A NEW PARTIAL UPDATE NLMS FOR STEREOPHONIC ACOUSTIC ECHO CANCELLATION

Mehdi Bekrani[1], *Andy W. H. Khong*[2], *and Mojtaba Lotfizad*[1]

[1]Tarbiat Modares University, Tehran, Iran
Email: {bekrani,lotfizad}@modares.ac.ir
[2]Nanyang Technological University, Singapore
Email: andykhong@ntu.edu.sg

ABSTRACT

In this paper, we present a new partial update NLMS adaptive filtering algorithm for improving the performance of stereophonic acoustic echo cancellers. The proposed partial update approach brings about a low interchannel coherence, independent of the location of the source in the transmission section which in turn increases the robustness to source positions as well as the convergence rate of adaptive filters in this application. Simulation results verify the increase in performance of the proposed algorithm over other partial update algorithms.

Index Terms— stereophonic acoustic echo cancellation, misalignment, partial update

1. INTRODUCTION

There is an increasing interest for employing multi-channel sound in audio communication systems in order to achieve better audio perception. Examples of such systems are teleconferencing systems and hands-free telecommunication hand-held devices. One of the problems that should be solved in such systems is the suppression of the multi-channel acoustic echo. Multi-channel acoustic echo cancellation has issues that make it considerably more difficult to overcome than the monophonic case. The fundamental problem is the existence of a mismatch between the impulse responses of the adaptive filters and those of the acoustic paths of the receiving room. This so-called misalignment problem [1] leads to a residual echo in the system as well reduces the robustness of the adaptive filters to the abrupt changes in the acoustic paths of the transmitting room, i.e., the paths from the talker to the microphones. As a result, such problems lead to a performance degradation of the conferencing systems that is in conflict with high quality communication requirements. It has been shown in [1] that the misalignment problem is due to the high interchannel coherence between the transmitted signals from the transmission to the receiving room.

To date, a variety of methods have been proposed to address the misalignment problem by reducing the interchannel coherence between the transmitted signals in case of stereophonic acoustic echo cancellation (SAEC). Such algorithms include methods that perform preprocessing on the input stereophonic signals, such as adding or modulating small quantities of independent noise to each input channel [2] and comb filtering [3]. It is important to note that although decorrelating the transmitted signal is important to achieve high convergence for the adaptive filters, algorithms proposed recently are aimed at minimizing the distortion introduced by such preprocessing methods. These algorithms include adding a nonlinear function of the signal of each channel to that of the same channel [1] [4] [5], time-varying all-pass filtering of the stereophonic signals [6], and the application of input-sliding technique to SAEC [7]. Yet one of the most popular technique to achieve decorrelation without adversely degrading the audio quality is the addition of a small nonlinearity to each channel [1]. Although the above methods yield improvements in the estimation of the true room impulse response (RIR), they have a key limitation in terms of audio degradation. This is because they act on the signals that are directly sent to the listeners in the receiving room, and hence in order to achieve a high degree of decorrelation, the stereophonic perception or quality is often sacrificed [8] [9].

More recently, new algorithms have been developed to achieve signal decorrelation in the weight updating process of adaptive filters, such as employing a two-channel adaptive lattice structure [10] and using an exclusive-maximum (XM) nonlinear tap-selection based adaptive algorithms [11]. Among the above methods, the XM tap-selection based algorithm [11] has shown to achieve high convergence with lower complexity compared with those techniques mentioned above. In this algorithm, coefficients update are performed only on a set of exclusive filter coefficients for which the total energy of their corresponding taps is maximized such that the coherence between the two channels then is minimized by the exclusivity constraint across both channels.

In this paper, we propose a new partial-update algorithm which improves the convergence performance of the adaptive filters in SAEC. The proposed algorithm exploits the advantages of the XM nonlinear (XMNL) tap-selection algorithm and we extend its development in order to have a better convergence performance. To achieve this, the proposed partial update algorithm replaces some of the XM selected taps with some specified unselected taps to reach to a lower level of interchannel coherence. This process is performed based on a proposed criterion that is a function of the interchannel coherence which in turn give rise to an improvement in convergence performance of adaptive filters.

2. STEREOPHONIC ACOUSTIC ECHO CANCELLATION

Figure 1 shows the stereophonic acoustic echo canceller for a typical teleconferencing application. For simplicity, we consider only one microphone in the receiving room, since a similar analysis can be applied to the other channel [1]. As can be seen, stereophonic signals $u_1(n)$ and $u_2(n)$ are received by the microphones of the transmission room. These signals are generated by the sound source $s(n)$ via RIRs $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$ in the transmission room. These received signals are then transmitted to the loudspeakers in the receiving room which produce an echo signal $y(n)$ given by

$$y(n) = \mathbf{h}_1^T \mathbf{u}_1(n) + \mathbf{h}_2^T \mathbf{u}_2(n) + w(n) , \tag{1}$$

This research is supported by the Singapore National Research Foundation Interactive Digital Media R&D Program, under research grant NRF2008IDM-IDM004-010.

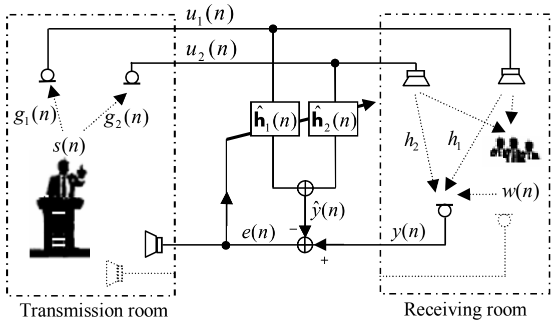

Fig. 1. Stereophonic acoustic echo cancellation for teleconferencing application.

where $\mathbf{h}_i = [h_{i,0}, h_{i,1}, \ldots, h_{i,L-1}]^T$ is the i^{th} channel receiving room RIR, $\mathbf{u}_i(n) = [u_i(n), u_i(n-1), \ldots, u_i(n-L+1)]^T$, i=1, 2 and $w(n)$ is the noise of the receiving room's microphone. To address the SAEC problem, two adaptive filters are employed to estimate $\mathbf{h}_1$ and $\mathbf{h}_2$ in order to reduce the stereophonic acoustic echo. The error signal between the echo signal and its estimation is thus given by

$$e(n) = y(n) - \left[\widehat{\mathbf{h}}_1^T(n)\mathbf{u}_1(n) + \widehat{\mathbf{h}}_2^T(n)\mathbf{u}_2(n)\right], \tag{2}$$

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 i^{th} channel.

2.1. Review of the Exclusive-maximum Nonlinear NLMS algorithm

One of the most recent algorithms proposed for SAEC is the exclusive-maximum nonlinear normalized least-mean-square algorithm (XMNL-NLMS) [11]. This algorithm incorporates the nonlinear preprocessor [1] and a tap-selection scheme that reduces the interchannel coherence by selecting an exclusive set of filter coefficients to update for each channel. The degradation due to this tap-selection is then minimized by jointly maximizing the L_2 norm of the selected tap-inputs across both channels.

The XMNL-NLMS algorithm can be described by an NL preprocessor operating on tap-input vectors $\mathbf{u}_1(n)$ and $\mathbf{u}_2(n)$ such that the transmitted signals $x_1(n)$ and $x_2(n)$ are given by

$$x_1(n) = u_1(n) + 0.5\alpha\{u_1(n) + |u_1(n)|\}, \tag{3}$$
$$x_2(n) = u_2(n) + 0.5\alpha\{u_2(n) - |u_2(n)|\}, \tag{4}$$

where α controls the amount of non-linearity and a value of $\alpha = 0.5$ offers a good compromise between speech quality and convergence rate of the NLMS algorithm [1]. The weight update for XMNL-NLMS is performed by

$$\widehat{\mathbf{h}}(n+1) = \widehat{\mathbf{h}}(n) + \frac{\mu}{\|\mathbf{x}(n)\|^2 + \varepsilon} e(n)\widetilde{\mathbf{x}}(n), \tag{5}$$

where $\widehat{\mathbf{h}}(n) = [\widehat{\mathbf{h}}_1^T(n)\ \widehat{\mathbf{h}}_2^T(n)]^T$ and $\widetilde{\mathbf{x}}(n) = [\widetilde{\mathbf{x}}_1^T(n)\ \widetilde{\mathbf{x}}_2^T(n)]^T$ such that the sub-selected tap-input vector $\widetilde{\mathbf{x}}_i(n)$ is defined by

$$\widetilde{\mathbf{x}}_i(n) = \mathbf{Q}_i(n)\mathbf{x}_i(n), \tag{6}$$

and $\mathbf{x}_i(n) = [x_i(n), x_i(n-1), \ldots, x_i(n-L+1)]^T$. In addition $\mathbf{Q}_i(n) = \text{diag}\{\mathbf{q}_i(n)\}$ is a $L \times L$ tap-selection matrix where elements in the $L \times 1$ vector $\mathbf{q}_i(n)$ are given by

$$q_{1,u}(n) = \begin{cases} 1 & p_u \in \{M \text{ maxima of } \mathbf{p}(n)\} \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

$$q_{2,v}(n) = \begin{cases} 1 & p_v \in \{M \text{ minima of } \mathbf{p}(n)\} \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

$$\mathbf{p}(n) = |\mathbf{x}_1(n)| - |\mathbf{x}_2(n)|, \tag{9}$$

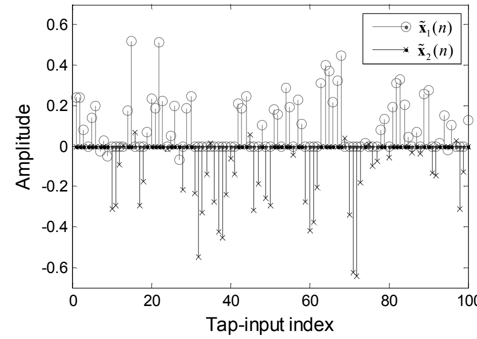

Fig. 2. Selected tap-inputs $\widetilde{\mathbf{x}}_1(n)$ and $\widetilde{\mathbf{x}}_2(n)$ for the XMNL-NLMS algorithm when the source is in $\{4.5, 1, 1.5\}$ m.

given that $u, v = 1,\ 2, \ldots, L$ represent element u of $\mathbf{q}_1(n)$ and element v of $\mathbf{q}_2(n)$, $|\mathbf{x}_i(n)| = [|x_i(n)|, |x_i(n-1)|, \ldots, |x_i(n-L+1)|]^T$ and $0 < M \leq L$, such that a good convergence rate is achieved when $M = 0.5L$.

2.2. Disadvantage of the XMNL-NLMS

One of the problems that has not been considered for XMNL-NLMS is its robustness to the source position in the transmission room. Simulations show that when the source is directly in-front-of the microphone pair centroid, XMNL-NLMS decorrelates the two stereophonic signals $\widetilde{\mathbf{x}}_1(n)$ and $\widetilde{\mathbf{x}}_2(n)$ efficiently and thus outperforms the full-update NL-NLMS. On the contrary, the convergence rate of XMNL-NLMS is reduced significantly when the source is located away from the centroid of the microphone pair. This is due to the reduction in its ability to decorrelate the stereophonic signals effectively. In order to gain further insights into the degradation in convergence performance of XMNL-NLMS with respect to the position of the source, we note that when the talker is in-front-of the microphone pair centroid, most of the selected taps in $\widetilde{\mathbf{x}}_1(n)$ are greater than zero whereas for the other channel, most of the taps in $\widetilde{\mathbf{x}}_2(n)$ are smaller than zero.

Figure 2 shows an illustrative example of the XM selected taps $\widetilde{\mathbf{x}}_1(n)$ and $\widetilde{\mathbf{x}}_2(n)$ when the source is in-front-of the microphone pair centroid. For clarity, we show only the first 100 samples of the two tap-input vectors $\widetilde{\mathbf{x}}_1(n)$ and $\widetilde{\mathbf{x}}_2(n)$ each of length 512 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 $\mathbf{x}_2(n)$ are smaller than zero. This is due to the intrinsic effect of NL preprocessing which increases the magnitude of the positive elements in $u_1(n)$ based on (3) and the negative elements in $u_2(n)$ based on (4). As a result of XM tap-selection on this NL preprocessed signals $\mathbf{x}_1(n)$ and $\mathbf{x}_2(n)$, the XM tap-selection criterion is utilized efficiently to decorrelate input vectors $\mathbf{x}_1(n)$ and $\mathbf{x}_2(n)$ and consequently, good convergence performance is achieved. We note that for this source location, the modest degradation in weight update due to tap-selection does not significantly offset the benefits brought about by the efficiently decorrelation due to the exclusivity criterion. On the other hand, when the source is away from the centroid of microphone pair, the effect of NL preprocessing on the XM tap-selection reduces and the similarity between $\widetilde{\mathbf{x}}_1(n)$ and $\widetilde{\mathbf{x}}_2(n)$ increases. This increases the cross-correlation between $\widetilde{\mathbf{x}}_1(n)$ and $\widetilde{\mathbf{x}}_2(n)$ which in turn reduces the convergence rate for the XMNL-NLMS algorithm compared to NL-NLMS.

3. PROPOSED PARTIAL UPDATE NLMS ALGORITHM

We propose a new partial update nonlinear NLMS (PUNL-NLMS) algorithm to further improve the convergence rate of XMNL-NLMS. We propose to achieve high convergence rate when the source is away from the microphone pair centroid by considering those tap-inputs that produce a low interchannel coherence while preventing a significant loss of the total tap-input energy to the weights. The weight update for our proposed PUNL-NLMS algorithm is performed by

$$\widehat{\mathbf{h}}(n+1) = \widehat{\mathbf{h}}(n) + \frac{\mu}{\left\|\mathbf{x}(n)\right\|^2 + \varepsilon} e(n)\breve{\mathbf{x}}(n), \tag{10}$$

where $\breve{\mathbf{x}}(n) = [\breve{\mathbf{x}}_1^T(n)\ \breve{\mathbf{x}}_2^T(n)]^T$ is the newly selected set of elements of the tap-input vectors $\mathbf{x}(n) = [\mathbf{x}_1^T(n)\ \mathbf{x}_2^T(n)]^T$ which are determined by a new criteria defined in the following.

In order to ease the evaluation, we define some preliminaries as follows

$$\begin{aligned} \widetilde{\mathbf{x}}_{i,p}(n) &= 0.5\{\widetilde{\mathbf{x}}_i(n) + |\widetilde{\mathbf{x}}_i(n)|\}\ , & (11)\\ \widetilde{\mathbf{x}}_{i,n}(n) &= 0.5\{\widetilde{\mathbf{x}}_i(n) - |\widetilde{\mathbf{x}}_i(n)|\}\ . & (12)\end{aligned}$$

As can be seen, the non-zero elements of $\widetilde{\mathbf{x}}_{i,p}(n)$ and $\widetilde{\mathbf{x}}_{i,n}(n)$ include only positive and negative elements of the tap-input vectors $\widetilde{\mathbf{x}}_i(n)$ respectively such that $\widetilde{\mathbf{x}}_i(n)$ is obtained using the XM tap-selection criterion described in (6). In a same manner we define the following for the NL processed received signals given by

$$\begin{aligned} \mathbf{x}_{i,p}(n) &= 0.5\{\mathbf{x}_i(n) + |\mathbf{x}_i(n)|\}\ , & (13)\\ \mathbf{x}_{i,n}(n) &= 0.5\{\mathbf{x}_i(n) - |\mathbf{x}_i(n)|\}\ . & (14)\end{aligned}$$

The aim of our partial update algorithm is therefore to reduce the interchannel coherence of the $\widetilde{\mathbf{x}}_i(n)$ efficiently while at the same time minimizing the energy loss to the weights during the tap-selection process for weight adaptation. We can achieve this by firstly removing elements of the XM tap-input vector $\widetilde{\mathbf{x}}_i(n)$ that have significant interchannel coherence and subsequently replacing these elements with other elements of the tap-input vector $\mathbf{x}_i(n)$ not selected by the XM preprocessor. In addition, these newly selected elements should not have any significant effect on the interchannel coherence in comparison to the rest of the unselected taps in $\mathbf{x}_i(n)$. The second procedure is done to compensate the loss of input energy to the weight update algorithm during the removal process described in the first step. These procedures in turn help the decorrelation effect of nonlinear preprocessing in the adaptive filtering process. To illustrate the above proposed approach, we first define

$k_1(n)$: the number of non-zero elements of $\widetilde{\mathbf{x}}_{1,n}(n)$
$k_2(n)$: the number of non-zero elements of $\widetilde{\mathbf{x}}_{2,p}(n)$
$r(n)$: the number of replaced tap-inputs from $\{\widetilde{\mathbf{x}}_{1,n}(n),\ \widetilde{\mathbf{x}}_{2,p}(n)\}$ in $\breve{\mathbf{x}}(n)$.

It should be noted that $r(n)$ is always less than or equal to $k_1(n) + k_2(n)$. We propose to employ a new criteria to reduce the interchannel coherence in comparison to full-update algorithm and XM tap-selection algorithm. This measure is defined as the ratio between the number of replaced tap-inputs from both channels in $\breve{\mathbf{x}}(n)$ and the number of non-zero elements in $\widetilde{\mathbf{x}}_{1,n}(n)$ and $\widetilde{\mathbf{x}}_{2,p}(n)$, i.e.,

$$\varphi = \frac{r(n)}{k_1(n) + k_2(n)}\ . \tag{15}$$

We also define the following

$$\begin{aligned} \gamma_1(n) &= k_1(n)\varphi\ , & (16)\\ \gamma_2(n) &= k_2(n)\varphi\ . & (17)\end{aligned}$$

We will use $\gamma_i(n)$ to represent the number of elements being removed from $\widetilde{\mathbf{x}}_i(n)$ in order to build $\breve{\mathbf{x}}_i(n)$. To achieve a reduction in interchannel coherence between $\widetilde{\mathbf{x}}_1(n)$ and $\widetilde{\mathbf{x}}_2(n)$, instead of considering M largest elements of $\mathbf{p}(n)$ to produce $\widetilde{\mathbf{x}}_1(n)$ and M smallest elements of $\mathbf{p}(n)$ to produce $\widetilde{\mathbf{x}}_2(n)$ - as defined in (6)-(9), we first obtain $M - \gamma_1(n)$ largest elements of $\mathbf{p}(n)$ to compute $\breve{\mathbf{x}}_1(n)$ and $M - \gamma_2(n)$ smallest elements of $\mathbf{p}(n)$ to compute $\breve{\mathbf{x}}_2(n)$. As such, we have used only a percentage of the elements of $\widetilde{\mathbf{x}}_i(n)$ to generate $\breve{\mathbf{x}}_i(n)$. This percentage depends on the amount of φ.

It is important, at this stage, to note that the rest of the elements in $\breve{\mathbf{x}}_i(n)$ are zero. Such null elements are expected to cause significant degradation in convergence performance due to the energy loss in the tap-input vector [11]. In order to address this, we define vectors $\mathbf{t}_1(n)$ and $\mathbf{t}_2(n)$ as

$$\begin{aligned} \mathbf{t}_1(n) &= \big(\mathbf{I}_{L\times L} - \mathbf{Q}_1(n)\big)\mathbf{x}_{1,p}(n)\ , & (18)\\ \mathbf{t}_2(n) &= \big(\mathbf{I}_{L\times L} - \mathbf{Q}_2(n)\big)\mathbf{x}_{2,n}(n)\ . & (19)\end{aligned}$$

such that $\mathbf{t}_1(n)$ is an $L\times 1$ vector containing positive elements that are not selected by the XM preprocessor of channel 1 and the rest of its elements are zero, while $\mathbf{t}_2(n)$ is an $L\times 1$ vector containing negative elements that are not selected by the XM preprocessor of channel 2 and the rest of its elements are also zero. Using vectors $\mathbf{t}_1(n)$ and $\mathbf{t}_2(n)$, we compensate for such energy loss by adding $\gamma_1(n)$ number of elements that belong to the set of the elements of $\mathbf{t}_1(n)$ to $\breve{\mathbf{x}}_1(n)$. In a similar manner to the second channel, we add $\gamma_2(n)$ number of elements that belong to the set of the elements of $\mathbf{t}_2(n)$ to $\breve{\mathbf{x}}_2(n)$. In order to maximize the tap-input energy to adaptive filters and at the same time to minimize the interchannel coherence between $\breve{\mathbf{x}}_1(n)$ and $\breve{\mathbf{x}}_2(n)$, the selected elements of the $\mathbf{t}_1(n)$ should correspond to elements having the largest values and the selected elements of the $\mathbf{t}_2(n)$ should correspond to elements having the smallest values.

Using these definitions, the tap-input vector to the NLMS algorithm $\breve{\mathbf{x}}_i(n)$ is achieved using

$$\breve{\mathbf{x}}_i(n) = \mathbf{B}_i(n)\mathbf{x}_i(n), \tag{20}$$

where i = 1, 2 is the channel index and $\mathbf{B}_i(n) = \mathrm{diag}\{\mathbf{b}_i(n)\}$ is a $L\times L$ tap-selection matrix such that elements in the $L\times 1$ vectors $\mathbf{b}_i(n)$ are given by (21) and (22) shown on the next page, given that $u, v = 1,\ 2,\ldots,L$ and $b_{1,u}(n)$ and $t_{1,u}(n)$ represent u^{th} element of $\mathbf{b}_1(n)$ and $\mathbf{t}_1(n)$ respectively and also $b_{2,v}(n)$ and $t_{2,v}(n)$ represent v^{th} element of $\mathbf{b}_2(n)$ and $\mathbf{t}_2(n)$ respectively.

It can be seen that with $\varphi = 0$, the proposed PUNL-NLMS algorithm is equivalent to XMNL-NLMS since $\gamma_1(n)$ and $\gamma_2(n)$ defined in (18) and (19) are equal to zero and hence we have $\breve{\mathbf{x}}_i(n) = \widetilde{\mathbf{x}}_i(n)$. On the other hand, with increasing φ, the effect of nonlinear preprocessing becomes more significant since elements in $\mathbf{x}_1(n)$ that are selected for generating $\widetilde{\mathbf{x}}_1(n)$ often have positive amplitudes. At the same time, elements in $\mathbf{x}_2(n)$ that are selected for generating $\widetilde{\mathbf{x}}_2(n)$ often have negative amplitudes. As we will show by simulation in the next section, through the use of φ, the amount of interchannel coherence is reduced and consequently the convergence of the adaptive filter is robust to each source location in the transmission room. In the next section we also show the effect of φ on the interchannel coherence and convergence rate of adaptive filter.

4. SIMULATION RESULTS

For evaluation purpose, we consider the specifications of the simulated environment in SAEC as follows. The dimensions of the receiving and transmission rooms were $5\times 7\times 3$ m and $6\times 5\times 4$ m respectively. The microphones were positioned at $\{4, 1.2, 1.5\}$ m and $\{5, 1.2, 1.5\}$ m in the transmission room. In addition, two loudspeakers were placed at $\{1, 6, 2.5\}$ m and $\{4, 6, 2.5\}$ m while the microphone was at $\{3, 4, 1.5\}$ m in the receiving room. We vary the position of the source starting from the front of the array centroid at coordinates $\{4.5, 1.0, 1.5\}$ m to the front of the right microphone at coordinates $\{5, 1.0, 1.5\}$ m. All the room RIRs are generated synthetically using the method of images [12] such that they are

$$b_{1,u}(n) \quad = \quad \begin{cases} 1 & p_u \in \{M-\gamma_1(n) \text{ maxima of } \mathbf{p}(n)\} \text{ or } t_{1,u}(n) \in \{\gamma_1(n) \text{ maxima of } \mathbf{t}_1(n)\} \\ 0 & \text{otherwise} \end{cases} , \tag{21}$$

$$b_{2,v}(n) \quad = \quad \begin{cases} 1 & p_v \in \{M-\gamma_2(n) \text{ minima of } \mathbf{p}(n)\} \text{ or } t_{2,v}(n) \in \{\gamma_2(n) \text{ minima of } \mathbf{t}_2(n)\} \\ 0 & \text{otherwise} \end{cases} . \tag{22}$$

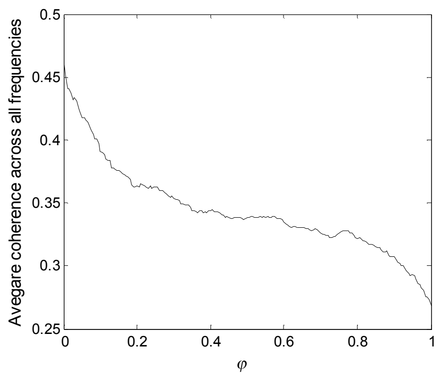

Fig. 3. Interchannel coherence versus φ when the source is at $\{4.6, 1.0, 1.5\}$ m.

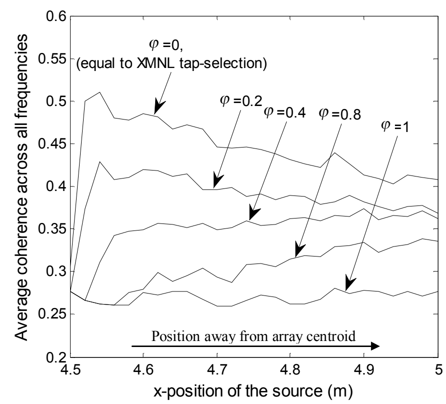

Fig. 4. Interchannel coherence versus x-position of the source.

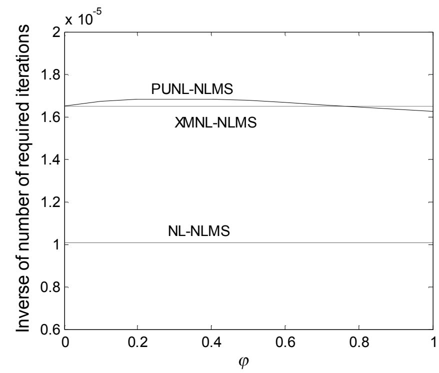

Fig. 5. Convergence rate for NL-NLMS, XMNL-NLMS, and PUNL-NLMS in terms of φ when the source is at $\{4.5, 1.0, 1.5\}$ m.

of length $L = 512$ and $L_g = 512$. We use of two kinds of input source signal, the first being a colored signal that is produced by filtering a wight Gaussian noise (WGN) through an FIR filter with coefficients [0.3574, 0.9, 0.3574] [13] and the second one is a male speech signal with sampling rate 22050 Hz. A white Gaussian noise $w(n)$ was added in all simulations to the desired signal with SNR=30 dB. We also consider $M = 0.5L$ in (7) and (8) for evaluating XMNL-NLMS throughout this paper. First we evaluate the effect of φ on interchannel coherence of the stereophonic signals. The interchannel coherence between $\mathbf{z}_1(n)$ and $\mathbf{z}_2(n)$ is defined by

$$C_{\mathbf{z}_1\mathbf{z}_2}(f) = \frac{|P_{\mathbf{z}_1\mathbf{z}_2}(f)|^2}{P_{\mathbf{z}_1\mathbf{z}_1}(f)P_{\mathbf{z}_2\mathbf{z}_2}(f)} , \tag{23}$$

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.

Figure 3 shows the mean interchannel coherence between $\mathbf{z}_1(n) = \breve{\mathbf{x}}_1(n)$ and $\mathbf{z}_2(n) = \breve{\mathbf{x}}_2(n)$ across different frequencies when the source is at coordinates $\{4.6, 1.0, 1.5\}$ m. In this example $\breve{\mathbf{x}}_1(n)$ and $\breve{\mathbf{x}}_2(n)$ are each of length 2000 samples containing the colored signal as described before. In addition, Fig. 4 shows the mean interchannel coherence across different frequencies for various source positions ranging from $\{4.5, 1.0, 1.5\}$ m to $\{5, 1.0, 1.5\}$ m for five different values of φ, using the above mentioned colored signal. As can be seen from Figs. 3 and 4, the coherence reduces with increasing φ. Note also that when the source is in the middle of the microphone pair at $\{4.5, 1.0, 1.5\}$ m, as shown in Fig. 4, the coherence is low and does not vary with φ. For this case, the total number of non-zero elements in $\widetilde{\mathbf{x}}_{1,n}(n)$ and $\widetilde{\mathbf{x}}_{2,p}(n)$ is negligible in comparison to the total number of $\widetilde{\mathbf{x}}_1(n)$ and $\widetilde{\mathbf{x}}_2(n)$ respectively. More importantly, we note that when the source is away from the centroid of the microphone pair, the number of non-zero elements of $\{\widetilde{\mathbf{x}}_{1,n}(n), \widetilde{\mathbf{x}}_{2,p}(n)\}$ is high and XM tap-selection method reduces the interchannel coherence. In such situations our proposed partial update procedure compensate the drawback of the XM tap-selection method giving a lower average interchannel coherence as shown in Fig. 4.

In order to quantify the convergence rate of adaptive filters in SAEC, we use the normalized misalignment as

$$\eta(n) = \frac{\|\widehat{\mathbf{h}}(n) - \mathbf{h}\|^2}{\|\mathbf{h}\|^2} . \tag{24}$$

We set the step-size for each algorithm such that the algorithms reach the same steady-state. For this we consider the step-size for NL-NLMS, XMNL-NLMS and the proposed PUNL-NLMS algorithm to be equal to 0.8, 0.58 and 0.62 respectively. We then define quantity T_{-10} as the inverse of the number of iteration required for normalized misalignment to reach -10 dB. Figures 5, 6, 7 show the convergence rate of NL-NLMS, XMNL-NLMS, and PUNL-NLMS corresponding to three source positions $\{4.5, 1.0, 1.5\}$ m, $\{4.6, 1.0, 1.5\}$ m and $\{5, 1.0, 1.5\}$ m with respect to φ.

As can be seen from Fig. 5, the convergence of XMNL-NLMS and the proposed PUNL-NLMS are comparable and are considerably higher than NL-NLMS. Figures 6 and 7 show that when the source is away from the centroid of microphone pair, the convergence of XMNL-NLMS reduces and that this degradation is significant when the source is at $\{5, 1.0, 1.5\}$ m, as shown in Fig. 7. Based on these three figures, we can see that in general for $\varphi >> 0$, we have a better convergence behavior for the proposed PUNL-NLMS

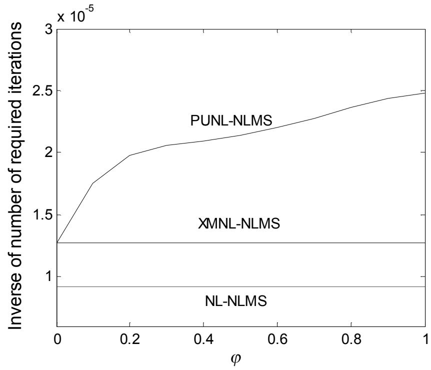

Fig. 6. Convergence rate for NL-NLMS, XMNL-NLMS, and PUNL-NLMS in terms of φ when the source is at $\{4.6, 1.0, 1.5\}$ m.

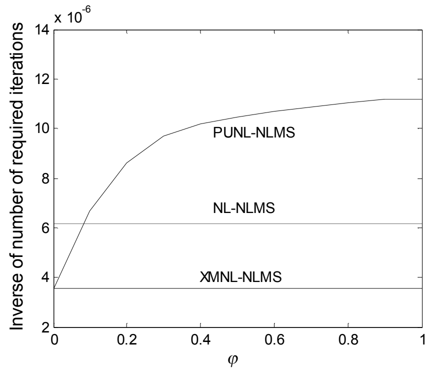

Fig. 7. Convergence rate for NL-NLMS, XMNL-NLMS, and PUNL-NLMS in terms of φ when the source is at $\{5, 1.0, 1.5\}$ m.

algorithm compared to both NL-NLMS and XMNL-NLMS and the best behavior occurs when $\varphi \approx 1$. Figures 8 and 9 compare the convergence rate of three algorithms for speech signals when the source is at $\{4.5, 1.0, 1.5\}$ m and $\{5, 1.0, 1.5\}$ m for $\varphi = 1$. As can be seen, our proposed PUNL-NLMS algorithm achieves a higher convergence rate for both stationary and non-stationary source signals achieving nearly 6 dB improvement of normalized misalignment over XMNL-NLMS when the source is in front of the right microphone.

5. CONCLUSION

In this paper, we presented a new approach for improving the convergence rate of adaptive filters for SAEC. This approach is a partial update method that selects taps with a criterion such that it maintains the interchannel coherence to a minimum level for various source positions in the transmission room. The advantage of the proposed approach is verified with simulation using the NLMS algorithm.

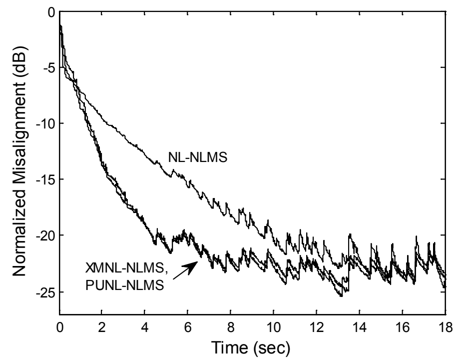

Fig. 8. Normalized misalignment for NL-NLMS, XMNL-NLMS, and PUNL-NLMS with φ=1 when the source is at $\{4.5, 1.0, 1.5\}$ m.

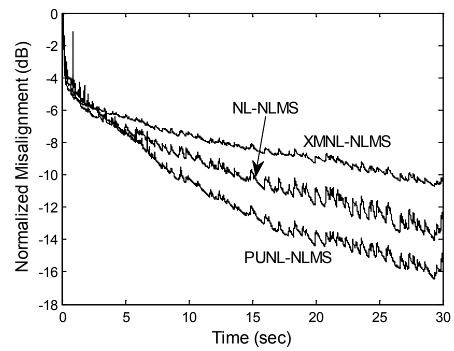

Fig. 9. Normalized misalignment for NL-NLMS, XMNL-NLMS, and PUNL-NLMS with φ=1 when the source is at $\{5, 1.0, 1.5\}$ m.

6. REFERENCES

[1] 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, pp. 156 –165, 1998.

[2] S. Shimauchi and S. Makino, “Stereo projection echo canceller with true echo path estimation,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing,*, 1995.

[3] J. Benesty, D. R. Morgan, J. L. Hall, and M. M. Sondhi, “Stereophonic acoustic echo cancellation using nonlinear transformations and comb filtering,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1998.

[4] ——, “Synthesized stereo combined with acoustic echo cancellation for desktop conferencing,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1999.

[5] 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, pp. 686 – 696, 2001.

[6] 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 *International Conference on Digital Signal Processing*, 2002.

[7] A. Sugiyama, Y. Joncour, and A. Hirano, “A stereo echo canceler with correct echo-path identification based on an input-sliding technique,” *IEEE Trans Signal Process*, vol. 49, p. 25772587, 2001.

[8] P. Surin, N. Tangsangiumvisai, and S. Aramvith, “An adaptive noise decorrelation technique for stereophonic acoustic echo cancellation,” in *IEEE Tencon*, 2004.

[9] 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 Processing*, vol. 86, pp. 1157 –1167, 2006.

[10] K. Mayyas, “Stereophonic acoustic echo cancellation using lattice orthogonalization,” *IEEE Trans. Speech Audio Process*, vol. 10, pp. 517 –525, 2002.

[11] A. W. H. Khong and P. A. Naylor, “Stereophonic acoustic echo cancellation employing selective-tap adaptive algorithms,” *IEEE Trans. Speech Audio Process*, vol. 14, pp. 785–796, 2006.

[12] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Amer*, vol. 65, pp. 943 – 950, 1979.

[13] S. Attallah, “The wavelet transform-domain LMS adaptive filter with partial subband-coefficient updating,” *IEEE Trans. Circuits Sys. II: Express briefs*, vol. 53, pp. 8– 12, 2006.