A Subspace-based Adaptive Approach for Multichannel Equalization of Room Acoustics

Di Liu *, Rajan S. Rashobh *, Andy W. H. Khong * and Masahiro Yukawa †
* Nanyang Technological University, Singapore
E-mail:{liudi, rajansr, andykhong}@ntu.edu.sg
† Niigata University, Niigata, Japan
E-mail: yukawa@eng.niigata-u.ac.jp

***Abstract*— Multichannel equalization (MCEQ) techniques have been proposed for the inversion of acoustic impulse responses (AIRs) with a wide range of applications such as speech dereverberation. Adaptive MCEQ techniques can address the problem of high computational complexity required for the inversion of high order AIRs using non-adaptive algorithms based on the multiple-input/output inverse theorem (MINT). In this paper, we propose a subspace-based adaptive algorithm for MCEQ in the time domain. Unlike conventional techniques, instead of using a single set of inverse filters, multiple sets of inverse filters, each designated for each subspace, are adaptively estimated in our algorithm and equalization is achieved using these filters. Simulation and experimental results show that the proposed adaptive algorithm not only achieves a high convergence rate but also yields better equalization in terms of quality. In addition, the proposed algorithm achieves higher robustness to AIR estimation errors.**

I. Introduction

Reverberation is caused by multi-path propagation of an acoustic signal from its source to one or more microphones within an enclosed environment. A speech signal can be distorted by room reverberation, resulting in reduced intelligibility as well as a reduction in performance of, for example, automatic speech recognizers, particularly for hands-free devices and hearing aids [1]. It is therefore necessary to apply dereverberation techniques so as to improve the speech quality. One method to achieve dereverberation is to estimate the acoustic impulse responses (AIRs) using blind system identification (BSI) algorithms such as proposed in [2]. The estimated AIRs are then inverse filtered with the received signals to achieve dereverberated speech. In this work, we focus on the estimation of inverse filters for equalization.

Inverse systems can be obtained, for the single-channel case, using least squares and homomorphic equalization [3]. However, since AIRs are often non-minimum in phase [1], it is challenging to achieve perfect equalization using such single-channel techniques. When multiple microphones are deployed, inverse systems can be obtained by employing the multiple-input/output inverse theorem (MINT) [4]. In theory, channel equalization can be achieved using MINT if the AIRs do not contain any common zeros [4]. However, high computational costs required for the inversion of the convolution matrix prevents MINT from being directly applied in practical scenarios where the length of a typical AIR is of several hundred samples.

Adaptive inverse filtering algorithms such as [5], [6] have been proposed to reduce the computational complexity incurred in the close-form solution of MINT. The adaptive-MINT (A-MINT) approach estimates the inverse filters iteratively by minimizing a cost function that is formulated similar to that of MINT. Although it can achieve a reasonable inverse system [7], A-MINT suffers from slow convergence. To achieve faster convergence, the sparsity of the signal corresponding to the convolution between the inverse filters and the estimated AIRs has been incorporated to A-MINT [8]. Although this approach can increase the convergence rate for adaptive equalization to some extent, it does not improve the equalization performance significantly.

In this work, we propose an adaptive multichannel equalization (MCEQ) algorithm using a time-domain adaptive (TDA) framework. In this proposed TDA-MCEQ algorithm, a set of inverse filters are estimated iteratively for each subspace. The cost function, for each subspace, is formulated to minimize the error between a Kronecker delta function with a delay corresponding to each subspace and the convolution between the inverse filters for each subspace and the AIRs. As will be shown, the desired Kronecker delta functions with delays corresponding to the subspaces form a sparse identity matrix.

II. Multichannel Equalization

Consider a single-input and multiple-output finite impulse response equalization system under the noiseless condition as illustrated in Fig. 1. A clean speech signal $s(n)$ generated by a source within an enclosed space propagates through M acoustic channels denoted by $\mathbf{h}_m = [h_{m,0}, h_{m,1}, \ldots, h_{m,L_h-1}]^T$, where $m = 1, \ldots, M$ is the channel index, L_h is the length of AIRs and $\{\cdot\}^T$ denotes the transpose operation. The reverberant signal $x_m(n)$ received by the mth microphone is the convolution between the original signal $s(n)$ and the mth channel impulse response h_m, i.e.,

$$x_m(n) = h_m * s(n), \qquad m = 1, \ldots, M, \tag{1}$$

This work is supported by the Singapore National Research Foundation Interactive Digital Media R&D Program, under research grant NRF2008IDMIDM004-010.

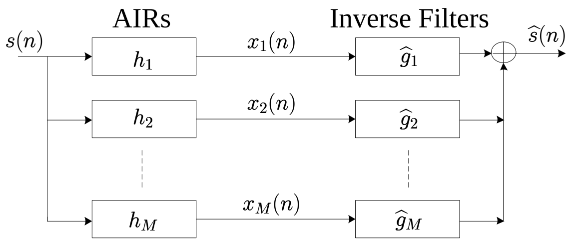


Fig. 1. Multichannel equalization system model.

where $*$ denotes the convolution operation. The relationship in (1) can be represented in vector form, as

$$\mathbf{x}_m(n) = \mathbf{H}_m \mathbf{s}(n), \tag{2}$$

where

$$\mathbf{x}_m(n) = [x_m(n), x_m(n-1), \ldots, x_m(n-L_g+1)]^T, \tag{3}$$
$$\mathbf{s}(n) = [s(n), s(n-1), \ldots, s(n-L_g-L_h+1)]^T \tag{4}$$

are $L_g \times 1$ and $(L_g + L_h - 1) \times 1$ vectors, respectively, and

$$\mathbf{H}_m = \begin{bmatrix} h_{m,0} & \ldots & h_{m,L_h-1} & 0 & 0 & \ldots & 0 \\ 0 & h_{m,0} & \ldots & h_{m,L_h-1} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & h_{m,0} & \ldots & h_{m,L_h-1} \end{bmatrix} = [\mathbf{h}_m^{(1)}, \mathbf{h}_m^{(2)}, \ldots, \mathbf{h}_m^{(L)}] \tag{5}$$

is an $L_g \times (L_g + L_h - 1)$ convolution matrix corresponding to $\mathbf{h}_m$. We have also defined, in (5), $\mathbf{h}_m^{(l)}$ of dimension $L_g \times 1$ as the lth column of $\mathbf{H}_m$.

Given a set of M AIRs, the aim of MCEQ is to estimate M inverse filters $\mathbf{g}_m = [g_{m,0}, g_{m,1}, \ldots, g_{m,L_g-1}]^T$, where L_g is the length of inverse filters such that

$$\sum_{m=1}^{M} \mathbf{H}_m^T \mathbf{g}_m = \mathbf{H}^T \mathbf{g} = \mathbf{d}. \tag{6}$$

Here the $ML_g \times (L_g + L_h - 1)$ matrix $\mathbf{H}$ and the $ML_g \times 1$ vector $\mathbf{g}$ are defined as

$$\mathbf{H} = [\mathbf{H}_1^T, \mathbf{H}_2^T, \ldots, \mathbf{H}_M^T]^T, \tag{7}$$
$$\mathbf{g} = [\mathbf{g}_1^T, \mathbf{g}_2^T, \ldots, \mathbf{g}_M^T]^T \tag{8}$$

while

$$\mathbf{d} = [\underbrace{0, 0, \ldots, 0}_{\tau}, 1, 0, \ldots, 0, 0]^T \tag{9}$$

is a Kronecker delta function of length $L_g + L_h - 1$ with a modeling delay τ.

Assuming that $\mathbf{H}$ is estimated using BSI techniques such as presented in [2], an estimated of $\mathbf{g}$ can be achieved by formulating the equalization problem as

$$\widehat{\mathbf{g}} = \arg\min_{\mathbf{g}} ||\mathbf{H}^T \mathbf{g} - \mathbf{d}||_2^2, \tag{10}$$

where $\|\cdot\|_2$ is the l_2-norm, $\widehat{\mathbf{g}} = [\widehat{\mathbf{g}}_1^T, \widehat{\mathbf{g}}_2^T, \ldots, \widehat{\mathbf{g}}_M^T]^T$ is the estimate of $\mathbf{g}$. The MCEQ filters can be computed according to [9],

$$\widehat{\mathbf{g}} = [\mathbf{H}^T]^+ \mathbf{d}, \tag{11}$$

where $\{\cdot\}^+$ is the matrix pseudo-inverse [10].

If $\mathbf{H}$ is a full-rank square matrix, i.e., $ML_g = L_h + L_g - 1$, the pseudo-inverse in (11) reduces to a square matrix inverse. The solution $\widehat{\mathbf{g}}$ is then unique and equivalent to that of MINT [4]. In general however, $ML_g > L_h + L_g - 1$ giving rise to an underdetermined system. Selection of the inverse filter length L_g and the delay τ in $\mathbf{d}$ plays an important role in equalization performance [11], [12]. For $M \geq 2$, $L_g \geq (L_h - 1)/(M - 1)$ is selected for the minimum norm solution. In addition, computation of $\widehat{\mathbf{g}}$ using (11) requires matrix inversion. It is therefore foreseeable that, in a practical scenario where L_h is of several hundreds, estimation of $\mathbf{g}$ is computationally expensive. We therefore propose an efficient approach for estimating $\mathbf{g}$ adaptively.

III. The proposed TDA-MCEQ algorithm

In this section, we propose an adaptive algorithm (TDA-MCEQ) based on MINT by estimating inverse filters for each subspace. Being an adaptive algorithm, TDA-MCEQ avoids matrix inversions in the process of estimation reducing the computational complexity significantly as compared to MINT.

A transform domain may be characterized by a unitary transform matrix $\mathbf{U} \in \mathbb{R}^{L \times L}$ consisting of a set of basis vectors which defines the subspaces in that transform domain. The basis vectors of a transform matrix are orthogonal to each other and have unit norm. In the time domain, the identity matrix $\mathbf{I} = [\mathbf{i}_1, \mathbf{i}_2, \ldots, \mathbf{i}_L] \in \mathbb{R}^{L \times L}$ defines a possible transform matrix, where each $L \times 1$ column vector

$$\mathbf{i}_l = [\underbrace{0, 0, \ldots, 0}_{l-1}, 1, 0, \ldots, 0, 0]^T \tag{12}$$

is the basis vector defining a subspace. While the unit-norm vectors are orthogonal to each other, it is useful to note that each vector $\mathbf{i}_l$, in fact, corresponds to a delay of $l-1$ samples in the time domain.

The proposed TDA-MCEQ algorithm operates by minimizing a cost function that is constructed for each subspace. In view of this, we first define, for each mth channel, the inverse filter of length L_g, for the lth subspace, given by

$$\mathbf{g}_m^{(l)} = [g_{m,0}^{(l)}, g_{m,1}^{(l)}, \ldots, g_{m,L_g-1}^{(l)}]^T, \tag{13}$$

where the superscript $l = 1, \ldots, L$ is defined as the subspace index while $L = L_g + L_h - 1$ is the number of subspaces. Concatenating these M channels, we obtain, for the lth subspace, the inverse filters of length ML_g given by

$$\mathbf{g}^{(l)} = [(\mathbf{g}_1^{(l)})^T, (\mathbf{g}_2^{(l)})^T, \ldots, (\mathbf{g}_M^{(l)})^T]^T. \tag{14}$$

The aim of the proposed TDA-MCEQ algorithm is therefore to estimate L sets of inverse filters for L subspaces defined by the $ML_g \times L$ matrix

$$\mathbf{G} = [\mathbf{g}^{(1)}, \mathbf{g}^{(2)}, \ldots, \mathbf{g}^{(L)}] \tag{15}$$

such that, for each subspace, the corresponding inverse filters $\mathbf{g}^{(l)}$ should satisfy $\mathbf{H}^T \mathbf{g}^{(l)} = \mathbf{d}_l$, where the $L \times 1$ desired response $\mathbf{d}_l$ is the Kronecker delta function with a

TABLE I

THE PROPOSED TDA-MCEQ ALGORITHM

Initialization:

Initialize $\widehat{\mathbf{G}} = \mathbf{0}_{ML_g \times L}$

$\mathbf{D} = \mathbf{I}_{L\times L}$

$\boldsymbol{P} = [\underbrace{\boldsymbol{\rho}, \boldsymbol{\rho}, \ldots, \boldsymbol{\rho}}_{L}]$

where $\boldsymbol{\rho} = [\rho(0), \rho(1), \ldots, \rho(ML_g - 1)]^T$

and $\rho(j) = \dfrac{1}{2(M\sum_{l=1}^{L}|h^{(l)}(j)|^2 + \delta)}, \quad j = 0, \ldots, ML_g - 1$

Computation:

for $n = 0, 1, \ldots$

$\mathbf{E}(n) = \mathbf{D} - \mathbf{H}^T\widehat{\mathbf{G}}(n)$

where $\mathbf{E}(n) = [\mathbf{e}_1(n), \mathbf{e}_2(n), \ldots, \mathbf{e}_L(n)]$

$\boldsymbol{\nabla} J(n) = -2\mathbf{H}\mathbf{E}$

where $\boldsymbol{\nabla} J(n) = [\boldsymbol{\nabla} J_1(n), \boldsymbol{\nabla} J_2(n), \ldots, \boldsymbol{\nabla} J_L(n)]$

$\widehat{\mathbf{G}}(n+1) = \widehat{\mathbf{G}}(n) - \mu\boldsymbol{P} \odot \boldsymbol{\nabla} J(n)|_{\mathbf{G}=\widehat{\mathbf{G}}(n)}$

where $\widehat{\mathbf{G}}(n) = [\widehat{\mathbf{g}}^{(1)}(n), \widehat{\mathbf{g}}^{(2)}(n), \ldots, \widehat{\mathbf{g}}^{(L)}(n)]$

end for

modeling delay equal to $l-1$ defined in (9). Hence, the desired response for $\mathbf{G}$ corresponds to a sparse $L \times L$ matrix $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_L]$ that is equivalent to an identity matrix $\mathbf{I}$. The TDA-MCEQ algorithm is proposed to estimate $\mathbf{G}$ adaptively with the aim of equalizing $\mathbf{H}$ so that we can recover $\mathbf{s}(n)$ from $\mathbf{x}_m(n)$.

A. Adaptive Estimation of Inverse filters $\mathbf{G}$

In TDA-MCEQ, given $\mathbf{H}$, an optimization problem

$$\min_{\mathbf{G}} ||\mathbf{D} - \mathbf{H}^T\mathbf{G}||_{\mathrm{F}} = \min_{\mathbf{G}} \sum_{l=1}^{L} ||\mathbf{d}_l - \mathbf{H}^T\mathbf{g}^{(l)}||_2^2 \qquad (16)$$

is formulated, where $||\cdot||_{\mathrm{F}}$ denotes the Frobenius norm of a matrix. Since the terms in (16) corresponding to L subspaces are independent of each other, we can define L cost functions, one for each subspace, such that minimizing them gives rise to the inverse filters. Hence, $\mathbf{g}^{(l)}$ for the lth subspace is estimated iteratively in the least-square sense by minimizing

$$J_l(n) = ||\mathbf{d}_l - \mathbf{H}^T\widehat{\mathbf{g}}^{(l)}(n)||_2^2, \qquad (17)$$

where $\widehat{\mathbf{g}}^{(l)}$ is the estimate of $\mathbf{g}^{(l)}$ for $l = 1, \ldots, L$. The gradient of this cost function is given by the $ML_g \times 1$ vector

$$\begin{aligned}\boldsymbol{\nabla} J_l(n) &= \frac{\partial J_l(n)}{\partial \widehat{\mathbf{g}}^{(l)}(n)} \\ &= -2\mathbf{H}\mathbf{e}_l(n), \end{aligned} \qquad (18)$$

where the error for the lth subspace is defined by the $L \times 1$ vector

$$\mathbf{e}_l(n) = \mathbf{d}_l - \mathbf{H}^T\widehat{\mathbf{g}}^{(l)}(n). \qquad (19)$$

Subsequently, the gradient-based update equation of TDA-MCEQ to estimate the equalizers for the lth subspace is given by

$$\widehat{\mathbf{g}}^{(l)}(n+1) = \widehat{\mathbf{g}}^{(l)}(n) - \frac{\mu}{2}\boldsymbol{\nabla} J_l(n)|_{\widehat{\mathbf{g}}^{(l)}=\widehat{\mathbf{g}}^{(l)}(n)}, \qquad (20)$$

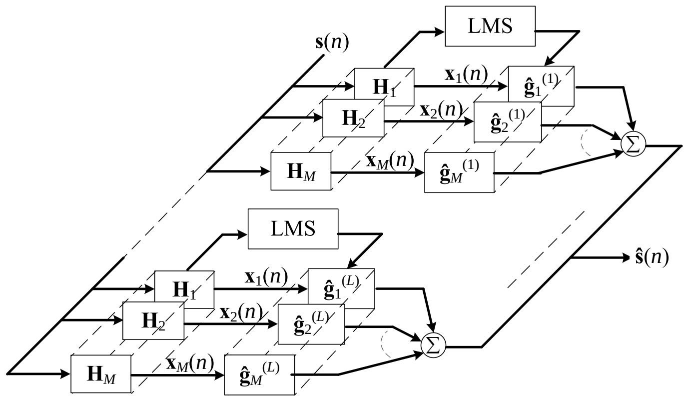


Fig. 2. An illustration of speech dereverberation using TDA-MCEQ framework.

where μ is the step-size.

For a constant μ, the convergence rate of an adaptive algorithm depends on the statistical characteristics of the input sequence [13]. In order to achieve faster convergence, we introduce the squared norm of the $ML_g \times 1$ "input" sequence,

$$\begin{aligned}\mathbf{h}^{(l)} &= [(\mathbf{h}_1^{(l)})^T, (\mathbf{h}_2^{(l)})^T, \ldots, (\mathbf{h}_M^{(l)})^T]^T \\ &= [h^{(l)}(0), h^{(l)}(1), \ldots, h^{(l)}(ML_g - 1)]^T \end{aligned} \qquad (21)$$

of dimension $ML_g \times 1$, into the update equation in (20) giving

$$\widehat{\mathbf{g}}^{(l)}(n+1) = \widehat{\mathbf{g}}^{(l)}(n) - \mu\boldsymbol{\rho} \odot \boldsymbol{\nabla} J_l(n)|_{\widehat{\mathbf{g}}^{(l)}=\widehat{\mathbf{g}}^{(l)}(n)}, \qquad (22)$$

where $\odot$ denotes the Hadamard product and the jth element of the vector $\boldsymbol{\rho} = [\rho(0), \rho(1), \ldots, \rho(j), \ldots, \rho(ML_g - 1)]^T$ is

$$\rho(j) = \frac{1}{2(M\sum_{l=1}^{L}|h^{(l)}(j)|^2 + \delta)}, \qquad (23)$$

while δ is a small value introduced to avoid the numerical errors when $\sum_{l=1}^{L}|h^{(l)}(j)|^2 \approx 0$. The proposed TDA-MCEQ algorithm for estimating $\mathbf{G}$ is summarized in Table I.

B. Speech dereverberation using TDA-MCEQ

We provide an example of equalization using TDA-MCEQ from which we achieve speech dereverberation on a frame-by-frame basis as illustrated in Fig. 2. Assuming that the AIRs have been estimated using, for example, the normalized multichannel fast least-mean-square (NMCFLMS) algorithm [2], the received reverberant speech signals are first partitioned into overlapping frames, each of length equal to L_g, with a frame shift of N samples, such that $1 \leq N \leq L$. For each frame $\mathbf{x}_m(n)$ defined in (3), the corresponding desired frame $\mathbf{s}(n)$ in (4) is recovered using

$$\widehat{\mathbf{s}}(n) = \widehat{\mathbf{G}}^T[\mathbf{x}_1^T(n), \mathbf{x}_2^T(n), \ldots, \mathbf{x}_M^T(n)]^T, \qquad (24)$$

where $\widehat{\mathbf{s}}(n)$ is the estimate of $\mathbf{s}(n)$. It is common to employ widowing and overlapping frames to achieve better reconstruction [14]. If a rectangular window with a frame shift N samples is used, reconstruction of the clean speech may be achieved using only N samples in each frame $\widehat{\mathbf{s}}(n)$ thereby avoiding the need to process the redundant $L - N$ samples which is a direct consequence of frame overlapping. This in turn can be implemented by recovering only N samples per frame using N corresponding sets of inverse filters $\mathbf{g}^{(l)}$. For the special case where $N = 1$, TDA-MCEQ is equivalent to A-MINT with the same modeling delay, since the entire signal is recovered using a single set of M inverse filters.

We further discuss the computational complexity of MINT, A-MINT and TDA-MCEQ in terms of the number of computation in floating point operations (flops). For MINT, (11) requires approximately $(ML_g)^2L + (ML_g)^3/3$ flops [10]. Defining $\mathcal{B}$ as the number of iterations, the number of flops required for A-MINT is approximately equivalent to $\mathcal{B}[2ML_g(2L + 1)]$, while the computational complexity of TDA-MCEQ given in (22) is approximately equivalent to $\mathcal{B}[2ML_g(2L(N + 0.5) + N)]$, where N is the number of subspaces being used. The complexity of TDA-MCEQ increases with the number of subspaces. As can be seen, both A-MINT and TDA-MCEQ are computationally more efficient than MINT. In addition, because of its fast convergence, the proposed TDA-MCEQ is still computationally comparable with A-MINT.

Compared to MINT, the proposed TDA-MCEQ reduces the computational load by avoiding the need to solve the close-form equation directly. Moreover, since it is desirable to provide a modeling delay τ in the desired response in order to compensate for the non causality of inverse filters [11], [12], TDA-MCEQ can yield higher equalization performance than MINT and A-MINT which operate with zero or improper modeling delay. The improved performance of TDA-MCEQ is achieved by utilizing multiple delays in the desired delta functions.

IV. Simulation and Experimental Results

We illustrate the performance of our proposed TDA-MCEQ algorithm by comparing it with MINT [4] and its adaptive version A-MINT [6]. The convergence performance of the algorithms is shown in Section IV-A while robustness to AIR estimation errors and speech dereverberation performance are illustrated in Sections IV-B and IV-C, respectively.

A. Convergence performance

To evaluate the convergence performance, we use the normalized misalignment defined by

$$\eta = \begin{cases} 20\log_{10}\dfrac{||\mathbf{d} - \widehat{\mathbf{d}}(n)||_2}{||\mathbf{d}||_2}\ \mathrm{dB} & \text{for A}-\text{MINT}, \\ \frac{1}{L}\sum\limits_{l=1}^{L} 20\log_{10}\dfrac{||\mathbf{d}_l - \widehat{\mathbf{d}}_l(n)||_2}{||\mathbf{d}_l||_2}\ \mathrm{dB} & \text{for TDA}-\text{MCEQ}, \end{cases}$$

where $\widehat{\mathbf{d}}$ is the equalized response obtained with A-MINT while $\widehat{\mathbf{d}}_l$ is the equalized responses obtained with TDA-MCEQ for the lth subspace and η for TDA-MCEQ is the averaged misalignment across all the L subspaces.

In this simulation, AIRs were generated using the method of images [15]. A virtual room with dimension $6\ \mathrm{m} \times 5\ \mathrm{m} \times 3\ \mathrm{m}$

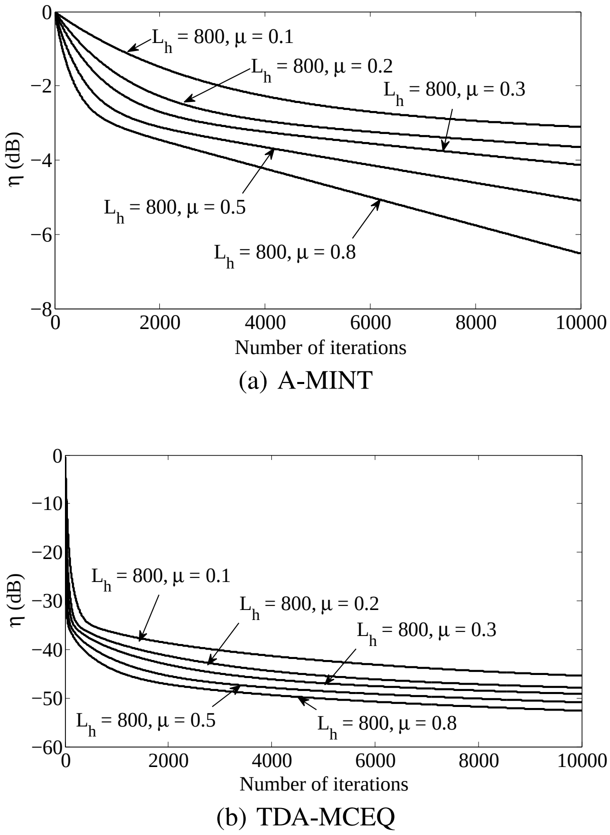


Fig. 3. Convergence performance for (a) A-MINT and (b) TDA-MCEQ with different step size μ.

was simulated with a linear microphone array of $M = 5$ microphones spaced 0.1 m apart at a height 1.6 m. The microphone array centroid was located at $(2, 2, 1.6)$ m. The AIRs were synthetically generated with a sampling frequency $f_s = 16$ kHz while $L_h = T_{60} \times f_s$, where T_{60} is the reverberation time. The performance of MINT and A-MINT may degrade in the presence of the bulk channel delay which is common across all the channels if the modeling delay is not properly selected. For all results presented in Section IV, we have used AIRs without this common bulk delay so as to avoid such performance degradation for MINT and A-MINT which are sensitive to such delays. In this simulation we used $L_g = L_h/2$ and a frame shift of $N = 10$ samples while the frame length is equal to L_g. The synthetic AIRs were equalized using A-MINT and TDA-MCEQ and the convergence performance comparisons in terms of η are illustrated in Fig. 3 and Fig. 4.

Figures 3(a) and 3(b) show the convergence performance of A-MINT and TDA-MCEQ, respectively, for different step sizes μ with $L_h = 800$. As can be seen, the convergence rate increases with μ for both algorithms as expected. We note from Fig. 3(b) that, for TDA-MCEQ, η reduces significantly to less than -20 dB within tens of iterations when $\mu = 0.5$. The normalized misalignment is reduced further to less than

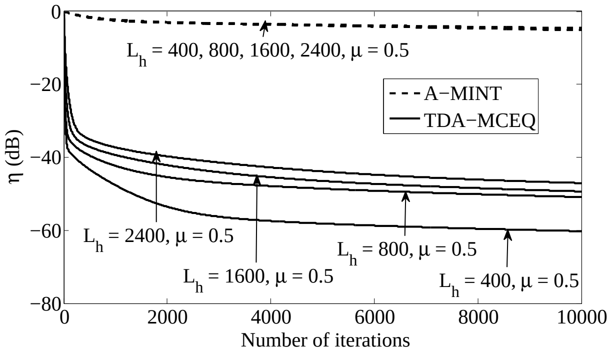


Fig. 4. Convergence performance comparisons for A-MINT and TDA-MCEQ with different AIR length L_h.

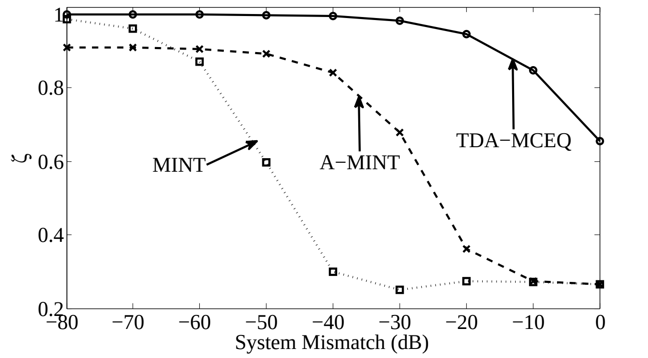


Fig. 5. Robustness comparison against system mismatches for MINT, A-MINT and TDA-MCEQ.

-50 dB within 1×10^4 iterations. On the other hand, the convergence of A-MINT is comparatively slower as can be seen from Fig. 3(a). Even after 1×10^4 iterations, η for A-MINT achieves approximately -4 dB for the same step-size $\mu = 0.5$. Comparing the convergence curves in Figs. 3(a) and 3(b), we note that TDA-MCEQ converges faster than A-MINT.

We next show the convergence performance of the algorithms for various L_h using $\mu = 0.5$. Since L_h increases with reverberation time T_{60}, this simulation is important to evaluate the performance of the algorithms under different reverberation times. Convergence performance for the algorithms with $L_h = 400,\ 800,\ 1600$ and 2400 corresponding to $T_{60} = 0.025$ s, 0.05 s, 0.1 s and 0.15 s, respectively, are shown in Fig. 4. As can be seen from Fig. 4, TDA-MCEQ significantly outperforms A-MINT for all the above cases.

B. Robustness against AIR inaccuracies

In practice, AIR estimation algorithms often provide inexact AIRs $\widehat{\mathbf{h}}_m \neq \mathbf{h}_m$ due to estimation errors. Hence it is important to examine the robustness of equalization algorithms to AIR estimation errors. In this section, we perform a comparison between the robustness of MINT, A-MINT and TDA-MCEQ to inaccurate AIRs estimates assuming knowledge of the order L_h of AIRs.

An inexact system impulse response $\widehat{\mathbf{h}}_m = [\widehat{h}_{m,0}, \widehat{h}_{m,1}, \ldots, \widehat{h}_{m,L_h-1}]^T$ is first generated by introducing zero mean white Gaussian noise to the unknown AIRs $\mathbf{h}_m$. Accordingly, the system mismatch denoted by $\mathcal{M}_m$ is quantified using

$$\mathcal{M}_m = 20 \log 10 \left(\frac{||\mathbf{h}_m - \widehat{\mathbf{h}}_m||_2}{||\mathbf{h}_m||_2} \right) \text{dB}. \tag{25}$$

We next designed an experiment in which the true AIRs $\mathbf{h}_m$ were equalized with the inverse filters estimated using the inaccurate AIRs $\widehat{\mathbf{h}}_m$. To evaluate the robustness of the algorithms to such estimation error, we use the sparseness measure of the equalized response $\widehat{\mathbf{d}}$ defined by [16]

$$\zeta = \frac{L_h}{L_h - \sqrt{L_h}} \left(1 - \frac{||\widehat{\mathbf{d}}||_1}{\sqrt{L_h} ||\widehat{\mathbf{d}}||_2} \right). \tag{26}$$

It is noted that $0 \leq \zeta \leq 1$ and that when $\zeta = 1$, $\widehat{\mathbf{d}}$ is a Kronecker delta function which is perfectly sparse while $\zeta = 0$ if all elements in $\widehat{\mathbf{d}}$ have the same magnitude. Therefore, for perfect equalization, $\widehat{\mathbf{d}}$ should be a Kronecker delta function $\mathbf{d}$ defined in (9) with $\zeta = 1$.

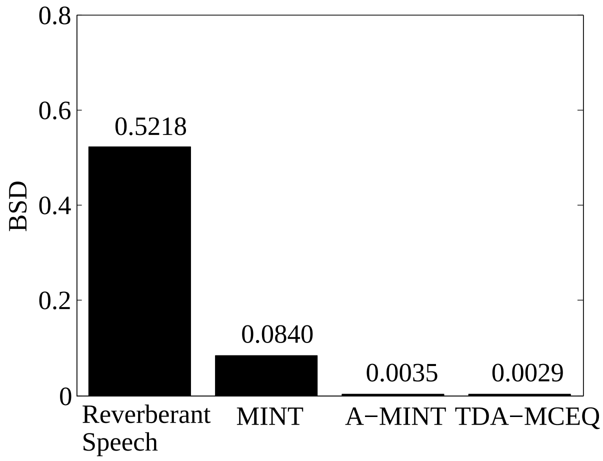
In this experiment, we used a set of five measured AIRs obtained from the MARDY database [17]. These AIRs were re-sampled to 16 kHz and truncated to length $L_h = 500$. We next introduce system mismatches ranging from -80 dB to 0 dB in steps of 10 dB to this set of measured AIRs $\mathbf{h}_m$ so as to generate $\widehat{\mathbf{h}}_m$. We have chosen $L_g = L_h = 500$ for the algorithms while the step size μ for the adaptive algorithms is 0.5. For TDA-MCEQ, framing was done using a frame length equal to L_g and a frame shift $N = 100$ samples. The equalizers were then estimated using these inexact AIRs for each $\mathcal{M}_m$ and equalization was performed using MINT, A-MINT and TDA-MCEQ. The sparseness ζ of $\widehat{\mathbf{d}}$ obtained with each algorithm was computed using (26) to quantify the robustness of each algorithm. It is important to note, in this experiment, that while inverse filters were estimated using erroneous AIRs $\widehat{\mathbf{h}}_m$, the Kronecker delta function $\widehat{\mathbf{d}}$ was generated using the estimated inverse filters and $\mathbf{h}_m$. This then allows one to evaluate the robustness of the algorithms to BSI errors.

Figure 5 shows how ζ varies with $\mathcal{M}_m$ for the three algorithms described above. In this experiment, A-MINT did not achieve its steady-state convergence to give $\zeta = 1$ when $\widehat{\mathbf{h}}_m = \mathbf{h}_m$ whereas MINT and TDA-MCEQ have achieved equalization with $\eta = 1$ under such perfectly matched condition. It can be observed from Fig. 5 that, for MINT, the performance degrades significantly with increasing AIR mismatch. The sparseness measure ζ of MINT reduces from its maximum value of 1 at $\mathcal{M}_m = -80$ dB to 0.3 for $\mathcal{M}_m \geq -40$ dB. The A-MINT algorithm shows almost stable performance up to $\mathcal{M}_m = -50$ dB after which its performance reduces considerably. On the contrary, TDA-MCEQ exhibits higher robustness to AIR mismatch compared to the other two algorithms. It achieves relatively stable performance up to $\mathcal{M}_m = -20$ dB after which its performance reduces modestly.

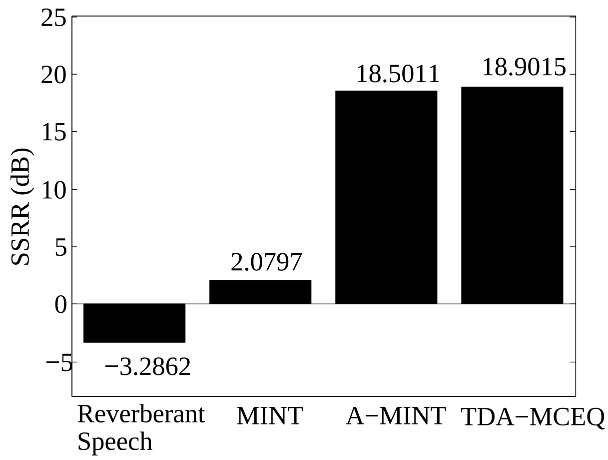
C. Speech dereverberation accuracy

We next compare the dereverberation performance of MINT, A-MINT and TDA-MCEQ. For this experiment, we adopted the method of speech dereverberation via BSI and inverse filtering. The segmental signal-to-reverberation ratio (SSRR) [1] and the Bark spectral distortion (BSD) [18] are used as the evaluation metrics for the dereverberation quality. Therefore a high value of SSRR and a low value of BSD translate to a good performance of the algorithm under evaluation.

In this experiment, we used the AIRs recorded inside a $6.5\ \mathrm{m} \times 8.75\ \mathrm{m}$ rectangular room using an array of five equally spaced microphones with a spacing of 0.14 m. The centroid of the array was placed at a height of 1.2 m. We convolved a male speech signal with these recorded AIRs to generate reverberant signals $x_m(n)$ from which the AIRs were estimated using the well-known NMCFLMS algorithm [2]. The average mismatch measured across all the channels was -9.64 dB. Equalization was then performed on the reverberant signals using the estimated AIRs applying MINT, A-MINT and TDA-MCEQ. The AIRs were truncated to $L_h = 800$ and the inverse filter length was chosen as $L_g = \lceil \frac{L_h - 1}{M-1} \rceil$ for all the algorithms, where $\lceil \cdot \rceil$ denotes the ceiling operator. Both the shift factor N and the frame length were set to L_g for TDA-MCEQ. A step size $\mu = 0.5$ was chosen for the adaptive algorithms and the SSRR and BSD were computed at the $10,000$th and $1,000$th iteration for A-MINT and TDA-MCEQ, respectively. The SSRRs and BSDs obtained are shown in Fig. 6. As can be seen, in the presence of system estimation mismatch, the improvement of BSD and SSRR of the recovered speech with respect to the reverberant speech is not significant for MINT compared to the other two algorithms. On the other hand TDA-MCEQ and A-MINT could improve the SSRR by approximately 21 dB. Although A-MINT achieves almost similar result as TDA-MCEQ, it required 1×10^4 iterations while TDA-MCEQ requires only 1000 iterations. This shows that TDA-MCEQ can converge faster than A-MINT while achieving better equalization performance than MINT.



(a) BSD



(b) SSRR

Fig. 6. Evaluation of dereverberation performance using (a) BSD and (b) SSRR.

V. Conclusions

Adaptive multichannel equalization of room acoustics in the time domain is proposed. The proposed TDA-MCEQ algorithm can equalize AIRs with better quality by designing dedicated inverse filters for each subspace. In addition, TDA-MCEQ exhibits higher robustness to AIR inaccuracies. The proposed algorithm can also achieve higher convergence rate when compared with A-MINT.

References

[1] P. A. Naylor, and N. D. Gaubitch, "Speech dereverberation," *Int. Workshop Acoust. Echo Noise Control,* September 2005.

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

[3] J. Mourjopoulos, P. Clarkson, and J. Hammond, "A comparative study of least-squares and homomorphic techniques for inversion of mixed phase signals," *in Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.,* vol. 7, pp. 1958–1961, May 1982.

[4] M. Miyoshi and K. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, and Signal Process.,* vol. 36, pp. 145–152, 1988.

[5] S. J. Elliott and P. A. Nelson, "Algorithm for multichannel LMS adaptive filtering," *Electronics Letters,* vol. 21, no. 21, pp. 979–981, October 1985.

[6] W. Zhang, A. W. H. Khong, and P. A. Naylor, "Adaptive inverse filtering of room acoustics," *in Proc. Asilomar Conf. on Signals, Systems and Computers,* pp. 788–792, 2008

[7] P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Inverse filter design and equalization zones in multichannel sound reproduction," *IEEE Trans. Speech Audio Process.,* vol. 3, no. 3, pp. 185–192, May 1995.

[8] L. Liao and A. W. H. Khong, "Adaptive channel equalization of room acoustics exploiting sparseness constraint," *IEEE Signal Process. Letters,* vol. 18, no. 4, pp. 275–278, April 2011.

[9] 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, September 2005.

[10] G. H. Golub and C. F. van Loan, *Matrix Computations,* ser. John Hopkins series in the mathematical sciences, 3rd ed. London, U.K.: John Hopkins Univ. Press, 1996.

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

[12] J. C. Claerbout and E. A. Robinson, "The error in least squares inverse filtering," *Geophysics,* vol. 29, pp. 118–120, February 1964.

[13] S. Haykin, *Adaptive Filter Theory,* 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

[14] L. Rabiner and R. Schafer, *Digital Processing of Speech Signals,* Englewood Cliffs: Prentice Hall, 1978.

[15] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," *J. Acoust. Soc. Am.,* vol. 65, no. 4, pp. 943–905, April 1979.

[16] A. W. H. Khong and P. A. Naylor, "Efficient use of sparse adaptive filters," *in Proc. Int. Conf. Signals, Systems and Computers*, pp. 1375–1379, October 2006.

[17] 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*, September 2006.

[18] S. Wang, A. Skey, and A. Gersho, "An objective measure for predicting subjective quality of speech coders," *IEEE J. Select. Areas Commun.*, vol. 10, no. 5, pp. 819–829, June 1992.