LOW DISTORTION DECOUPLED CROSSTALK RESISTANT ADAPTIVE NOISE CANCELLER

Ludovick Lepauloux[1], *Pascal Scalart*[2] *and Claude Marro*[1]

[1]Orange Labs - TECH/SSTP, 2 Avenue Pierre Marzin, 22307 Lannion Cedex, France
[2]ENSSAT - IRISA/CAIRN, 6 Rue de Kerampont, B.P. 447, 22305 Lannion Cedex, France
E-mail: ludovick.lepauloux,claude.marro@orange-ftgroup.com; pascal.scalart@enssat.fr

ABSTRACT

This paper addresses the problem of adaptive noise cancelling with low signal-distortion. The proposed canceller consists in adding an adaptive decoupling filter to the well-known crosstalk resistant adaptive noise canceller (CTRANC) while maintaining the benefits of the cross-coupled structure. It uses a new cross-coupled structure named decoupled CTRANC (DCTRANC) suitable for situations where the noise reference sensor is closely spaced relative to the primary one. This method is analyzed and results are given in voice communication context.

***Index Terms*—** noise canceller, noise reduction, distortion, crosstalk, adaptive filter, speech enhancement

1. INTRODUCTION

The quality of signal enhancement systems, based on adaptive filtering, is highly dependent on the ”quality of the noise reference”. Any amount of signal leakage into the noise reference quickly results in signal distortion and poor noise cancellation. Many algorithms have been proposed in order to deal with crosstalk in adaptive noise canceller (ANC). Each of them can be characterized according to its structure. In fact, we can differentiate *forward* from *backward* methods. In this work we focus on the second one and propose both a new structure and an algorithm which reduce signal distortion and improve noise cancellation by reducing gradient estimation noise and misadjustment errors at the same time. In their work [1], Al-Kindi and Dunlop have shown that the gradient noise decreases the performance of the cross-coupled canceller known as CTRANC, for crosstalk resistant adaptive noise canceller defined in [2], especially when the filters adaptation is done jointly. In addition, as proved by Ikeda et al.[3, 4], misadjustment errors in the adaptive filters may introduce reverberation. To cope with this problem, they proposed an ANC (Figure 1) based on the estimation of the signal-to-noise-ratio (SNR) so as to control the step sizes of the two main adaptive filters (W1, W2). Further development regarding signal distortion were given by Sato et al. in [5] with introduction of time variable subfilters step sizes (μ_{SAF1},μ_{SAF2}), dedicated to SNR estimation.
In this paper we address the problem of reducing both reverberation and signal distortion caused by crosstalk and improving noise cancellation by adding a front-end adaptive decoupling filter. This new structure is called decoupled CTRANC (DCTRANC). In Section 2, the basic concept of the proposed DCTRANC with the normalized LMS (NLMS) algorithm is explained. In Section 3, performance of this new noise canceller is evaluated by computer simulations in comparison with previous works proposed by Ikeda (method A) and Sato (method B).

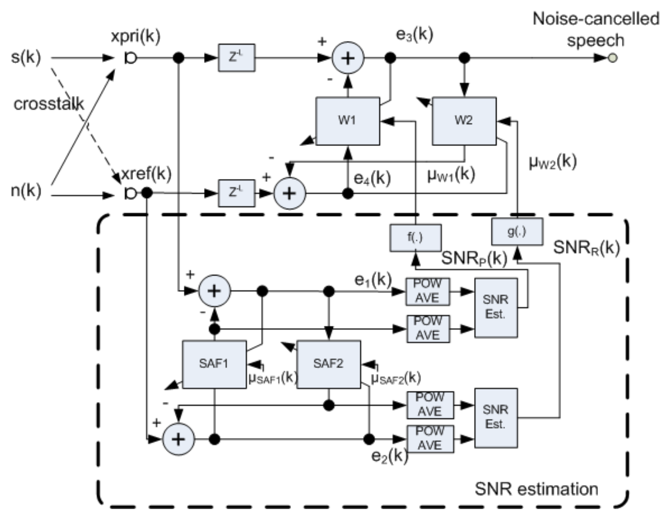

Fig. 1. Block diagram of ANC proposed by Ikeda.

2. PROPOSED CTRANC

2.1. Classical CTRANC

The classical CTRANC algorithm consists in a dual joint process estimator, where the primary signal $xpri(k)$ and the reference signal $xref(k)$ can be written as

$$\begin{aligned} xpri(k) &= s(k) + n_P(k) \\ &= s(k) + \sum_{j=0}^{N_1-1} h_j(k) n(k-j) \end{aligned} \quad (1)$$

$$\begin{aligned} xref(k) &= n(k) + c_R(k) \\ &= n(k) + \sum_{j=0}^{N_2-1} g_j(k) s(k-j) \end{aligned} \quad (2)$$

where k is the time index, N_1 and N_2 are the lengths of the noise and crosstalk paths (resp. h and g). The crosstalk signal on the reference is represented by $c_R(k)$ and the noise signal on the primary signal correlated with the noise on the reference is $n_P(k)$. The relations between the inputs and the outputs of the CTRANC system are

$$e_4(k) = xref(k) - \sum_{j=0}^{L_2-1} w_j^{(2)}(k) e_3(k-j) \quad (3)$$

where $w_j^{(2)}(k)$ represents the jth coefficient of the adaptive filter $W2$ at time k and L_2 is the corresponding length. These coefficients are updated according to the equation

$$w_j^{(2)}(k+1) = w_j^{(2)}(k) + \mu_{W2}(k) * \frac{e_4(k)\mathbf{E}_3(k)}{\mathbf{E}_3^T(k)\mathbf{E}_3(k)} \tag{4}$$

where μ_{W2} is the step size and $(.)^T$ denotes the transpose operator. In the same way, the output signal is written as

$$e_3(k) = xpri(k) - \sum_{j=0}^{L_1-1} w_j^{(1)}(k)e_4(k-j) \tag{5}$$

where L_1 is the number of taps of $W1$ which are updated according to

$$w_j^{(1)}(k+1) = w_j^{(1)}(k) + \mu_{W1}(k) * \frac{e_3(k)\mathbf{E}_4(k)}{\mathbf{E}_4^T(k)\mathbf{E}_4(k)} \tag{6}$$

where

$$\mathbf{E}_4^T(k) = [e_4(k)\ e_4(k-1)\ldots e_4(k-L2+1)] \tag{7}$$

$$\mathbf{E}_3^T(k) = [e_3(k)\ e_3(k-1)\ldots e_3(k-L1+1)] \tag{8}$$

are the delayed input sequences to the two filters.

2.2. Gradient estimation noise control

As an instructive example we consider in the following the case where the sensors SNRs are different (35 dB and 15 dB). The input SNRs are computed using the ITU-T recommendation P.56 speech voltmeter (SV56). Babble noise was used as a noise source, and clean male voice as a signal source with a sample rate of 8 kHz. Synthetic noise path and crosstalk path impulse responses are shown on Figure 2. The noise component is generated by convolution of the

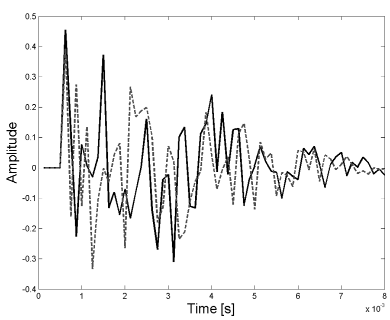

Fig. 2. Impulse responses of noise path h (solid line) and crosstalk path g (dashed line).

noise source with the noise path, and then added to the speech signal to create a noise-contaminated signal. The reference signal is generated by adding the noise to the crosstalk generated by convolution of the speech signal and the crosstalk path. Other parameters are shown in Table 1. The drawback of all these methods is the need to adjust precisely the different thresholds (SNR_{Wx}) of the two main adaptive filters ($W1$, $W2$) for the gradient step sizes control. Parameters given in Table 1 have been chosen in order to avoid divergence at the output and ensure optimal pursuit of the time-varying noise for each of these methods for our test conditions.

As said before, gradient estimation noise is crucial for cross-coupled canceller. To reduce it, we proposed to add a constraint that disable adaptation of μ_{W1} during periods when the adaptation of μ_{W2} is

Parameter	Value	Parameter	Value
N	64	M	128
L	64	Q	512
μ_{SAF1}	0.01	μ_{SAF2}	0.001
$SNR_{SF1/2min}$	-7 dB	$SNR_{SF1/2max}$	5 dB
SNR_{W1min}	5 dB	SNR_{W1max}	15 dB
μ_{W1min}	0.002	μ_{W1max}	0.1
SNR_{W2min}	-8 dB	SNR_{W2max}	14 dB
μ_{W2min}	0.002	μ_{W2max}	0.1
L_1	64	L_2	64

Table 1. Parameters settings.

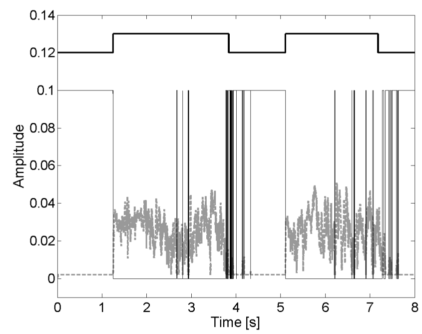

Fig. 3. Step sizes behavior: μ_{W1} (solid line) and μ_{W2} (dashed line).

allowed as shown on Figure 3. Hence, it works as a vocal activity detection[1] (VAD) which ensures that $W2$ is adapted only during period when crosstalk is important, thereby eliminating gradient noise in the adaptation process. In order to evaluate the behavior of the different methods, we use the following objective measures:

- Normalized energy system mismatch:

$$\Delta W(k) = 10log_{10}\left[\frac{\sum_{j=0}^{N-1}(w_j(k)-h_j(k))^2}{\sum_{j=0}^{N-1}h_j^2(k)}\right]$$

- Normalized output:

$$R(k) = 10log_{10}\left[\frac{\sum_{j=0}^{Q-1}e_3^2(k-j)}{\sum_{j=0}^{Q-1}xpri^2(k-j)}\right]$$

- Output distortion:

$$D(k) = 10log_{10}\left[\frac{\sum_{j=0}^{Q-1}(e_3(k-j)-s(k-j))^2}{\sum_{j=0}^{Q-1}s^2(k-j)}\right]$$

From Figure 4, we can see that the convergence of $W1$ is more stable when the proposed adaptation rule is introduced. In addition, we can see that all methods fail in the estimation of the crosstalk path since no one stand below 0 dB. Nevertheless, method A keeps on going far away the optimal solution while the proposed algorithmic constraint gives a better solution than method B. The average distortion is displayed on Figure 5. We can see that the proposed method provides a distortion of -17.6 dB thus improving the performance by 9.5 dB and 14.6 dB in comparison with method B and A respectively. However, the normalized output is more disturbed and peaks especially

[1]Note that in the whole paper the upper curve is drawn only to show speech activity and it is not used in the algorithm.

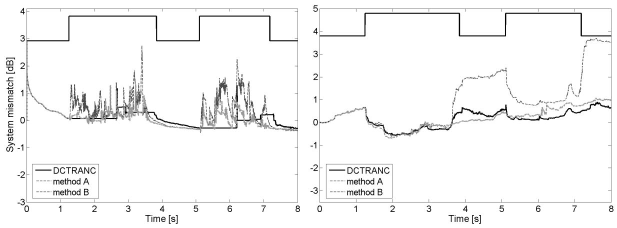

Fig. 4. Normalized energy system mismatch of W1 (left) and W2 (right).

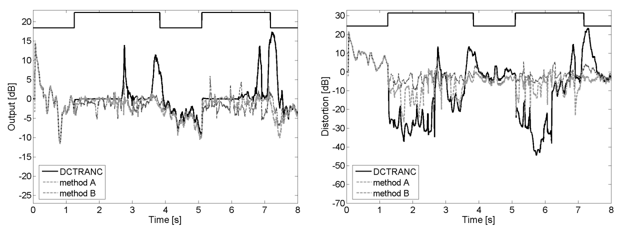

Fig. 5. Normalized output (left) and distortion (right).

at the end may cause some impulsive amplification. This is due to the fact that the filter adaptation is frozen, thus rapid change caused by a strong signal component may introduce misadjustment. At this point, it is obvious that we need to speed up convergence and reduce even more misadjustment. That is realized thanks to an additional filter, which is explained in the following section.

2.3. Mismatch compensation thanks to decoupling filter

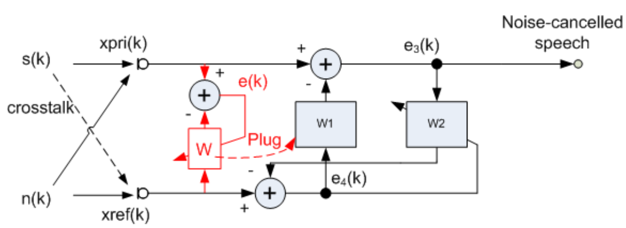

Fig. 6. Block diagram of DCTRANC.

In contrast with the usual CTRANC's equations (3,5), the output signals of the DCTRANC at time k are given by (see Figure 6)

$$e(k) = xpri(k) - \sum_{j=0}^{L_1-1} w_j(k)xref(k-j) \tag{9}$$

$$e_4(k) = xref(k) - \sum_{j=0}^{L_2-1} w_j^{(2)}(k)e_3(k-j) \tag{10}$$

Consequently, the time update equations are

$$w_j(k+1) = w_j(k) + \mu_{W1}(k) * \frac{e(k)\mathbf{E}(k)}{\mathbf{E}^T(k)\mathbf{E}(k)} \tag{11}$$

$$w_j^{(2)}(k+1) = w_j^{(2)}(k) + \mu_{W2}(k) * \frac{e_4(k)\mathbf{E}_3(k)}{\mathbf{E}_3^T(k)\mathbf{E}_3(k)} \tag{12}$$

where

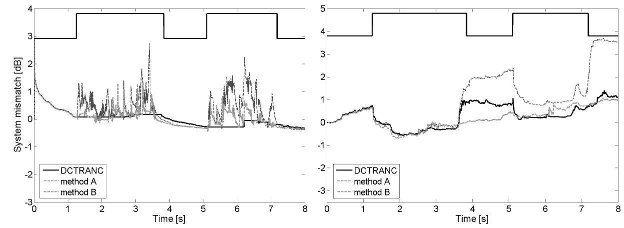

Fig. 7. Normalized energy system mismatch of W1 (left) and W2 (right).

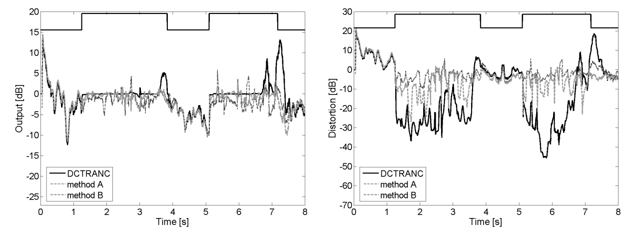

Fig. 8. Normalized output (left) and distortion (right).

$$\mathbf{E}^T(k) = [xref(k)\ xref(k-1) \dots xref(k-L1+1)] \tag{13}$$

$$\mathbf{E}_3^T(k) = [e_3(k)\ e_3(k-1) \dots e_3(k-L2+1)] \tag{14}$$

are the delayed input sequences to the two filters. For each time sample k, $W1$ is a duplicate version of the decoupling filter W, in other words

$$W1(k) = W(k) \tag{15}$$

Equations (9, 10, 11, 12 and 15) also define the new algorithm. On Figure 7 properties of the LMS algorithm are easily visible during speech period where the normalized energy system mismatch of $W1$ is disturbed. Indeed, for a given filter length, small misadjustment and large adaptation are conflicting design objectives. Figure 8 shows how the decoupling filter W helps to solve misadjustment problem. We can see that peaks in the normalized output are reduced and average distortion is further improved and reached -22.3 dB. Simulations on speech signal of 16 seconds duration prove that impulse amplification is highly reduced. Here, we can note that $W2$ is not well learned whatever the method used.

3. MEAN BEHAVIOR

Performance of the proposed ANC was evaluated by computer simulations from the viewpoints of noise reduction and speech distortion in comparison with method A and method B for different layouts. We use car noise and babble noise as noise sources. Figures 9 to 11 show the normalized output and the distortion for various SNR in car noise. For all the considered cases, we can see that DCTRANC is more efficient than methods A and B. Nevertheless, the solution presented here may suffer from instability with more adverse noise conditions. In contrast with the two other methods we need to adjust thresholds, for example when SNR_{pri}= 10 dB and SNR_{ref}= -10 dB. Basically, the same behavior occurs for method A and method B when crosstalk is significant, more precisely when SNR_{ref} is higher than 15 dB. This has been overcame thanks to the thresholds we have chosen.

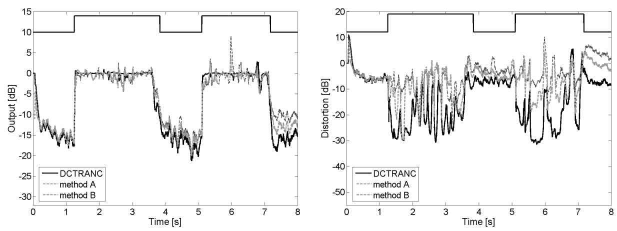

Fig. 9. Normalized output (left) and distortion (right), SNR=(25,5).

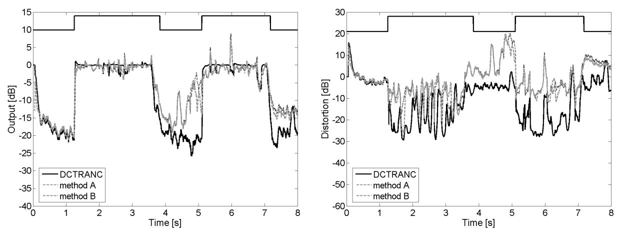

Fig. 10. Normalized output (left) and distortion (right), SNR=(20,0).

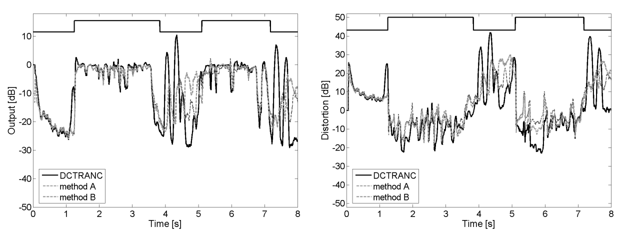

Fig. 11. Normalized output (left) and distortion (right), SNR=(10,0).

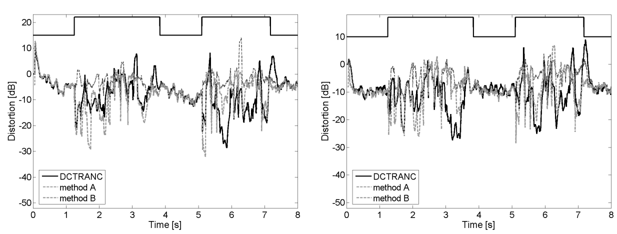

Fig. 12. Distortion: babble noise (left) and car noise (right), SNR=(35,25).

Figure 12 shows the ability of our algorithm to deal with severe crosstalk. Finally, to evaluate the robustness of the proposed method we give in Tables 2-3 distortion values for the two noise cases chosen and two specific configurations that simulate difficult noise and crosstalk conditions at the same time. For those conditions, DCTRANC remains the more efficient.

Babble noise	**Distortion [dB]**		
SNR (Pri,Ref) [dB]	DCTRANC	method A	method B
(15,5) min.	-29	-15.3	-20
(15,5) ave.	-11.6	-3.3	-6.7
(15,10) min.	-26.5	-13	-19.6
(15,10) ave.	-8.6	-3.5	-6

Table 2. Minimal (min.) and average (ave.) distortion for different configurations in babble noise.

Car noise	**Distortion [dB]**		
SNR (Pri,Ref) [dB]	DCTRANC	method A	method B
(15,5) min.	-25.7	-20.8	-23.2
(15,5) ave.	-13.7	-5.9	-8
(15,10) min.	-24.8	-19.3	-23.9
(15,10) ave.	-12	-5.2	-7.8

Table 3. Minimal (min.) and average (ave.) distortion for different configurations in car noise.

4. CONCLUSION

A new adaptive noise canceller (ANC) with low signal distortion named DCTRANC has been proposed. It has a front-end adaptive decoupling filter which coupled with an algorithmic constraint reduces gradient estimation noise and misadjustment. Results of computer simulations show that the proposed ANC improves signal distortion by up to 13 dB in car noise. However, two main drawbacks have been pointed out. In case of difficult noise conditions, instability can be observed without change in threshold values for the gradient step sizes control. In addition, none of these methods correctly identify crosstalk path. This point will be address in future research.

5. ACKNOWLEDGMENT

The authors express gratitude to the anonymous reviewers for their helpful comments.

6. REFERENCES

[1] M.J. Al-Kindi and J. Dunlop, "Improved adaptive noise cancellation in the presence of signal leakage on the noise reference channel," *Signal Processing*, vol. 17, no. 3, pp. 241–250, 1989.

[2] Richard L. Zinser and Joseph B. Evans, "Some experimental and theoretical results using a new adaptive filter structure for noise cancellation in the presence of crosstalk," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, April 1985, vol. 10, pp. 1253–1256.

[3] Shigeji Ikeda and Akihiko Sugiyama, "An adaptive noise canceller with low signal-distorsion in the presence of crosstalk," *IEICE Transactions on Fundamentals*, vol. E82-A, no. 8, pp. 1517–1525, August 1999.

[4] Shigeji Ikeda and Akihiko Sugiyama, "An adaptive noise canceller with low signal distortion for speech codecs," *IEEE Transactions on Signal Processing*, vol. 47, no. 3, pp. 665–674, March 1999.

[5] Miki Sato, Akihiko Sugiyama, and Shin'ichi Ohnaka, "An adaptive noise canceller with low signal-distortion based on variable stepsize subfilters for human-robot communication," *IEICE Transactions on Fundamentals*, vol. E88-A, no. 8, pp. 2055–2061, August 2005.