Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 9 (4): 651-661, 2009
ISSN 1812-5654

Transform Domain Based Multi-Channel Noise Cancellation Based on Adaptive Decorrelation and Least Mean Mixed-Norm Algorithm

M. Geravanchizadeh and T.Y. Rezaii
Department of Electrical and Computer Engineering, Tabriz University, Tabriz, Iran

Abstract: In this study, a transform domain based adaptive noise cancellation algorithm is proposed to enhance noise carrying speech signals. The algorithm deals with situations where the microphones should locate in close proximity, as they cancel out the crosstalk effects. In other words, the source of the noise signal is not available separately and is independent of the desired speech signal. This is the case in mobile phones and hands-free systems, where the smallness of the dimension of the applied speech enhancement system is desirable. In the proposed algorithm the Discrete Sine Transform (DST) is used as self orthogonalizing transform to address the eigen-spread problem of adaptive filter, whereas Least Mean Mixed-Norm (LMMN) adaptation algorithm and Symmetric Adaptive Decorrelation (SAD) structure are applied to improve the convergence rate of the adaptive filter and make a considerable improvement in the performance of the noise cancellation procedure. Also, the Voice Activity Detection (VAD) is used to reduce the computational costs and decrease the execution time. However in this study, there was an utmost attempt to consider all of the practical problems, while the minimum simplifying assumptions are made. The simulation results have proven the robustness of this algorithm compared with commonly used algorithms, in the sense of SNR and MSE improvement and speech intelligibility.

Key words: Noise canceller, adaptive filter, symmetric adaptive decorrelation, LMMN algorithm, discrete sine transform

INTRODUCTION

It can be mathematically shown that the convergence rate of the steepest descent based adaptive algorithms such as LMS and LMMN is highly dependent of the eigen-structure of the correlation matrix (Cowan, 1987). The problem of a poor eigen-structure is common in noisy acoustic environments because of the presence of multi-path signals (Tinati *et al*., 2008). In fact, in such cases, the correlation matrix may even be ill-conditioned. It has been shown that a self-orthogonalizing transformation of the input speech signal to adaptive filter can result in a gradient algorithm whose convergence rate is essentially independent of the underlying eigen-structure of the signal correlation matrix (Beaufays, 1995). For this reason the Discrete Sine Transform is used as orthogonalizing function to enhance the LMMN algorithm in our adaptive noise canceller system.

In general, standard Adaptive Noise Cancellation (ANC) mechanism is composed of two main elements: the primary microphone and the reference microphones (microphone). The primary microphone is the microphone, which contains the desired speech signal that is corrupted by a distorted version of the noise signal of reference microphone. The noise cancellation procedure involves using the reference noise signal, to estimate the noise component of the primary signal. This estimation is then used to cancel out the noise component of the desired speech signal at the primary microphone. In this mechanism, it is assumed that the reference microphones are far enough from the primary microphone so that the desired speech signal of the primary microphone does not leak into the reference microphone. The block-diagram of the conventional adaptive noise canceller is illustrated in Fig. 1.

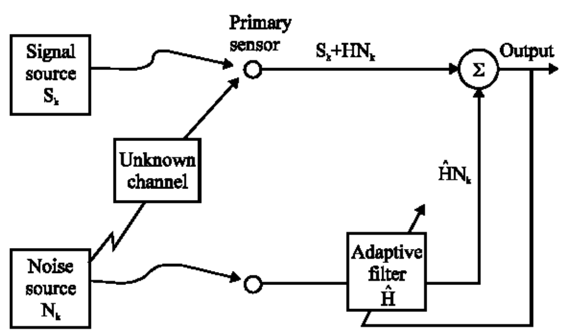

Fig. 1: Block-diagram of the conventional adaptive noise canceller

Corresponding Author: Tohid Yousefi Rezaii, Department of Electrical and Computer Engineering, Tabriz University, Tabriz, Iran Tel: +98 411 3300832 Fax: +98 411 3300819

In many applications, such as in mobile and hands-free phones, speech enhancement systems are expected to be small in size (Bouquin, 1996, Martin, 2001). Hence the distance between reference and primary microphones should be very small (in the range of mm). But microphones located in close proximity, undergo serious crosstalk effects due to signal leakage (Zeng and Abdulla, 2006). In other words, the performance of Multi-Channel Adaptive Noise Canceller (MANC) is highly dependent on signal leakage, from primary microphone to reference microphones, so that, the higher the signal leakage intensity, the lower the performance and convergence rate of MANC. Therefore, the necessity to develop such a robust algorithm that depends less on the amount of signal leakage and crosstalk effect between microphones is obvious.

So far, several two-channel Crosstalk Resistant ANC (CRANC) methods have been investigated by Mirchandani *et al*. (1992), Kuo and Peng (1999) and Madhavan and Bruin (1990), but they are relatively computationally expensive and somewhat unstable. Zeng and Abdulla (2006) proposed a new Multi-Channel Crosstalk Resistant ANC (MCRANC) that extends the two-channel CRANC method to Multi-channel processing. Although results of the given conditions are good, but it assumes that the environment remains unchanged during detected Voice Periods (VP), which is not a correct assumption in non-stationary environments. Furthermore, it assumes that the adaptation algorithm converges to its steady-state values during Non-Voice Periods (NVP). However, this assumption would not hold in practical applications of ANC, specifically when the NVP is so small and therefore the adaptive mechanism would not have enough time (not sufficient iterations) to get converged.

In this study, a transform domain robust multi-channel crosstalk resistant ANC based on Symmetric Adaptive Decorrelation and Least Mean Mixed-Norm adaptation algorithms is proposed. It has been shown that using SAD algorithm in the proposed method would improve the MSE performance and the output speech signal quality, while applying LMMN algorithm in the adaptation mechanism would significantly improve the convergence rate of the adaptation procedure, even if the NVPs are very small in length. Furthermore, the Voice Activity Detection is used to control the noise cancellation procedure in the manner described later, so that computational cost will decrease significantly.

MATERIALS AND METHODS

Transform domain adaptive filtering: Transform domain adaptive filtering can lead to an improvement in the convergence properties of the standard LMS algorithm. Let s be the input the vector to LMS FIR filter. In what follows, the orthogonal transform of s is indicated by $\tilde{s}$ and is computed as:

$$\tilde{s} = Ts \tag{1}$$

where, T is the orthogonal transform. The optimal weights in transform domain are given by (Beaufays, 1995).

$$W_{topt} = (T^t)^{-1} W_{opt} \tag{2}$$

where, W_{topt} and W_{opt} are the optimum weight vectors in transform domain and time domain, respectively.

It can be mathematically shown that the convergence rate of the steepest descent based algorithms such as LMS and LMMN are highly dependent of the eigen-structure of the correlation matrix (Cowan, 1987). The problem of a poor eigen-structure arises because of the correlation between signals arriving at the microphone array. On the other hand the self-orthogonalizing transformation of the input array vector can result in a gradient algorithm whose convergence rate is essentially independent of the underlying eigen-structure of the signal correlation matrix. Since discrete sine transform, is real valued and provides good orthogonalization property, we use this transform in the proposed structure.

Defining x as the received signal (output of microphone array), its DST can be calculated by:

$$\tilde{s}(n) = \sum_{n=1}^{N} s(n)\sin\left(\frac{kn\pi}{N+1}\right), \quad k = 1, 2, ..., N \tag{3}$$

where, N is the No. of microphone elements.

Symmetric adaptive decorrelation: The performance of speech enhancement systems, based on adaptive filtering, is highly dependent on the quality of the noise reference. In other words, the noise signal in the reference and primary microphones must be sufficiently coherent to obtain desired noise reduction. Furthermore, any leakage from the primary speech signal into the reference microphone must be avoided, since it results in signal distortion and poor noise cancellation (Gerven and Compernolle, 1995). To illustrate the problem, the block diagram of a system with signal leakage followed by the feedback implementation of SAD structure, is shown in Fig. 2. In Fig. 2, the signal leakage problem from channel 1 to channel 2 and vice versa is shown within the red dashed box. In this box, h_1 and h_2 are the column vectors of the corresponding impulse responses of the unknown channels, which need to be estimated by the adaptive mechanism. In continue for the sake of estimating this

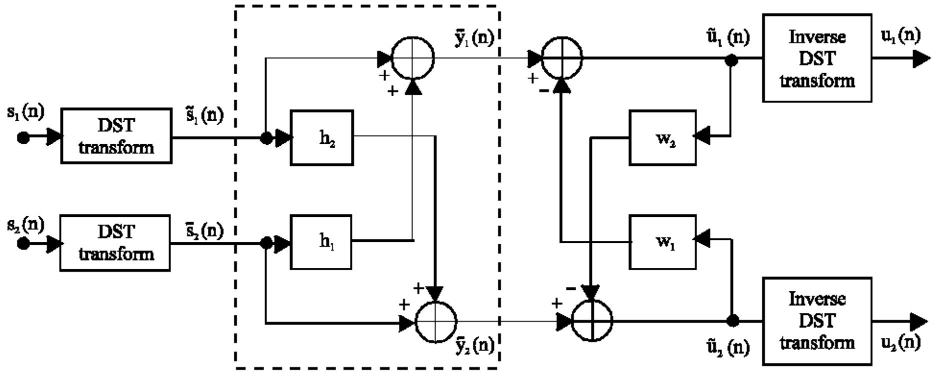

Fig. 2: Block diagram of a system with leakage followed by the SAD structure

unknown leakage channels, the feedback implementation of SAD structure is obtained by placing two adaptive filters w_1 and w_2 in a feedback loop as shown in Fig. 2. These filters are also column vectors. The purpose of the SAD structure is to obtain the clean signals s_1 and s_2 in its output, by processing on the corrupted and mixed version of these two signals in transform domain, namely $\tilde{y}_1(n)$ and $\tilde{y}_2(n)$. So, the input signals of SAD structure, $\tilde{y}_1(n)$ and $\tilde{y}_2(n)$ are defined as follows:

$$\tilde{y}_2(n) = \tilde{s}_2(n) + \tilde{s}_1^T h_2 \tag{4}$$

$$\tilde{y}_2(n) = \tilde{s}_2(n) + \tilde{s}_1^T h_2 \tag{5}$$

The symbol T is used as the transposition operator. The corresponding outputs in transform domain are $\tilde{u}_1(n)$ and $\tilde{u}_2(n)$ and defined as:

$$\tilde{u}_1(n) = \tilde{y}_1(n) - \tilde{u}_2^T w_1 \tag{6}$$

$$\tilde{u}_2(n) = \tilde{y}_2(n) - \tilde{u}_1^T w_2 \tag{7}$$

where, $\tilde{u}_1$ and $\tilde{u}_2$ are column vectors of the same dimension as w_1 and w_2. By replacing Eq. 7 into Eq. 6 we obtain:

$$\tilde{u}_1(n) = \frac{1}{1 - w_1^T w_2}(\tilde{y}_1(n) - \tilde{y}_2^T w_1) \tag{8}$$

where, $\tilde{y}_2$ is a column vector of the same size as w_1. The relations for error estimate and tap weight adaptation formula based on NLMS algorithm for filter w_1, are obtained as follows:

$$\tilde{e}_1(n) = \tilde{u}_1(n) = \tilde{y}_1(n) - \tilde{u}_2^T w_1 \tag{9}$$

$$w_1(n+1) = w_1(n) + K\tilde{u}_1(n)\tilde{u}_2 \tag{10}$$

Where:

$$K = \frac{\tilde{\mu}_1}{c + \|\tilde{u}_2\|^2} \quad 0 \le \tilde{\mu}_1 \le 2$$

In the same manner, we obtain the relations for error estimate and tap weight adaptation formula for filter w_2.

Least mean mixed-norm algorithm: The LMMN adaptation algorithm is proposed by Chambers *et al.* (1994) in order to improve the convergence rate of the conventional LMS (or NLMS) adaptation algorithm. This method is based on the modified cost function to be minimized. The modified cost function is defined as follows:

$$J(n) = \gamma E\{e^2(n)\} + (1-\gamma)E\{e^4(n)\} \tag{11}$$

where, the mixing parameter γ, lies in the interval $0 \le \gamma \le 1$. For operation in a statistically non-stationary environment, γ may be adapted to match appropriately the properties of the measured signal. Considering the cost function defined in Eq. 9, the following recursion formula for tap weight adaptation of the LMMN algorithm is obtained:

$$w(n+1) = w(n) + 2\mu_{LMMN}[\gamma e(n) + 2(1-\gamma)e^3(n)]x(n) \tag{12}$$

where, w (n) is the column vector of the tap weights, x (n) is the column vector of the same size as w (n) of the input samples, e (n) is the error signal and μ_{LMMN} is the adaptation gain of the LMMN algorithm.

PROPOSED STRUCTURE

Describing multi-path propagation environment: Suppose the speech signal s (n) and the noise signal n (n) are generated by independent sources. As shown in

Fig. 3, these signals arrive at microphone array $\{M_0, M_1,...,M_N\}$ through multi-path and are acquired as $s_i(n)$ and $n_i(n)$.

In Fig. 3, H_{si} and H_{ni} are the transfer functions of the intermediate media between the speech and noise sources and the acquiring microphones, respectively. Hence, the signal at microphone M_i is:

$$x_i(n) = s_i(n) + n_i(n) \tag{13}$$

where, i = 0, 1, ..,N and N is the No. of used microphones.

According to Fig. 3, M_0 is assumed to be the primary microphone and the signal at this microphone is denoted as the primary signal. The remaining signals $x_i(n)$, i = 1, 2, ..,N are considered as the N referential signals, obtained from N referential microphones. From Fig. 3 we have:

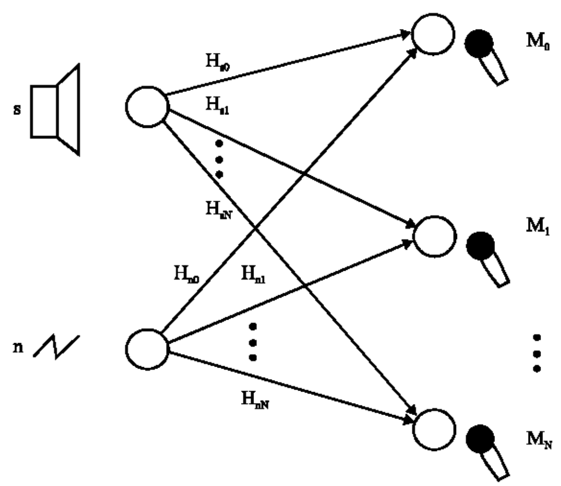

Fig. 3: Speech and noise propagation model

$$s_i(n) = h_{si}(n) \times s(n) \tag{14}$$

$$n_i(n) = h_{ni}(n) \times n(n) \tag{15}$$

where, $h_{si}(n)$ and $h_{ni}(n)$ are the impulse responses corresponding to the transfer functions, H_{si} and H_{ni}, respectively.

NLMS-based algorithm: The block diagram of the proposed structure in transform domain, using NLMS adaptive filter is shown in Fig. 4. As shown in Fig. 4, the proposed structure employs a conventional adaptive FIR filter, namely A and two SAD-based adaptive FIR filters, namely w_1 and w_2, in which the conventional NLMS adaptation algorithm is applied for the estimation purpose.

The VAD algorithm is applied to detect the NVP and VP states of the primary signal x_0 and then to drive the adaptation algorithms of filter A and the filters of SAD structure. In simple words, this algorithm detects the voiced and silence frames of the input speech signal using Zero Crossing Rate (ZCR), energy and the value of autocorrelation function of each frame. For the sake of clarity, an illustrative example of the input and output signals of this algorithm is shown in Fig. 5.

Figure 6 shows that how the VAD algorithm is applied to derive adaptive filter A and the adaptive filters of SAD structure. According to Fig. 6, in each branch of the VAD output, the active blocks are marked as grey blocks. If the VAD output is zero, this means that the primary signal x_0 contains silence frame and then the coefficients of filter A will be updated and the SAD structure is inactive. In the same way, if the VAD output equals one, this means that the primary signal x_0 contains voice frame and then the tap weights of the filters of the SAD structure will be updated and filter A is inactive. The output of the system is denoted as y_2.

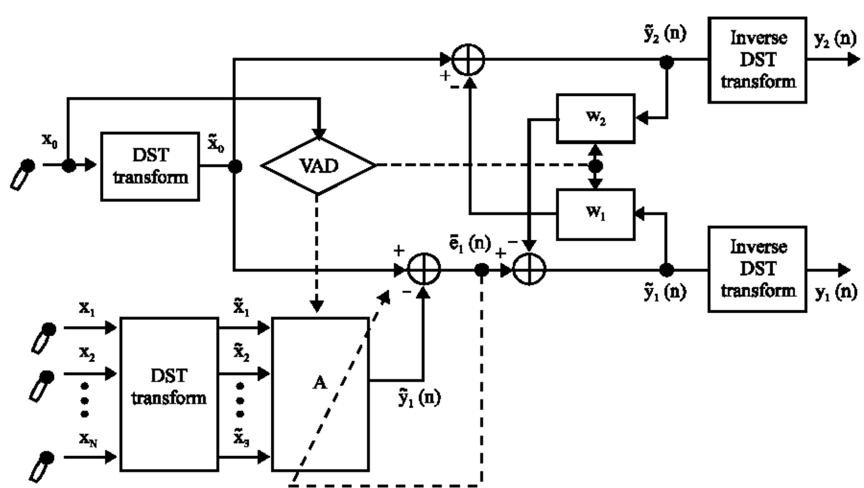

Fig. 4: Block diagram of the proposed structure

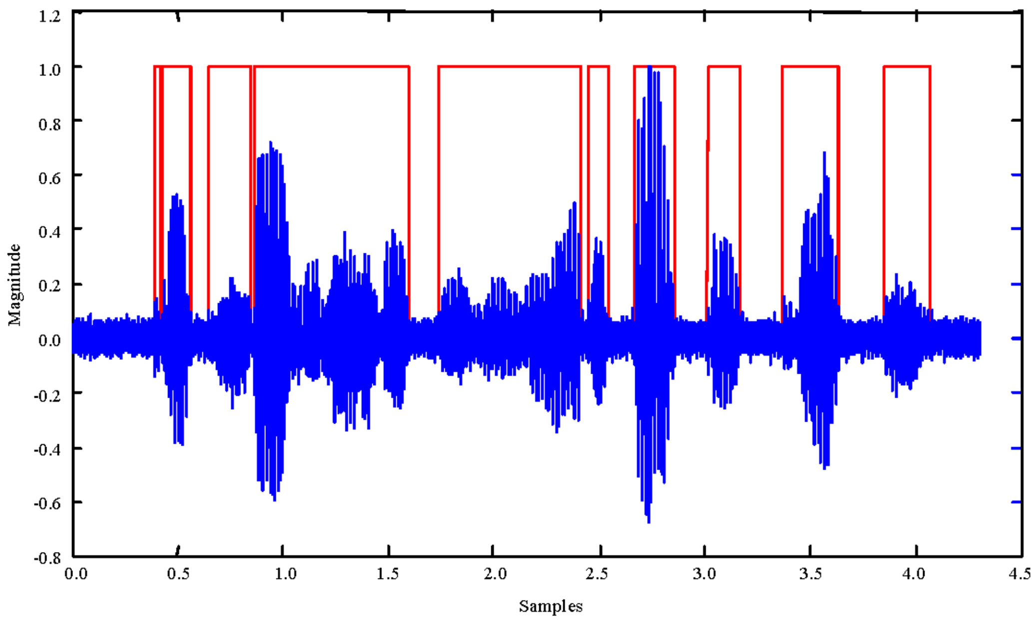

Fig. 5: An illustrative example of the input and output signals of the VAD algorithm

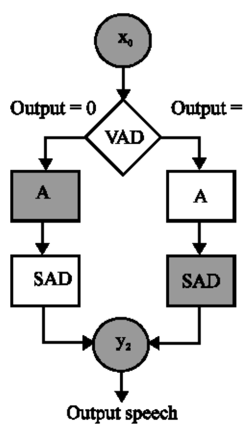

Fig. 6: Operation of the VAD structure (Grey blocks are active)

So, during NVP of the signal x_0, it is assumed that $s_0(n)$ and $s_i(n)$; i = 1, 2,...,N are zero. Therefore, the referential signals are used to cancel the primary signal. So, we have:

$$\tilde{n}_0(n) = w_A \tilde{\tilde{n}}(n) + \tilde{e}(n) \tag{16}$$

where, w_A is the weight vector of adaptive filter A, i.e.,

$$w_A = [w_{A1}, w_{A2}, \ldots, w_{AN}]^T \tag{17}$$

with

$$w_{Ai} = [w_{Ai_1}, w_{Ai_2}, \ldots, w_{Ai_L}],\ i = 1, 2, \ldots, N \tag{18}$$

and $\tilde{\tilde{n}}_i(n)$ is the vector of referential noise signals.

$$\tilde{\tilde{n}}(n) = [\tilde{\tilde{n}}_1(n), \tilde{\tilde{n}}_2(n), \ldots, \tilde{\tilde{n}}_N(n)]^T \tag{19}$$

Where:

$$\tilde{\tilde{n}}_i(n) = [\tilde{n}_i(n), \tilde{n}_i(n-1), \ldots, \tilde{n}_i(n-L+1)]^T,\ i = 1, 2, \ldots, N \tag{20}$$

We need to update the weights of filter A to minimize the square sum of error signal $\tilde{e}(n)$ in Fig. 4. Here, the NLMS algorithm is used to update the weight vector of filter A. Also, it is assumed that the NVP is long enough so that the NLMS algorithm has enough time to converge to its optimum weight vector w^o, so the error signal $\tilde{e}(n)$ reaches its optimum value $\tilde{e}^o(n)$. Thus, we will have:

$$\tilde{e}^o(n) = \tilde{n}_0(n) - w_A^{oT} \tilde{\tilde{n}}(n) \tag{21}$$

During VP, which just follows non-voiced period, the adaptive algorithm of SAD structure will start to update and the weights of filter A will be frozen. So, the error signal $\tilde{e}(n)$ becomes:

$$\tilde{e}(n) = B(n) + \tilde{s}_0(n) \tag{22}$$

Where:

$$B(n) = -w_A^{oT} \tilde{\tilde{s}}(n) + \tilde{e}^o(n) \tag{23}$$

In practice, due to non-stationary noise sources, the acoustic environment is dynamic and changes with time.

So, the error signal $\tilde{e}(n)$ may have some coherent component with primary noise signal $\tilde{n}_0(n)$. In other words, due to non-stationary characteristics of the noise source, the adaptive filter fails to efficiently cancel out the noise from the primary signal. This can be interpreted as the signal leakage from the primary signal $\tilde{x}_0(n)$ to the error signal $\tilde{e}(n)$. We name this coherent component $\tilde{e}_{leak}(n)$. So, the error signal e becomes:

$$\tilde{e}(n) = B(n) + \tilde{e}^{o}(n) + \tilde{e}_{leak}(n) \quad (24)$$

and the primary signal is:

$$\tilde{x}_0(n) = \tilde{s}_0(n) + \tilde{n}_0(n) \quad (25)$$

As described above, in one hand the error component $\tilde{e}_{leak}(n)$ is coherent with the primary noise component $\tilde{n}_0(n)$. In the other hand, the error component B(n) is coherent with the primary signal component $\tilde{s}_0(n)$. To resolve this issue, we have used the SAD algorithm to efficiently remove the error components from the primary speech signal.

LMMN-based algorithm: It is assumed that the NVP is long enough so that the NLMS algorithm has enough time to converge to its optimum weight vector. However, this is not a correct assumption in practice, since the NVP might have so small value in some situations, so that NLMS adaptation algorithm would not reach its optimum tap weights. For this reason, the LMMN algorithm is used to update the tap weights of filter A. As it was mentioned previously, the LMMN algorithm has higher convergence rate compared with the conventional NLMS algorithm.

RESULTS AND DISCUSSION

Here, the simulation results for our proposed transform-domain NLMS-based and LMMN-based algorithms are obtained and compared with the time-domain MCRANC algorithm given in (Zeng and Abdulla, 2006). The simulations are performed for three reference microphones (N = 3).

Furthermore, tap lengths of filters A, w_1 and w_2 are set to 60, 30 and 30, respectively. Also tap lengths of filters A and B for MCRANC algorithm are set to 60 and 50, respectively. The step size parameter for all NLMS algorithms is set to 2 and the LMMN mixing parameter is set to 0.75. The clean speech signal and also the corrupted speech signal (the signal at primary microphone) with AWGN noise of variance 0.1 are shown in Fig. 7. The output signals of three mentioned

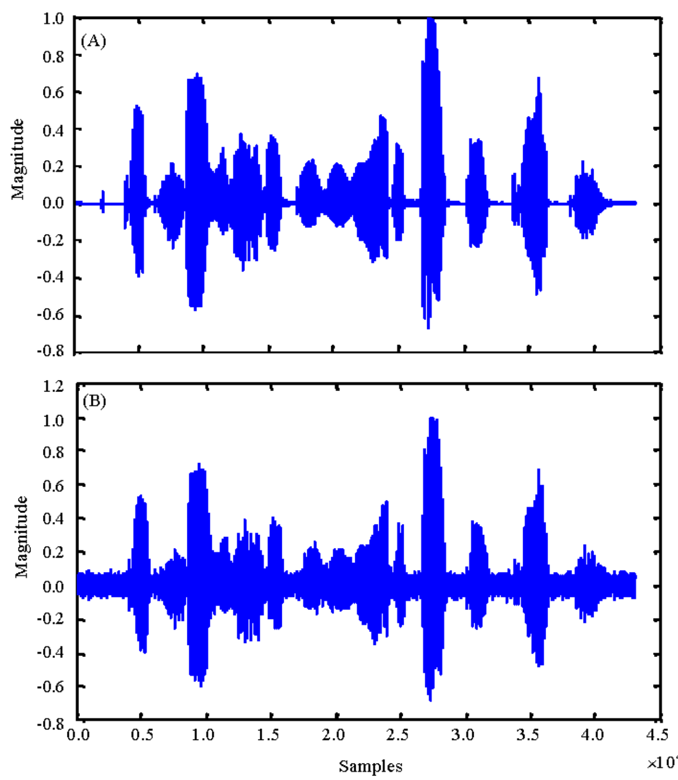

Fig. 7: (A) clean speech signal and (B) Noisy speech signal at primary microphone

algorithms, including time-domain MCRANC algorithm and transform-domain NLMS-based and LMMN-based proposed algorithms are shown in Fig. 8. Also the mean squared error plots for these algorithms are shown in Fig. 9. The error is defined as the residual noise component after the noise cancellation procedure. As it is shown form Fig. 8 and 9, the NLMS-based algorithm has better error performance than MCRANC, while the LMMN-based algorithm outperforms these two algorithms.

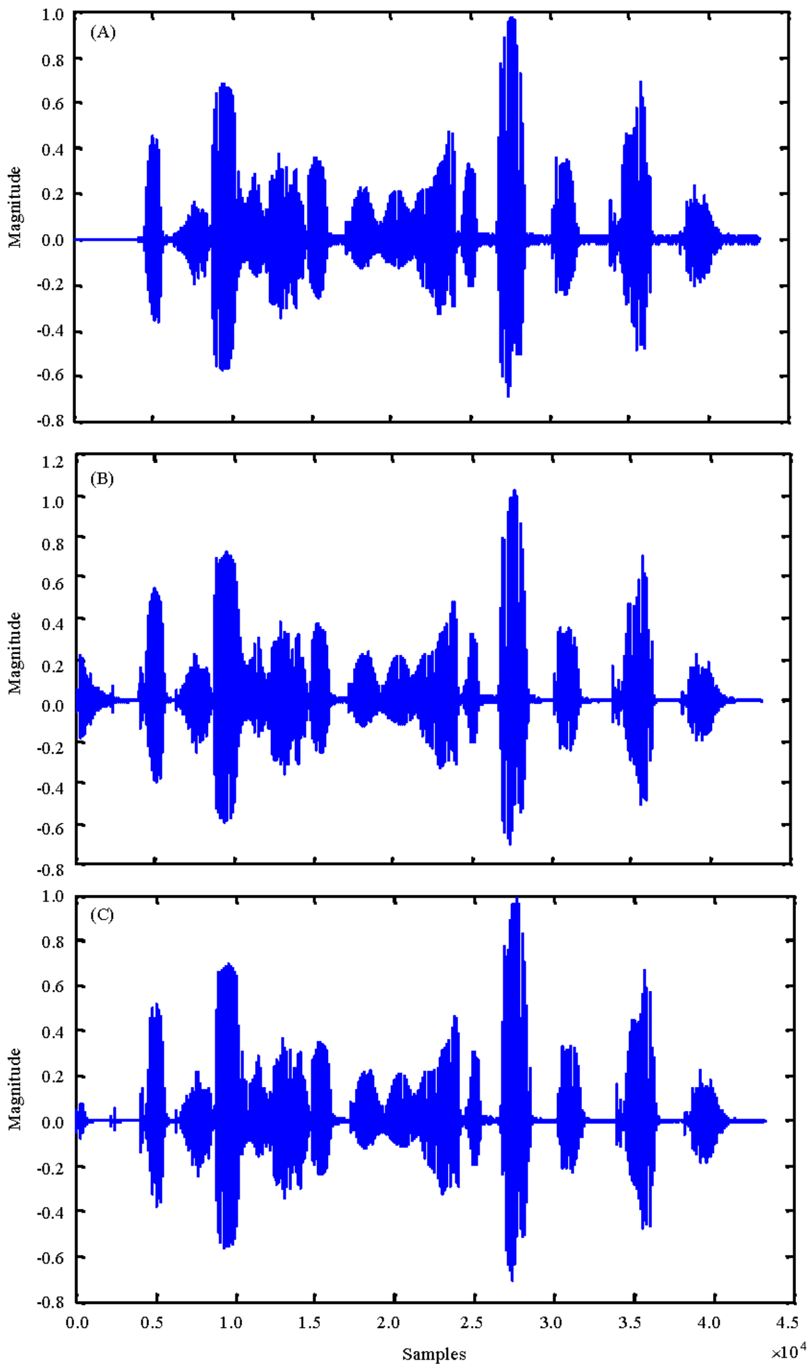

Fig. 8: Output speech signals for (A) Time domain MCRANC algorithm, (B) Transform domain NLMSA-based proposed algorithm and (C) Transform domain LMMN-based proposed algorithm

For better comparison of the steady state performance of these algorithms, the MSE plots for the steady state of them are shown in Fig. 10. It is clearly seen that the residual error of the proposed LMMN-based algorithm is less than the others and so the cancellation ability of this algorithm is the best in both transition and steady states. As mentioned earlier, the LMMN algorithm has better convergence rate than the conventional NLMS algorithm. As it is shown from Fig. 11, in the LMMN-based proposed algorithm the squared error reaches its

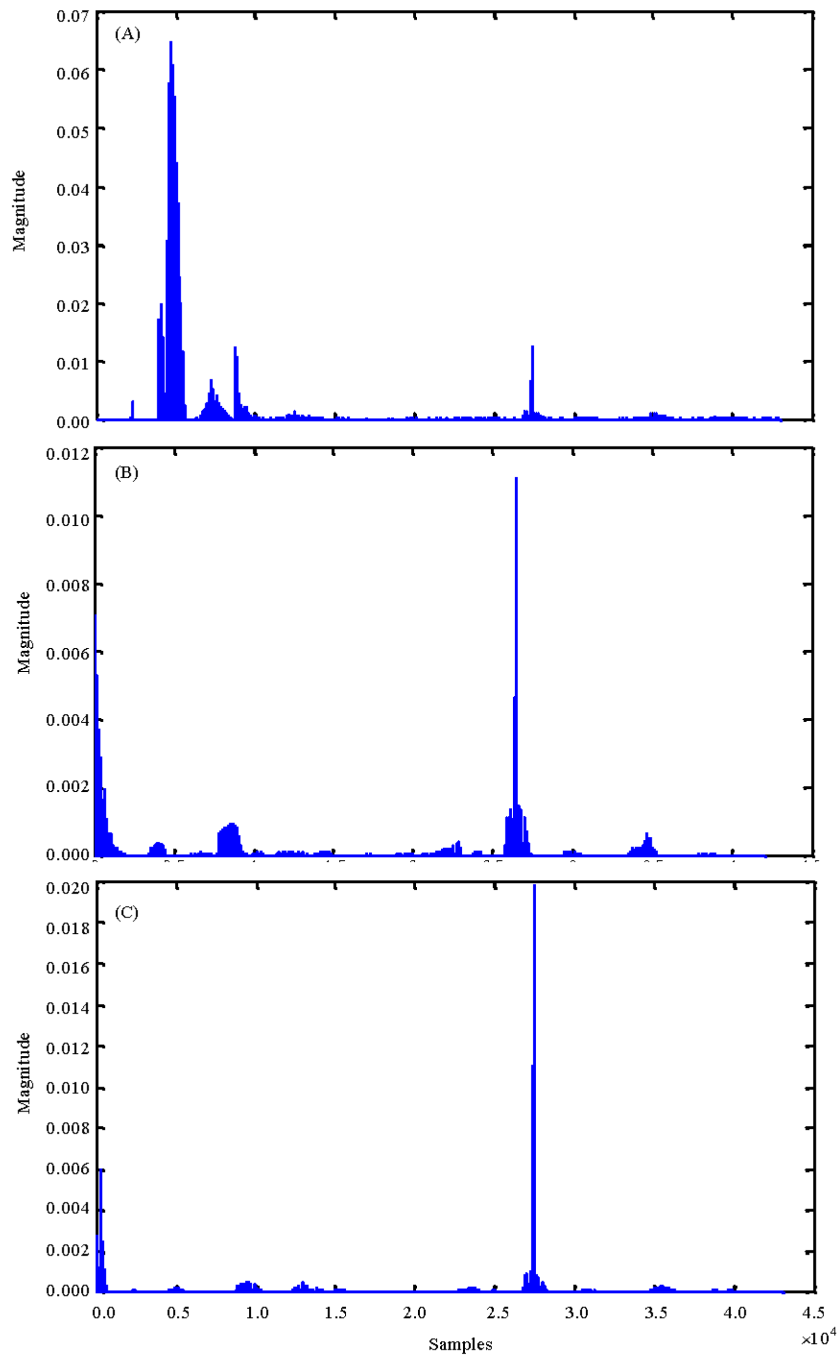

Fig. 9: MSE of residual noise for (A) Time domain MCRANC algorithm, (B) Transform domain NLMSA-based proposed algorithm and (C) Transform domain LMMN-based proposed algorithm

steady state value very fast compared to the MCRANC algorithm. Figure 11 shows the squared error of the two algorithms at the beginning of a NVP. So, in situations that the NVP is too small (for example smaller than about 500 samples), the MCRANC algorithm fails to efficiently remove the noise component, while the LMMN-based proposed algorithm successfully cancels out the noise component, even if the NVP is too small. This property will reduce the error propagation in the subsequence frames.

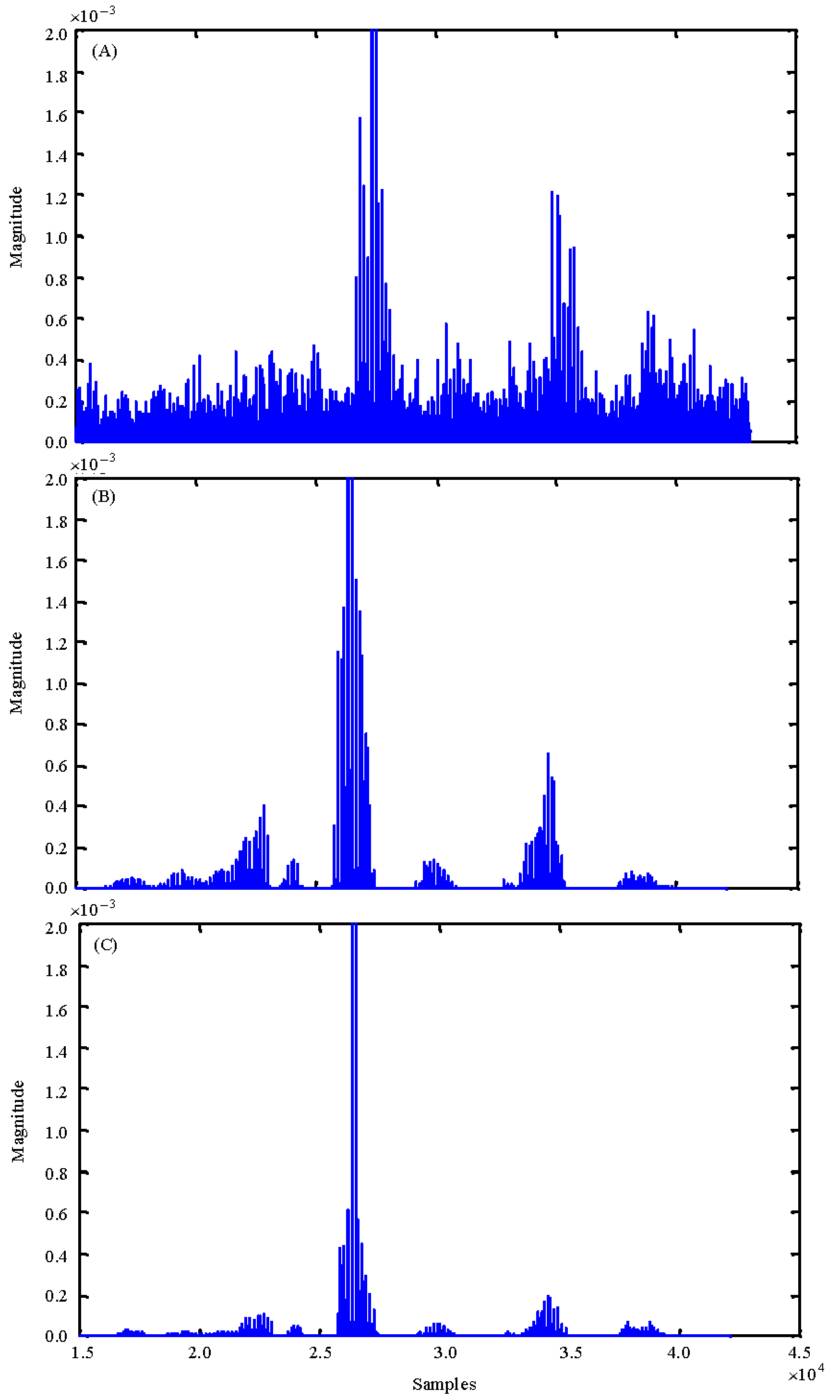

Fig. 10: MSE of residual noise in steady state for (A) Time domain MCRANC algorithm, (B) Transform domain NLMSA-based proposed algorithm and (C) Transform domain LMMN-based proposed algorithm

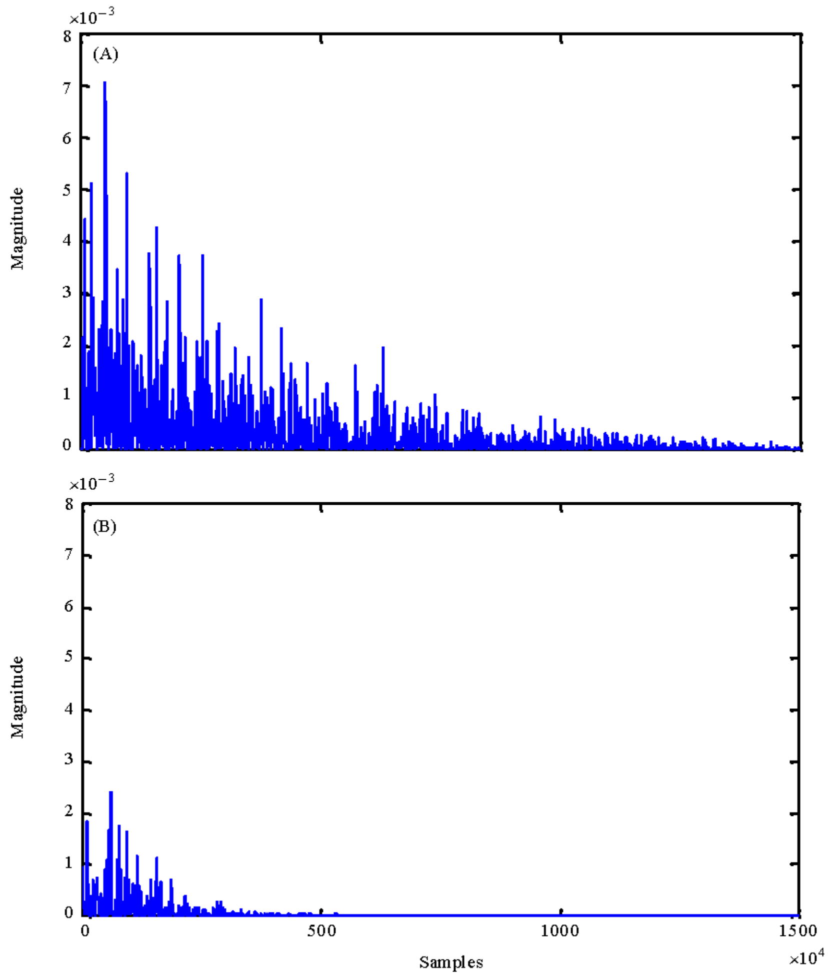

Fig. 11: Convergence rate of (A) NLMS-based proposed algorithm and (B) LMMN-based proposed algorithm

CONCLUSION

In this study, a transform domain adaptive noise cancellation system based on SAD structure and LMMN adaptation algorithm implemented on microphone array, is proposed. The presented noise canceller can be used in situations in which the primary and reference microphones need to be located in very close proximity. Mobile phones and hands-free systems are the examples of such applications in which the smallness of the dimensions is desirable.

Also, the proposed algorithm is robust against environmental variations such as noise source with non-stationary characteristics. Finally, the simulations are presented to evaluate the performance and reliability of the proposed system. The results show the robustness and reliability of the proposed multi-channel crosstalk resistant noise canceller.

REFERENCES

Beaufays, F., 1995. Transform domain adaptive filters: An analytical approach. IEEE Trans. Signal Process., 43: 422-431.

Bouquin, R.L., 1996. Enhancement of noise speech signals: Application to mobile radio communications. J. Speech Commun., 18: 3-19.

Chambers, J.A., O. Tanrikulu and A.G. Constantinides, 1994. Least mean mixed-norm adaptive filtering. IEEE Elect. Lett., 30: 1574-1575.

Cowan, C.F.N., 1987. Performance comparisons of finite linear adaptive filters. IEE Proc., 134: 211-216.

Gerven, S.V. and D.V. Compernolle, 1995. Signal separation by symmetric adaptive decorrelation: Stability, convergence and uniqueness. IEEE Trans. Signal Process., 43: 1602-1612.

Kuo, S.M. and W.M. Peng, 1999. Asymmetric crosstalk-resistant adaptive noise canceller. Proceedings of IEEE Workshop on Signal Processing Systems, October 20-22, Taipei, Taiwan.

Madhavan, G. and H.D. Bruin, 1990. Crosstalk resistant adaptive noise cancellation. Ann. Biomed. Eng., 18: 57-67.

Martin, R., M. Brandstein and D. Ward, 2001. Small microphone arrays with Postfilters for noise and acoustic echo reduction. Microphone Arrays, 12: 255-279.

Mirchandani, G., R.L. Zinser and J.B. Evans, 1992. A new adaptive noise cancellation scheme in the presence of crosstalk. IEEE Trans. Circ. Syst., 39: 681-694.

Tinati, M.A., A. Rastegarnia and T. Yousefi Rezaii, 2008. A transform domain based Least Mean Mixed-Norm algorithm to improve adaptive beamforming. Proceedings of 2nd International IEEE Conference on Electrical Engineering, March 25-26, Lahore, Pakistan, pp: 1-4.

Zeng, Q. and W.H. Abdulla, 2006. Speech enhancement by multi-channel crosstalk resistant adaptive noise cancellation. Proceedings of ICASSP Conference, May 14-19, IEEE Computer Society, pp: I-484-I-488.