COMPARATIVE STUDY OF NEW BLIND SOURCE SEPARATION STRUCTURES FOR TWO-CHANNEL ACOUSTIC NOISE CANCELLATION

Mohamed Djendi [1, 2], Pascal Scalart [1] and André Gilloire [3]

[1] ENSSAT – IRISA/CAIRN, 6 Rue de Kerampont, ENSSAT, B.P. 447, 22305, Lannion, Cedex, France.
[2] Univ. Saad Dahleb, Signal Proc. and Image Laboratory (LATSI), Route de Soumaa B.P. 270, Blida 09000, Algeria
[3] Orange-Labs - TECH/SSTP, 2 Avenue Pierre Marzin, 22307 Lannion Cedex, France

Emails: {m_djendi@yahoo.fr, pascal.scalart@enssat.fr, andre.gilloire@wanadoo.fr}

ABSTRACT

This paper addresses the problem of Blind Source Separation (BSS) applied to Acoustic Noise Cancellation (ANC) schemes when two microphones are used for the sound pick-up. The proposed approach is based on an improved Forward BSS structure combined with a post-filter [1] in order to correct the inherent speech distortion brought by the Forward BSS structure. The performance of the proposed algorithm is compared to the performance of two Backward BSS structures, namely the classical Backward structure [2] and the adaptive solution proposed in [3]. Performances of the proposed algorithm are evaluated by the output SNR and the cepstral distance under various environments. Experimental results indicate that the proposed method outperforms the classical Backward structure and the adaptive one, especially in the critical case of closely spaced microphones.

1. INTRODUCTION

In the classical noise canceling structure with a noise-free reference sensor, a filter is used to approximate the transfer function between the noise source and the primary sensor. The noise from the reference sensor is then filtered and the output of the filter is subtracted from the output produced by the primary sensor. Unfortunately, when the primary and reference sensors are closely spaced, significant leakage of the primary signal can occur onto the noise reference. This reduces the effectiveness of the noise cancellation and also produces distortion of the signal components in the output. The maximum SNR obtained at the output of such a canceler is equal to the noise to signal ratio present on the reference input [4]. Some improvement is possible if the primary signal is intermittent and the filter is adapted only during periods when the primary signal is absent, but this relies on an efficient primary signal detector. Furthermore, a post-processing stage may be required to reduce signal distortion [2]. To overcome these problems, two suitable types of BSS structures, named Forward and Backward, are available. In this paper, three structures are detailed and compared. Two of them are of Backward type [2, 3] and the third one is of Forward type [1]. The paper is organized as follows. Section 2 presents the used mixing model for generating the test signals. In section 3, we describe the proposed realization of the Forward BSS structure with post-filtering. Section 4 is dedicated to the description of two Backward structures which are compared to the Forward one described in section 3. In the last section, the three structures are experimentally compared in two configurations, namely loosely and closely spaced microphones.

2. MIXING MODEL

We consider the described mixing model in Fig.1. It involves two convolutive mixtures of two uncorrelated point sources with impulse responses $h_{11}(n), h_{22}(n)$, $h_{12}(n)$ and $h_{21}(n)$. $n_1(n)$ and $n_2(n)$ represent the non-coherent parts of the diffuse acoustic (background) noise in the vicinity of the microphones plus the electronic noise in the sensors circuits.

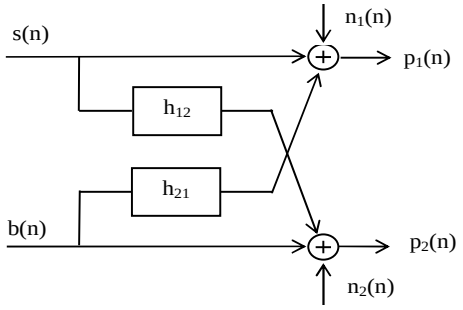

Figure 1 – The mixing model.

The model is defined as follows in the frequency domain:

$$\begin{pmatrix} P_1(\omega) \\ P_2(\omega) \end{pmatrix} = \begin{pmatrix} H_{11}(\omega) & H_{21}(\omega) \\ H_{12}(\omega) & H_{22}(\omega) \end{pmatrix} \begin{pmatrix} S(\omega) \\ B(\omega) \end{pmatrix} + \begin{pmatrix} N_1(\omega) \\ N_2(\omega) \end{pmatrix} \quad (1)$$

One of the two point sources (S) corresponds to speech (the useful signal), and the second one (B) can represent either the car noise or far-end speech that we want to cancel. $H_{11}(\omega)$ and $H_{22}(\omega)$ represent the frequency responses of each direct channel separately, and $H_{12}(\omega)$ and $H_{21}(\omega)$ represent the cross-coupling effects between the channels. $N_1(\omega)$ and $N_2(\omega)$ represent the Fourier transforms of the diffuse noise components. In this work, $h_{11}(n)$ and $h_{22}(n)$ are assumed to be identity; this assumption does not impact the practical usefulness of the model as noted in [4]. Moreover, we do not take into account the non-coherent components of the diffuse acoustic noise in the microphones vicinity (*i.e.* we assume $n_1(n) = n_2(n) = 0$).

3. FORWARD BSS STRUCTURE (FBSS)

The FBSS structure that we have investigated is shown in Fig.2. The theoretical solution of the problem is given by setting $w_{21}(n)=h_{21}(n)$ and $w_{12}(n)=h_{12}(n)$ [2]. The Least Squares solution of the problem is obtained by minimizing the MSE of $u_1(n)$ and $u_2(n)$, or equivalently in the Fourier domain:

$$\begin{pmatrix} U_1(\omega) \\ U_2(\omega) \end{pmatrix} = \begin{pmatrix} 1 & -W_{21}(\omega) \\ -W_{12}(\omega) & 1 \end{pmatrix} \begin{pmatrix} P_1(\omega) \\ P_2(\omega) \end{pmatrix} \quad (2)$$

where $W_{21}(\omega)$ and $W_{12}(\omega)$ represent the frequency responses of the separating filters $w_{12}(n)$ and $w_{21}(n)$ respectively. Inserting equation (1) in equation (2), we get the input-output relationship:

$$\begin{pmatrix} U_1(\omega) \\ U_2(\omega) \end{pmatrix} = \begin{pmatrix} F_1(\omega) & H_{21}(\omega)-W_{21}(n) \\ H_{12}(\omega)-W_{12}(n) & F_2(\omega) \end{pmatrix} \begin{pmatrix} S(\omega) \\ B(\omega) \end{pmatrix} \quad (3)$$

where $$F_1(\omega)=1-H_{12}(\omega)W_{21}(\omega) \quad (4)$$

$$F_2(\omega)=1-H_{21}(\omega)W_{12}(\omega) \quad (5)$$

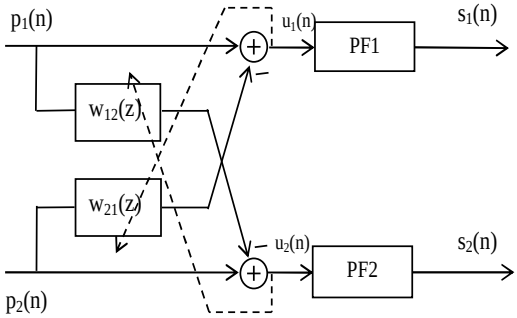

Figure 2 – Forward BSS Structure with two adaptive filters and post-filters.

3.1. Optimal Solution

To retrieve the original signals from u_1 and u_2 (minimum distortion solution) we should have:

$$\begin{pmatrix} S_1(\omega) \\ S_2(\omega) \end{pmatrix} = \begin{pmatrix} U_1(\omega)\left(1-H_{12}(\omega)W_{21}(\omega)\right)^{-1} \\ U_2(\omega)\left(1-H_{21}(\omega)W_{12}(\omega)\right)^{-1} \end{pmatrix} \quad (6)$$

Using post-filters at the output of the Forward BSS structure, as shown in Fig.2, permits to approximate that solution. From (6), the two post-filters PF1 and PF2 are ideally given by:

$$\begin{pmatrix} PF1(\omega) \\ PF2(\omega) \end{pmatrix} = \begin{pmatrix} \left(1-H_{12}(\omega)W_{21}(\omega)\right)^{-1} \\ \left(1-H_{21}(\omega)W_{12}(\omega)\right)^{-1} \end{pmatrix} \quad (7)$$

In practice, the filters $w_{12}(n)$ and $w_{21}(n)$ are adjusted by using adaptive algorithms. Assuming that the two adaptive filters tend asymptotically to the theoretical solutions, the two post-filters PF1 and PF2 lead to the same ideal solution:

$$PF1^*(\omega)=PF2^*(\omega)=\left[1-H_{12}(\omega)H_{21}(\omega)\right]^{-1} \quad (8)$$

Since we are interested in the reduction of the speech distortion, we focus our interest on the output $s_1(n)$ which corresponds to the denoised speech signal.

3.2 Frequency Domain Post-Filter (FDPF)

The FDPF is shown in Fig.3. This new proposed structure [1] is based on a frequency domain implementation of the equalizing post-filter deduced from (4), which is updated by an adaptive algorithm on a frame-by-frame basis.

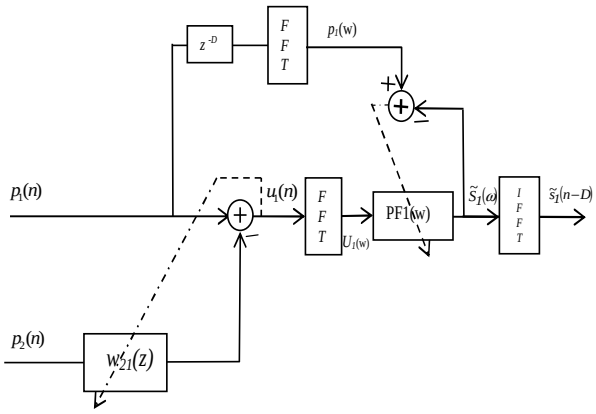

Figure 3 – Forward BSS with closed-loop frequency domain implementation of the post-filter.

The frequency gain $PF_1(\omega,k)$ is used to correct the output $u_1(n)$ of the original Forward BSS structure of Fig.2. This gain is updated in the frequency domain by using the FLMS algorithm [5]. For each frame *k*, we propagate the following equations:

$$PF_1(\omega,k)=PF_1(\omega,k-1)+\mu(\omega,k)E^*(\omega,k)U_1(\omega,k) \quad (9)$$

$$\text{with} \quad E(\omega,k)=P_1(\omega,k)-PF_1(\omega,k-1)U_1(\omega,k) \quad (10)$$

$E(\omega,k)$ represents the filtering error. Parameters $P_1(\omega,k)$ and $U_1(\omega,k)$ represent respectively the frequency components of the mixture signal and of the output of the FBSS structure without post-filtering. In order to obtain a robust denoising system, the step size $\mu(\omega,k)$ is made dependent on the signal to noise ratio SNR in each frequency bin, according to a rule similar to the Wiener filter:

$$\mu(\omega,k)=\left(\alpha\frac{SNR(\omega,k)}{1+SNR(\omega,k)}\right)\frac{1}{\Phi_{U_1U_1}(\omega,k).NFFT} \quad (11)$$

$$=\mu_0(\omega,k)\frac{1}{\Phi_{U_1U_1}(\omega,k).NFFT}$$

where $\Phi_{U_1U_1}(\omega,k)$ is a running estimate of the power spectral density of the signal u_1 and NFFT represents the size of the discrete Fourier transform. The parameter α

is used as a control parameter for the adaptive step-size $\mu_0(\omega,k)$. To estimate the SNR, we have used the decision-directed *a priori* estimation combined with the two step noise reduction technique as described in [6]. The proposed adaptive step-size (11) provides specific properties for the adaptive frequency domain implementation of the post-filter. Indeed, it is known that the NLMS algorithm exhibits a mean square deviation (MSD) in the filter coefficients given by:

$$\mathrm{MSD}(\omega,k)=\frac{\mu}{2-\mu}\frac{1}{\mathrm{SNR}(\omega,k)} \qquad (12)$$

By replacing in this relation μ by the value $\mu(\omega,k)$ given in (11), we get

$$\mathrm{MSD}(\omega,k)=\frac{\alpha'}{2+(2-\alpha')\,\mathrm{SNR}(\omega,k)} \qquad (13)$$

with $\alpha'=\alpha/\left(\Phi_{U_1U_1}(\omega,k)\times \mathrm{NFFT}\right)$. Fig. 4 gives the MSD as a function of the signal to noise ratio for different values of the control parameter α'. To get the same asymptotic behaviour of (12) and (13) for high values of the SNR, we have made the specific choice $\alpha'=\mu$. We see on Fig. 4 that the MSD computed from (13) with the SNR-dependent step-size $\mu(\omega,k)$ is always lower than the one obtained from (12) with a non-SNR dependent step-size.

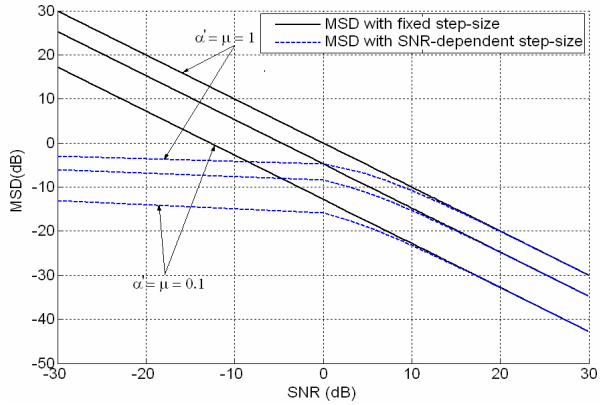

Figure 4 – Comparison of MSD curves. $\mu=\alpha'$ =0.1, 0.5 and 1.

Once the frequency gain $\mathrm{PF}_1(\omega,k)$ is calculated, the resulting speech spectrum is estimated as follows:

$$\tilde{S}_1(\omega,k)=\mathrm{PF}_1(\omega,k)U_1(\omega,k) \qquad (14)$$

To reconstruct the speech signal at the output $\tilde{s}_1(n-D)$, we have used the overlap-save method as described in [6].

4. BACKWARD BSS STRUCTURE (BS)

The classical form of the BS structure is shown in Fig.5. We note that the de-noised outputs of this structure are used as inputs of the cross-coupled adaptive filters $w_{12}(n)$ and $w_{21}(n)$.

4.1. Optimal Solution

The theoretical solution of the problem (*i.e.* complete signal separation) is obtained when: $w_{21}(n)=h_{21}(n)$ and $w_{12}(n)=h_{12}(n)$ [4]. The outputs of the structure shown in Fig.5 are given by:

$$\begin{pmatrix}S_1(\omega)\\S_2(\omega)\end{pmatrix}=\frac{1}{\Delta}\begin{pmatrix}1 & -W_{21}(\omega)\\-W_{12}(\omega) & 1\end{pmatrix}\begin{pmatrix}P_1(\omega)\\P_2(\omega)\end{pmatrix} \qquad (15)$$

Or equivalently

$$\begin{pmatrix}S_1(\omega)\\S_2(\omega)\end{pmatrix}=\frac{1}{\Delta}\begin{pmatrix}1-W_{21}(\omega)H_{12}(\omega) & H_{21}(\omega)-W_{21}(\omega)\\H_{12}(\omega)-W_{12}(\omega) & 1-H_{21}(\omega)W_{12}(\omega)\end{pmatrix}\begin{pmatrix}S(\omega)\\B(\omega)\end{pmatrix} \qquad (16)$$

where: $$\Delta=1-W_{12}(\omega)W_{21}(\omega) \qquad (17)$$

The MSE solution for this structure allows to retrieve the original signals directly from $S_1(\omega)$ and $S_2(\omega)$ without the need for post-filters, thus we should have ideally $S_1(\omega)=S(\omega)$ and $S_2(\omega)=B(\omega)$. Note that to cancel the cross-talk components, the non diagonal elements of (16) must be equal to zero (*i.e.* $H_{21}(\omega)=W_{21}(\omega)$ and $H_{12}(\omega)=W_{12}(\omega)$).

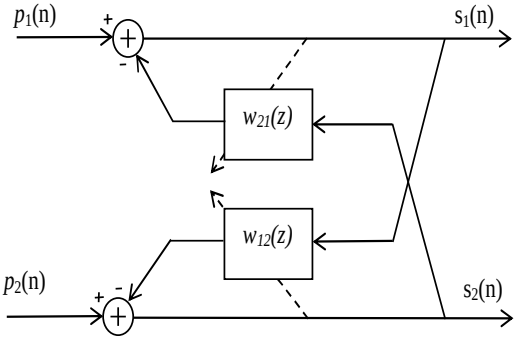

Figure 5 – Backward BSS structure (BS).

4.2 Classical implementation of the BS (CBS)

In this method, we used the scheme of Fig.5 and we used two adaptive algorithms to adapt the two cross-filters $w_{12}(n)$ and $w_{21}(n)$ as described in [7]. In our case we have used the NLMS algorithm to update the coefficients of the FIR filters so as to minimize MSE between the adaptive filters outputs $w_{12}(n)$ and $w_{21}(n)$ and the desired-response signals $p_1(n)$ and $p_2(n)$. The update is made in the time domain [3].

4.3 Robust implementation of the BS (RBS)

The RBS scheme that we consider [3] is shown on Fig.6. This block-diagram corresponds to the ANC Backward BSS structure with variable step size sub-filters [3]. Four adaptive filters, namely, the main adaptive filters (MAF1, MAF2) and the sub adaptive filters (SAF1, SAF2) generate noise and crosstalk replicas. Coefficients in the main and sub adaptive filters are updated by the NLMS algorithm [3]. To reduce signal distortion in the output, the step sizes for coefficients adaptation in the MAFs filter are controlled according to estimated signal-to-noise ratios (SNRs) of the input signal. This SNR estimation is carried out using SAF output signals. The SAF1 output

$y_1(n)$ and the subtraction result $e_1(n)$ are used to estimate a more precise SNR at the primary input. This error $e_1(n)$ serves as an approximation to the target speech, and $y_1(n)$ is used as that to the noise.

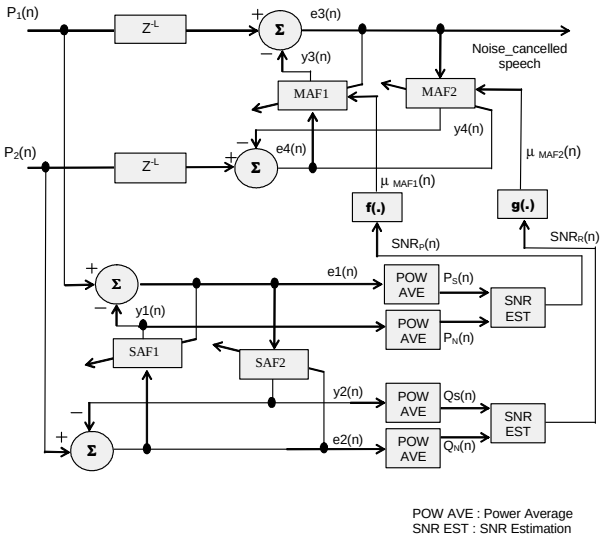

Figure 6 – Robust Implementation of the Backward BSS (RBS) with controlled stepsizes.

The stepsize for MAF1 is controlled by the estimated SNR calculated from SAF1 output signals. SAF2 works for crosstalk instead of the noise in a similar way to that of SAF1. The resulting SNR estimate from SAF2 output signals is used to control the MAF2 stepsize (we use $y_2(n)$ and $e_2(n)$ to estimate the SNR for the SAF2 filter). All the details of the parameters of this structure are given in [3].

5. ANALYSIS OF SIMULATION RESULTS

In this section, we analyse the behaviour of each method that has been presented in the previous sections. Also, we compare our FDPF method with the two backward methods, *i.e.* CBS and the RBS, in two cases. The first corresponds to the configuration when the microphones are loosely spaced and the second one is when the microphones are closely spaced. To represent appropriately the effect of the distance between the two microphones on the characteristics of the signals, we have used the specific model proposed in [8] which yields simulated impulse responses $h_{12}(n)$ and $h_{21}(n)$ [The sampling frequency is f_s = 8 kHz; the corresponding reverberation time is 30.8ms; the length of the impulse responses is $L = 100$]. The speech signal is a sentence of about 4s and the point-source noise signal is stationary white noise. The SNRs (speech-to-noise ratios) are chosen equal to 3dB at the input (p_1) and equal to 0dB at the other input (p_2).

5.1. Simulations with loosely spaced microphones

In this simulations the length of the adaptive filters (LMS algorithms) $w_{12}(n)$ and $w_{21}(n)$ is equal to L=100 (L is the length of the generated impulse response). The frequency-domain processing uses frames of size 256 with 50% overlapping. The simulations show that the three methods detailed above perform well: the speech signal is almost completely denoised (see Fig.6 in [1]). A comparison in terms of the averaged cepstral distance (CD) between the original speech signal and those obtained respectively, at the output of each of the three methods is shown in Fig.7. On this figure, each point corresponds to a smoothing of 256 consecutive frames.

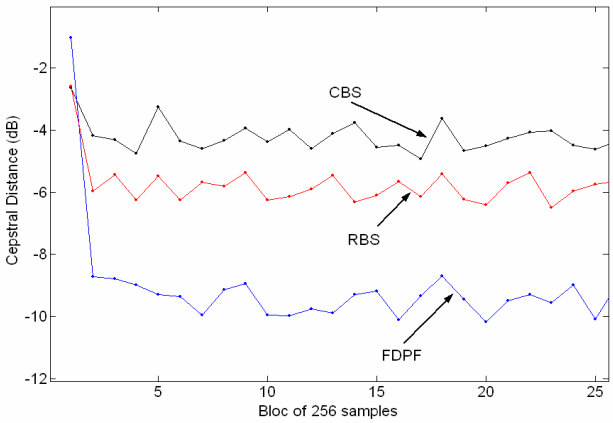

Figure 7 – Comparison of the CD for the three methods with loosely spaced microphones

The good performance of the FDPF method appears clearly with an average CD of –9.82 dB. One can also see on Fig.8 that the RBS method has a superior behaviour over the CBS one, (-6.12 dB for the RBS and –3.50 dB for the CBS). In Fig.8, we have evaluated the SNR criterion for the three methods (FDPF, RBS and CBS methods). Each point on the figure corresponds to a smoothing of 1024 consecutive frames. The mean value of the SNR of the RBS method is about 27.25 dB, 8.85 dB for the CBS method and 44.85 dB for the FDPF method. It means that there is a gain of 17.60 dB between the FDPF method and the RBS one and a gain of 36 dB for the FDPF over the CBS method. This confirms the superiority of the FDPF method over the CBS and RBS one's.

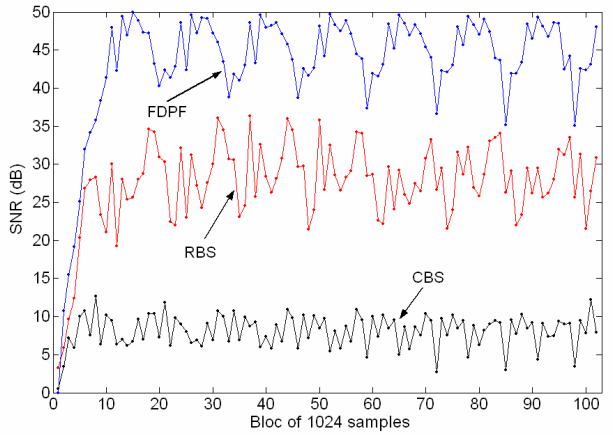

Figure 8 – Output SNR evolutions of the FDPF, RBS and CBS methods in the case of loosely spaced microphones.

5.2. Simulations with closely spaced microphones

In this experiment, the two adaptive filters $w_{12}(n)$ and $w_{21}(n)$ are close to $\delta(n)$. In this case, one can see in (the middle of Fig.8 in [1]) that in the FDPF method the signal u_1 is strongly attenuated, whereas the attenuation is compensated at the output $s_1(n)$ thanks to the post-filter and the original speech signal is restored. Furthermore, a very poor behaviour has been observed with the CBS structure. This is due to the high misadjustment of $w_{12}(n)$ which is always adapted in the presence of the mixing signal (speech plus noise). We have also noted the good performance of the RBS method but it remains inferior to the performance of the proposed FDPF method. We have evaluated the CD criteria for the three methods in Fig.9.

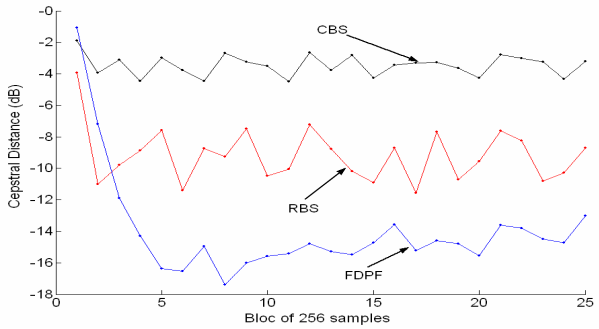

Figure 9 – Comparison of the CD of the three methods with closely spaced microphones.

We observe the good performance of the FDPF method and the fairly good one of the RBS method (CD=-15.56 dB for the FDPF method and -9.46 dB for the RBS one). On the other hand, we have observed a poor behaviour of the CBS method (-3.05 dB). This is due to the large misadjustment of the filter $w_{12}(n)$. In the end, we have confirmed the better behaviour of the FDPF method of section 3.2 *vs.* the two other methods through informal listening tests. We have evaluated the SNR behaviour for each method on Fig.10.

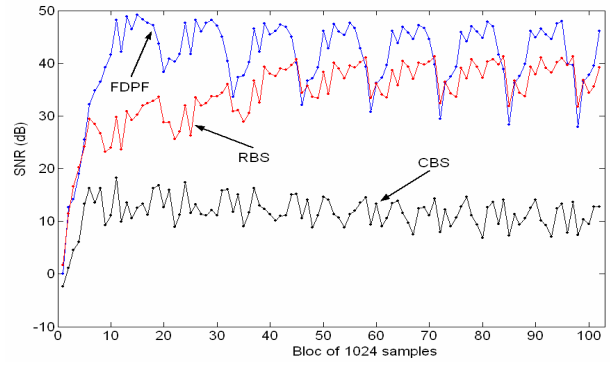

Figure 10 – Output SNR evolutions of the FDPF, RBS and CBS methods in the case of closely spaced microphones.

We observe that the mean values of the SNR are about 44.95 dB for the FDPF, 33.91 dB for the RBS and about 11.25 dB for the CBS one. We have noted a gain value of about 11,04dB for the FDPF method over the RBS method and 33.70 dB over the CBS one. This shows the good behaviour of the proposed FDPF method and its superiority in term of SNR even in critical situation when the microphones are closely spaced.

6. CONCLUSION

In this paper, we have presented and compared three methods to extract the speech signal from noisy observations. The three methods use two microphones, either loosely (first case) or closely (second case) spaced. The FDPF method has given good simulation results for the two cases. The good performance of this method and its superiority over the CBS methods and the RBS one is confirmed by the CD criterion, SNR values and by informal listening tests. We have also noted a fairly good performance of the RBS method when the microphones are loosely or closely spaced. A very poor behaviour of the CBS method is obtained when the microphones are closely spaced. The CD criterion and the informal listening tests have shown the superiority of the FDPF method over the RBS one. We note also that the RBS method has a higher complexity than the other ones; moreover, it needs the adjustment of many important parameters. According to all those results and considerations in both tested cases, we recommend the FDPF method to be used in practice. Further work on the problem of the diffuse noise is carried out and adequate solutions for this problem are under development.

REFERENCES

[1] M. Djendi, A. Gilloire, P. Scalart, New frequency domain post-filters for noise cancellation using two closely spaced microphones, Proc. EURASIP Conference EUSIPCO, Poznan, Poland, 3-8 September 2007, vol.1, pp.218-221.

[2] S. Van Gerven and D. Van Compernolle, Signal separation by symmetric adaptive decorrelation: stability, convergence, and uniqueness, IEEE Trans. Signal Proc. (July 1995), vol.74, no.3, 1602-1612.

[3] S. Ikeda and A. Sugiyama., An adaptive noise canceller with low signal-distortion in the presence of crosstalk, IEICE Trans. Fundamentals (Aug. 1999), vol. E.82-A, no.8, 1517-1525.

[4] M.J. Al-Kindi and J. Dunlop, Improved adaptive noise cancellation in the presence of signal leakage on the noise reference channel, Signal Processing (July 1989), vol. 17, no.3, 241-250.

[5] E. R. Ferrara, Fast implementation of LMS adaptive filter, IEEE Trans. Sig. Proc, vol.28, 474-475, Aug. 1980.

[6] C. Plapous, C. Marro, P. Scalart, L. Mauuary, A two-step noise reduction technique, IEEE ICASSP, Montreal, Canada, vol. 1, pp. 289–292, May 2004.

[7] M. Gabrea, Double affine projection algorithm-based speech enhancement algorithm, Proc. IEEE. ICASSP Montréal, Canada, vol.2, pp. 904-907, April 2003,

[8] M. Djendi, A. Gilloire, P. Scalart, Noise cancellation using two closely spaced microphones: experimental study with a specific model and two adaptive algorithms, IEEE Int. Conf. ICASSP, Toulouse, France, 14-19 May 2006, vol.3, pp. 744-747.