ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/224218410

A Clipping-Based Selective-Tap Adaptive Filtering Approach to Stereophonic Acoustic Echo Cancellation

Article *in* IEEE Transactions on Audio Speech and Language Processing · September 2011

DOI: 10.1109/TASL.2010.2102752 · Source: IEEE Xplore

CITATIONS
8

READS
182

3 authors:

Mehdi Bekrani
Qom University Of Technology
18 PUBLICATIONS 55 CITATIONS

SEE PROFILE

Andy Khong
Nanyang Technological University
155 PUBLICATIONS 1,247 CITATIONS

SEE PROFILE

Mojtaba Lotfizad
Tarbiat Modares University
49 PUBLICATIONS 323 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Adaptive Blind Multichannel System Identification View project

All content following this page was uploaded by Mojtaba Lotfizad on 29 May 2014.

The user has requested enhancement of the downloaded file.

A Clipping-Based Selective-Tap Adaptive Filtering Approach to Stereophonic Acoustic Echo Cancellation

Mehdi Bekrani, Andy W. H. Khong, *Member, IEEE*, and Mojtaba Lotfizad

***Abstract*—Stereophonic acoustic echo cancellation remains one of the challenging areas for tele/video-conferencing applications. However, the existence of high interchannel coherence between the two input signals for such systems leads to considerable degradation in misalignment convergence of the adaptive filters. We propose a new algorithm for improving the convergence performance and steady-state misalignment by considering robustness to the source position in the transmission room. We achieve this by exploiting the inherent decorrelating properties of selective-tap adaptive filtering as well as employing a variable clipping threshold for the unselected taps. Simulation results using colored noise and speech signals show an improvement over existing algorithms both in terms of convergence rate as well as steady-state normalized misalignment.**

***Index Terms*—Center clipping, convergence rate, interchannel coherence, partial updating, selective-tap, stereophonic acoustic echo cancellation (SAEC).**

I. Introduction

THERE has been increasing interest in employing stereophonic audio communication systems for video/tele-conferencing, home entertainment, and E-learning applications in order to achieve better perception of sound [1], [2]. Such systems have become increasingly popular since a stereophonic audio system provides spatial information, leading to better perception of the transmitted speech as well as improving the ambience of the transmission room. These systems mitigate, to a certain extent, the cocktail party problem that exists in a multiparty conferencing scenario [2]. One of the problems that should be addressed in such systems is the cancellation of stereophonic acoustic echo using a pair of adaptive filters. Stereophonic acoustic echo cancellation (SAEC) has issues that are considerably more challenging to overcome than the monophonic case [3]. The fundamental problem is the poor mismatch between adaptive filter coefficients and the receiving room acoustic impulse responses. It has been shown [4] that, in a practical scenario, the adaptive filter misalignment converges poorly, leading to a performance degradation. The misalignment problem is caused by the high interchannel coherence that exists between the two transmitted stereo signals [4].

A variety of methods have been proposed to address the misalignment problem. These methods revolve around reducing the interchannel coherence between the two transmitted signals. One of the first methods involved adding controlled quantities of independent noise to each input channel [5], or modulating them [6]. These two approaches are however not feasible since distortion can be heard even when the added noise level is very low [7]. Another approach involves the use of comb filtering [8], where frequency components of the left and right channels are separated in order to reduce the interchannel coherence. Although this method improves the performance of the adaptive filters, it degrades the quality of the stereophonic sound, especially at lower frequencies.

To address the disadvantages of the methods described above, a preprocessor has been proposed to add a nonlinear (NL) function of the transmitted signal in each channel to the signal itself [4], [9]. This method employs a half-wave rectifier and is attractive in terms of improving the misalignment behavior of the adaptive filters as well as its simplicity in implementation. Although less distortion was introduced compared to the other techniques discussed above, this distortion is still found to be objectionable in some cases for music applications [4], [10]. An adaptive nonlinearity control was subsequently proposed to maintain the desired level of misalignment and to minimize the audio distortion [11].

It is apparent by now that algorithms proposed for SAEC need to decorrelate the transmitted signals without degrading the quality of the speech signals or destroying the stereophonic image of the transmission room. In view of this, the use of time-varying all-pass filtering of the stereophonic signals has been proposed [12] with the aim of signal decorrelation while maintaining the stereophonic perception. The use of psychoacoustic properties to reduce perceived distortion while achieving signal decorrelation has also been proposed, including the use of spectrally shaped random noise [7], [13], gain controlled phase distortion [14] and the combination of comb filtering and all-pass filtering with respect to the masking effect [15]. These methods exploit a perceptual property of the human auditory system, called "noise masking."

Manuscript received October 09, 2009; revised June 18, 2010; accepted November 29, 2010. Date of publication February 10, 2011; date of current version June 03, 2011. This work was supported in part by the Singapore National Research Foundation Interactive Digital Media R&D Program under research grant NRF2008IDM-IDM004-010 and in part by the Research Institute for ICT. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sharon Gannot.

M. Bekrani was with Tarbiat Modares University, Tehran, Iran. He is now with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: mbekrani@ntu.edu.sg).

A. W. H. Khong is with Nanyang Technological University, Nanyang Technological University, Singapore 639798 (e-mail: andykhong@ntu.edu.sg).

M. Lotfizad is with the Department of Electrical and Computer Engineering, Tarbiat Modares University, P.O. Box 14115-143, Tehran, Iran (e-mail: lotfizad@modares.ac.ir).

Digital Object Identifier 10.1109/TASL.2010.2102752

1558-7916/$26.00 © 2010 IEEE

More recent advances in SAEC research involve a decorrelation procedure for the adaptive weight update [10], [16]–[19]. These approaches decorrelates the tap-input vectors of the adaptive filters as opposed to decorrelating the transmitted signals. Among these methods, the exclusive-maximum (XM) tap-selection algorithm proposed in [17] appeared to be an attractive solution. This algorithm achieves update signal decorrelation by ensuring that only an exclusive set of filter coefficients from each channel is selected for adaptation. In order to reduce any degradation in convergence performance due to this subselection procedure, the XM tap-selection strategy further ensures that the energies of these exclusive tap inputs are maximized. The XM tap selection has been incorporated with the NL half-wave rectifier and the resulting XMNL normalized least-mean-square (XMNL-NLMS) algorithm has been shown to achieve a higher rate of misalignment convergence compared to that of nonlinear NLMS (NL-NLMS) [17].

We propose to further improve the convergence performance of the XM tap-selection algorithm. This motivation is derived from the degradation in convergence performance of the XM tap-selection algorithm when the interchannel coherence between the two tap-input vectors is relatively low. This can occur, for example, when the source in the transmission room is located away from the centroid of the stereophonic microphone pair. We note that the robustness issue of XMNL-NLMS to the source position has not been investigated and that, in this work, we present insight into this problem. Utilizing this new knowledge, we propose to improve the misalignment convergence of XMNL-NLMS by employing a center-clipping algorithm so that the low interchannel coherence and maximization of tap-input energy criteria can be jointly optimized. The proposed algorithm ensures that the misalignment convergence and steady-state misalignment of the adaptive filters will be robust to the source position in the transmission room.

In [7] and references therein, the authors evaluated the adaptive signal decorrelation filter as a preprocessor and reported that *complete* decorrelation in the frequency domain cannot be achieved unless one or both of the stereophonic signals are zero at every frequency. This process is undesirable since it destroys the stereophonic image of the transmitted signals, which is important to the listeners in the receiving room. Therefore, they concluded that complete decorrelation is not applicable in practice. Our proposed method, as opposed to the technique discussed in [7], does not apply decorrelation filtering to the transmitted signals. We instead operate on the tap-input vector of the adaptive filters. Therefore, similar to XMNL-NLMS, the stereophonic image is preserved. We also note that the decorrelated tap-input vectors of both algorithms may prevent some of the adaptive filter coefficients from adaptation in some iterations. However, this effect will not significantly degrade the convergence of weights since any significant reduction in the interchannel coherence due to our center-clipping approach will bring about an improvement in convergence rate. It is also important to note that, similar to the approach in [17], the use of the NL preprocessor is required to provide a solution to the ill-posed SAEC problem, while our proposed center-clipping approach improves the convergence rate and robustness of XMNL-NLMS to the source position.

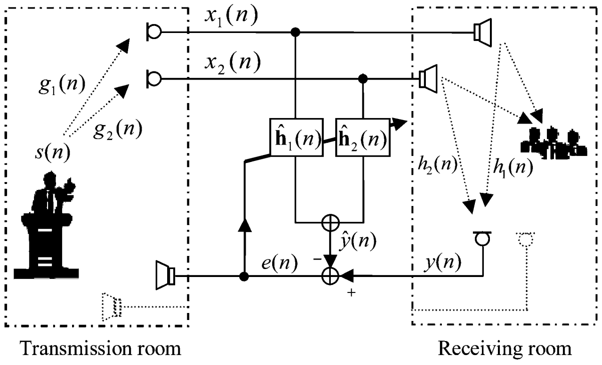

Fig. 1. Stereophonic acoustic echo cancellation for teleconferencing application.

II. Review of Stereophonic Acoustic Echo Cancellation

Fig. 1 shows the stereophonic acoustic echo canceller. For simplicity, we consider only one microphone in the receiving room, since similar analysis can be applied to the other channel [4]. Microphones in the transmission room receive signals produced by the sound source $s(n)$ via acoustic impulse responses $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$ giving transmitted signals $x_1(n)$ and $x_2(n)$, respectively, where $\mathbf{g}_i(n) = [g_{i,0}(n), g_{i,1}(n), \ldots, g_{i,L_g-1}(n)]^T$ for Channel i and L_g is the length of the transmission room impulse responses, while $[\cdot]^T$ is defined as the transpose operator. The transmitted signal to the receiving room for Channel i can then be expressed as

$$x_i(n) = \mathbf{g}_i^T(n)\mathbf{s}(n) \tag{1}$$

where $\mathbf{s}(n) = [s(n), s(n-1), \ldots, s(n-L_g+1)]^T$. These signals produce an echo $y(n)$ in the receiving room given by

$$y(n) = \mathbf{h}_1^T(n)\mathbf{x}_1(n) + \mathbf{h}_2^T(n)\mathbf{x}_2(n) \tag{2}$$

where $\mathbf{h}_i(n) = [h_{i,0}(n), h_{i,1}(n), \ldots, h_{i,L-1}(n)]^T$ is the ith channel receiving room impulse response and $\mathbf{x}_i(n) = [x_i(n), x_i(n-1), \ldots, x_i(n-L+1)]^T$ is the ith channel tap-input vector while L is the length of $\mathbf{h}_i(n)$.

Similar to single-channel AEC, adaptive filters are employed to estimate $\mathbf{h}_1(n)$ and $\mathbf{h}_2(n)$. In this paper, we assume, similar to that of [4], [17], that the adaptive filters are each of length L, which is of the same length as that of $\mathbf{h}_1(n)$ and $\mathbf{h}_2(n)$. For realistic applications where the adaptive filters are shorter than that of $\mathbf{h}_i(n)$, residual echo will be transmitted back to the transmission room due to the unmodeled "tails" of $\mathbf{h}_i(n)$.

The error between the echo and its estimate can then be expressed as

$$e(n) = y(n) - \left[\widehat{\mathbf{h}}_1^T(n)\mathbf{x}_1(n) + \widehat{\mathbf{h}}_2^T(n)\mathbf{x}_2(n)\right] \tag{3}$$

where $\widehat{\mathbf{h}}_i(n) = [\widehat{h}_{i,0}(n), \widehat{h}_{i,1}(n), \ldots, \widehat{h}_{i,L-1}(n)]^T$, i =1, 2, is the vector of adaptive filter coefficients for the ith channel.

A. Nonlinear Normalized Least-Mean-Square Algorithm

In order to efficiently reduce $e(n)$, adaptive algorithms are employed for SAEC. Defining $E\{\cdot\}$ as the expectation operator, the NLMS algorithm [20] is the result of minimizing $E\{e^2(n)\}$, and is popular because of its simplicity in computational efficiency and ease of implementation. The two-channel NLMS algorithm is expressed as

$$\widehat{\mathbf{h}}(n+1) = \widehat{\mathbf{h}}(n) + \frac{\mu}{\|\mathbf{x}(n)\|_2^2 + \epsilon} e(n)\mathbf{x}(n) \tag{4}$$

where $\mathbf{x}(n) = [\mathbf{x}_1^T(n) \quad \mathbf{x}_2^T(n)]^T$ and $\widehat{\mathbf{h}}(n) = [\widehat{\mathbf{h}}_1^T(n) \quad \widehat{\mathbf{h}}_2^T(n)]^T$ are the concatenated two-channel tap-input vector and filter coefficient vector, respectively, while μ is the step-size, which controls the rate of convergence, and ϵ is a regularization parameter to prevent division by zero.

Unlike single-channel AEC, estimation of $\mathbf{h}_1(n)$ and $\mathbf{h}_2(n)$ for the stereophonic case is challenging. As shown in [4], when $L_g \leq L$, the adaptive filters coefficients is of the form

$$\begin{bmatrix} \widehat{\mathbf{h}}_1(n) \\ \widehat{\mathbf{h}}_2(n) \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1(n) \\ \mathbf{h}_2(n) \end{bmatrix} + \kappa(n) \begin{bmatrix} \mathbf{g}_2'(n) \\ -\mathbf{g}_1'(n) \end{bmatrix} \tag{5}$$

where $\mathbf{g}_i'(n) = [\mathbf{g}_i^T(n), 0, 0, \ldots, 0]^T$, i =1, 2 are vectors given that $\mathbf{g}_i^T(n)$ is appended with $L - L_g$ zeros and $\kappa(n)$ is a scalar quantity. Equation (5) indicates that multiple solutions exist and that $\widehat{\mathbf{h}}_1(n)$ and $\widehat{\mathbf{h}}_2(n)$ are linearly related to $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$. The dependency of $\widehat{\mathbf{h}}_1(n)$ and $\widehat{\mathbf{h}}_2(n)$ on these multiple solutions due to $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$ is known as the non-uniqueness problem. In practical cases where $L < L_g$, the non-uniqueness problem is mitigated [4]. However, even in such cases, direct application of the standard adaptive filtering is normally not successful because $\mathbf{x}_1(n)$ and $\mathbf{x}_2(n)$ are highly correlated, giving rise to an ill-conditioned system identification problem. As a result, the misalignment convergence of the adaptive filters is impaired significantly. This degradation is known as the misalignment problem [4].

In order to address the misalignment problem, a nonlinear (NL) preprocessor is proposed [4], [21]. This preprocessor operates on $\mathbf{x}_1(n)$ and $\mathbf{x}_2(n)$ such that the modified transmitted signals $\mathbf{x}_1'(n)$ and $\mathbf{x}_2'(n)$ are given by

$$\mathbf{x}_1'(n) = \mathbf{x}_1(n) + 0.5\alpha[\mathbf{x}_1(n) + |\mathbf{x}_1(n)|] \tag{6}$$

$$\mathbf{x}_2'(n) = \mathbf{x}_2(n) + 0.5\alpha[\mathbf{x}_2(n) - |\mathbf{x}_2(n)|] \tag{7}$$

where α controls the amount of nonlinearity to be added. It has been shown in [4] that a value of $\alpha = 0.5$ is a good compromise between speech quality and misalignment convergence of the NLMS algorithm.

Due to its simplicity in implementation and the low distortion introduced, the NL preprocessor has become an intrinsic part of SAEC and has been incorporated into several recently proposed algorithms for SAEC [10], [17]. For the remainder of this paper, the NLMS algorithm employing this NL preprocessing will be referred to as NL-NLMS.

B. Exclusive-Maximum (XM) Tap-Selection Algorithm

The XMNL-NLMS update [17] can be expressed as

$$\widehat{\mathbf{h}}_i(n+1) = \widehat{\mathbf{h}}_i(n) + \frac{\mu}{\|\mathbf{x}'(n)\|_2^2 + \epsilon} e(n)\mathbf{Q}_i(n)\mathbf{x}_i'(n) \tag{8}$$

where i =1, 2 is the channel index, $\mathbf{x}_i'(n) = [x_i'(n), x_i'(n-1), \ldots, x_i'(n-L+1)]^T$ is the NL-preprocessed tap-input vector of the ith channel defined by (6) and (7), $\mathbf{x}'(n) = [\mathbf{x}_1'^T(n) \quad \mathbf{x}_2'^T(n)]^T$, while $\mathbf{Q}_i(n) = \mathrm{diag}\{\mathbf{q}_i(n)\}$ is a $L \times L$ tap-selection matrix and $\mathbf{q}_i(n) = [q_{i,0}(n), q_{i,1}(n), \ldots, q_{i,L-1}(n)]^T$ with elements given by

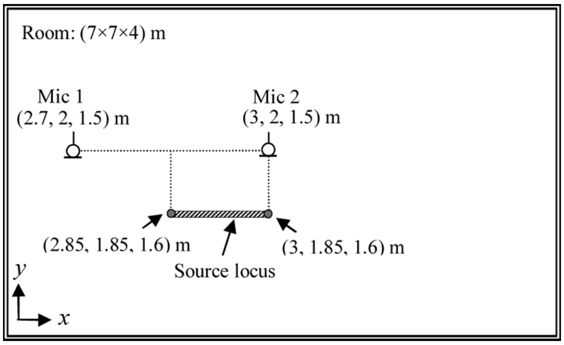

Fig. 2. Locations of the source and the microphone pair in the transmission room.

$$q_{1,u}(n) = \begin{cases} 1, & p_u \in \{0.5L \text{ largest components of } \mathbf{p}(n)\} \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

$$q_{2,v}(n) = \begin{cases} 1, & p_v \in \{0.5L \text{ smallest components of } \mathbf{p}(n)\} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where $u, v = 0, 1, \ldots, L-1$ denote the elemental indices of $\mathbf{q}_1(n)$ and $\mathbf{q}_2(n)$, respectively, and $\mathbf{p}(n) = |\mathbf{x}_1'(n)| - |\mathbf{x}_2'(n)|$, $|\mathbf{x}_i'(n)| = [|x_i'(n)|, |x_i'(n-1)|, \ldots, |x_i'(n-L+1)|]^T$.

As can be seen, the XM algorithm [17] incorporates a tap-selection scheme that reduces the interchannel coherence by selecting exclusive filter coefficients for updating in each channel. It is important to note that the selected tap inputs are only used for updating the coefficients and hence no distortion is introduced. However, with any tap-selection updating strategy, convergence performance of the adaptive filters will be reduced. This degradation is then minimized by jointly maximizing the L_2 norm of the selected tap inputs across both channels. The use of the NL preprocessor is required to provide a solution to the ill-posed SAEC problem, while the XM approach improves the convergence rate of NL-NLMS. As a result of this combination, which we refer to as XMNL, better misalignment convergence of the adaptive filter can be achieved. Alternatively, the XM tap selection can be seen as an effective approach to achieve good misalignment convergence with lower distortion brought about by a smaller nonlinearity factor α.

III. Effect of Source Position on Misalignment Convergence

One of the problems that has yet been considered for XMNL-NLMS is its robustness to the position of the source in the transmission room. We now illustrate the misalignment convergence of XMNL-NLMS by considering different source positions. Fig. 2 shows an experimental setup where two microphones are placed at (x, y, z) positions (2.7,2,1.5) m and (3,2,1.5) m in a room with dimensions 7 m × 7 m × 4 m. We vary the x position of the source starting from the front of the

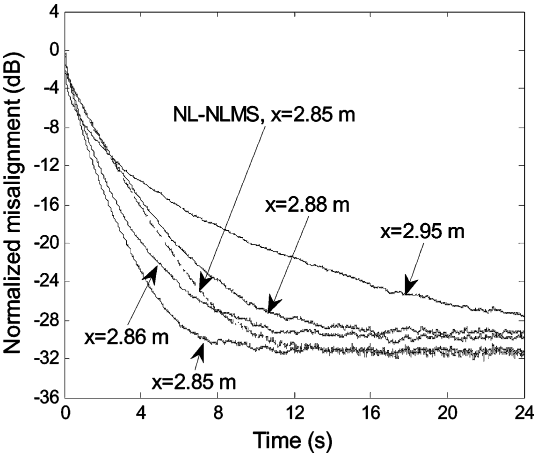

Fig. 3. Misalignment convergence of the XMNL-NLMS algorithm (solid) for different positions of the source, as compared to NL-NLMS (dashed).

array centroid at $x = 2.85$ m to the front of the right microphone at $x = 3$ m. The positions in the receiving room are (3, 2, 1.5) m for the microphone and (2.85, 1.8, 1.6) m and (2.4, 1.1, 1.7) m for the two loudspeakers. For the purpose of illustration, all room impulse responses are generated synthetically using the method of images [22] such that $L = 256$ and $L_g = 400$ samples and these synthetic impulse responses have lengths that correspond to their reverberation times. At a sampling rate of $f_s = 11\,025$ Hz, this corresponds to 23 ms and 36 ms, respectively. A stationary colored noise source signal $s(n)$ is obtained by filtering white Gaussian noise through a low-pass finite impulse response (FIR) filter with coefficients given by $[0.3574, 0.9, 0.3574]$ [23] which was chosen to generate a speech-like spectrum. The convergence of the algorithms is quantified by the normalized misalignment

$$\eta(n) = \frac{||\widehat{\mathbf{h}}(n) - \mathbf{h}(n)||_2^2}{||\mathbf{h}(n)||_2^2} \tag{11}$$

where $\mathbf{h}(n) = \left[\mathbf{h}_1^T(n) \quad \mathbf{h}_2^T(n)\right]^T$. Fig. 3 shows the misalignment convergence of XMNL-NLMS averaged over ten independent trials, for the different source positions. The convergence performance of NL-NLMS for $x = 2.85$ m, when it is directly in front of the microphone pair centroid, has been included for comparison. Additional tests conducted have shown that the misalignment convergence of NL-NLMS for various source positions is comparable to that shown in Fig. 3. A step-size of $\mu = 0.6$ was used for XMNL-NLMS, while the step-size of NL-NLMS was adjusted to $\mu = 0.8$ so that its steady-state misalignment reaches that of XMNL-NLMS. As shown in Fig. 3, XMNL-NLMS outperforms the full-update NL-NLMS when the source is directly in front of the microphone pair centroid at $x = 2.85$ m. On the contrary, the convergence rate of XMNL-NLMS is reduced significantly when the source is located away from the microphone pair centroid such as when $x = 2.95$ m.

To gain further insight into the degradation in convergence rate of XMNL-NLMS with respect to the source position, we consider both the interchannel coherence as well as the ratio of selected tap-input energy to the total tap-input energy.

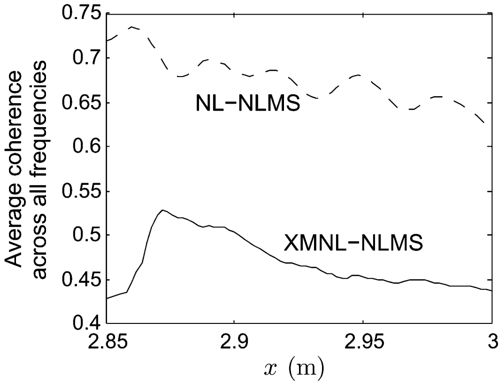

Fig. 4. Average interchannel coherence of the NL-NLMS (dashed) and XMNL-NLMS (solid) algorithms for various positions of the source.

A. Effect of XM Tap Selection on Interchannel Coherence

We first investigate the effect of XM tap-selection on the interchannel coherence for various source positions. We denote the XM subselected tap-input vector in (8) as

$$\mathbf{z}_i(n) = \mathbf{Q}_i(n)\mathbf{x}'_i(n). \tag{12}$$

The interchannel coherence between $\mathbf{z}_1(n)$ and $\mathbf{z}_2(n)$ is then defined by

$$C_{\mathbf{z}_1\mathbf{z}_2}(f) = \frac{|P_{\mathbf{z}_1\mathbf{z}_2}(f)|^2}{P_{\mathbf{z}_1\mathbf{z}_1}(f)P_{\mathbf{z}_2\mathbf{z}_2}(f)} \tag{13}$$

where $P_{\mathbf{z}_1\mathbf{z}_2}(f)$ is the cross power spectrum between $\mathbf{z}_1(n)$ and $\mathbf{z}_2(n)$ while f is the normalized frequency. For the same condition as in Figs. 3 and 4 shows the mean interchannel coherence between $\mathbf{z}_1(n)$ and $\mathbf{z}_2(n)$, across different frequencies for various source positions, obtained by averaging over ten independent trials.

As can be seen, when the source is in front of the microphone pair centroid ($x = 2.85$ m), the XM tap-selection criterion is utilized efficiently to decorrelate input vectors $\mathbf{x}'_1(n)$ and $\mathbf{x}'_2(n)$, giving a low interchannel coherence of 0.43. Due to this efficient decorrelation, a good misalignment convergence shown in Fig. 3 is achieved. For this source location, the modest amount of degradation due to tap selection does not significantly outweigh the benefits brought about by the reduction in interchannel coherence due to the exclusivity criterion. Fig. 5(a) shows an example of the XM selected taps $\mathbf{z}_1(n)$ and $\mathbf{z}_2(n)$ in this case. For clarity, we show only the first 80 samples of $\mathbf{z}_1(n)$ and $\mathbf{z}_2(n)$, each of length $L = 256$ samples. As can be seen, most of the selected taps in the first channel correspond to elements in $\mathbf{x}'_1(n)$ being greater than zero, whereas for the second channel, most of the active taps correspond to elements in $\mathbf{x}'_2(n)$ being less than zero. This is due to the intrinsic effect of NL preprocessing, which increases the magnitude of the positive elements in the first channel and the negative elements in the second channel. As a result of XM tap selection on $\mathbf{x}'_1(n)$ and $\mathbf{x}'_2(n)$, low interchannel coherence of 0.43 is achieved as shown in Fig. 4.

As shown in Fig. 4, the interchannel coherence between $\mathbf{z}_1(n)$ and $\mathbf{z}_2(n)$ employing XM tap selection increases from $x = 2.85$ m to approximately $x = 2.9$ m. To understand this increase in interchannel coherence even after XM tap selection is applied, Fig. 5(b) shows a plot of $\mathbf{z}_1(n)$ and $\mathbf{z}_2(n)$ for $x = 2.88$ m.

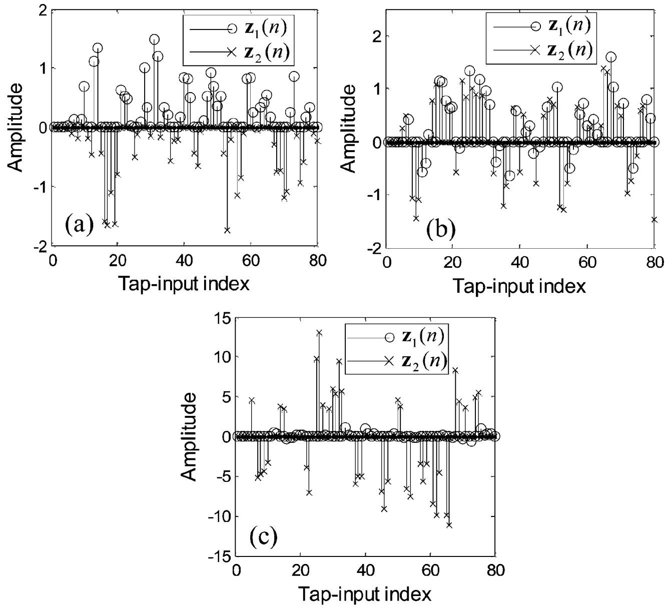

Fig. 5. Selected tap inputs $\mathbf{z}_1(n)$ and $\mathbf{z}_2(n)$ for the XMNL-NLMS algorithm for various source position x. (a) $x = 2.85$ m. (b) $x = 2.88$ m. (c) $x = 2.95$ m.

As can be seen, the effect of NL preprocessing on XM tap selection reduces and the similarity between $\mathbf{z}_1(n)$ and $\mathbf{z}_2(n)$ increases. The increase in interchannel coherence between $\mathbf{z}_1(n)$ and $\mathbf{z}_2(n)$ in turn reduces the convergence rate for XMNL-NLMS, as can be seen from Fig. 3.

Fig. 5(c) illustrates $\mathbf{z}_1(n)$ and $\mathbf{z}_2(n)$ when the source is in front of the right microphone at $x = 2.95$ m. Due to the source being further away from the left microphone, elements in $\mathbf{z}_1(n)$ have magnitudes much lower than those of $\mathbf{z}_2(n)$. In addition, it is expected that $\mathbf{g}_1(n)$ differs from $\mathbf{g}_2(n)$. As a result of this difference, the interchannel coherence is lower for $x = 2.88$ m compared to $x = 3$ m, as shown in Fig. 4. It is therefore expected that the misalignment convergence of XMNL-NLMS should increase when the source moves from $x = 2.88$ m to $x = 3$ m. On the contrary, however, the XMNL-NLMS convergence rate continues to reduce with increasing x position, as can be seen from Fig. 3. In Section IV, we gain better insight into this contradictory behavior by studying the effect of XM tap selection on the energies of the active tap inputs for different source positions.

B. Effect of XM Tap Selection on Tap-Input Energies

In order to further illustrate how XM tap selection affects the energies of the tap inputs, we employ the $\mathcal{M}$-ratio criterion [17] defined as

$$\mathcal{M} = \frac{\|\mathbf{Q}(n)\mathbf{x}'(n)\|_2^2}{\|\mathbf{x}'(n)\|_2^2} \tag{14}$$

where $\mathbf{Q}(n) = \operatorname{diag}\{[\mathbf{q}_1^T(n) \quad \mathbf{q}_2^T(n)]\}$ for the XM tap selection, with elements defined by (9) and (10). Our intention is not to show the exact relationship between $\mathcal{M}$ and the misalignment convergence rate, but to illustrate that the loss of tap-input energy has an undesirable effect on the convergence rate of the adaptive filters. Fig. 6 illustrates how $\mathcal{M}$ varies with source position for XMNL-NLMS and NL-NLMS, obtained by averaging over all frames in the signal where each frame is calculated using (14). As can be seen, NL-NLMS has $\mathcal{M} = 1$ for all source positions since all tap-inputs are used for weight update. On the other hand, for XMNL-NLMS, $\mathcal{M} < 1$ and increases with x position of the source.

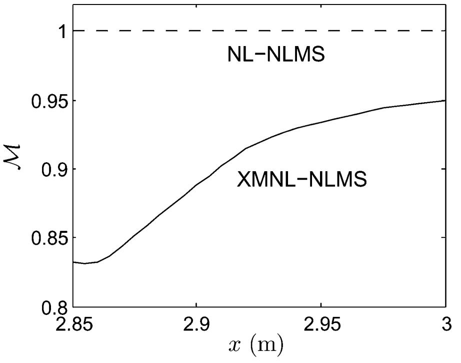

Fig. 6. Variation of $\mathcal{M}$ against source position for NL-NLMS (dashed) and XMNL-NLMS (solid).

We can now see from Figs. 4 and 6 why XMNL-NLMS achieves poorer convergence performance when the source is far from the centroid of the microphone pair: although the interchannel coherence is relatively low, the $\mathcal{M}$-ratio is not sufficiently high to reduce the degradation in convergence rate due to tap selection. As a consequence of this conflict between the need to reduce interchannel coherence and maximization of tap-input energies, the overall result is a reduction in convergence rate of XMNL-NLMS, as can be seen from Fig. 3. On the other hand, when $x = 2.85$ m, the degradation in convergence rate due to a reduction in $\mathcal{M}$ is offset by a significant reduction in interchannel coherence, as shown in Figs. 6 and 4, respectively. As a result of this joint effect, good overall convergence performance can be obtained for XMNL-NLMS, as can be seen in Fig. 3.

As an additional note, the position of the source affects not only the misalignment convergence of XMNL-NLMS, but also its steady-state value. As can be seen from Fig. 3, the steady-state normalized misalignment is higher than that of NL-NLMS for increasing x position since the weight update is performed using only a fraction of the tap inputs. This causes an additional error in the weight update, resulting in an increase in the steady-state normalized misalignment.

IV. Center-Clipping Approach to SAEC

We now propose a center-clipping algorithm that has the ability to reduce the interchannel coherence according to the source location in the transmission room. We exploit the decorrelation properties of XM tap selection, similar to that of XMNL-NLMS. In addition we propose an error-based compensation technique that addresses the additional steady-state normalized misalignment resulting from XM tap selection.

A. Center-Clipping Exclusive Maximum Tap Selection

We propose to apply center-clipping to the tap-input vectors in order to increase the energies of the "inactive" tap inputs when the interchannel coherence between $\mathbf{z}_1(n)$ and $\mathbf{z}_2(n)$ is relatively low, such as when the source is in front of one of the microphones. This is to reduce the degradation in misalignment convergence brought about by the XM tap-selection process. As

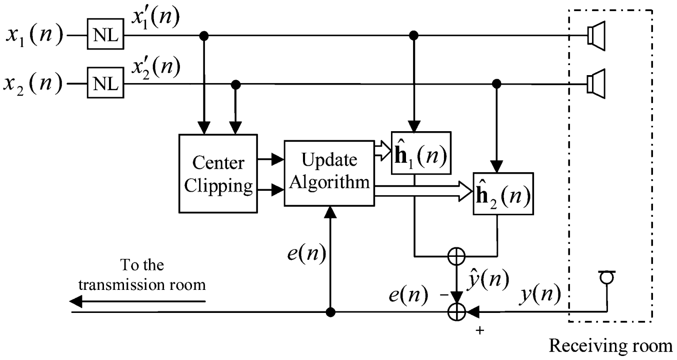

Fig. 7. Schematic diagram of the proposed structure.

will be shown in Section IV-D, the clipping threshold is based on indirect estimation of the similarity between the energies of $\mathbf{x}'_1(n)$ and $\mathbf{x}'_2(n)$. This similarity reflects how close the source is to the microphone centroid, which in turn affects the convergence behavior. The proposed approach ensures a soft-optimization constraint, which makes it robust to source position. A schematic diagram of the proposed method is as shown in Fig. 7.

The proposed clipping-based XMNL-NLMS algorithm (cXMNL-NLMS) updates the filter coefficients using

$$\widehat{\mathbf{h}}_i(n+1) = \widehat{\mathbf{h}}_i(n) + \frac{\mu}{\|\mathbf{x}'(n)\|_2^2 + \epsilon} e(n)\widetilde{\mathbf{x}}'_i(n) \quad (15)$$

$$\widetilde{\mathbf{x}}'_i(n) = \mathbf{Q}_i(n)\mathbf{x}'_i(n) + \overline{\mathbf{Q}}_i(n)\breve{\mathbf{x}}'_i(n) \quad (16)$$

where $\mathbf{x}'_i(n)$ is defined in (6) and (7), $\mathbf{Q}_i(n)$ is the XM tap-selection matrix with diagonal elements defined in (9) and (10), and the $L \times L$ matrix $\overline{\mathbf{Q}}_i(n)$ is defined by

$$\overline{\mathbf{Q}}_i(n) = \mathbf{I}_{L\times L} - \mathbf{Q}_i(n). \quad (17)$$

The matrix $\overline{\mathbf{Q}}_i(n)$ is used to identify the tap-input elements not selected by XM. The purpose of the proposed center-clipping strategy is to increase the energies of tap inputs corresponding to the "inactive" (unselected) taps. We achieve this by first defining $\breve{\mathbf{x}}'_i(n) = [\breve{x}'_i(n), \breve{x}'_i(n-1), \ldots, \breve{x}'_i(n-L+1)]^T$ as the clipped vector whose elements are computed by (18), as shown at the bottom of the page, where $\gamma_i(n)$ controls the amount of clipping for $\breve{\mathbf{x}}_i(n)$. In Section V, we discuss how this clipping threshold can be determined for our SAEC application.

B. Effect of $\gamma_i(n)$ on Tap-Input Energy

The range of $\gamma_i(n)$ for each tap-input vector $\mathbf{x}'_i(n)$ can be bounded between zero and the maximum magnitude of any element within that vector, i.e., $0 \le \gamma_i(n) \le \gamma_{i,\max}(n)$, where

$$\gamma_{i,\max}(n) = \max_{x'_i} |\mathbf{x}'_i(n)|. \quad (19)$$

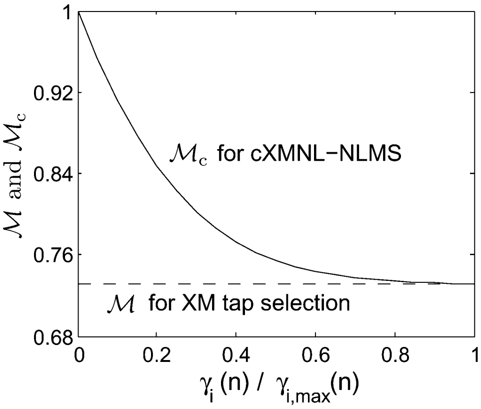

Fig. 8. Variation of $\mathcal{M}$ and $\mathcal{M}_c$ against $\gamma_i(n)/\gamma_{i,\max}(n)$ for the XM tap-selection algorithm and center-clipping algorithm, respectively.

It can be seen from (18) that when $\gamma_i(n) = 0$, we have $\breve{x}'_i(n-k) = x'_i(n-k)$, which results in the second term of (16) having values equivalent to the unselected tap inputs, so that $\widetilde{\mathbf{x}}'_i(n) = \mathbf{x}'_i(n)$ and the proposed cXMNL-NLMS algorithm becomes the full-update NL-NLMS algorithm. On the other hand, when $\gamma_i(n) = \gamma_{i,\max}(n)$, we have $\breve{x}_i(n-k) = 0$, causing the second term of (16) to vanish, hence reducing cXMNL-NLMS to XMNL-NLMS.

In order to illustrate how the clipping threshold $\gamma_i(n)$ affects the energies of the tap-input vectors $\widetilde{\mathbf{x}}'_1(n)$ and $\widetilde{\mathbf{x}}'_2(n)$, we employ a $\mathcal{M}$-ratio criterion similar to the one defined by (14)

$$\mathcal{M}_c = \frac{\|\widetilde{\mathbf{x}}'(n)\|_2^2}{\|\mathbf{x}'(n)\|_2^2} \quad (20)$$

where the subscript c in $\mathcal{M}_c$ denotes center-clipped signals. Fig. 8 illustrates how $\mathcal{M}$ and $\mathcal{M}_c$ vary with $\gamma_i(n)/\gamma_{i,\max}(n)$, using signals generated by convolving the previously defined colored speech-like noise sequence with $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$ when the source position is at (2.85,1.85,1.6) m. If $\gamma_i(n) = 0$ for both channels, $\mathcal{M}_c = 1$ and therefore the convergence behavior of the center-clipped cXMNL-NLMS algorithm will be equivalent to NL-NLMS. On the other hand, when $\gamma_i(n) = \gamma_{i,\max}(n)$, we have $\mathcal{M}_c = \mathcal{M}$ and hence the performance of the proposed cXMNL-NLMS algorithm will be equivalent to XMNL-NLMS.

C. Effect of Clipping Threshold on Interchannel Coherence

As illustrated in Section III, the misalignment convergence of XMNL-NLMS depends on both the tap-input energy as well as the interchannel coherence between $\widetilde{\mathbf{x}}'_1(n)$ and $\widetilde{\mathbf{x}}'_2(n)$. As such, we investigate the effect of $\gamma_i(n)$ on the interchannel coherence between $\widetilde{\mathbf{x}}'_1(n)$ and $\widetilde{\mathbf{x}}'_2(n)$. We convolve the same colored Gaussian noise sequence with $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$, where $L_g = 1024$, with the source positioned at coordinates (2.85,1.85,1.6) m. A total of ten

$$\breve{x}'_i(n-k) = \begin{cases} [|x'_i(n-k)| - \gamma_i(n)]\,\mathrm{sign}\,[x'_i(n-k)], & |x'_i(n-k)| > \gamma_i(n) \\ 0, & |x'_i(n-k)| \le \gamma_i(n) \end{cases}, 0 \le k \le L-1 \quad (18)$$

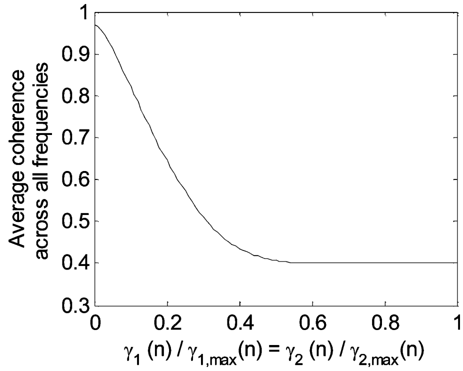

Fig. 9. Interchannel coherence versus clipping threshold for colored noise signal.

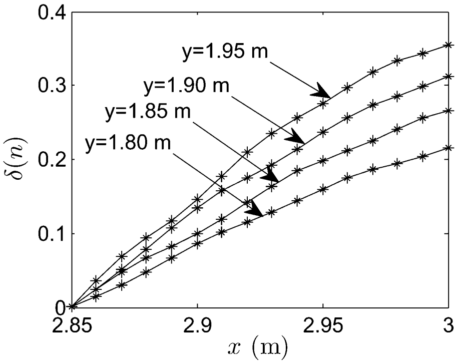

Fig. 10. Variation of $\delta(n)$ against horizontal position x of the source in the transmission room for four cases of vertical positions y.

independent trials are averaged and Fig. 9 illustrates how the mean interchannel coherence across frequency varies with $\gamma_i(n)/\gamma_{i,\max}(n)$, where we constrain $\gamma_1(n)$ and $\gamma_2(n)$ such that $\gamma_1(n)/\gamma_{1,\max}(n) = \gamma_2(n)/\gamma_{2,\max}(n)$. As can be observed, the interchannel coherence reduces with increasing values of clipping threshold. This is as expected since with increasing $\gamma_i(n)$, less energy will be allocated to the unselected taps brought about by the XM tap-selection criterion, thus reducing the similarity between $\mathbf{x}'_1(n)$ and $\mathbf{x}'_2(n)$, which as a consequence causes a reduction in interchannel coherence.

D. Soft-Decision Rule for Clipping Threshold $\gamma_i(n)$

As shown in Figs. 8 and 9, a high value of $\gamma_i(n)$ will reduce both $\mathcal{M}_c$ and interchannel coherence. As described in Section III, a reduction in interchannel coherence is crucial when the source position is near the centroid of the microphone array pair, while the need to increase $\mathcal{M}_c$ becomes important when the source is nearer to one microphone. Hence, $\gamma_i(n)$ enables a tradeoff between interchannel decorrelation and degradation of misalignment convergence due to tap selection. Therefore, high $\gamma_i(n)$ is desirable when the source is near the centroid of the microphone pair, while low $\gamma_i(n)$ is desirable when the source is near one of the microphones. As was pointed out, the similarity between the energies of $\mathbf{x}'_1(n)$ and $\mathbf{x}'_2(n)$ contributes to the high convergence rate of XMNL-NLMS when the source is near the microphone pair centroid. On the other hand, considerable difference in the energies as well as low interchannel coherence between $\mathbf{x}'_1(n)$ and $\mathbf{x}'_2(n)$ cause the convergence rate of the XMNL-NLMS to reduce significantly.

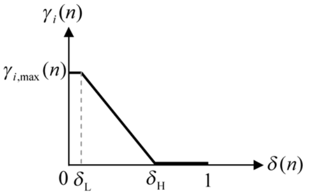

Fig. 11. Clipping threshold $\gamma_i(n)$ versus $\delta(n)$.

We therefore propose to use the difference between absolute values of the two channels as a measure of energy dissimilarity. It is foreseeable that when the source is near the microphone array centroid, the relative absolute values of $\mathbf{x}'_1(n)$ and $\mathbf{x}'_2(n)$ are approximately equal. On the contrary, when the source is nearer to one microphone, the relative absolute values of these tap-input vectors differ from each other. Thus, the dissimilarity measure is defined as

$$\delta(n) = \frac{|\mathrm{mean}\{|\mathbf{x}'_1(n)|\} - \mathrm{mean}\{|\mathbf{x}'_2(n)|\}|}{\mathrm{mean}\{|\mathbf{x}'_1(n)|\} + \mathrm{mean}\{|\mathbf{x}'_2(n)|\}} \tag{21}$$

where $\mathrm{mean}\{|\mathbf{x}'_i(n)|\}$, i =1, 2, is a moving average of the absolute values of the input elements given by

$$\mathrm{mean}\{|\mathbf{x}'_i(n)|\} = \frac{1}{L'} \sum_{j=0}^{L'-1} |x'_i(n-j)|\,. \tag{22}$$

Here, we use $L' = 5L$ to smooth $\delta(n)$ so as to avoid the effects of instantaneous changes of $\delta(n)$ on the value of the clipping threshold. We see that $0 \le \delta(n) < 1$ and that $\delta(n)$ is close to 0 when the received energy of the two microphones are approximately equal and approaches 1 when the received energy from one microphone is much larger than that of the other microphone.

Fig. 10 shows four illustrative examples of how $\delta(n)$ varies with different x and y positions of the source in the transmission room. In these examples, the room dimensions and the coordinates of microphones are given as shown in Fig. 2. As expected, the value of $\delta(n)$ is small when the source is close to the centroid at $x = 2.85$ m. On the contrary, $\delta(n)$ is large when the source is near Microphone 2 at $x = 3$ m.

We now incorporate $\delta(n)$ into cXMNL-NLMS by varying the value of $\gamma_i(n)$ as a function of $\delta(n)$. Since we desire a low $\gamma_i(n)$ when the source is nearer to one microphone and vice versa, $\gamma_i(n)$ should reduce with increasing $\delta(n)$. We therefore propose a piecewise linear mapping

$$\gamma_i(n) = \begin{cases} \gamma_{i,\max}(n), & \delta(n) < \delta_L \\ \gamma_{i,\max}(n)\frac{\delta(n)-\delta_H}{\delta_L-\delta_H}, & \delta_L \le \delta(n) < \delta_H \\ 0, & \delta(n) \ge \delta_H. \end{cases} \tag{23}$$

The relation between $\gamma_i(n)$ and $\delta(n)$ is plotted in Fig. 11. For speech signals, we use values $\delta_L = 0.1$ and $\delta_H = 0.4$ which were determined empirically.

We note from (23) that $\gamma_i(n)$ is independently determined for each channel, and indirectly depends on the relative position of the source and the microphones. In addition, when $\delta(n) < \delta_H$, such as when the source is near the microphone pair centroid, $\gamma_i(n)$ increases with reducing $\delta(n)$. This reduces the effect of the second term in (16) and as a result, the proposed algorithm converges in the same manner as XMNL-NLMS. On the other

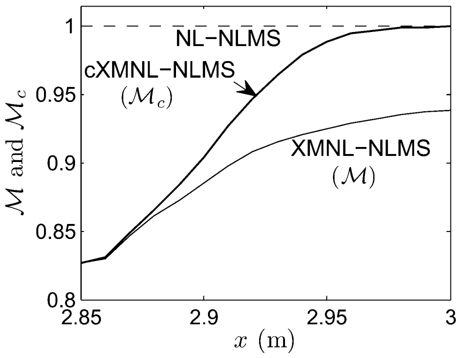

Fig. 12. Comparison of $\mathcal{M}$ and $\mathcal{M}_c$ for NL-NLMS, XMNL-NLMS, and cXMNL-NLMS.

hand, when $\delta(n)$ increases, such as when the source is near one of the microphones, $\gamma_i(n)$ reduces to zero. As shown in Fig. 8, this has the effect of increasing the energy of the unselected taps, which in turn reduces the degradation in misalignment convergence due to XM tap selection in situations where reducing the interchannel coherence cannot further improve the convergence rate of the adaptive algorithm.

Fig. 12 further illustrates how degradation in $\mathcal{M}$, and consequently the convergence performance due to tap selection, can be reduced by incorporating $\gamma_i(n)$ into the proposed cXMNL-NLMS algorithm. As described earlier in Section III-B, the degradation of convergence performance for XMNL-NLMS when the source is far away from the microphone pair centroid ($x \gg 2.85$) is due to a reduction of $\mathcal{M}$ compared to NL-NLMS. In this scenario, the proposed cXMNL-NLMS algorithm ensures that $\mathcal{M}_c$ is closer to the value 1 achieved by NL-NLMS. On the other hand, when the source is near the centroid ($x = 2.85$ m), cXMNL-NLMS attains the beneficial properties of the XM tap-selection strategy to maximally decorrelate the tap-input vectors. The overall joint result is a fast converging cXMNL-NLMS that is robust to the source position in the transmission room.

V. Enhancement of the Steady-State Performance

As noted from Fig. 3 and Section III, the steady-state normalized misalignment of XMNL-NLMS is higher than that of NL-NLMS. This is due to the unselected filter coefficients introducing an error during adaptation since there is now a mismatch between $\mathbf{x}'_i(n)$, which drives the unknown system, and the selective tap-input vector $\mathbf{Q}_i(n)\mathbf{x}'_i(n)$. This error occurs regardless of the source position. It is therefore expected that cXMNL-NLMS also suffers from increased steady-state misalignment since the proposed clipping method generates a signal $\widetilde{\mathbf{x}}'_i(n)$ that is different from $\mathbf{x}'_i(n)$.

To illustrate this, we consider input vectors $\widetilde{\mathbf{x}}'_i(n)$ defined by (17) each of length $L = 128$. A colored noise source signal generated as described in Section III, is positioned at coordinates (2.89,1.85,1.6) m. As before, the room is of dimension as shown in Fig. 2. This steady-state normalized misalignment is achieved by allowing the algorithm to reach its steady-state and averaging over the last 5000 samples. Fig. 13 shows how the steady-state normalized misalignment varies with normalized clipping thresholds $\gamma_i(n)/\gamma_{i,\max}(n)$.

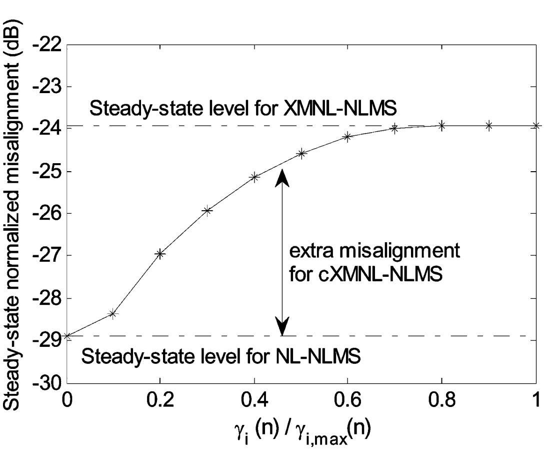

Fig. 13. Relation between steady-state misalignment and normalized clipping threshold in NL-NLMS, XMNL-NLMS, and cXMNL-NLMS, when the source is at (2.89,1.85,1.6) m.

As can be seen, the steady-state normalized misalignment for NL-NLMS is -29 dB, and -24 dB for XMNL-NLMS. If we employ $\gamma_i(n)$ defined by (23) for the above source position, we obtain $\gamma_i(n)/\gamma_{i,\max}(n) \approx 0.42$, so the proposed cXMNL-NLMS algorithm gives an additional 1 dB of steady-state normalized misalignment improvement over XMNL-NLMS.

As a final improvement, we propose to enhance the steady-state normalized misalignment performance of cXMNL-NLMS. We note that the steady-state performance of cXMNL-NLMS depends on the source position and therefore when $\gamma_i(n) > 0$, we need to reduce the additional steady-state normalized misalignment. Hence, we propose to reduce $\gamma_i(n)$ to zero after convergence of the mean-square error (MSE). To estimate the convergence of the algorithm, we employ the following recursive relation for estimating the MSE [24]

$$\varepsilon(n) = \lambda\varepsilon(n-1) + (1-\lambda)e^2(n) \tag{24}$$

where $0 < \lambda < 1$ is related to the time-constant of the averaging process. We consider $\lambda = 0.99$ for our experiments. Hence, when $\varepsilon(n)$ reaches below a lower limit, $\gamma_i(n)$ will be set to zero. In this case, the normalized misalignment reduces towards the steady-state misalignment of NL-NLMS. We therefore propose to incorporate $\varepsilon(n)$ into (23) giving (25),

$$\gamma_i(n) = \begin{cases} \gamma_{i,\max}(n), & \delta(n) < \delta_L, \varepsilon(n) > \nu \\ \gamma_{i,\max}(n)\frac{\delta(n)-\delta_H}{\delta_L-\delta_H}, & \delta_L \le \delta(n) < \delta_H, \varepsilon(n) > \nu \\ 0, & \text{otherwise} \end{cases} \tag{25}$$

where ν is an empirically-derived lower limit that aims to achieve a low level of MSE after convergence. Fig. 14 shows an example of MSE and misalignment convergence for a single trial when the source is at coordinates (2.87,1.85,1.6) m. As can be seen, setting $\gamma_i(n)$ based on (23) brings about a higher initial convergence rate than NL-NLMS, while reducing $\gamma_i(n)$ to zero using (25) after MSE convergence will bring about additional reduction in steady-state normalized misalignment. Additional tests revealed that although the convergence of MSE occurs before convergence of misalignment, exact knowledge of MSE is not required. The proposed cXMNL-NLMS algorithm is summarized in Table I.

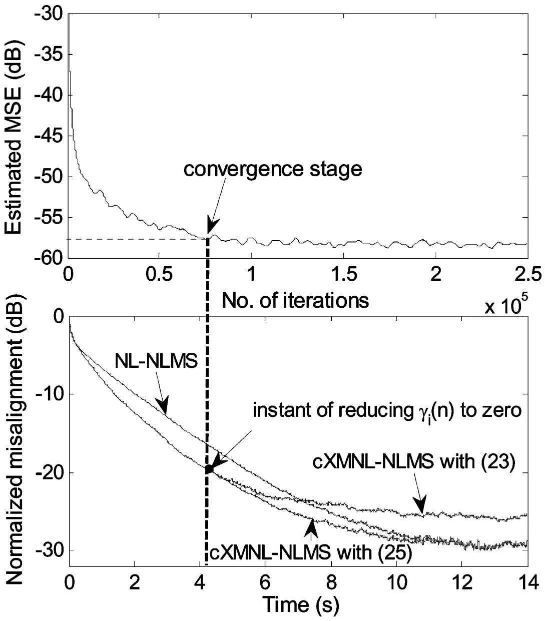

Fig. 14. Estimated MSE and misalignment for the cXMNL-NLMS algorithm.

TABLE I
THE cXMNL-NLMS ALGORITHM

$$
\begin{aligned}
x'_1(n) &= x_1(n) + 0.5\alpha\,[x_1(n) + |x_1(n)|] \\
x'_2(n) &= x_2(n) + 0.5\alpha\,[x_2(n) - |x_2(n)|] \\
\mathbf{Q}_i(n) &= \mathrm{diag}\{\mathbf{q}_i(n)\} \\
q_{1,u}(n) &= \begin{cases} 1, & p_u \in \{0.5L \text{ largest components of } \mathbf{p}(n)\} \\ 0, & \text{otherwise} \end{cases} \\
q_{2,v}(n) &= \begin{cases} 1, & p_v \in \{0.5L \text{ smallest components of } \mathbf{p}(n)\} \\ 0, & \text{otherwise} \end{cases} \\
\mathbf{p}(n) &= |\mathbf{x}'_1(n)| - |\mathbf{x}'_2(n)| \\
\overline{\mathbf{Q}}_i(n) &= \mathbf{I}_{L\times L} - \mathbf{Q}_i(n) \\
\delta(n) &= \frac{\left|\mathrm{mean}\{|\mathbf{x}'_1(n)|\} - \mathrm{mean}\{|\mathbf{x}'_2(n)|\}\right|}{\mathrm{mean}\{|\mathbf{x}'_1(n)|\} + \mathrm{mean}\{|\mathbf{x}'_2(n)|\}} \\
\varepsilon(n) &= \lambda\varepsilon(n-1) + (1-\lambda)e^2(n) \\
\gamma_i(n) &= \begin{cases} \gamma_{i,\max}(n), & \delta(n) < \delta_{\mathrm{L}},\ \varepsilon(n) > \nu \\ \gamma_{i,\max}(n)\frac{\delta(n)-\delta_{\mathrm{H}}}{\delta_{\mathrm{L}}-\delta_{\mathrm{H}}}, & \delta_{\mathrm{L}} \le \delta(n) < \delta_{\mathrm{H}},\ \varepsilon(n) > \nu \\ 0, & \text{otherwise} \end{cases} \\
\breve{x}'_i(n-k) &= \begin{cases} \mathrm{sign}[x'_i(n-k)][|x'_i(n-k)| - \gamma_i(n)], & |x'_i(n-k)| > \gamma_i(n) \\ 0, & |x'_i(n-k)| \le \gamma_i(n) \end{cases} \\
\widetilde{\mathbf{x}}'_i(n) &= \mathbf{Q}_i(n)\mathbf{x}'_i(n) + \overline{\mathbf{Q}}_i(n)\breve{\mathbf{x}}'_i(n) \\
\widehat{\mathbf{h}}_i(n+1) &= \widehat{\mathbf{h}}_i(n) + \frac{\mu}{||\mathbf{x}'(n)||_2^2+\epsilon}e(n)\widetilde{\mathbf{x}}'_i(n)
\end{aligned}
$$

VI. FURTHER SIMULATION AND EXPERIMENTAL RESULTS

TABLE II
SPECIFICATIONS OF THE SIMULATED ENVIRONMENT IN SAEC

Transmission room	
Dimensions	$7 \times 7 \times 4$ m
Coordinates of Microphone 1	$(3, 2, 1.5)$ m
Coordinates of Microphone 2	$(2.7, 2, 1.5)$ m
Coordinates of source, Case 1	$(3, 1.9, 1.55)$ m
Coordinates of source, Case 2	$(2.88, 1.85, 1.6)$ m
Coordinates of source, Case 3	$(2.85, 1.85, 1.6)$ m
L_g	1024
Receiving room	
Dimensions	$6.3 \times 4 \times 3.5$ m
Coordinates of microphone	$(3, 2, 1.5)$ m
Coordinates of loudspeaker 1	$(2.85, 1.8, 1.6)$ m
Coordinates of loudspeaker 2	$(2.4, 1.1, 1.7)$ m
L	512

We evaluate, by way of further simulation, the performance of cXMNL-NLMS under different source positions. In order to simulate the SAEC system, impulse responses $\mathbf{h}_1(n)$, $\mathbf{h}_2(n)$, $\mathbf{g}_1(n)$, and $\mathbf{g}_2(n)$ were generated using the method of images [22]. To evaluate the robustness of the algorithms, we fixed the location of the microphones while the source position in the transmission room was varied across three cases shown in Table II. A sampling rate of $f_s = 11\,025$ Hz was used throughout the experiment. The source signal was generated by filtering a white Gaussian noise signal through a low-pass finite impulse response (FIR) filter with coefficients $[0.3574, 0.9, 0.3574]$, as was used in Section III.

We compare the convergence performance of the proposed cXMNL-NLMS algorithm with NL-NLMS and XMNL-NLMS. Since the steady-state normalized misalignment for XMNL-NLMS varies with the source position, we chose its step-size so that its steady-state normalized misalignment reaches that of NL-NLMS and cXMNL-NLMS when the source position is in front of the microphone array centroid at (2.85,1.85,1.6) m. This corresponds to $\mu = 0.8$ for both NL-NLMS and cXMNL-NLMS and $\mu = 0.6$ for XMNL-NLMS. White Gaussian noise (WGN) is added to $y(n)$ to achieve an $\mathrm{SNR} = 30$ dB. For all simulations, we have used $\nu = -58$ dB for cXMNL-NLMS. The normalized misalignment curves, obtained by averaging over ten independent trials, are plotted for Cases 1, 2, and 3 (Table II) in Fig. 15(a)–(c) respectively.

Fig. 15(a) shows the convergence performance of the algorithms where the source is directly in front of the right microphone. In this case, $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$ are significantly different and hence a high value of $\delta(n)$ defined in (21) is expected. As shown in Fig. 11, this translates to a low $\gamma_i(n)$, and consequently, as shown in (18), $\breve{x}'_i(n-k) = x'_i(n-k)$. As a result, the convergence performance of cXMNL-NLMS is equivalent to that of NL-NLMS. The proposed cXMNL-NLMS algorithm thus achieves an initial convergence of nearly 8 dB better than XMNL-NLMS and reaches a steady-state normalized misalignment of 4 dB lower as expected.

Fig. 15(b) shows convergence results when the source position is mid-way between the microphone pair centroid and the right microphone at (2.88,1.85,1.6) m. Now, the interchannel coherence increases relative to the previous case and as can be seen from this result, cXMNL-NLMS achieves the highest rate

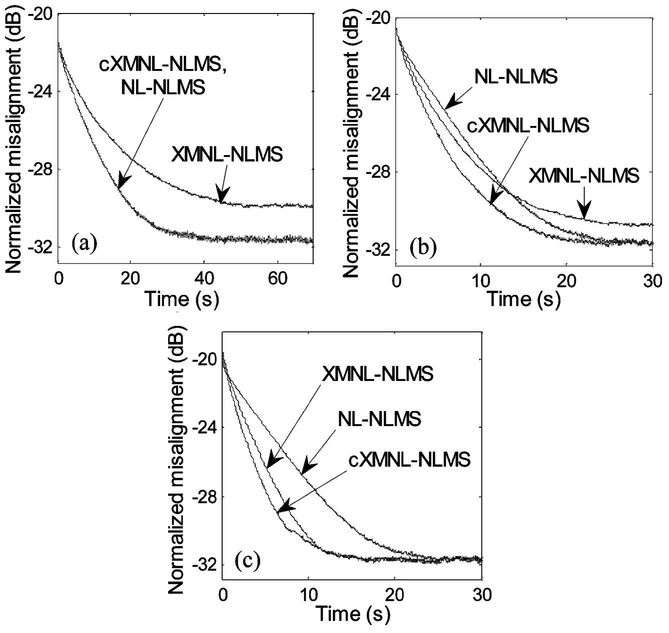

Fig. 15. Normalized misalignment of the NL-NLMS, XMNL-NLMS, and cXMNL-NLMS algorithms for a colored noise source signal. (a) Source directly in front of right microphone at (3,1.9,1.55) m. (b) Source at (2.88,1.85,1.6) m. (c) Source in the center of microphone pair at (2.85,1.85,1.6) m.

of initial convergence, improving that of NL-NLMS by nearly 4 dB during initial convergence. We note that when compared to XMNL-NLMS, cXMNL-NLMS achieves approximately 3 dB improvement during initial convergence and about 2 dB lower steady-state normalized misalignment.

Finally, when the source position is in front of the microphone pair centroid at coordinates (2.85,1.85,1.6) m, $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$ are similar and the interchannel coherence between $\mathbf{x}_1'(n)$ and $\mathbf{x}_2'(n)$ is high. As can be seen from Fig. 15(c), the convergence of cXMNL-NLMS achieves the highest rate of convergence with an improvement of approximately 4 dB over that of XMNL-NLMS and nearly 10 dB over that of NL-NLMS. In terms of steady-state normalized misalignment, the NL-NLMS algorithm requires nearly 10 s more than that of cXMNL-NLMS to reach -30 dB.

To further illustrate the convergence performance of the proposed cXMNL-NLMS algorithm, we simulated the SAEC system using a speech signal as shown in Fig. 16. In this example, the speech signal is sampled at 11 025 Hz and a WGN is added to $y(n)$ to achieve an SNR $= 30$ dB. The position of the source in the transmission room is (2.88,1.85,1.6) m. As can be seen from this result, cXMNL-NLMS achieves approximately 6 dB lower misalignment than NL-NLMS and 4 dB lower than XMNL-NLMS during initial convergence.

We consider using recorded impulse responses, where the dimensions of the transmission room is 6.5 m × 8.75 m × 2.65 m, the source was positioned at (3.25,4.37,1.15) m while the two microphones were placed at (3.11,2.37,1.2) m and (3.39,2.37,1.2) m for Case 1 and at (3.25,2.37,1.2) m and (2.83,2.37,1.2) m for Case 2. The estimated reverberation time was 280 ms. These impulse responses were of length 3087 samples and subsequently truncated to 512 samples. Fig. 17 shows one of the measured impulse responses in the transmission room. For this experiment, the sampling frequency, step-sizes as well as the SNRs were the same as that of the previous simulations. The results are shown in Fig. 18. As can be seen from Fig. 18(a), cXMNL-NLMS achieves nearly 3 dB improvement in convergence performance compared to XMNL-NLMS when the source is in front of the microphone centroid. In Fig. 18(b), when the source is in front of the right microphone, the proposed algorithm achieves nearly 6 dB improvement in convergence compared to XMNL-NLMS.

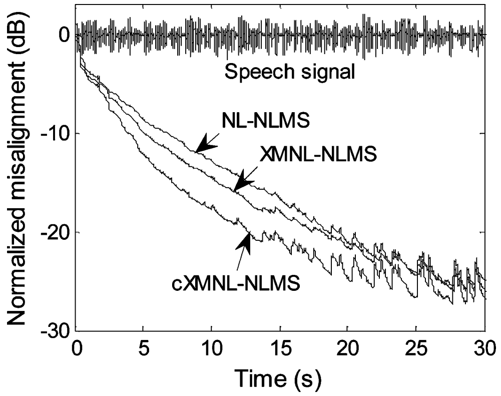

Fig. 16. Normalized misalignment for the NL-NLMS, XMNL-NLMS, and cXMNL-NLMS algorithms when the source is at (2.88,1.85,1.6) m for a speech signal.

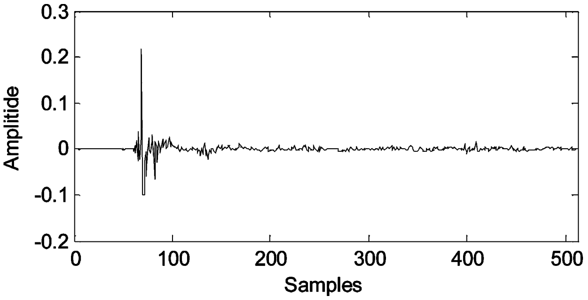

Fig. 17. Illustration of measured transmission room impulse response $\mathbf{g}_1$.

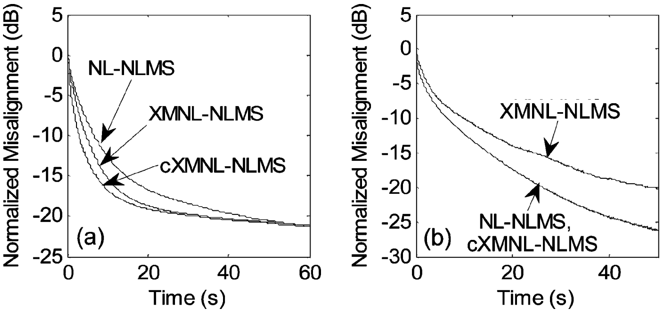

Fig. 18. Normalized misalignment of the NL-NLMS, XMNL-NLMS, and cXMNL-NLMS algorithms for real room impulse responses. (a) Case 1: source in the center of microphone pair. (b) Case 2: source approximately in front of the right microphone.

VII. Conclusion

We presented a new approach to improve the misalignment convergence as well as the steady-state performance and robustness of adaptive filters for SAEC. This approach retains the decorrelation properties of the XM selective-tap algorithm when the source is located near the microphone centroid, but employs a variable center-clipping threshold whose value is derived based on the absolute values of the received microphone signals in order to work better, when the source is located closer to one of the microphones. The proposed approach achieves better convergence performance for different source positions in comparison to both NL-NLMS and XMNL-NLMS.

REFERENCES

[1] *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Y. Huang and J. Benesty, Eds.. Norwell, MA: Kluwer, 2004.

[2] J. Benesty, M. M. Sondhi, and Y. Huang, *Handbook of Speech Processing*. Secaucus, NJ: Springer-Verlag, 2008.

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

[4] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 156–165, Mar. 1998.

[5] M. M. Sondhi and D. R. Morgan, "Acoustic echo cancellation for stereophonic teleconferencing," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust.*, 1991, pp. 141–142.

[6] S. Shimauchi and S. Makino, "Stereo projection echo canceller with true echo path estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1995, pp. 3059–3062.

[7] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic acoustic echo cancellation-An overview of the fundamental problem," *IEEE Signal Process. Lett.*, vol. 2, no. 8, pp. 148–151, Aug. 1995.

[8] J. Benesty, D. R. Morgan, J. L. Hall, and M. M. Sondhi, "Stereophonic acoustic echo cancellation using nonlinear transformations and comb filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1998, pp. 3673–3676.

[9] J. Benesty, D. R. Morgan, J. L. Hall, and M. M. Sondhi, "Synthesized stereo combined with acoustic echo cancellation for desktop conferencing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1999, pp. 148–158.

[10] K. Mayyas, "Stereophonic acoustic echo cancellation using lattice orthogonalization," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 7, pp. 517–525, Oct. 2002.

[11] T. Gänsler and J. Benesty, "New insights into the stereophonic acoustic echo cancellation problem and an adaptive nonlinearity solution," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 257–267, Jul. 2002.

[12] M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1998, pp. 3689–3692.

[13] T. Tangsangiumvisai, J. A. Chambers, and A. G. Constantinides, "Time varying all-pass filters using spectral-shaped noise for signal decorrelation in stereophonic acoustic echo cancellation," in *Proc. Int. Conf. Digital Signal Process.*, 2002, pp. 87–92.

[14] J. Herre, H. Buchner, and W. Kellermann, "Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2007, pp. I-17–I-20.

[15] J. M. Valin, "Perceptually-motivated nonlinear channel decorrelation for stereo acoustic echo cancellation," in *Proc. Hands-Free Speech Commun. Microphone Arrays (HSCMA)*, 2008, pp. 188–191.

[16] S. Emura, Y. Haneda, A. Kataoka, and S. Makino, "Stereo echo cancellation algorithm using adaptive update on the basis of enhanced input-signal vector," *Signal Process.*, vol. 86, pp. 1157–1167, Jun. 2006.

[17] A. W. H. Khong and P. A. Naylor, "Stereophonic acoustic echo cancellation employing selective-tap adaptive algorithms," *IEEE Trans. Speech Audio Process*, vol. 14, no. 3, pp. 785–796, May 2006.

[18] M. Bekrani, A. W. H. Khong, and M. Lotfizad, "Neural network based adaptive echo cancellation for stereophonic teleconferencing application," in *Proc. Int. Conf. Multimedia Expo*, 2010, pp. 1172–1177.

[19] M. Bekrani, M. Lotfizad, and A. W. H. Khong, "An efficient quasi LMS/newton adaptive algorithm for stereophonic acoustic echo cancellation," in *Proc. IEEE Asia Pacific Conf. Circuits Syst.*, 2010.

[20] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 2001.

[21] D. R. Morgan, J. L. Hall, and J. Benesty, "Investigation of several types of nonlinearities for use in stereo acoustic echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 6, pp. 686–696, Sep. 2001.

[22] 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.

[23] S. Attallah, "The wavelet transform-domain LMS adaptive filter with partial subband-coefficient updating," *IEEE Trans. Circuits Syst. II: Express Briefs*, vol. 53, no. 1, pp. 8–12, Jan. 2006.

[24] K. Mayyas, "New transform-domain adaptive algorithms for acoustic echo cancellation," *Digital Signal Process.*, vol. 13, no. 3, pp. 415–432, Jul. 2003.

Mehdi Bekrani was born in Gorgan, Iran, in 1979. He received the B.Sc. degree from Ferdowsi University of Mashhad, Mashad, Iran, in 2002, and the M.Sc. and Ph.D. degrees from Tarbiat Modares University, Tehran, Iran, in 2004 and 2010, respectively, all in electrical engineering.

He is currently a Research Fellow at Nanyang Technological University, Singapore. His current research interests include acoustic signal processing and their applications.

Andy W. H. Khong (M'06) received the B.Eng. degree from Nanyang Technological University, Singapore, in 2002 and the Ph.D. degree from the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K., in 2005. His Ph.D. research was mainly on partial-update and selective-tap adaptive algorithms with applications to mono- and multi-channel acoustic echo cancellation for hands-free telephony.

He is currently an Assistant Professor in the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Prior to that, he served as a Research Associate in the Department of Electrical and Electronic Engineering, Imperial College London, from 2005 to 2008. His postdoctoral research involved the development of signal processing algorithms for vehicle destination inference as well as the design and implementation of acoustic array and seismic fusion algorithms for perimeter security systems. He has also published works on acoustic blind channel identification and equalization for speech dereverberation. His other research interests include human-computer interfaces, source localization, speech enhancement, and blind deconvolution.

Mojtaba Lotfizad was born in Tehran, Iran, in 1955. He received the B.S. degree in electrical engineering from AmirKabir University of Technology, Tehran, in 1980, and the M.S. and Ph.D. degrees from the University of Wales, Cardiff, U.K., in 1985 and 1988, respectively.

He joined the Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran. He has also been a Consultant to several industrial and governmental organizations. His current research interests are in signal processing, adaptive filtering, speech processing, and specialized processors.

View publication stats