IJICT International Journal of Information & Communication Technology

ITRC

Volume 1- Number 1- March 2009 (pp. 29-37)

A Modified Wavelet-domain Adaptive Filtering Scheme for Stereophonic Acoustic Echo Cancellation in the Teleconferencing Application

Mehdi Bekrani
Department of Electrical Engineering
Tarbiat Modares University
Tehran, Iran
bekrani@modares.ac.ir

Mojtaba Lotfizad
Department of Electrical Engineering
Tarbiat Modares University
Tehran, Iran
lotfizad@modares.ac.ir

Abstract— **Stereophonic acoustic echo cancellers are being used in high quality audio communication systems to reduce the echoes that arise from coupling between the loudspeakers and microphones. These echo cancellers have problems more severe than their monophonic counterparts. The most important problem is the mismatch between the impulse responses of adaptive filters and those of the acoustic paths of the receiving section and therefore a severe divergence of filters in the event of abrupt changes of the transmission section acoustic paths. The main reason for such mismatching is the two-channel nature of the system and strong cross-correlation between these two channels. In this paper, with inspiration from the lattice predictor structure, we present a novel approach for reduction of the interchannel cross-correlation of the input vectors in the wavelet domain that leads to faster reduction of the misalignment error.**

Keywords- stereophonic acoustic echo cancellation; wavelet transform; lattice structure; misalignment; correlation.

I. INTRODUCTION

Today there is an increasing interest for employing the stereophonic systems to deliver the audio information and establish a good speech communication, because a stereophonic audio system gives a 3-dimentional sense of the sound, leading to a better understanding of it. Examples of such systems are teleconferencing, virtual reality (such as E-learning), multimedia workstations, televideo gaming, hands-free telephones, etc. One of the problems that should be solved in such systems is the suppression of the stereo acoustic echo.

Stereophonic acoustic echo cancellation (SAEC) has issues that make it more difficult to overcome in comparison to the monophonic case. The most important problem is the mismatch between the impulse responses of the adaptive filters and those of the acoustic paths of the receiving room, even after convergence. This leads to a dependency of estimated impulse response to acoustic paths of the transmission section; namely the paths from the human-speakers to the microphones.

Therefore, in the event of a sudden change in these paths (it is a common event in multi talker situations), the filter coefficients should be changed to cancel the error (echo). However, in most situations these filters cannot track rapid changes and as a result part of or the whole echo remains within a time span and thus degrades the sound quality. The rapid changes occur when the talker changes his/her location fast or one talker finishes his/her talking and another person starts to talk from a different location in the same room.

The existence of a high correlation between the two channels causes a considerable increase to the misalignment. Therefore, finding some ways to decorrelating these two signals without any degradation in the sound quality may solve the misalignment problem to some extent.

So far, there has been proposed a variety of methods for reduction of the misalignment based on reducing the interchannel coherence. In general, it is possible to divide the approaches proposed for reducing the correlation between the two channels into two categories; one of these methods that performs a preprocessing on the input stereo signals, such as , such as adding or modulating little quantities of an independent noises to each channel [1], [2], comb filtering [3], frequency shifting of one channel relative to another [4], adding a nonlinear function of the signal of each channel to that same channel [5], [6], time-variable all-pass filtering of stereo signals [7] and input-sliding [8] ,[9]. The most famous of which is adding a small nonlinearity to each channel [5]. Although the above mentioned approaches give rise to improvements in the misalignment behavior, they have a key limitation in performance, that is because they act on the signals that are directly sent to the listener in the receiving section, and hence a high degree of decorrelation can easily degrade the stereophonic perception or quality of the stereo sound [10], [11]. Recently some authors have proposed a number of approaches utilizing properties of human auditory system, such as [12], [13]; however they have a high computational complexity.

Another category includes methods that try to decorrelate the stereo signals in that branch of the input paths to the adaptive filters, in order to improve the convergence behavior of misalignment. Khasawneh, et al. [14] used a combination of time-domain and frequency-domain transforms in adaptive filters, namely DCT transform and the gradient adaptive lattice algorithm [15]. This procedure reduces the eigenvalue spread of the cross-correlation matrix, thereby leading to a better convergence compared to its pure counterparts, namely GAL, DCT, and NLMS. In [16] the authors proposed a stereo canceller based on Gabor expansion and showed that the performance of the canceller is improved because of the whitening effect as well as reducing the cross-correlation of this expansion, and can represents a better performance compared with the SAEC with nonlinear preprocessing [5]. In [17] a hybrid mono/stereo structure between fast affine projection (FAP) and frequency domain normalized least mean square (F-NLMS) algorithms is introduced, based on subband processing. This technique indicates improved performance in terms of convergence rate and computational complexity reduction, compared to the fullband structure.

In [18] it was proposed that the updating of adaptive coefficients may be accomplished based on a changed form of the input; namely, adding two uncorrelated and independent signals to the input channels and then increasing the contribution of these signals by a multiplying factor at the inputs of the two adaptive filters. This process leads to an increase in the convergence rate of the algorithm and a better decrease in the misalignment. Khong et al. [19] introduced a tap-selection based algorithm as a means to reduce the interchannel coherence and thus improving the condition of input correlation matrix in SAEC. In this algorithm, in each update, only those weights are selected that the total energy of corresponding taps is maximum whereas the coherence between the two channels is minimum. This algorithm has a better convergence of misalignment with respect to traditional algorithms with preprocessing [5].

Recently we have proposed a wavelet thresholding approach for signal decorrelation [20]; it could increase the convergence rate compared to the conventional wavelet transform based adaptive filters, but at the same time, the final level of MSE error and consequently final level of misalignment would increase. Mayyas [21] has more recently proposed a selective coefficient update rule for LMS-type adaptive algorithms with a low complexity, but it has the same problem of [20], that is the increasing in the final MSE.

It should be noted that in practice, the perfect alignment is not possible due to the finite length of impulse responses of the modeling filters (the adaptive filters) as well as due to the noise present in the signal [7].

In this paper, we introduce a novel adaptive filtering approach to reduce the auto- and cross-correlation, based on a combination of channels information in the wavelet transform domain. Using the wavelet transform (or more generally, orthogonal transforms, such as DCT, DFT, etc.) in adaptive filtering leads to a better convergence property [15]. Among the applicable transform domain adaptive filtering, the discrete wavelet transform (DWT), a powerful tool for analysis of non-stationary signals, has recently received considerable attention in adaptive filtering [22]. This is because of its two very good properties: the time and spectral localization [23]. The former can result in a computational cost reduction, and the latter can lead to the input decorrelation so as to speed up the convergence.

The paper is organized as follows: Section II introduces the SAEC system and the misalignment problem. We will consider in our discussion the teleconferencing application. In Section III, we propose a novel method for improving the convergence behavior of misalignment. Section IV presents simulation results, and finally Section V draws conclusions from our work.

II. Stereophonic Acoustic Echo Cancellation in Teleconferencing Application

Fig. 1 shows the stereo echo canceller in the teleconferencing application. Here, for simplicity, we consider only one microphone at the receiving room, since a similar analysis can be applied to the other channel.

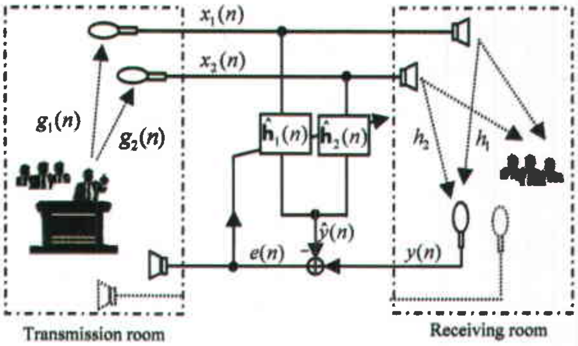

Figure 1. The stereophonic acoustic echo cancellation in the teleconferencing system

An actual setup of transmission section of teleconferencing system is brought in Fig. 2.

For analysis behavior of the system, we could see from Fig. 1 that firstly, the microphones in the transmission room receive the signal produced by source $s(n)$ (e.g. speech from a human-speaker) due to acoustic paths represented by impulse responses $g_1(n)$ and $g_2(n)$, thereby creating $x_1(n)$ and $x_2(n)$. These signals are then fed to the speakers in the receiving room and are then transmitted to the microphone through the acoustic paths h_1 and h_2, and make the echo signal $y(n)$ in the microphone, which is given by

$$y(n) = \mathbf{h}_1^T(n)\mathbf{x}_1(n) + \mathbf{h}_2^T(n)\mathbf{x}_2(n), \qquad (1)$$

where $\mathbf{h}_i^T(\mathrm{n}) = [h_{i,0}(\mathrm{n}) \quad h_{i,1}(\mathrm{n}) \cdots h_{i,L-1}(\mathrm{n})]^{\mathrm{T}}$, and $\mathbf{x}_i^T(\mathrm{n}) = [x_i(\mathrm{n}) \quad x_i(\mathrm{n-1}) \cdots x_i(\mathrm{n-L+1})]^{\mathrm{T}}$.

On the other hand, the adaptive filters try to simulate h_1 and h_2 so as to make a copy of the echo signal in order to reduce the acoustic feedback as much as possible. The error signal between the echo signal and its estimate is given by

$$e(n) = y(n) - [\hat{\mathbf{h}}_1^T(n)\mathbf{x}_1(n) + \hat{\mathbf{h}}_2^T(n)\mathbf{x}_2(n)], \qquad (2)$$

where $\hat{\mathbf{h}}_i^T(n)$, $i = 1, 2$, is the vector of adaptive filter coefficients for channel i.

Figure 2. A typical setup of the transmission section of the teleconferencing system

The aim of the adaptive system is to minimize the echo or error $e(n)$. However, in practice it is not possible to align the impulse responses of adaptive filters to h_1 and h_2. In fact, it is shown in [5] that when the length of the impulse response of the transmission room (M) is less than or equal to that of the adaptive filters (L), the solution of the adaptive filters coefficients is in the form

$$\begin{bmatrix} \hat{\mathbf{h}}_1(\mathrm{n}) \\ \hat{\mathbf{h}}_2(\mathrm{n}) \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1(\mathrm{n}) \\ \mathbf{h}_2(\mathrm{n}) \end{bmatrix} + \gamma(n) \begin{bmatrix} \mathrm{g}_2(\mathrm{n}) \\ -\mathrm{g}_1(\mathrm{n}) \end{bmatrix}, \qquad (3)$$

where $\gamma(n)$ is a scalar quantity. Equation (3) indicates that the solutions for $\hat{\mathbf{h}}_i^T(n)$ are nonunique. In practical cases where $L < M$, the joint–input correlation matrix is ill-conditioned because the input signals $\mathbf{x}_1(n)$ and $\mathbf{x}_2(n)$ are highly correlated [5]. This, in turn, reduces the convergence of weights, as well as the MSE, considerably. The above problem is referred to in literatures as the "misalignment" problem.

III. The Proposed Method

In this section we present our modified wavelet-domain adaptive filtering scheme which could reduce the eigenvalue spread of the joint-input correlation matrix. In a conventional transform domain adaptive algorithm the input signal vector is transformed by an orthogonal transform before being used in the adaptive algorithm. The transformation aims at reducing the eigenvalue spread of the transformed signal correlation matrix, thereby improving the convergence behavior of the time-domain algorithm in the transform domain [24].

In a two-channel adaptive filtering scheme, used for SAEC, the joint-input correlation matrix in addition to having elements containing the cross-correlation, consists of the elements containing the auto correlation of each channel x_1 and x_2. Therefore to improve the convergence behavior of a two-channel adaptive filter, it is preferable to not only reduce the interchannel cross-correlation (thereby substantially increasing the weight convergence rate of adaptive filters), but also reduce the autocorrelation of each individual channel (i.e., equivalent to a whitening procedure). In this section, we propose a novel scheme for adaptive filtering which has the ability of the above both cases. It employs the lattice structure, in conjunction with an orthogonal wavelet transform that could reduce the auto correlation of each channel as well as the interchannel cross-correlation.

We assume that the effective length of the impulse response of the acoustic path from the microphone to the loudspeaker in the transmission room is L. Hence we have (Proof in Appendix1)

$$x_1(n) = \sum_{k=0}^{L-1} \alpha_k(n).x_2(n-k) + e_1(n), \qquad (4)$$

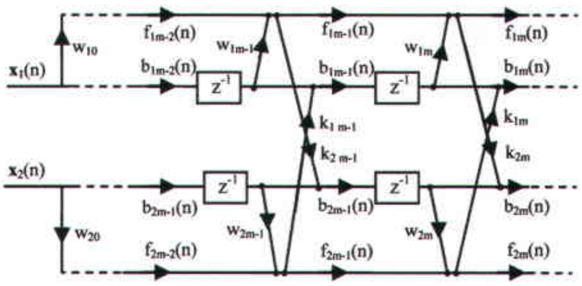

Figure 3. The proposed structure for reducing the cross-correlation

and

$$x_2(n)=\sum_{k=0}^{L-1}\beta_k(n).x_1(n-k)+e_2(n), \qquad (5)$$

where $\alpha_k(n)$ and $\beta_k(n)$ are coefficients depending on acoustic paths $g_1(n)$ and $g_2(n)$, and $e_1(n)$ and $e_2(n)$ are considered as the estimation error. The values of all these parameters are calculated in appendix1.

With respect to (4) and (5) we can say that $\alpha_k(n)$ and $\beta_k(n)$ are the main cause of the cross-correlation and therefore we may consider a model for the signals of the two channels in which the present sample of $x_1(n)$ can be obtained from the present and past samples of $x_2(n)$ and in the same manner the present sample of $x_2(n)$ can be obtained from the present and past samples of $x_1(n)$. Therefore, with an intuitive approach adopted from the lattice structures for autoregressive models [15], we propose the structure shown in Fig. 3 for the cross-correlation reduction.

In this structure, $m=0,1,\cdots,L-1$ and w_{1m}, w_{2m} are coefficients of the wavelet analysis filters which are of constant values. k_{im} are coefficients aimed to reduce the cross-correlation. This structure is a subband of several bands associated with wavelet transform that are combined with a lattice structure.

The set of these structures constitutes the decorrelation part of the proposed wavelet transform-based adaptive filter.

Hence, based on this structure we have

$$\begin{aligned}
b_{1m}(n)&=b_{1m-1}(n-1)+k_{1m}.f_{2m}(n)\\
b_{2m}(n)&=b_{2m-1}(n-1)+k_{2m}.f_{1m}(n)\\
f_{1m}(n)&=f_{1m-1}(n)+w_{1m}.b_{1m-1}(n-1)\\
f_{2m}(n)&=f_{2m-1}(n)+w_{2m}.b_{2m-1}(n-1).
\end{aligned} \qquad (6)$$

In order to reduce the correlation, the following cost functions $J_1(n)$ and $J_2(n)$ are proposed for channel 1 and channel 2 respectively as

$$\begin{aligned}
J_1(n)&=E\left[b_{1m}^2(n)\right],\\
J_2(n)&=E\left[b_{2m}^2(n)\right].
\end{aligned} \qquad (7)$$

The optimal values of k_{1m} and k_{2m} are obtained by minimizing $J_1(n)$ and $J_2(n)$ with respect to k_{1m} and k_{2m} respectively as (Appendix 2)

$$\Rightarrow k_{1m,opt}=-\frac{E\left[b_{1m-1}(n-1).f_{2m}(n)\right]}{E\left[(f_{2m}(n))^2\right]}. \qquad (8)$$

In the same manner

$$k_{2m,opt}=-\frac{E\left[b_{2m-1}(n-1).f_{1m}(n)\right]}{E\left[(f_{1m}(n))^2\right]}. \qquad (9)$$

With regard to the fact that in practice, we usually deal with nonstationary environments and also for reducing the computational complexity, it would be better to recursively calculate k_{im}, i=1, 2 [25, page 382].

Hence, these coefficients may be calculated with a gradient method, like the NLMS as (Proof in Appendix 3)

$$k_{1m}(n+1)=k_{1m}(n)-\mu_k\cdot\frac{2f_{2m}(n).b_{1m}(n)}{p_1(n)+\delta}, \qquad (10)$$

and

$$k_{2m}(n+1)=k_{2m}(n)-\mu_k\cdot\frac{2f_{1m}(n).b_{2m}(n)}{p_2(n)+\delta}. \qquad (11)$$

where $p_1(n)$ and $p_2(n)$ are used for normalizing the step size, μ_k to ensure fast convergence and are defined in appendix 3.

After the transformation, we perform the wavelet-domain adaptive filtering. For updating the weights, we use the following equation that was first introduced in [26] as a Wavelet-based adaptive filtering operation for system identification which we will extend to the two-channel system identification case,

$$\hat{\mathbf{h}}_{wj}(n+1)=\hat{\mathbf{h}}_{wj}(n)+\frac{\mu}{\sigma_{wj}^2(n)}e(n)\,\mathbf{z}_{wj}(n). \qquad (12)$$

In this equation, $j=0,1,\cdots M-1$, represents the number of wavelet subband, $\mathbf{z}_{wj}(n)$ is the wavelet coefficients vector, and $\sigma_{wj}^2(n)$ is the power estimation of j th subband signal which can be computed with a recursive method as:

$$\sigma^2_{w\,j}(n) = \rho\,\sigma^2_{w\,j}(n-1) + (1-\rho)\left\|\mathbf{z}_{w\,j}(n)\right\|^2, \quad (13)$$

where ρ is a positive constant less than or equal to unity[26].

The error signal is as follows,

$$e(n) = y(n) - \hat{y}(n). \quad (14)$$

In the two-channel case, we can extend (12) for each channel as

$$\hat{\mathbf{h}}_{w1\,j}(n+1) = \hat{\mathbf{h}}_{w1\,j}(n) + \frac{\mu}{\sigma^2_{w1\,j}(n) + \sigma^2_{w2\,j}(n)} e(n)\mathbf{z}_{w1\,j}(n), \quad (15)$$

and

$$\hat{\mathbf{h}}_{w2\,j}(n+1) = \hat{\mathbf{h}}_{w2\,j}(n) + \frac{\mu}{\sigma^2_{w1\,j}(n) + \sigma^2_{w2\,j}(n)} e(n)\mathbf{z}_{w2\,j}(n), \quad (16)$$

where indices 1 and 2 refer to channels 1 and 2 respectively, and $\mathbf{z}_{wi\,j}(n)$ is the modified wavelet coefficients vector in channel i and subband j.

A schematic diagram of the proposed structure is shown in Fig. 4.

In the structure shown in Fig. 4, the input signals to the branches toward the adaptive filters are first transformed into signals with a lower correlation and then are used by the weight update algorithm for computing the coefficients of the adaptive filters.

IV. Simulation Results

For evaluating the performance of the decorrelation part (Fig.3), three different types of signals, white and colored with different correlation values were considered. Then we obtained the eigenvalue spread of the output correlation matrix and compared with that of the input. For producing the correlated signals, two independent white Gaussian noise (WGN) were generated for each run and then the test (white) signals were made as $x_1(n) = WGN_1$ and $x_2(n) = \xi.WGN_1 + (1-\xi).WGN_2$. Colored signals were also produced by passing the signals WGN_1 and WGN_2 through a FIR filter that produce signals with a speech-like spectrum. The correlation matrix associated to the signals of the two channels was calculated as [27],

$$C_{X1,X2}(n) = E\left\{\begin{bmatrix}\mathbf{X}_1(n)\\ \mathbf{X}_2(n)\end{bmatrix}\begin{bmatrix}\mathbf{X}_1^T(n) & \mathbf{X}_2^T(n)\end{bmatrix}\right\}. \quad (17)$$

Table 1 shows an example of the eigenvalue spread ($\chi = \lambda_{\max}/\lambda_{\min}$) for the two-channel input correlation matrix and also for the corresponding matrix for the output transformed signal of the decor-

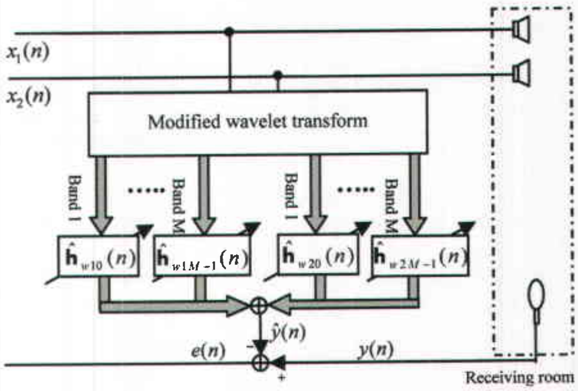

Figure 4. the schematic diagram of proposed structure

relating part of the proposed structure and a pure wavelet transform.

Here we employed the Daubechies wavelet family [28], which is a commonly used wavelet because of its orthogonality, compact support in time, and a high degree of regularity [29]. In addition, we used of these wavelet filters in a uniform subband tree structure in all simulations of this paper, because of its very good results in terms of decorrelation for such a class of signals [30].

As can be seen, the eigenvalue spread in our method is less than that of the wavelet transform and also than the initial data, especially in the case of strong correlation. In order to simulate this approach in the teleconferencing system, the impulse responses g_1, g_2 and h_1, h_2 were generated using the method of images[31].

Specifications of the simulated environment:

a) Transmission room:

Dimensions: $3m \times 4m \times 4m$

Coordinates of Mic1: $(1, 2, 1.5)$meter,

Mic2: $(2, 1.8, 1.5)$meter, Source: $(1, 1, 1.5)$meter,

Length of (truncated) g_1 , g_2: 256 points.

b) Receiving room:

Dimensions: $3.5m \times 4m \times 3.3m$

Coordinates of Mic: $(0.5, 1, 2)$meter,

Loudspeaker 1: $(3, 2, 1.5)$ meter,

Loudspeaker 2: $(2.8, 3.2, 1.2)$meter,

Length of (truncated) h_1 , h_2: 64 points.

TABLE I. Eigenvalue spread for vectors with length L=8 and using Daubechies 2 (db2)

Signal type	ξ	$C_{X1,X2}(n)$ *input*	$C_{Z1,Z2}(n)$ *wavelet transform*	$C_{U1,U2}(n)$ *proposed transform*
white	0.5	10.3166	5.8761	4.5843
	0.8	80.1678	7.7219	5.8897
speech-like	0.5	156.4022	12.2193	10.0586
	0.8	277.5251	18.3577	13.6433

For evaluation purpose, we have compared our proposed algorithm with the standard two-channel NLMS algorithm and two recently proposed algorithms (having the same configuration) namely, the selective-tap LMS [19] and DCT-GAL LMS [14] as well as the Wavelet-based LMS (WTLMS) [26] which we extended to the stereo case. In the selective-tap LMS algorithm, based on a criterion related to the convergence rate of misalignment, only those taps are selected for weight update that reduce the cross-correlation of the two input signals in the path of adaptive filters as much as possible and at the same time, the sum of the input energies to the weights is not decreased as much. The mentioned criterion is obtained with regard to the relation of the misalignment to the two factors of the cross-correlation and the total energy of the input vectors of the channels. In the DCT-GAL algorithm, in one channel, the gradient adaptive lattice is employed and the other channel uses the DCT transform.

This scheme causes a considerable reduction of the eigenvalue spread, thereby resulting to increasing in the speed of convergence.

In our simulations, the step size for each algorithm was chosen such that the maximum speed of convergence is achieved.

The chosen wavelet analysis scheme has a dyadic structure with 3 levels, each level with 2 band decomposition, based on Daubechies 2 (db2). In each level, both the HPF and LPF branches are decomposed; this is equivalent to a symmetrically decomposed wavelet packet. Assuming the length of h_1 and h_2 to be 64, we have 8 output subbands. The microphone signal of the receiving room is a stationary signal of type speech-like and is mixed with a white noise with SNR=45dB. The misalignment diagram for the colored input signal is shown in Fig. 5. This plot is obtained by averaging over 40 independent runs. It is obvious that the misalignment reduction rate of our method is better with respect to its counterparts.

The diagram of the MSE error for the same colored input signal is plotted in Fig. 6. This plot is also obtained with 40 runs. The curves are smoothed with a Gaussian window with a length of 60. It is evident that the proposed method has a better MSE convergence rate compared to the other algorithms.

Evaluation of the effect of the talker change:

As mentioned above, the misalignment leads to a dependence of the impulse responses of the two adaptive filters ($\hat{h}_1$, $\hat{h}_2$) on the acoustic paths of the transmitting room (g_1, g_2). With regard to the misalignment curves, we can find out that, the more the level of the misalignment (at different times) is decreased, the less would be the sensitivity of the adaptive filters to the changes in paths g_1 and g_2.

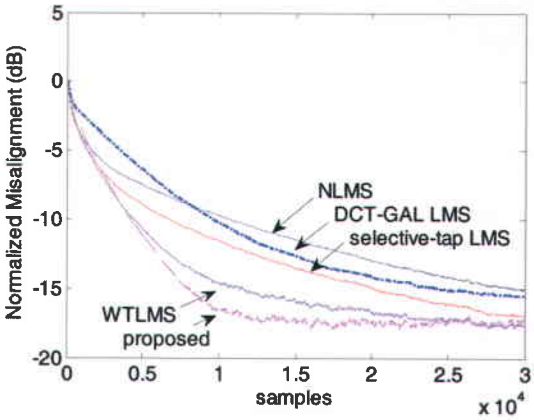

Figure 5. Comparison of the misalignment of various adaptive algorithms for a speech-like input signal.

In order to show the sensitivity of adaptive filters to such changes, we arrange an experiment as follows and evaluate the proposed, as well as the conventional NLMS-based stereophonic acoustic echo canceller.

In this experiment, the sound source model that was initially placed at the coordinates $(1,1,1.5)$ meter, in the transmitting room is moved to new coordinates $(1.8,1.5,1.6)$ meter, after passing 20000 samples of the input source signal, that is actually equal to the talker change at a particular time. Fig. 7 shows the amount of divergence of the adaptive filters (at the instant of the talker change) for the standard NLMS compared to our proposed approach, based on the MSE criterion.

As the above mentioned figure shows at the time of the talker change, the collective output error of the adaptive filters (that is, the same as the echo signal) is increased abruptly which is very high for the NLMS algorithm and a sufficiently long period of time is required for the adaptive filters to converge again and hence minimize the output error. During this time interval, the residual echo gets reflected in the transmitting room and thus degrades the quality of sound. As it can be seen, such residual echo is much less in our proposed approach.

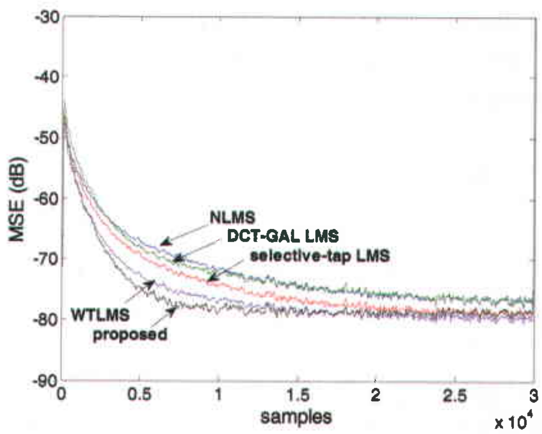

Figure 6. Comparison of the MSE of various adaptive algorithms for a speech-like input.

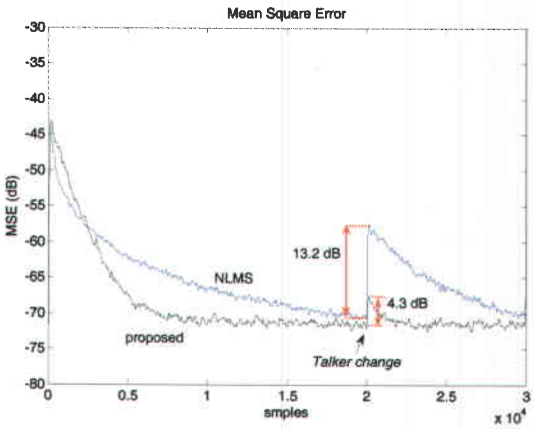

Figure 7. MSE Behavior of proposed and standard echo canceller with talker change.

V. CONCLUSION

In this paper, we have presented an approach for reduction of the misalignment problem of adaptive filters in stereophonic acoustic echo cancellers. This approach improves the convergence behavior of misalignment and MSE of adaptive filters in comparison to its counterparts. We evaluated this approach in the teleconferencing application and showed performance enhancement resulted by our proposed approach.

APPENDIX 1

Assume that the effective length of the impulse response of the acoustic path from the microphone to the loudspeaker in the transmission room is L. Hence we have:

$$x_1(n) = s(n) * g_1(n) = \sum_{k=0}^{L-1} s(n-k) g_{1k}(n), \quad (18)$$

$$x_2(n) = s(n) * g_2(n) = \sum_{k=0}^{L-1} s(n-k) g_{2k}(n), \quad (19)$$

where $g_{1k}(n)$ and $g_{2k}(n)$ are the k th elements of $g_1(n)$ and $g_2(n)$ respectively. Thus we can write in Z-transform:

$$X_1(z) = S(z) \cdot G_1(z), \quad (20)$$

$$X_2(z) = S(z) \cdot G_2(z). \quad (21)$$

So

$$\frac{X_1(z)}{X_2(z)} = \frac{G_1(z)}{G_2(z)} \Rightarrow X_1(z) \cdot G_2(z) = X_2(z) \cdot G_1(z) \quad (22)$$

$$\Rightarrow x_1(n) * g_2(n) = x_2(n) * g_1(n)$$

$$\Rightarrow \sum_{k=0}^{L-1} x_1(n-k) g_{2k}(n) = \sum_{k=0}^{L-1} x_2(n-k) g_{1k}(n). \quad (23)$$

Hence, there is a linear relation between $x_1(n)$ and $x_2(n)$ and we can change it to:

$$x_1(n) g_{20}(n) + \sum_{k=1}^{L-1} x_1(n-k) g_{2k}(n) = \sum_{k=0}^{L-1} x_2(n-k) g_{1k}(n), \quad (24)$$

$$x_2(n) g_{10}(n) + \sum_{k=1}^{L-1} x_2(n-k) g_{1k}(n) = \sum_{k=0}^{L-1} x_1(n-k) g_{2k}(n). \quad (25)$$

Hence, we have

$$x_1(n) = \sum_{k=0}^{L-1} \alpha_k(n) . x_2(n-k) + e_1(n), \quad (26)$$

and

$$x_2(n) = \sum_{k=0}^{L-1} \beta_k(n) . x_1(n-k) + e_2(n), \quad (27)$$

where

$$\alpha_k(n) = g_{1k}(n) / g_{20}(n), \quad (28)$$

$$\beta_k(n) = g_{2k}(n) / g_{10}(n), \quad k = 0, 1, \cdots, L-1, \quad (29)$$

and

$$e_1(n) = -\sum_{k=1}^{L-1} x_1(n-k) g_{2k}(n) / g_{20}(n), \quad (30)$$

and

$$e_2(n) = -\sum_{k=1}^{L-1} x_2(n-k) g_{1k}(n) / g_{10}(n). \quad (31)$$

So, in (26) and (27), we have an estimator that can estimate present samples of $x_1(n)$ and $x_2(n)$ with estimation errors $e_1(n)$ and $e_2(n)$.

APPENDIX 2

$$\frac{\partial J_1(n)}{\partial k_{1m}} = 2E\left[b_{1m-1}(n-1).\left(f_{2m-1}(n) + w_{2m}.b_{2m-1}(n-1)\right)\right]$$
$$+ 2k_{1m}E\left[\left(f_{2m-1}(n) + w_{2m}.b_{2m-1}(n-1)\right)^2\right] = 0 \tag{32}$$

$$\Rightarrow k_{1m,opt} = -\frac{E\left[b_{1m-1}(n-1).\left(f_{2m-1}(n) + w_{2m}.b_{2m-1}(n-1)\right)\right]}{E\left[\left(f_{2m-1}(n) + w_{2m}.b_{2m-1}(n-1)\right)^2\right]}$$
$$= -\frac{E\left[b_{1m-1}(n-1).f_{2m}(n)\right]}{E\left[\left(f_{2m}(n)\right)^2\right]}. \tag{33}$$

APPENDIX 3

$$k_{im}(n+1) = k_{im}(n) - \frac{\mu_k}{p_{im}(n) + \delta} . \frac{\partial J_i(n)}{\partial k_{im}}, \tag{34}$$

where μ_k is the (unnormalized) step size and δ is a small constant to avoid zeroing the denominator. The cost function is calculated approximately with $J_i(n) \approx b_{im}^2(n)$ and $\partial J_i(n)/\partial k_{im}$ is the slope of the surface of the cost function with respect to k_{im}. For ensuring a fast convergence, the step size is normalized to the estimated power of the signal inputting to the m th stage, $p_{im}(n)$, [25]. Ideally $p_1(n) = E\left[\left(f_{2m}(n)\right)^2\right]$ and $p_2(n) = E\left[\left(f_{1m}(n)\right)^2\right]$ and for practical computation, we can use the following recursive computation

$$p_1(n) = \lambda.p_1(n-1) + (1-\lambda) * \left(f_{2m}(n)\right)^2,$$
$$p_2(n) = \lambda.p_2(n-1) + (1-\lambda) * \left(f_{1m}(n)\right)^2, \tag{35}$$

where λ is a constant between 0 and 1.
Now, with regard to the above estimated cost function, we have

$$\frac{\partial J_1(n)}{\partial k_{1m}} = 2 f_{2m}(n).\left[b_{1m-1}(n-1) + k_{1m}.f_{2m}(n)\right] \tag{36}$$
$$= 2 f_{2m}(n).b_{1m}(n).$$

Similarly

$$\frac{\partial J_2(n)}{\partial k_{2m}} = 2 f_{1m}(n).b_{2m}(n). \tag{37}$$

Thus, we have

$$k_{1m}(n+1) = k_{1m}(n) - \mu_k . \frac{2 f_{2m}(n).b_{1m}(n)}{p_1(n) + \delta}, \tag{38}$$

and

$$k_{2m}(n+1) = k_{2m}(n) - \mu_k . \frac{2 f_{1m}(n).b_{2m}(n)}{p_2(n) + \delta}. \tag{39}$$

REFERENCES

[1] M. M. Sondhi, D. R. Morgan, "Acoustic echo cancellation for stereophonic teleconferencing," IEEE Workshop on Applications of Signal Processing to Audio Acoustics, pp. 141-142, Oct. 1991.

[2] S. Shimauchi and S. Makino, "Stereo projection echo canceller with true echo path estimation," IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.5, pp. 3059–3062, May 1995.

[3] J. Benesty, D. R. Morgan, J. L. Hall, M. Mohan Sondhi, "Stereophonic acoustic echo cancellation using nonlinear transformations and comb filtering," IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 6, pp. 3673-3676, May 1998.

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

[5] 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.

[6] 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 and Audio Processing, vol. 9, no. 6, pp. 686-696, Sep. 2001.

[7] M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation," in Proc. IEEE Int. Conf. Acoustics Speech Signal Processing, vol. 6, pp. 3689–3692, May 1998.

[8] 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 Processing, vol. 49, pp. 2577–2587, Nov. 2001.

[9] K. Ikeda, and R. Sakamoto, "Convergence analyses of stereo acoustic echo cancelers with preprocessing," IEEE Transactions on Signal Processing, vol. 51, no. 5, pp. 1324 – 1334, May 2003.

[10] P. Surin, N. Tangsangiumvisai, and S. Aramvith,"An adaptive noise decorrelation technique for stereophonic acoustic echo cancellation," IEEE Tencon, vol. 1, pp.112-115, Nov. 2004.

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

[12] J. Herre, H. Buchner, and W. Kellermann, "Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement," IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1, pp. I-17-I-20, Apr. 2007.

[13] J. M. Valin, "Perceptually-motivated nonlinear channel decorrelation for stereo acoustic echo cancellation, "Hands-Free Speech Communication and Microphone Arrays (HSCMA), pp. 188-191, May 2008.

[14] M. Khasawneh, K. Mayyas, R. M. Shalabi, and M. I. Haddad, "A combined TDA/FDA adaptive schema for stereophonic acoustic echo cancellation," IEEE International Symposium on Circuits and Systems, vol. 2, pp. 1686 - 1689, May 2005.

[15] S. Haykin, Adaptive filter theory, 4th ed. Prentice Hall, 2001.

[16] L. Parolini, S. Bartoloni, F. Piazza," Gabor expansion for stereophonic acoustic echo cancellation," IEEE International Symposium on Circuits and Systems, vol. 2, pp. 11-769 – 11-772, May 2001.

[17] B. Krittayanun, and N. Tangsangiumvisai, "On performance improvement of stereophonic acoustic echo cancellation through the use of hybrid structure," 48th Midwest

Symposium on Circuits and Systems, vol. 2, pp. 1863 – 1866, Aug. 2005.

[18] S. Emura, Y. Haneda, A. Kataoka, Sh. Makino,"Stereo echo cancellation algorithm using adaptive update on the basis of enhanced input-signal vector," Signal Processing, vol. 86, no. 6, pp. 1157–1167, 2006.

[19] 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.

[20] M. Bekrani, M. Lotfizad, "A wavelet thresholding approach for signal decorrelation in the stereophonic acoustic echo cancellation application", 9th International Conference on Signal Processing, pp. 2681-2683, Oct. 2008.

[21] K. Mayyas, "Low complexity LMS-type adaptive algorithm with selective coefficient update for stereophonic acoustic echo cancellation," Computers and Electrical Engineering, (2008), doi:10.1016/j.compeleceng.2008.11.004.

[22] Y. Zhou, S.C. Chan, K.L. Ho, "A wavelet based partial update fast LMS/Newton algorithm," Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication Systems, Issue, 13-16, pp. 817 – 820, Dec. 2005.

[23] S. Mallat, A Wavelet tour of signal processing, 2nd ed. San Diego, Academic Press, 1999.

[24] F. Beaufays, "Transform-domain adaptive filters: an analytical approach," IEEE. Trans. Signal Process. vol. 43, no. 32, pp. 422–431, 1995.

[25] B. Farhang-Boroujeny, Adaptive filters: theory and applications, Wiley, Chichester, UK, 1998.

[26] S. Attallah, "The wavelet transform-domain LMS algorithm: a more practical approach," IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process. vol. 47, no. 3, pp. 209–213, Mar. 2000.

[27] J. Benesty, F. Amand, A. Gilloire, Y. Grenier, "Adaptive filtering algorithms for stereophonic acoustic echo cancellation," Proceedings of International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 3099–3102, May 1995.

[28] I. Daubechies, "Orthonormal bases of compactly supported wavelets," Comm. in Pure and Applied Math., vol. 41, pp. 909–996, 1988.

[29] Q. Jin, Z. Q. Luo, and K. M. Wong, "Optimum filter banks for signal decomposition and its application in adaptive echo cancellation," IEEE. Trans. Signal Process, vol. 44, no. I, pp. 1669–1679, July 1996.

[30] 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.

[31] 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.

Mehdi Bekrani was born in Gorgan, Iran, in 1979. He received the B.Sc. degree from Ferdowsi University of Mashhad, Iran, and the M.Sc. degree from Tarbiat Modares University, both in electrical engineering. He is currently a Ph.D. student in Tarbiat Modares University.

Mojtaba Lotfizad was born in Tehran, Iran, in 1955. He received the B.S. degree in electrical engineering from Amir Kabir University of Technology, Iran, in 1980, and the M.S. and Ph.D. degrees from the University of Wales, UK, in 1985 and 1988, respectively. He joined the Engineering Faculty, Tarbiat Modares University of 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.