DOUBLE AFFINE PROJECTION ALGORITHM-BASED SPEECH ENHANCEMENT ALGORITHM

M. Gabrea

École de Technologie Supérieure
Electrical Engineering Department
1100, Notre-Dame West
Montreal, Quebec
H3C 1K3, Canada
mgabrea@ele.etsmtl.ca

ABSTRACT

In this paper a symmetric feedback implementation scheme of a two microphones speech enhancement is presented. We consider the coupling systems modeled as a linear time-invariant Finite Impulse Response (FIR) filters and propose a new recursive-based adaptive filter solution to enhance the noisy speech. The optimum filter weight adaptation is based on a Double Affine Projection Algorithm (DAPA). This approach can be extended for a subclass of signal separations where the direct link is stronger than the interference link in the both channels. A comparative study with other adaptive algorithms shows the superiority of the DAPA in performances improvement.

1. INTRODUCTION

Let us consider the system modelled by the diagram represented in the figure 1. The purpose is to recover the free noise speech signal $s(n)$ from the two available observations $p_1(n)$ and $p_2(n)$ in the presence of the noise signal $b(n)$.

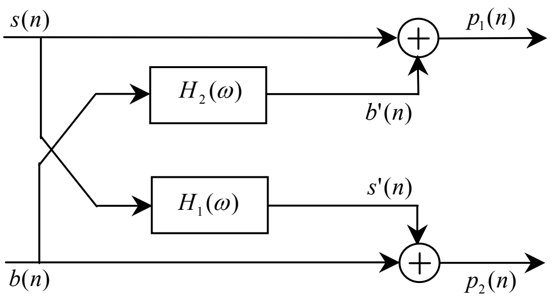

Figure 1. Signal model for noise cancellation

The general technique of adaptive noise canceling has been applied successfully to a number of problems. The initial work on adaptive noise canceling began in the 1960s. Adaptive noise canceling refers to a class of adaptive enhancement algorithms based on the availability of a primary input source and a secondary reference source ($H_1(\omega)=0$). The primary input source $p_1(n)$ is assumed to contain the speech signal $s(n)$ plus an additive noise $b'(n)$, and the secondary or the reference is assumed to contain only a realization of a stochastic process $b(n)$ that is correlated with the noise $b'(n)$ but not with the speech signal $s(n)$. The basic scheme of adaptive noise canceller given in [17] uses an adaptive filter based on the Least Mean Squares (LMS) algorithm for estimating the additive noise, which is then subtracted from the primary input (see Figure 2). One problem with the adaptive noise canceling algorithm is the need for the reference microphone to be well separated from the primary microphone, so that it picks up as little speech as possible. If the microphones are too close to one another, cross talk occurs and a typical adaptive filter will thereby suppress a portion of the input speech characteristics. One means of addressing this problem is to place a second adaptive filter in the feedback loop.

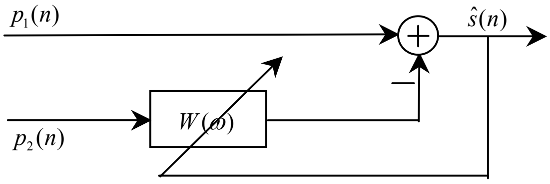

Figure 2. Basic scheme of adaptive noise canceller

In the simplified case where the filters $H_1(\omega)$ and $H_2(\omega)$ are assumed to be single tap another system called Symmetric Adaptive Decorrelation (SAD) using two adaptive filters, as an extension of the classical LMS acoustic noise canceller, has been presented in [2]. This result has been later generalized to a convolutive mixtures modeled by two FIR filters $H_1(\omega)$ and $H_2(\omega)$ [8].

Adaptive algorithms for separation of wide-band signal, under the condition of fourth-order white noise, for convolutive mixtures modeled by FIR filters has been proposed in [11][12] based on the cancellation of 4th-order output cross-cumulants.

The feedback implementation of an adaptive noise canceller (see Figure 3) has been proposed in [15] using Double Least Mean Squares (DLMS) algorithm. Other noise cancellers using two adaptive filters: feedforward and feedback symmetric adaptive noise canceller have been described in [4][3][16][14][10][9].

In this paper we present a new feedback implementation of a noise canceller based on the DAPA algorithm. We only suppose that the speech signal and the noise are statistically independents and we consider the coupling systems being FIR filters. This algorithm can also be used for a subclass of signal separations where the direct link must be stronger than the interference link in the both channels. A comparative performance study is presented in the framework of noise cancellation.

The remainder of the paper is organized as follows. In the next section we present the Double Fast Affine Projection Algorithm. A comparative experimental study of different schemes and algorithms is presented in section 3. We conclude by evaluating the performance of the proposed system.

2. THE DOUBLE FAST AFFINE PROJECTION (DAPA) ALGORITHM

Figure 3 shows the feedback implementation of the noise canceller. $W_1(\omega)$ and $W_2(\omega)$ are two adaptive filters. Each one has as input the output error signal of the other filter. $W_1(\omega)$ is an adaptive filter which has an input signal $s_1(n)$, a desired signal $p_2(n)$ and an error signal $s_2(n)$. $W_2(\omega)$ is an adaptive filter which has an input $s_2(n)$ and an error signal $s_1(n)$.

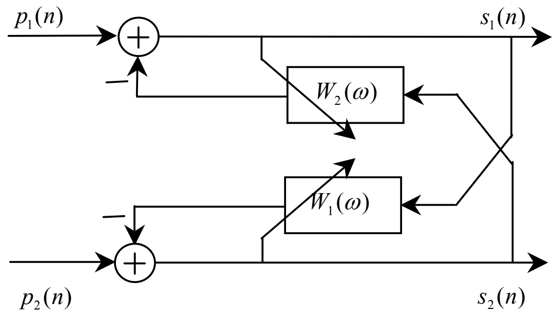

Figure 3. Feedback implementation of the noise canceller

The optimum values in the Wiener sense, in the case of wide sense stationary processes and in term of the power density spectrum, of the filters $W_1(\omega)$ and $W_2(\omega)$ are given by [13]:

$$W_1(\omega)=\frac{S_{p_2s_1}(\omega)}{S_{s_1s_1}(\omega)} \tag{1}$$

$$W_2(\omega)=\frac{S_{p_1s_2}(\omega)}{S_{s_2s_2}(\omega)} \tag{2}$$

and if we suppose that the speech signal $s(n)$ and the noise $b(n)$ are two uncorrelated processes we can rewrite (1) and (2) as follows:

$$\begin{aligned} S_{s_1s_1}(\omega)\left|1-W_1(\omega)W_2(\omega)\right|^2 &= S_{ss}(\omega)\left|1-W_2(\omega)H_1(\omega)\right|^2 \\ &+S_{bb}(\omega)\left|W_2(\omega)-H_2(\omega)\right|^2 \end{aligned} \tag{3}$$

$$\begin{aligned} S_{s_2s_2}(\omega)\left|1-W_1(\omega)W_2(\omega)\right|^2 &= S_{bb}(\omega)\left|1-W_1(\omega)H_2(\omega)\right|^2 \\ &+S_{ss}(\omega)\left|W_1(\omega)-H_1(\omega)\right|^2 \end{aligned} \tag{4}$$

or [5]:

$$\begin{aligned} &S_{ss}(\omega)\big(H_1(\omega)-W_1(\omega)\big)\big(1-H_1^*(\omega)W_2^*(\omega)\big)/D_i(\omega)+ \\ &S_{bb}(\omega)\big(H_2^*(\omega)-W_2^*(\omega)\big)\big(1-W_1(\omega)H_2(\omega)\big)/D_i(\omega)=0,\ i=1,2 \end{aligned} \tag{5}$$

$$D_1(\omega)=S_{ss}(\omega)\left|1-H_1(\omega)W_2(\omega)\right|^2+S_{bb}(\omega)\left|H_2(\omega)-W_2(\omega)\right|^2 \tag{6}$$

$$D_2(\omega)=S_{ss}(\omega)\left|H_1(\omega)-W_1(\omega)\right|^2+S_{bb}(\omega)\left|1-H_2(\omega)W_1(\omega)\right|^2 \tag{7}$$

We can see that the equations (5) provide multiple solutions. Among all these solutions we can find the "*desired solution*" $W_i(\omega)=H_i(\omega)$, $i=1,2$. In this case it is easy to verify that $s_1(n)=s(n)$ and $s_2(n)=b(n)$ and it is possible to recover the signals that would have been measured at each microphone in the absence of the other source signal.

If for each generating filter:

$$\sum h_i^2(n)<1,\ i=1,2 \tag{8}$$

then the filters $W_i(\omega)$ ($i=1,2$) converge to the desired solutions.

These desired solutions can be reached using a weight adaptive filters updating based on the LMS or RLS algorithm. We propose to use the APA algorithm for the following reasons. The affine projection algorithm is a generalization of the well-known Normalized Least Mean Square (NLMS) algorithm [6]. Under this interpretation, each tap weight vector update of NLMS is viewed as a one-dimensional affine projection. In APA the projections are made in multiple dimensions. As the projection dimension increases, so does the convergence speed of the tap weight vector, and unfortunately, the algorithm's computational complexity. Using techniques similar to those which led to fast recursive least squares [1], a fast version of APA may be derived [7]. The affine projection algorithm, is a relaxed and regularized form [7]. The Double Affine Projection Algorithm (DAPA) is defined as follows:

- *Filtering*:

$$\underline{s}_2(n)=\underline{p}_2(n)-\boldsymbol{S}_1^t(n)\underline{w}_1(n-1) \tag{9}$$

$$\underline{s}_1(n)=\underline{p}_1(n)-\boldsymbol{S}_2^t(n)\underline{w}_2(n-1) \tag{10}$$

where:

$$\underline{p}_1(n)=\left[p_1(n),\cdots,p_1(n-L+1)\right]^t$$

$$\underline{p}_2(n)=\left[p_2(n),\cdots,p_2(n-L+1)\right]^t$$

$$\underline{s}_1(n)=\left[s_1(n),\cdots,s_1(n-L+1)\right]^t$$

$$\underline{s}_2(n)=\left[s_2(n),\cdots,s_2(n-L+1)\right]^t$$

$$\boldsymbol{S}_1(n)=\left[\underline{s}_1(n),\cdots,\underline{s}_1(n-N_1+1)\right]^t$$

$$\boldsymbol{S}_2(n)=\left[\underline{s}_2(n),\cdots,\underline{s}_2(n-N_2+1)\right]^t$$

$$\underline{w}_1(n) = \left[w_{1,0}(n), \cdots, w_{1,N_1-1}(n)\right]^t$$

$$\underline{w}_2(n) = \left[w_{2,0}(n), \cdots, w_{2,N_2-1}(n)\right]^t$$

- *Filters update*:

$$\underline{w}_1(n) = \underline{w}_1(n-1) + \mu_1 \boldsymbol{S}_1(n)\left[\boldsymbol{S}_1^t(n)\boldsymbol{S}_1(n) + \delta_1 \boldsymbol{I}\right]^{-1} \underline{s}_2(n) \quad (11)$$

$$\underline{w}_2(n) = \underline{w}_2(n-1) + \mu_2 \boldsymbol{S}_2(n)\left[\boldsymbol{S}_2^t(n)\boldsymbol{S}_2(n) + \delta_2 \boldsymbol{I}\right]^{-1} \underline{s}_1(n) \quad (12)$$

The scalars δ_i ($i = 1,2$) is the regularization parameters for the sample autocorrelation matrix inverse used in (11) and (12). Where $\boldsymbol{S}_i^t(n)\boldsymbol{S}_i(n)$ may have eigenvalues close to zero, creating problems for the inverse, $\boldsymbol{S}_i^t(n)\boldsymbol{S}_i(n) + \delta_i \boldsymbol{I}$ has δ_i as its smallest eigenvalue which, if large enough, yield a well behaved inverse [7]. The step-size parameter, μ_i ($i = 1,2$) is the relaxation factor. As in NLMS, the algorithm is stable for $0 \le \mu_i < 2$.

3. SIMULATION RESULTS

The noise has been separately recorded in a car moving in five different conditions, the microphone is placed in front of the driver and the noises have been artificially added to the noise-free speech so that one would master the SNR input. The coupling systems are 10 taps two FIR filters with.

An example of one signal captured by the first microphone $p_1(n)$ and another by the second microphone $p_2(n)$ is respectively shown in figure 4a and 4b. In this case the SNR of $p_1(n)$ and $p_2(n)$ are respectively 3.09 dB and 3.79 dB.

The output signal of the noise canceller system using the DAPA algorithm ($\mu_1 = \mu_2 = 1$) is shown in the figure 5. The first original desired speech signal $s_1(n)$ is shown in the figure 6.

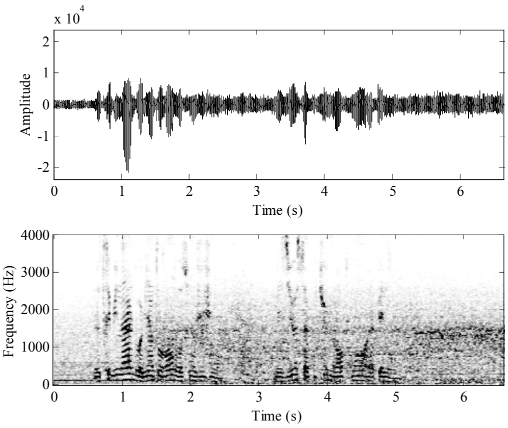

Figure 4a: The signal $p_1(n)$ captured by the first microphone and its spectrogram (SNR = 3.09 dB)

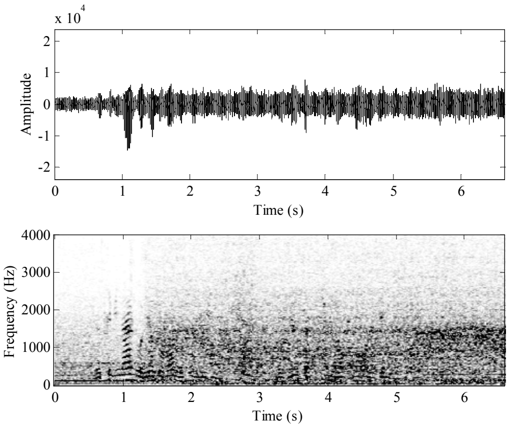

Figure 4b: The signal $p_2(n)$ captured by the second microphone and its spectrogram (SNR = 3.79 dB)

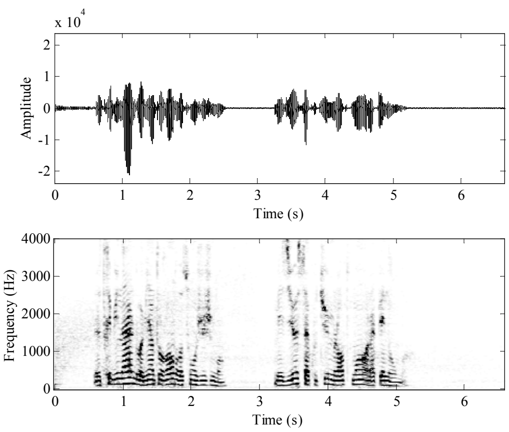

Figure 5: Enhanced speech $s_1(n)$ obtained with the noise canceller system based DAPA algorithm and its spectrogram (L = 4, $\mu_1 = \mu_2 = 1$, SNR = 15.70 dB)

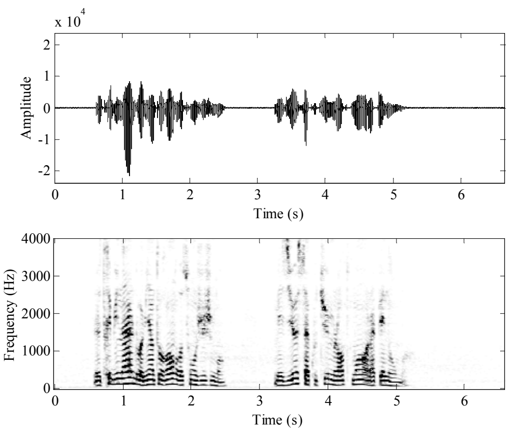

Figure 6: The original speech signal $s(n)$

A comparative SNR output gain between the Double SAD [8], the Extended LMS [16], the Double LMS [15] and the DAPA algorithms is provided in table 1. This table shows the superiority of the noise canceller DAPA based algorithm.

Case	Input SNR (dB) $p_1(n)$	Gain SNR (dB) $s_1(n)$				
		SAD [8]	ELMS [16]	DLMS [15]	DAPA *L*=2	DAPA *L*=4
1	10.48	7.65	9.10	9.77	10.73	10.86
2	2.45	13.54	13.48	13.22	14.32	14.43
3	11.40	8.30	8.36	7.93	8.86	8.98
4	9.97	11.18	11.15	12.35	13.59	13.72
5	3.09	13.05	12.06	14.74	15.70	15.87

Table 1: The SNR gain of $s_1(n)$ for different algorithms

This global performance behaviour is confirmed also by the frame by frame SNR output. The SAD algorithm and the ELMS algorithms take more time before handling the noise field after which its segmental SNR behaviour is close to the segmental behaviour of the DAPA algorithm. As for the Normalized LMS [17], its segmental SNR behaviour is always lower since it is penalized by its slow convergence and therefore can't track the statistical change of the noise between two successive frames. However, during high-energy regions, its behaviour is close to the two other algorithms. The reason is that noise is masked by the high-energy speech regions, and hence does not require complex treatment.

4. CONCLUSION

In conclusion, in this paper we have presented a noise canceller system based on the Double Affine Projection algorithm. Different aspects, such as the convergence, global and segmental SNR and subjective quality, have been considered. We have shown the superiority of the presented algorithm compared to the Double SAD, the Extended LMS and the Double LMS algorithm. Furthermore, the structure based on the coupling FIR filters permits the DAPA algorithm to be also used as signal separators or signal deconvolvers rather than only a simple noise canceller.

Informal quality and intelligibility tests indicate also significant superiority of such algorithm to enhance speech signal.

We should remark that the discussions about a complete mathematical convergence analysis are given in [5]. In this short paper we have preferred to focalise the presentation on the case where the physical solutions of the equations (1) and (2) are possible.

5. REFERENCES

[1] Benallal A. and Gilloire A. "A New Method to Stabilize Fast RLS Algorithms Based on a First-Order Model of the Propagation of Numerical Errors," *IEEE International Conference on Acoustics, Speech, and Signal Processing 1992,* New York, NY, April 1988.

[2] Compernolle D.V. and Gerven S.V. "Blind Separation of Sources: Signal Separation in a Symmetric Adaptive Noise Canceller by Output Decorrelation", *IEEE International Conference on Acoustics, Speech, and Signal Processing 1992*, vol. IV, pp. 221-224.

[3] Gabrea M., E. Mandridake, M. Menez, M. Najim and A. Vallauri, "Two Microphones Speach Enhancement System Based on a Double Fast Recursive Least Squares (DFRLS) Algorithm", *Proc. of EUSIPCO 1996*, Trieste, Italy, September 1996, vol. 2, pp. 983-986.

[4] Gabrea M., "A Double RLS Noise Canceller and Some Related Schemes", *Proc. of ICECS 1996*, Rodos, Greece, October 1996.

[5] Gabrea M. "Rehaussement de la parole en ambience bruitée. Méthodes monovoie et bivoie". *Ph.D. thesis*, University of Bordeaux 1, France (in French).

[6] Gay S.L. "Fast Projection Algorithms with Application in Voice Excited Echo Cancellers", *Ph.D. thesis*, Rutgers University, Piscataway, New Jersey, October 1994.

[7] Gay S.L. and Tavathia S. "The Fast Affine Projection Algorithm", *IEEE International Conference on Acoustics, Speech, and Signal Processing 1995*, Detroit, MI, vol. 5, pp. 3023-3026, May 1995.

[8] Gerven S. V. and Compernolle D. V. "Signal Separation by Symmetric Adaptive Decorrelation: Stability, Convergence, and Uniqueness". *IEEE Transactions on Signal Processing*, vol. 43, no. 7, pp. 1602-1612, July 1995.

[9] Gustafsson T., Lindgren U. and Sahlin H. "Statistical Analysis of a Signal Separation Method Based on Second-Order Statistics," *IEEE Transactions on Signal Processing*, vol. 49, no.2, pp. 441-444 , February 2001.

[10] Lindgren U.A. and Broman H. "Source Separation Using a Criterion Based on Second-Order Statistics," *IEEE Transactions on Signal Processing*, vol. 46, no. 7, pp 1837-1850, July 1998.

[11] Nguyen-Thi H.-L. and Jutten C. "Blind Source Separation for Convolutive Mixtures," *Signal Processing*, vol. 45, pp. 209-229, 1995.

[12] Nguyen-Thi H.-L., Gerven S.V., Jutten C. and Compernolle D.V. "Stability Study for Source Separation in Convolutive Mixtures of Two Sources," *Signal Processing*, vol. 62, pp. 163-171, 1997.

[13] Therrien C. W. *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall, New Jersey 1992.

[14] Yellin D. and Weinstein E., "Criteria for Multi-Channel Signal Separation," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 2158-2168, August 1994.

[15] Vallauri A. "L'étude et le développement de méthodes de reconnaissance de la parole et de réduction du bruit, et application". *Ph.D. thesis*, Univ. Nice, France, December 1992 (in French).

[16] Weinstein E., Feder M. and Oppenheim A.V. "Multi-Channel Signal Separation by Decorrelation," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 4, pp. 405-413, October 1993.

[17] Widrow B. and al, "Adaptive Noise Canceling: Principles and Applications," *Proceedings of IEEE*, vol. 63, no. 3, pp. 1692-1716, Dec. 1975.