf{g}_m = [g_{m,0}, \dots, g_{m,L_g-1}]^T$ is estimated corresponding to $\mathbf{h}_m = [h_{m,0}, \dots, h_{m,L_h-1}]^T$, where L_g and L_h are the length of $\mathbf{g}_m$ and $\mathbf{h}_m$, respectively. Given a set of M AIRs, the MINT algorithm estimates the exact inverse filters, in theory, if the channels are co-prime [6] satisfying the relation

$$\sum_{m=1}^{M} \mathbf{H}_m^T \mathbf{g}_m = \mathbf{d}, \tag{1}$$

This research, which is carried out at BeingThere Centre, is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

where $\mathbf{H}_m$ is a $L_g \times L_d$ filtering matrix constructed from $\mathbf{h}_m$ with $L_d = L_g + L_h - 1$. The $L_d \times 1$ vector $\mathbf{d} = [\mathbf{0}_{1\times\tau}, 1, \mathbf{0}_{1\times(L_d-\tau-1)}]^T$ is a Kronecker delta function where an important parameter τ, defined as the modeling delay, has been incorporated so as to compensate for any non-causality of the MCEQ filter $\mathbf{g} = [\mathbf{g}_1^T, \ldots, \mathbf{g}_M^T]^T$ [9] and $\mathbf{0}_{1\times\tau}$ defines a $1 \times \tau$ null vector. From (1), the MCEQ filter can be estimated using [17]

$$\widehat{\mathbf{g}} = \arg\min_{\mathbf{g}} \|\mathbf{H}^T\mathbf{g} - \mathbf{d}\|_2^2 = (\mathbf{H}^T)^+\mathbf{d}, \tag{2}$$

where $(.)^+$ is the matrix pseudo-inverse operator, $\widehat{\mathbf{g}}$ is an estimate of $\mathbf{g}$, $\mathbf{H} = [\mathbf{H}_1^T, \ldots, \mathbf{H}_M^T]^T$. In practice, $\widehat{\mathbf{g}}$ is computed using an estimate $\widehat{\mathbf{H}} = [\widehat{\mathbf{H}}_1^T, \widehat{\mathbf{H}}_2^T, \ldots, \widehat{\mathbf{H}}_M^T]^T$ of $\mathbf{H}$ which is constructed from $\widehat{\mathbf{h}}_m, m = 1, \ldots, M$, obtained using BSI algorithms such as presented in [1]–[4]. It is therefore important for equalization algorithms to be robust to BSI errors.

The inverse of a mixed-phase sequence is, in theory, infinite in length and non-causal [18]. In practice however, to achieve minimum-norm solutions for the case of $M \geq 2$, one often choose a finite filter length $L_g \geq L_c$, where $L_c = \lceil \frac{L_h-1}{M-1} \rceil$ [19]. It has further been shown in [9], [13] that the equalization error is proportional to the energy of $\widehat{\mathbf{g}}$ and hence an appropriate choice of L_g will result in minimum-energy filters, increasing the robustness to BSI errors and noise. However, for large L_h, we often choose $L_c \leq L_g < L_h$ for practicality [19]. With $L_g < L_h$, the selection of an adequate τ is important to ensure good equalization performance [9].

To elaborate the importance of τ and its effect on MINT, we evaluated the equalization performance using $\widehat{\mathbf{g}}$ computed for $0 \leq \tau \leq L_d - 1$ where $M = 5$ synthetic AIRs, each of length $L_h = 512$, were generated using the method of images [20] at $f_s = 8$ kHz. Three sets of AIRs were generated inside a room of dimension $6 \times 5 \times 3$ m^3 for source-sensor distance $r = 0.3, 3.2$ and 7.1 m. Estimated AIRs were then simulated by perturbing $\mathbf{h}_m$ as

$$\widehat{\mathbf{h}}_m = (\mathbf{I}_{L_h} + \boldsymbol{\epsilon}_m)\mathbf{h}_m, \tag{3}$$

where $\boldsymbol{\epsilon}_m = \mathrm{diag}(\epsilon_{m,0}, \ldots, \epsilon_{m,L_h-1})$ such that $\mathcal{M}_m = 10\log_{10}\sigma_\epsilon^2 = -40$ dB is the system mismatch and $\epsilon_{m,i}$ is a zero-mean white Gaussian random variable of variance σ_ϵ^2. In this illustrative example, we have used $L_g{=}L_c$ giving $L_d{=}639$ and $\widehat{\mathbf{g}}$ was computed from $\widehat{\mathbf{H}}$ using $\mathbf{d}$ with a modeling delay τ and (2). For each set of AIRs, we then computed $e(\tau){=}10\log_{10}\|\mathbf{d} - \mathbf{H}^T\widehat{\mathbf{g}}\|_2^2$ dB for $0 \leq \tau \leq L_d - 1$ in order to quantify the equalization of $\mathbf{H}$. We note from Fig. 1(a) that the optimal value of τ is dependant on the AIRs. For the set of AIRs with $r = 0.3$ m, $e(\tau) = -45$ dB is the lowest when $\tau = 29$. However this choice of τ will increase $e(\tau)$ by 70 and 85 dB when $r = 3.2$ and 7.1 m, respectively. In addition, for an arbitrarily chosen $\tau = L_h/2$, the set of AIRs for $r = 7.1$ m gives a lower $e(\tau) \approx -4$ dB compared to $e(\tau) \approx 26$ and 30 dB for the case of $r = 0.3$ and 3.2 m, respectively. These results explicitly show that the selection of τ is crucial in the process of AIR equalization when $L_g < L_h$. In practice, $\mathbf{H}$ is unknown and hence evaluating $e(\tau)$ to obtain an optimal τ corresponding to the lowest $e(\tau)$ is not possible.

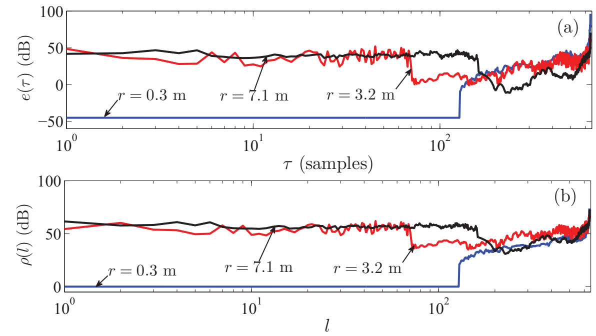

Fig. 1. a) Variation of $e(\tau)$ with τ, and b) variation of $\rho(l)$ with l, for different AIR sets.

3. THE PROPOSED MCEQ ALGORITHMS

3.1. Equalization using multiple filters (MCEQ-MF)

We propose to achieve frame-by-frame equalization where the received signals are first partitioned into K overlapping frames, each of length L_x, with a frame shift of N samples, where $1 \leq N \leq L$ and $L = L_x + L_h - 1$. This time-domain algorithm can be characterized by an identity matrix $\mathbf{I}_L = [\mathbf{i}_1, \ldots, \mathbf{i}_L]$, where each basis vector $\mathbf{i}_l = [\mathbf{0}_{1\times(l-1)}, 1, \mathbf{0}_{1\times(L-l)}]^T$ is used to define a subspace in the time domain. Therefore, to describe the proposed MCEQ multiple filter (MCEQ-MF) algorithm, we first express the kth clean speech frame $\mathbf{s}(k) = [s(kN + L_x), \ldots, s(kN + L_x - L + 1)]^T$, where $s(n) = 0\,\forall\, n < 0$. We note that since

$$\mathbf{s}(k) = \mathbf{I}_L\mathbf{s}(k), \tag{4}$$

the lth element corresponding to the lth subspace of $\mathbf{s}(k)$ can be expressed using the lth basis vector as

$$s(kN + L_x - l + 1) = \mathbf{i}_l^T\mathbf{s}(k), \qquad l = 1, \ldots, L. \tag{5}$$

The basis vectors therefore allow one to define time-delayed elements of a signal vector. In addition, $\mathbf{s}(k) = \sum_{l=1}^{L} s(kN + L_x - l + 1)\mathbf{i}_l$. With $\mathbf{H}_m = \mathbf{I}_{L_x}\mathbf{H}_m\mathbf{I}_L$ and (4) we can express, similar to (4), the kth frame of the received signal by

$$\begin{aligned}\mathbf{x}_m(k) &= [x_m(kN + L_x), \ldots, x_m(kN + 1)]^T = \mathbf{H}_m\mathbf{s}(k) \\ &= \mathbf{I}_{L_x}\mathbf{x}_m(k), \quad m = 1, \ldots, M.\end{aligned} \tag{6}$$

Defining $\mathbf{g}_{m,l} = [g_{m,l}(1), \ldots, g_{m,l}(L_x)]^T$ as the mth channel inverse filter for the lth subspace, we can estimate, using these subspace filters, the lth subspace of $\mathbf{s}(k)$ given by

$$\widehat{s}(kN{+}L_x{-}l{+}1) = \sum_{m=1}^{M}\mathbf{g}_{m,l}^T\mathbf{x}_m(k) = \left(\sum_{m=1}^{M}\mathbf{g}_{m,l}^T\mathbf{H}_m\right)\mathbf{s}(k), \tag{7}$$

where we have utilized (6). If the AIRs are co-prime, perfect estimate of $s(kN + L_x - l + 1)$ is theoretically possible if

$$\sum_{m=1}^{M}\mathbf{g}_{m,l}^T\mathbf{H}_m = \mathbf{d}_l^T, \tag{8}$$

where $\mathbf{d}_l$ is a Kronecker delta function corresponding to the lth subspace such that $\tau = l - 1$ defines the relationship between modeling delay and the subspace index. Therefore, in practice, each of these inverse filters, which corresponds to the lth subspace and comprises of a set of inverse filters (across M channels), can be estimated using $\widehat{\mathbf{H}}$ as

$$\widehat{\mathbf{g}}_l = [\widehat{\mathbf{g}}_{1,l}^T, \ldots, \widehat{\mathbf{g}}_{M,l}^T]^T = (\widehat{\mathbf{H}}^T)^{+}\mathbf{d}_l. \tag{9}$$

Defining $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_L]$, the entire set of inverse filters can be obtained across all L subspaces using

$$\widehat{\mathbf{G}} = [\widehat{\mathbf{g}}_1, \ldots, \widehat{\mathbf{g}}_L] = [\widehat{\mathbf{H}}^T]^{+}\mathbf{D}, \tag{10}$$

where $\mathbf{D}$ is also equivalent to $\mathbf{I}_L$ which corresponds to L possible delays.

Employing the filters obtained by the proposed MCEQ-MF algorithm in (10), the lth coefficient of the kth frame source signal is estimated using $\mathbf{x}_m(k)$ given by

$$\widehat{s}(kN + L_x - l + 1) = \widehat{\mathbf{g}}_l^T[\mathbf{x}_1^T(k), \ldots, \mathbf{x}_M^T(k)]^T. \tag{11}$$

The kth frame of the recovered speech is related to the above by

$$\widehat{\mathbf{s}}(k) = \sum_{l=1}^{L} \widehat{s}(kN + L_x - l + 1)\mathbf{i}_l. \tag{12}$$

The above implies that L sets of M equalization filters, each of length L_x, are used for recovering $\mathbf{s}(k)$. For better reconstruction, overlapped windows can be applied for data partitioning and only a segment $[\widehat{s}(kN + L_x + (\beta - 1)L + 1), \ldots, \widehat{s}(kN + L_x + (\beta - 1)L + N)]$ from $\widehat{\mathbf{s}}(k)$ is required to construct $\widehat{s}(n)$, where $\beta = (l_{\mathrm{s}} - 1)/L$ with $1 \leq l_{\mathrm{s}} \leq L - N + 1$. Hence only N corresponding sets of filters $[\widehat{\mathbf{g}}_{l_{\mathrm{s}}}, \ldots, \widehat{\mathbf{g}}_{l_{\mathrm{e}}}] = [\mathbf{H}^T]^{+}[\mathbf{d}_{l_{\mathrm{s}}}, \ldots, \mathbf{d}_{l_{\mathrm{e}}}]$, where $l_{\mathrm{e}} = l_{\mathrm{s}} + N - 1$, need to be estimated. To prevent data loss due to truncation, βL and $L(1 - \beta) - N$ zeros are padded to the beginning and the end of $\mathbf{x}_m(n)$, respectively.

3.2. Insights into the MCEQ-MF algorithm

The relationship between MCEQ-MF and MINT can be explained by noting that the subspace operator $\mathbf{i}_l$ contains delay information of the lth coefficient in $\mathbf{s}(k)$. Therefore, while MINT employs a single filter $\widehat{\mathbf{g}}_m$ per channel to recover all samples in $\mathbf{s}(n)$, MCEQ-MF estimates and employs a dedicated filter $\widehat{\mathbf{g}}_{m,l}$ per channel for only the lth coefficient in $\widehat{\mathbf{s}}(k)$ described by (11). The link between MCEQ-MF, applying $\widehat{\mathbf{g}}_l$ estimated using (9) and MINT with $\tau = l - 1$, can be described by letting the frame shift $N = 1$ in (11) and recovering only the lth sample in each frame. Signal $\widehat{\mathbf{s}}(n)$ can be obtained by concatenating $\widehat{s}(k + L_x - l + 1)$, $k = 0, \ldots, K - 1$, where K is the length of the zero padded $\mathbf{x}_m(n)$ since $N = 1$.

We explain why MCEQ-MF outperforms MINT with an inadequate τ by first noting that $\mathbf{D}$ consists of all L possible delays and $\widehat{\mathbf{s}}(k)$ is obtained using multiple sets of filters $\widehat{\mathbf{g}}_l, l = l_{\mathrm{s}}, \ldots, l_{\mathrm{e}}$, with each set corresponding to $\tau = l - 1$ across the M channels. Application of multiple delays in MCEQ-MF results in a smaller mean-square error (MSE) than MINT with an inadequate choice of τ since, in MCEQ-MF, each coefficient of $\mathbf{s}(k)$ is estimated using a dedicated set of inverse filters. While the estimation error is common across all the elements in $\mathbf{s}(k)$ for MINT due to the application of a single set of inverse filters, the error associated with each element is different for MCEQ-MF.

To quantify the improvement of MCEQ-MF over MINT using an inadequate τ in the presence of BSI error, we let $\widehat{\mathbf{H}} = \mathbf{H} + \boldsymbol{\mathcal{E}}$, where $\boldsymbol{\mathcal{E}} = [\boldsymbol{\mathcal{E}}_1^T, \ldots, \boldsymbol{\mathcal{E}}_M^T]^T$ is a $ML_g \times L$ matrix constructed from the estimation error. For MINT, with $N = 1$ and defining $p = k + L_x - l + 1$, we can express (11) as

$$\widehat{s}(p) = \sum_{m=1}^{M} \widehat{\mathbf{g}}_{m,l}^T[(\widehat{\mathbf{H}}_m - \boldsymbol{\mathcal{E}}_m)\mathbf{s}(k)] = s(p) - \delta(p), \tag{13}$$

where $\delta(p) = \sum_{m=1}^{M} \widehat{\mathbf{g}}_{m,l}^T \boldsymbol{\mathcal{E}}_m \mathbf{s}(k) = \widehat{\mathbf{g}}_l^T \boldsymbol{\eta}(k)$ denotes the reconstruction error, $\boldsymbol{\eta}(k) = [\boldsymbol{\eta}_1^T(k), \ldots, \boldsymbol{\eta}_M^T(k)]^T$ and $\boldsymbol{\eta}_m(k) = \boldsymbol{\mathcal{E}}_m \mathbf{s}(k)$. For MINT, the MSE corresponding to equalization of $\mathbf{s}(n)$ using a specific τ, which results in a $\widehat{\mathbf{g}}_l$ where $l = \tau + 1$, is given by $\xi_{\mathrm{MINT}} = \widehat{\mathbf{g}}_{\tau+1}^T \mathbf{R}_\eta \widehat{\mathbf{g}}_{\tau+1}$, where $\mathbf{R}_\eta = E[\boldsymbol{\eta}^T(k)\boldsymbol{\eta}(k)]$ and $E(.)$ denotes the expectation operator. As opposed to the above, for MCEQ-MF with $N = l_{\mathrm{e}} - l_{\mathrm{s}} + 1$ sets of inverse filters, $\xi_{\mathrm{MF}} = \frac{1}{N}\sum_{l=l_{\mathrm{s}}}^{l_{\mathrm{e}}} \widehat{\mathbf{g}}_l^T \mathbf{R}_\eta \widehat{\mathbf{g}}_l$. We note that, for an inappropriate choice of τ, $\xi_{\mathrm{MF}} < \xi_{\mathrm{MINT}}$ as a consequence of averaging. Therefore, MCEQ-MF ensures a lower MSE compared to MINT with an inappropriate choice of τ.

3.3. Equalization using optimal filter (MCEQ-F$_{\mathrm{opt}}$)

Using the inherent property of MCEQ-MF, it is also possible to find a *single* set (out of the L sets) of M inverse filters that yields the minimum equalization error, corresponding to the optimal delay τ_{opt}, from $\widehat{\mathbf{G}}$. Compared to performing an iterative search for the optimal τ using MINT, the proposed algorithm is a better implementation.

We first show that the optimal filter $\widehat{\mathbf{g}}_l'$ corresponding to the τ_{opt} achieves the smallest energy which is in accordance with results shown in [9], [15]. Letting $\mathbf{d}_{l'}$ denote the optimal desired response, the least-square error obtained with an arbitrarily chosen $\widehat{\mathbf{g}}_l$ is then given by

$$\|\mathbf{H}^T\widehat{\mathbf{g}}_l - \mathbf{d}_{l'}\|_2^2 = \|\mathbf{H}^T(\widehat{\mathbf{g}}_l - \widehat{\mathbf{g}}_{l'}) + (\mathbf{H}^T\widehat{\mathbf{g}}_{l'} - \mathbf{d}_{l'})\|_2^2. \tag{14}$$

For the optimal solution $\widehat{\mathbf{g}}_{l'}$ we achieve the least squared error $\mathbf{H}^T\widehat{\mathbf{g}}_{l'} - \mathbf{d}_{l'}$, which is in the null space of $\mathbf{H}$ giving $\mathbf{H}(\mathbf{H}^T\widehat{\mathbf{g}}_{l'} - \mathbf{d}_{l'}) = \mathbf{0}_{ML_g \times 1}$ [21]. Hence, $[\mathbf{H}^T(\widehat{\mathbf{g}}_l - \widehat{\mathbf{g}}_{l'})]^T(\mathbf{H}^T\widehat{\mathbf{g}}_{l'} - \mathbf{d}_{l'}) = 0$, which implies that $\mathbf{H}^T(\widehat{\mathbf{g}}_l - \widehat{\mathbf{g}}_{l'}) \perp (\mathbf{H}^T\widehat{\mathbf{g}}_{l'} - \mathbf{d}_{l'})$. Therefore $\|\mathbf{H}^T\widehat{\mathbf{g}}_l - \mathbf{d}_{l'}\|_2^2 = \|\mathbf{H}^T(\widehat{\mathbf{g}}_l - \widehat{\mathbf{g}}_{l'})\|_2^2 + \|\mathbf{H}^T\widehat{\mathbf{g}}_{l'} - \mathbf{d}_{l'}\|_2^2$. For $\widehat{\mathbf{g}}_l \neq \widehat{\mathbf{g}}_{l'}$, we have

$$\|\mathbf{H}^T\widehat{\mathbf{g}}_{l'} - \mathbf{d}_{l'}\|_2^2 < \|\mathbf{H}^T\widehat{\mathbf{g}}_l - \mathbf{d}_{l'}\|_2^2 \tag{15}$$

since $\mathbf{H}^T(\widehat{\mathbf{g}}_l - \widehat{\mathbf{g}}_{l'}) \neq \mathbf{0}_{L\times 1}$. From (15) we can deduce that, for the optimal least-square solution, $\|\widehat{\mathbf{g}}_{l'}\|_2^2 < \|\widehat{\mathbf{g}}_l\|_2^2, l = 1, \ldots, L, l \neq l'$. Hence the inverse filter corresponding to the τ_{opt} has the lowest energy and it is therefore expected that $\widehat{\mathbf{g}}_{l'}$ can achieve the smallest equalization error.

We investigate the relationship between the estimation error and filter energy by plotting the filter-norm $\rho(l) = \|\widehat{\mathbf{g}}_l\|_2$ for the same sets of AIRs described in Section 2 as shown in Fig. 1(b). Comparing Figs. 1(a) and 1(b), we can see that the plots look almost identical, except for the scale, i.e., the range

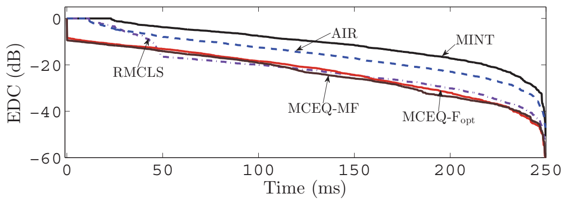

Fig. 2. Equalization performance measured using EDCs for various equalization algorithms.

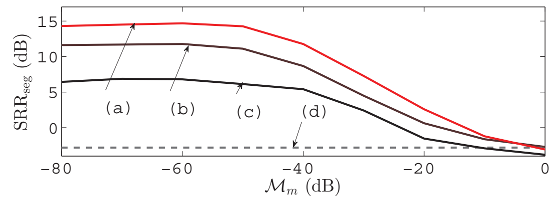

Fig. 3. The $\mathrm{SRR_{seg}}$ achieved for speech equalized with (a) MCEQ-F$_{\rm opt}$, (b) MCEQ-MF, (c) MINT, and (d) noisy reverberant data.

of τ for which $e(\tau)$ and $\rho(l)$ is minimum is the same. This implies that the delay associated with the lth subspace giving the smallest estimated error $e(\tau)$ is equivalent to that giving the smallest $\rho(l)$, and the corresponding inverse filter set $\widehat{\mathbf{g}}_{l'} = [\widehat{\mathbf{g}}_{1,l'}^T, \ldots, \widehat{\mathbf{g}}_{M,l'}^T]^T$ achieves the best equalization performance. In MCEQ-F$_{\rm opt}$, as opposed to MCEQ-MF where each element of $\mathbf{s}(k)$ is estimated using a dedicated filter, the full length $\mathbf{s}(n)$ is estimated using a single set of filters $\widehat{\mathbf{g}}_{l'}$ as $\widehat{\mathbf{s}}(n) = \sum_{m=1}^{M} \widehat{\mathbf{G}}_{m,l'}^T \mathbf{x}_m(n)$ where $\widehat{\mathbf{G}}_{m,l'}$ is the filtering matrix of $\widehat{\mathbf{g}}_{m,l'}$.

4. SIMULATION RESULTS

We first compare the equalization performance of the proposed MCEQ-MF and MCEQ-F$_{\rm opt}$ algorithms with MINT, and the relaxed multichannel least squares (RMCLS) algorithm [11], which is a CS-based approach. In this paper we have chosen $L_g = L_{\rm c}$ [13] and $l = \lceil L_g/4 \rceil, \ldots, \lceil L_g/2 \rceil$ for MCEQ-MF. We used two recorded AIRs from the MARDY database [22], which were re-sampled to 8 kHz and truncated to $L_h = 2000$. Equalization was then achieved using the inverse filters estimated from $\widehat{\mathbf{h}}_m, m = 1, 2$, simulated using (3) with $\mathcal{M}_m = -20$ dB. We used a commonly chosen value $\tau = 0$ for MINT, and a relaxation window length of $L_w = 50$ ms for RMCLS [12]. For MCEQ-MF, to compute the desired response $\widehat{\mathbf{d}}$, each $\mathbf{h}_m$ was partitioned into frames of length $L_x = L_g$ after zero padding to a length $(\lceil L_h/N \rceil - 1)N + L_g$, where in this example $N = 501$ is equal to the number of subspaces used for reconstruction. Frame-wise equalization was then performed using $\widehat{\mathbf{d}}(k) = \sum_{l=l_{\rm s}}^{l_{\rm e}} \widehat{d}(k,l)\mathbf{i}_l$, where $\widehat{d}(k,l) = \widehat{\mathbf{g}}_l^T [\mathbf{h}_1^T(k), \ldots, \mathbf{h}_M^T(k)]^T$, $\quad l = 500, \ldots, 1000$, while $\mathbf{h}_m(k)$ and $\widehat{\mathbf{d}}(k)$ denote the kth frame of $\mathbf{h}_m$ and $\widehat{\mathbf{d}}$, respectively. The full-length $\widehat{\mathbf{d}}$ was obtained using $\widehat{\mathbf{d}}(k)$.

In Fig. 2, we show the energy decay curve (EDC) [23] of $\widehat{\mathbf{d}}$ achieved with each algorithm, truncated to L_h samples, along with the EDC of the recorded AIRs (averaged across the channels). This result shows that, in the presence of system mismatch, MINT achieves a lower decay rate of the EDC than that of the AIRs, which implies that MINT with an arbitrary choice of $\tau = 0$ fails to achieve equalization. The RMCLS algorithm achieves better suppression of the late reverberation, while the early reflections are not fully equalized. On the other hand, in addition to achieving better equalization of the early reverberation than existing techniques, the proposed algorithms achieve comparable suppression of the late reflections to RMCLS.

Table 1. PESQ scores.

Method	$\mathcal{M}_m = -40$ dB	$\mathcal{M}_m = -20$ dB
MINT	2.48	2.19
RMCLS	2.86	2.73
MCEQ-MF	3.3	2.74
MCEQ-F$_{\rm opt}$	3.36	3.1

Next, to compare the robustness of the algorithms to system mismatches in terms of segmental signal-to-reverberation ratio ($\mathrm{SRR_{seg}}$) [23], we generated a set of five synthetic AIRs each at twenty different locations with the same setup as described in Section 2 where $L_h = 2000$. With an SNR $=$ 40 dB, equalization was then achieved for $-80 \le \mathcal{M}_m \le$ 0 dB. We have not included RMCLS in this comparison since CS algorithms, which deal with perceptual quality only [11], result in a lower $\mathrm{SRR_{seg}}$ as a result of partial equalization. From the $\mathrm{SRR_{seg}}$, averaged across the twenty sets of AIRs, shown in Fig. 3, we note that the proposed MCEQ-MF and MCEQ-F$_{\rm opt}$ algorithms outperform MINT exhibiting higher robustness to additive noise and BSI errors.

We also compare the perceptual quality of the equalized signals with the clean speech signal as the reference using PESQ [24] for the same set of recorded AIRs used in the first simulation. Equalization was performed on the reverberant data with $\mathcal{M}_m = -40$ and -20 dB at SNR $=$ 40 dB with $L_w = 50$ ms for RMCLS. The PESQ for signal corresponding to the first channel was computed as 2.52. From Table 1 we note that MINT results in the lowest PESQ scores indicating further degradation in the perceptual quality. On the other hand, the proposed MCEQ-MF and MCEQ-F$_{\rm opt}$ algorithms achieve the highest PESQ scores as a result of better equalization performance.

5. CONCLUSION

By applying multiple sets of inverse filters, the proposed MCEQ-MF outperforms MINT using an inadequate modeling delay. In the proposed MCEQ-F$_{\rm opt}$ algorithm, equalization is performed using an optimal set of inverse filters thereby achieving higher robustness to BSI errors and better performance. Unlike CS algorithms, which perform partial equalization to achieve robustness, the proposed algorithms attain robust full equalization of the AIRs. Simulation results show that the proposed algorithms exhibit higher robustness to AIR estimation errors and achieve better equalization.

6. REFERENCES

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

[2] 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.

[3] R. Ahmad, A. W. H. Khong, and P. A. Naylor, "Proportionate frequency domain adaptive algorithms for blind channel identification," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, May 2006, vol. 5.

[4] S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," *EURASIP J. Applied Signal Process.*, vol. 11, pp. 1074–1090, 2003.

[5] J. Mourjopoulos, P. Clarkson, and J. Hammond, "A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, May 1982, vol. 7, pp. 1858–1861.

[6] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 36, no. 2, pp. 145–152, Feb. 1988.

[7] L. Liao and A. W. H. Khong, "An adaptive sub-system based algorithm for channel equalization in a SIMO system," *IEEE Trans. Circuits and Systems I*, to appear.

[8] A. W. H. Khong, X. (Shawn) Lin, and P. A. Naylor, "Algorithms for identifying clusteres of near-common zeros in multichannel blind system identification and equalization," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, Mar. 2008.

[9] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations," *EURASIP J. Advances in Signal Process.*, vol. 2007:34013, no. 1, pp. 1–12, 2007.

[10] M. Kallinger and A. Mertins, "Multi-channel room impulse response shaping - a study," in *Proc. Int. Conf. Acoust., Speech, and Signal Process.*, May 2006, vol. 5.

[11] W. Zhang, E. A. P. Habets, and P. A. Naylor, "On the use of channel shortening in multichannel acoustic system equalization," in *Proc. Int. Workshop Acoust. Echo Noise Control*. Sep. 2010.

[12] F. Lim and P. A. Naylor, "Relaxed multichannel least squares with constrained initial taps for multichannel dereverberation," in *Proc. Int. Workshop Acoust. Signal Enhancement*, Sep. 2012.

[13] I. Kodrasi, S. Goetze, and S. Doclo, "Increasing the robustenss of acoustic multichannel equalization by means of regularization," in *Proc. Int. Workshop Acoust. Signal Enhancement*. Sep. 2012.

[14] R. S. Rashobh and A. W. H. Khong, "A fast frequency-domain algorithm for equalizing acoustic impulse responses," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 797–800, Dec. 2012.

[15] T. Hikichi, M. Delcroix, and M. Moiyoshi, "On robust invesrse filter design for room transfer function fluctuations," in *Proc. European Signal Process. Conf.*, Sep. 2006.

[16] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Sep. 2006, pp. 1–4.

[17] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 882–895, Sep. 2005.

[18] S. Treitel and E. Robinson, "The design of high resolution digital filters," *IEEE Trans. Geosci. Electron.*, vol. VGE-4, pp. 25–38, Jun. 1966.

[19] G. Harikumar and Y. Bresler, "FIR perfect signal reconstruction from multiple convolutions: minimum deconvolver orders," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 215–218, 1998.

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

[21] S. Haykin, *Adaptive filter theory*, Prentice-Hall, Upper Saddle River, NJ, 4th edition, 2002.

[22] J. Y. C. Wen, N. D. Gaubitch, E. A. P. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Sep. 2006.

[23] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*, Springer Publishing Company Inc., 1st edition, 2010.

[24] "ITU-T Recommendation p.862: Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narroband telephone networks and speech codecs," International Telecommunication Union (ITU-T), Feb. 2001.