A New Dual-Microphone Speech Enhancement Method for Oriented Noises

H. R. Abutalebi, M. Pourahmadi, M. R. Aghabozorgi

Signal Processing Research Lab
Department of Electrical Engineering, Yazd University
Yazd, Iran
habutalebi@yazduni.ac.ir, aghabozorgi@yazduni.ac.ir

Abstract

In this paper, we examine a modified Cross-Talk Resistant Adaptive Noise Canceller (CTRANC) structure that contains a delay unit on the primary channel to solve the causality constraint of conventional CTRANC. This Asymmetric CTRANC (ACTRANC) structure allows flexible alignment of noise source with the sensors array when the speech source is fixed in its position. Inserting delay unit on the primary path releases the placement of the noise sources and consequently, improves the capability of the structure in noise reduction. This can also be interpreted as the releasing causality constraint on the adaptive filters. We have shown that the best delay value in the structure is proportional to the propagation delay between reference and primary microphones (in terms of sample). Objective evaluations and informal listening tests demonstrate superior performance of the proposed method rather than conventional CTRANC.

Index Terms: speech enhancement, adaptive noise cancellation, cross talk

1. Introduction

Many speech communication systems have to work under the noisy conditions, leading to need to a front-end speech enhancement/noise cancellation block. Among various speech enhancement methods, Adaptive Noise Cancellation (ANC) is the most widely used dual-microphone technique. Adaptive noise canceller exploits a reference noise microphone besides the main (primary) one. Using gradient-based iterative algorithms, ANC estimates the transfer function between these two input microphones and results in enhanced speech at the output. It has been shown that the efficiency of the ANC is highly based on the independence of the reference input and the speech [1-4]. However, in real-life situations, near placement of two input microphones causes the speech leakage (or cross-talk) in the reference input and finally results in "signal cancellation" phenomenon [2]. The left half side of Fig. 1 shows the signal production model for the primary and reference input microphones in which S' is the leakage of source S in the reference microphone. In order to combat the cross-talk (or leakage) problem, the Cross-Talk Resistant ANC (CTRANC) system (see right half side of Fig. 1 has been already introduced [1-4]. In this structure, a second adaptive filter (named V in Fig. 1) is used to eliminate the cross-talk from the reference input. This new adaptive filter has to model the transfer function between the main (primary) speech and its leakage. Two adaptive filters are adapted separately via adaptation algorithms. In (feedback) CTRANC structure, the output of each filter is the input of another one. This structure works well when the signal

source is close to signal (primary) microphone and the noise source is near to the noise (reference) microphone [1].

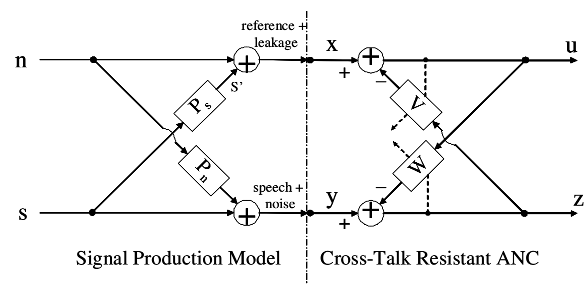

Fig. 1. Signal production model and CTRANC structure.

Due to the causality constraint of the system, the CTRANC structure cancels the later signal in each channel and retains the earlier one [5]; therefore, performance of the CTRANC degrades drastically when the orientation of the noise source changes. This means that the noise source can travels between the primary and reference microphones. Under such situations, the most important parameter that affects the system performance is the propagation delay difference from the source to the primary and reference microphones, For example, if the noise source
be closer to the primary microphone than the reference one, the noise component in the primary signal will be earlier than one in the reference signal and the first adaptive filter will not be causal. This results in the signal (speech) cancellation in both channels.

In this research, we examine a modified CTRANC structure that contains a delay unit on the primary channel to solve the causality constraint of conventional CTRANC. Furthermore, we introduce Double Affine Projection Algorithm as the adaptation technique for modified CTRANC.

The paper is organized as follows. In section 2, we introduce the modified CTRANC structure. The implemented adaptation technique is explained in section 3. Section 4 covers the simulation and evaluation results and finally, some concluding remarks are explained in section 5.

2. THE ASYMMETRIC CTRANC STRUCTURE

To release the causality constraint of conventional CTRANC, a modified structure has been proposed by inserting a delay unit on the primary input channel. This modified CTRANC structure is named Asymmetric CTRANC (ACTRANC) and shown in Fig. 2.

The effect and the proper value of inserted delay are explained as follows. Consider δ as the difference between propagation delay from the noise source to the primary and reference sensors. Since in ANC structure the noise component in the reference microphone should be earlier than the primary one, so the Δ delay on the primary channel must be greater than δ. On the other hand, according to the causality constraint on the filter V, the speech signal in the primary input (after the delay unit) should be earlier than the speech in the reference one; therefore, Δ must be smaller than δ_{max}, where δ_{max} is the propagation delay between reference and primary microphones (in terms of sample). Above discussion can be summarized as:

$$\delta < \Delta < \delta_{max} \tag{1}$$

$$\delta_{max} = \frac{d}{c} f_s \tag{2}$$

where d is the distance between the primary and reference sensors (in meter), c is the sound velocity (in meter/sec), and f_s is the sampling frequency (in sample/sec). When the noise source is near to the reference microphone δ is δ_{max} and by moving it away from the reference microphone, δ decreases until it reaches to the primary sensor, where δ comes to $-\delta_{max}$. As it can be seen in Fig. 3, δ can change between positive and negative values according to its position in the space. Since signal source is fixed in its position, based on the beamforming concepts, insertion the delay unit is justified as targeting the primary microphone to the signal source. This eliminates the crosstalk (leakage) effect in the primary microphone. According to this interpretation and by considering that difference between propagation delay from the signal source to the primary and reference sensors is $\delta = \delta_{max}$ so $\Delta = \delta_{max}$ is the perfect alignment of the signal source with the targeting direction.

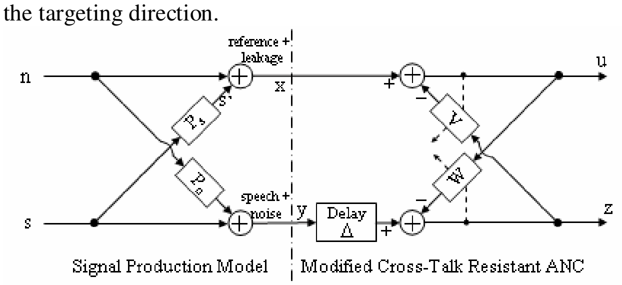

Fig. 2. *Asymmetric CTRANC structure.*

To relax this condition for practical implementation, we usually set $\Delta = \delta_{max} - 1$. This selection is also appropriate from equation (1) point of view.

So insertion of delay unit on the primary path releases the placement of the noise sources and consequently, improves the capability of the structure in noise reduction. This can also be interpreted as the releasing causality constraint on the adaptive filters. Computer simulations show better performance of ACTRANC for various noise orientations compared to the conventional CTRANC.

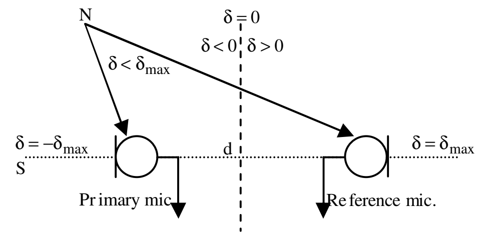

Fig. 3. *orientation of the noise source*

3. DOUBLE AFFINE PROJECTION ALGORITHM

In this paper, we employ a Double Affine Projection Algorithm (DAPA) to recursively reach to the optimum weights. The APA is a generalization of the well-known Normalized Lease Mean Square (NLMS) algorithm [6,7]. Under this interpretation, NLMS is viewed as a one dimensional APA. In APA(P) the projections are made in P dimensions. It means that P vectors are used for adaptation the filters. By increasing the projection dimension of APA, P, the convergence speed of the tap weight vector will be increased by the cost of more computational complexity of the algorithm.

The DAPA(P) is defined as follows:

- Filtering

$$\begin{aligned} \underline{u}(k) &= \underline{x}(k-\Delta) - Z'(k)\underline{V}^*(k) \\ \underline{z}(k) &= \underline{y}(k) - U'(k)\underline{W}^*(k) \end{aligned} \tag{3}$$

where:

$$\begin{aligned} \underline{x}(k) &= [x(k),....,x(k-P+1)] \\ \underline{y}(k) &= [y(k),....,y(k-P+1)] \\ Z(k) &= [\underline{z}(k),....,\underline{z}(k-P+1)] \\ U(k) &= [\underline{u}(k),...,\underline{u}(k-P+1)] \\ \underline{V}(k) &= [V_0(k),....,V_{L1}(k)] \\ \underline{W}(k) &= [W_0(k),....,W_{L2}(k)] \end{aligned} \tag{4}$$

and ' denote to the transpose operation.

- Filter updating

$$\underline{V}(k+1) = \underline{V}(k) + \mu_1 Z(k)\left[Z^H(k)Z(k) + \delta_1 I\right]^{-1} \underline{u}^*(k)$$
$$\underline{W}(k+1) = \underline{W}(k) + \mu_2 U(k)\left[U^H(k)U(k) + \delta_2 I\right]^{-1} \underline{z}^*(k) \quad (5)$$

The scalars $\delta_i (i = 1,2)$ are the regularization parameters for the sample autocorrelation matrix inverse $\left[Z^H(k)Z(k)+\delta_1 I\right]^{-1}$ or $\left[U^H(k)U(k)+\delta_2 I\right]^{-1}$ used to cure the ill-conditioning of $Z^H(k)Z(k)$ or $U^H(k)U(k)$ and 'I' is unitary matrix. For large enough δ_1 and δ_2, $Z^H(k)Z(k)+\delta_1 I$ and $U^H(k)U(k)+\delta_2 I$ always yield well-behaved inverses. The step size parameter, μ, is the relaxation factor. As in NLMS, the algorithm is stable for $0 < \mu < 2$.

4. COMPUTER SIMULATIONS

To evaluate the proposed system, we have used a male speech and a low-pass noise both samples at 16 kHz. The distance between two microphones is assumed about 10 cm (5 samples). In DAPA, $\mu_1 = \mu_2 = 0.1$ and $\delta_1 = \delta_2 = 0.01$. In all tests, we assume that the speech source is near to the primary microphone and noise source can travel between two microphones. Filter between sources and microphones are FIR filters. By change in the position of the noise source and SNR of the primary channel, we have good results that verify theoretical assumption. Fig. 4 shows SNR improvement of CTRANC and ACTRANC systems as a function of the propagation delay difference δ (distance difference from the noise source to the reference and primary sensors). It is obvious that for negative values of δ, where the noise source is closer to the primary microphone than the reference one, CTRANC has a very bad performance and its output is worse than input signal. On the other side, ACTRANC has a good result for both negative and positive values of δ. To further evaluation performance of the system, we have employed LAR-distance. It has been shown that among various spectral distance measures, LAR-distance has the highest correlation with the subjective evaluations [8]. Considering K(m,q) (q=1,…,Q) as Q reflection coefficients of mth frame, Area Ratio (AR) factors is defined as:

$$AR(m,q) = \frac{1+K(m,q)}{1-K(m,q)} \quad (6)$$

If $AR_s(m,q)$ and $AR_z(m,q)$ be AR parameters of primary clean speech and output speech of the system respectively, the LAR-distance for mth frame is evaluated as:

$$LAR_{sz}(m) = \left\{ \frac{1}{Q} \sum_{q=1}^{Q} \left| 20\log_{10}\left[\frac{AR_s(m,q)}{AR_z(m,q)}\right] \right|^2 \right\}^{\frac{1}{2}} \quad (7)$$

In order to remove frames with unrealistically high LAR-distances, we have computed the overall LAR-distance by first discarding frames with the top 5% LAR values, and then averaging over the remaining frames (as suggested in [4]). In this evaluation, we have used DAPA(4) (APA of order 4) as the adaptation technique and assumed SNR = -6 dB in the reference channel. Also, a 200-samples Hamming window has been employed in AR computations. To compare the performance of the systems in the case of noise source near to primary microphone, we have considered $\delta = -1$.

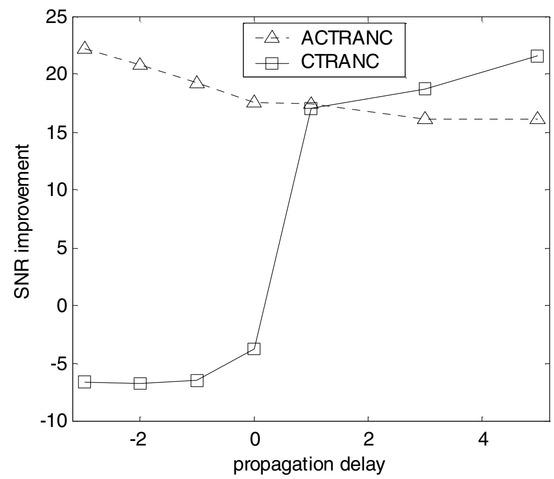

Fig. 4. *SNR improvement for various propagation delays.*

Fig. 5 shows the overall LAR-distance between the clean speech and the outputs of CTRANC and ACTRANC system. For comparison, the LAR-distance between the clean speech and the noisy speech at the primary channel is also plotted. Superior performance of ACTRANC over the conventional CTRANC is obvious in this figure. This demonstrates the effect of inserted delay unit in relaxing the sensitivity of the system to the noise source placement. In addition, we see that CTRANC output has LAR-distances larger than noisy speech. This can also be justified by considering the sensitivity of CTRANC to the placement of noise source.

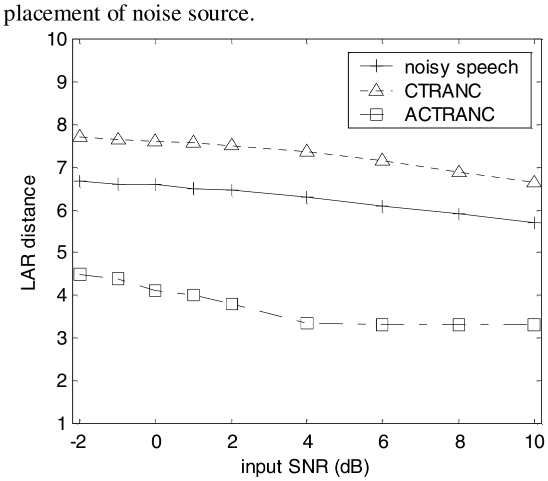

Fig. 5. *LAR distance for noisy speech, outputs of CTRANC and ACTRANC.*

To demonstrate superior performance of ACTRANC over conventional CTRANC, we have also plotted the sample waveforms of clean speech, noisy speech (at SNR = 2 dB), and the outputs of ACTRANC and CTRANC systems in Figs. 6-a to 6-d, respectively. Obviously, the ACTRANC has reduced the noise and resulted in an output similar to the clean speech.

In the last experiment, the effect of delay value (Δ) on the performance of the ACTRANC system was examined. This issue is directly related to the causality constraint on adaptive filters. In this test, we assume that $\delta_{max} = 5, \delta = -1$ and SNRs of the primary channel and reference one are 2 dB and -6 dB, respectively. Table 1 shows the output SNR for different values of Δ. It is obvious that for $\Delta > \delta_{max}$ or $\Delta < \delta$, the system requires non-causal filters. This drastically degrades the performance of ACTRANC system.

Table 1. *The output SNR of ACTRANC for different delay values.*

Δ	Output SNR
-3	-16.5
0	17.6
10	-4.5

5. CONCLUSION

In this research, we examined an asymmetric cross-talk resistant adaptive noise canceller. It was shown that in the presence of oriented noises, the asymmetric structure has very better results rather than the conventional symmetric one.

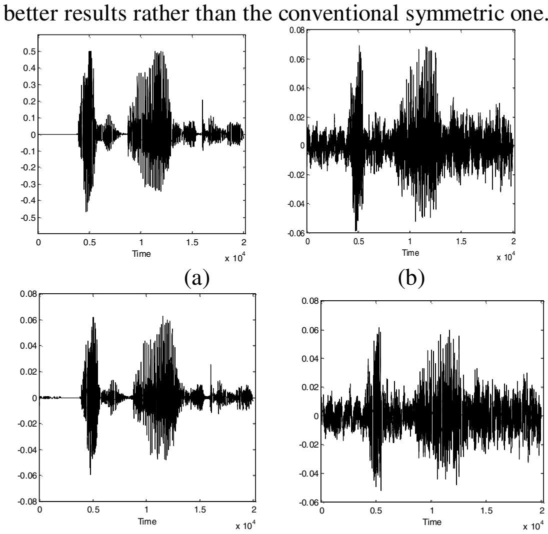

Fig. 6. *Sample waveforms of (a) clean speech, (b) noisy speech (at SNR = 2 dB), (c) output of ACTRANC system, and (d) output of CTRANC system.*

The presented system can also be interpreted as a Blind Source Separation (BSS) technique [10]. Currently, we are investigating the use of more complicated BSS methods to produce enhanced speech from two noisy recordings in a reflective environment [10]. It is expected that this system to have better performance rather than current ACTRANC [11,12].

6. References

[1] E. Weinstein, M. Feder, and A. V. Oppenheim, “Multi-channel signal separation by decorrelation,” *IEEE Trans. on Speech and Audio Processing,* vol. 1, no. 4, 1993.

[2] G. Mirchandani, R. L. Zinser, Jr., and J. B. Evans, “A new adaptive noise cancellation scheme in the presence of Crosstalk,” *IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 39, no. 10, pp. 681-694, Oct. 1992.

[3] S. V. Gerven, and D. V. Compernole, “Signal separation by symmetric adaptive decorrelation: Stability, convergence and uniqueness,” *IEEE Trans. on Signal Processing*, vol. 43, no. 7, pp. 1602-1612, July 1995.

[4] D. V. Compernolle, S. V. Gerven, “Signal separation in a symmetric adaptive noise canceler by output decorrelation,” in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 1992.

[5] S. M. Kou, and W. M. Peng, “Principle and application of asymmetric crosstalk resistant adaptive noise canceler,” in *Proc. of the IEEE Workshop on Signal Processing Systems*, Oct. 1999.

[6] M. Gabrea, “Double affine projection algorithm-based speech enhancement algorithm” in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Hong Kong, China, Apr. 2003.

[7] H. R. Abutalebi, H. Sheikhzadeh, R. L. Brennan, and G. H. Freeman, “Affine projection algorithm for oversampled subband adaptive filters,” in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Hong Kong, China, Apr. 2003.

[8] S. R. Quackenbush, T. P. Banwell and M. A. Clements, *Objective Measures of Speech Quality*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[9] L. Parra, C. Spence, “Convolutive blind separation of nonstationary sources,” *IEEE Trans. on speech and Audio Processing,* vol. 8, no. 3, pp.320-327, MAY. 2000.

[10] D. T. Pham, C. Serviere, H. Boumaraf, “Blind separation of convolutive audio mixtures using nonstationarity,” in Proc. of 4th Internationa Symposium on Independent Component Analysis and Blind Signal Separation (ICA), pp. 981-986, Apr. 2003.

[11] L. Parra, C. Spence, B. D. Vries, “Convolutive blind source separation based on multiple decorrelation,” *IEEE Workshop Neural Networks Signal Processing*, Cambridge, UK, Sept. 1998.

[12] E. Visser, M. Otsuka, T. W. Lee, “A spatio-temporal speech enhancement scheme for robust speech recognition in noisy environments,” *Speech Comunication*, vol.41, pp. 393-407, Dec. 2002.