

NEW FREQUENCY DOMAIN POST-FILTERS FOR NOISE CANCELLATION USING TWO CLOSELY SPACED MICROPHONES

Mohamed DJENDI¹, André GILLOIRE² and Pascal SCALART¹

¹ ENSSAT - LASTI, 6 Rue de Kerampont, B.P. 447, 22305 Lannion Cedex, France
Email: {mohamed.djendi,pascal.scalart}@enssat.fr, m_djendi@yahoo.fr

² France Telecom R&D - TECH/SSTP, 2 Avenue Pierre Marzin, 22307 Lannion Cedex, France
Email: andre.gilloire@francetelecom.com

ABSTRACT

This paper addresses the problem of speech enhancement in a moving car through a blind source separation (BSS) scheme involving two closely spaced microphones. We propose two frequency domain methods to reduce the distortion caused by the forward BSS structure, which is most important when microphones are closely spaced. Both methods aim at estimating post-filters to compensate for the distortion by equalization. The first method is based on an open-loop estimation. The second one uses a closed-loop adaptive algorithm to make the estimation recursively. Simulation results show that the second method is more efficient than the first one, and both of them allow to correct the signal distortion caused by the BSS forward structure.

1. INTRODUCTION

The quality of signal enhancement systems, based on adaptive filtering, is highly dependent on the quality of the "noise reference". It is well known that any amount of signal leakage into the noise reference results in signal distortion and poor noise cancellation [1]. In the literature, we find two proposed structures, conceptually simple. These structures allow to carry out the reduction of noise by sources separation. They are respectively called Backward [2] and Forward structures [3]. The Forward structure can be regarded as an extension of the basic structure of the ANC (Adaptive Noise Canceller) [1]. This structure presents the disadvantage of distorting the output signals. It was shown theoretically that the correction of the distortions is possible thanks to the equalization of the output signals by post-filtering [3]. We have also shown in [4], that the distortion becomes more important when the inputs are recorded with closely spaced microphones. The work presented herein is an extension of [4] where the focus was put on the adaptation of the separation structure. We concentrate here on the post-filters implementation, a problem for which, to our knowledge, no satisfactory solution in practice was proposed, especially in the case of closely spaced microphones.

In this paper, we propose two efficient methods to estimate the post-filters in the frequency domain. We consider the idealized case of BSS as in [4] (number of sources equal to number of observations). In section 2 we present the mixing model; the Forward BSS structure is detailed in section 3. In section 4, we describe the proposed methods for the computation of the equalizing post-filters. Finally,

experimental results obtained by simulations and a comparison between the two methods are given and discussed in section 5.

2. MIXING MODEL

The mixing model that we consider is described in Fig.1 [4]. This model involves two convolutive mixtures of two uncorrelated point sources, defined as:

$$\begin{aligned} p_1(n) &= \mathbf{h}_{11} * s(n) + \mathbf{h}_{21} * b(n) + n_1(n) \\ p_2(n) &= \mathbf{h}_{22} * b(n) + \mathbf{h}_{12} * s(n) + n_2(n) \end{aligned} \quad (1)$$

One of the two point sources is speech (the useful signal), and the second one can represent either the car engine noise or far-end speech that we want to cancel. \mathbf{h}_{11} and \mathbf{h}_{22} represent the impulse responses of each channel separately, and \mathbf{h}_{12} and \mathbf{h}_{21} represent the cross-coupling effects between the channels; n_1 and n_2 represent non-coherent part of the diffuse acoustic (background) noise in the vicinity of the microphones. * represents convolution. In this work, \mathbf{h}_{11} and \mathbf{h}_{22} are assumed to be identity; this assumption does not impact the practical usefulness of the model as noted in [4].

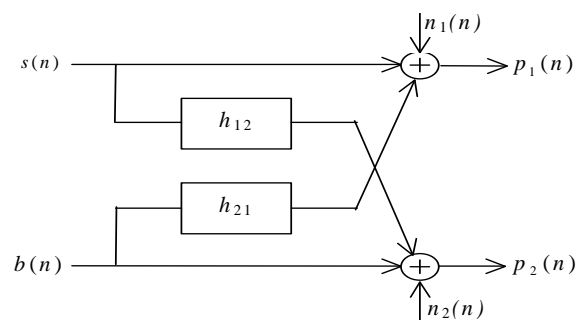


Fig. 1. Mixing model

3. FORWARD BSS STRUCTURE

The considered BSS structure is the Forward one shown in Fig.2. The evident theoretical solution of the problem is given by setting $\mathbf{w}_{21} = \mathbf{h}_{21}$ and $\mathbf{w}_{12} = \mathbf{h}_{12}$ [3]. The Least Squares solution of the problem is given by minimizing the *a priori* errors $\mathbf{u}_1(n) = p_1(n) - p_2(n) * \mathbf{w}_{21}(n)$ and $\mathbf{u}_2(n) = p_2(n) - p_1(n) * \mathbf{w}_{12}(n)$ or equivalently:

$$\begin{aligned} u_1(n) &= s(n) * [\delta(n) - \mathbf{h}_{12} * \mathbf{w}_{21}(n)] + n_1(n) - n_2(n) * \mathbf{w}_{21}(n) \\ u_2(n) &= b(n) * [\delta(n) - \mathbf{h}_{21} * \mathbf{w}_{12}(n)] + n_2(n) - n_1(n) * \mathbf{w}_{12}(n) \end{aligned} \quad (2)$$

The filters w_{12} and w_{21} are obtained by using adaptive algorithms.

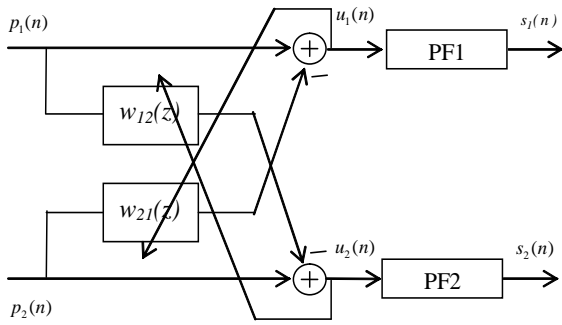


Fig. 2 : Forward BSS based system scheme

The two post-filters PF1 and PF2 are estimated by the proposed methods described in section 4. In noiseless case ($n_1 = n_2 = 0$), to retrieve the original signals from u_1 and u_2 , we should have :

$$s(n) = u_1(n) * [\delta(n) - h_{12} * w_{21}(n)]^{-1} \quad (3)$$

$$b(n) = u_2(n) * [\delta(n) - h_{21} * w_{12}(n)]^{-1}$$

Therefore, the two post-filters PF1 and PF2 are ideally given by:

$$PF1 = [\delta(n) - h_{12} * w_{21}(n)]^{-1} \quad (4)$$

$$PF2 = [\delta(n) - h_{21} * w_{12}(n)]^{-1}$$

When the two adaptive filters $w_{12}(n)$ and $w_{21}(n)$ tend to the theoretical solutions i.e. $w_{21}^* = h_{21}$ and $w_{12}^* = h_{12}$, the two post-filters PF1 and PF2 will have the same ideal solution:

$$PF1^* = PF2^* = [\delta(n) - h_{12} * h_{21}]^{-1} \quad (5)$$

4. DESCRIPTION OF THE PROPOSED METHODS FOR THE COMPUTATION OF THE POST-FILTER

As 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. The two proposed methods are based on a frequency domain implementation of the equalizing post-filter.

4.1. Open-loop estimation of the post-filter

The first new method is shown in Fig.3. In this implementation, the frequency gain $G(\omega, k)$ is used to correct the output $u_1(n)$ of the BSS structure by amplitude equalization. The equalizer gain is computed using two signals: the distorted speech signal $u_1(n)$ and the signal mixture $p_1(n)$. We aim at restoring for each frequency bin of the output signal the same amplitude of the speech signal contained in the mixture $p_1(n)$. The power spectral densities (PSD) of the signals $u_1(n)$ and $p_1(n)$ are estimated by means of first order recursive equations using the FFT computation of each signal.

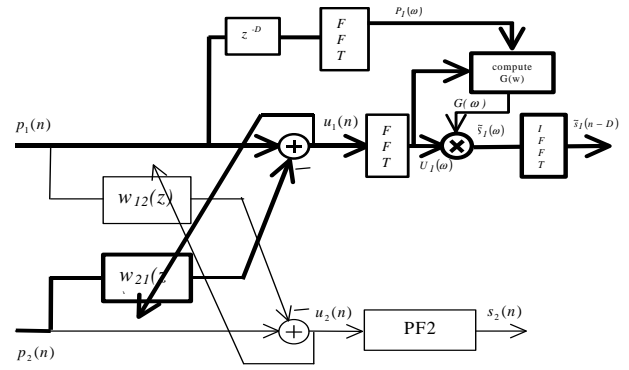


Fig. 3 : Scheme with open-loop post-filter

For each frame k , the calculation of the frequency gain $G(\omega, k)$ is carried out by the following equation:

$$G(\omega, k) = \sqrt{\frac{\gamma_{s_1 s_1}(\omega, k)}{\gamma_{u_1 u_1}(\omega, k)}} = \sqrt{\frac{\text{Max}[\gamma_{P_1 P_1}(\omega, k) - \gamma_{P_1 P_1}(\omega, k), 0]}{\gamma_{u_1 u_1}(\omega, k)}} \quad (6)$$

where $\gamma_{P_1 P_1}(\omega, k)$ and $\gamma_{u_1 u_1}(\omega, k)$ are respectively the PSD of $p_1(n)$ and $u_1(n)$. $\gamma_{s_1 s_1}(\omega, k)$ is computed in the Voice Activity periods (VAP) by subtracting the PSD of noise $\gamma_{P_1 P_1}(\omega, k)$, which is estimated in the non VAP, from the

PSD of the mixture signal $p_1(n)$. Therefore, we exploit the intermittence of the speech signal to estimate all the PSDs of this structure. To reconstruct the output, we use an overlap-save method. We point out that the correct adjustment of the delay D is very important for the correct operation of this structure. This parameter must be larger than the group-delay of the product $h_{12} * h_{21}$.

4.2. Closed-loop estimation of the post-filter

This new method shown by Fig.4 is based on the use of a frequency domain adaptive algorithm to update the coefficients of the correcting gain $G(\omega, k)$. We have used the frequency domain LMS algorithm (FLMS) [5].

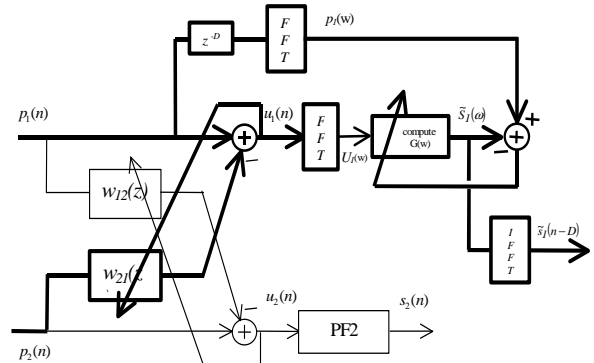


Fig. 4 : Scheme with closed-loop post-filter

For each frame k , we propagate the following equations:

$$G(\omega, k) = G(\omega, k-1) + \mu(\omega, k) E^*(\omega, k) U_1(\omega, k), \quad (7)$$

where $E(\omega, k)$ is the filtering error given by:

$$E(\omega, k) = P_1(\omega, k) - G(\omega, k)U_1(\omega, k) \quad (8)$$

where $P_1(\omega, k)$ et $U_1(\omega, k)$ represent respectively, the frequency components of the mixture signal and the output of the forward BSS structure without post-filtering. Practically in our implementation, the error is computed in the time domain according to [5]. In order to design a robust denoising system, the adaptation step size $\mu(\omega, k)$ is made dependent on the signal to noise ratio SNR in each frequency bin of $P_1(\omega, k)$. In this paper we have used the following step size equation:

$$\mu(\omega, k) = \frac{SNR(\omega, k)}{1 + SNR(\omega, k)}, \quad (9)$$

To estimate this SNR, we have used the *a priori* estimation combined with the two step noise reduction technique described in [7]. The resulting speech spectrum is estimated as follows:

$$S_1(\omega, k) = G(\omega, k)U_1(\omega, k), \quad (10)$$

We have used the overlap-save method as described in [5], to reconstruct the speech signal at the output $\tilde{s}_1(n - D)$.

5. ANALYSIS OF SIMULATION RESULTS

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 [4]. We have done two kinds of experiments with this model according to the microphones spacing (experiments with widely or closely spaced microphones).

The simulated impulse responses have been constructed as follows: $h_{12} = \delta(n) + h'_{12}$ and $h_{21} = \delta(n) + h'_{21}$, where $\delta(n)$ is a unit pulse localized at the beginning of the impulse response which represents the direct acoustic path from each source to the cross-coupled microphone. h' is an exponentially weighted tail representing the room effect [4]. Fig.5 shows examples of such impulse responses h_{21} (Fig.5, (a)) and the post-filter $\delta(n) - h_{12} * h_{21}$ (Fig.5, (b)), which correspond to closely spaced microphones; with a sampling period $T_s = 125 \mu s$, the corresponding reverberation time is 30.8 ms. The size of the impulse responses is $L = 100$. For the spaced microphones configuration, we have used the impulse responses given in [4].

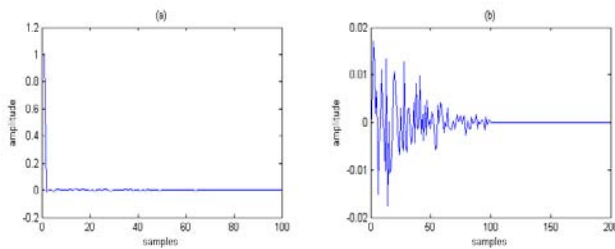


Fig. 5. Simulated impulse responses in closely spaced microphones case, (a): h_{21} , (b): $\delta(n) - h_{12} * h_{21}$

All the simulations have been performed at the sampling frequency $f_s = 8$ kHz. The speech signal is a sentence of about 4s and the point-source noise signal is stationary

with average speech spectrum (USASI noise). The SNR (speech-to-noise ratio) at the inputs of the forward structure is 3dB. We note that the length of the adaptive filters (LMS algorithms) $w_{12}(n)$ and $w_{21}(n)$ are equal to L . The frequency-domain processing uses frames of size 256 with 50% overlap.

5.1. Simulations with loosely spaced microphones

One can see from inspection of Fig.6 that the first proposed method (open-loop post-filter, see Fig.3), performs well: the speech signal is completely denoised and the distortion appears corrected.

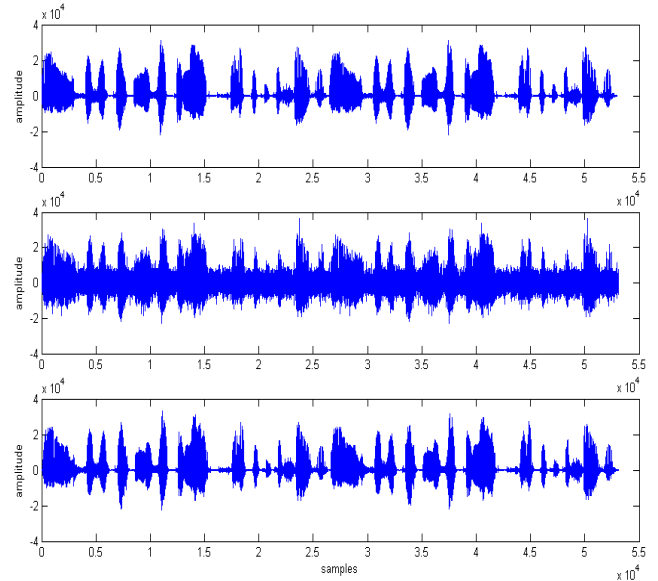


Fig. 6. Source signal (top), mixture (middle) and noise canceller outputs_l (bottom) obtained with the open-loop method.

We have compared the averaged cepstral distance between the original speech signal and those obtained respectively, at the input and the output of the equalizer filter $G(\omega, k)$. We have obtained -14.25 dB without the equalizer and -18.58 dB with the equalizer filter as shown in Fig7. On this figure, each point corresponds to a smoothing of 256 consecutive frames.

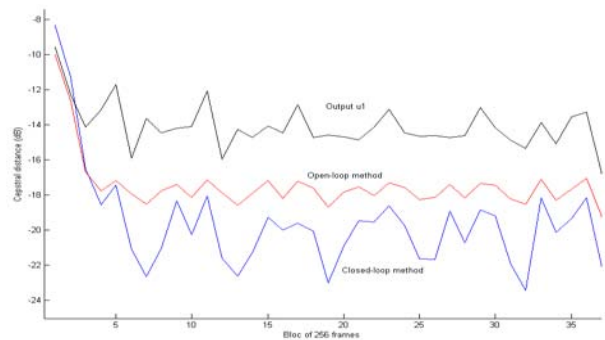


Fig.7. Comparison of the CD for the two methods with spaced microphones

With the second proposed method, we have obtained at the output of the adaptive filter $G(\omega, k)$ similar waveforms as those shown on Fig.6. The main difference between the two methods is on the cepstral distance (CD) values

(Fig.7). In this experiment, we have obtained, respectively, -18.08 dB and -21.56 dB, respectively for the first method and the second one. This result clearly shows that the spectral distortion of the basic forward structure is corrected when using the two proposed frequency-domain methods and we note also the superior performance of the second method, which was confirmed by informal listening tests.

5.2. Simulations with closely spaced microphones

The original speech signal, the output signal of the BSS structure u_l at the input of the equalizer $G(\omega, k)$ and the output of the open-loop post-filter are given by Fig.8.

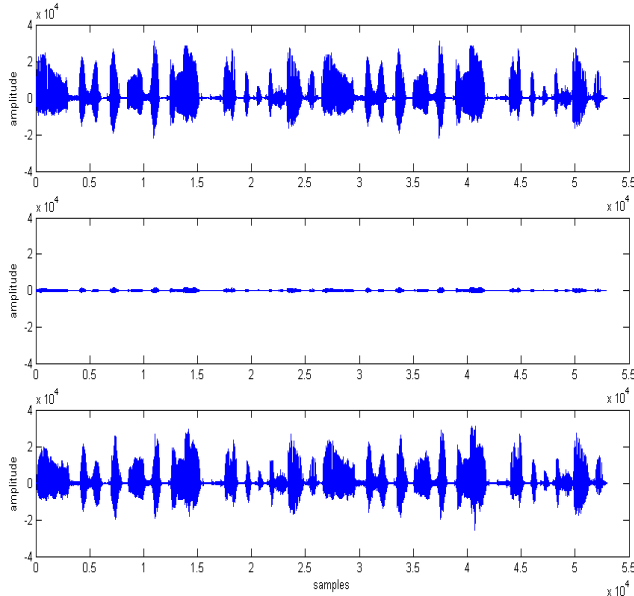


Fig.8. Source signal (top), Output u_l (middle) and noise canceller output s_l (bottom) obtained with the open-loop method

In this experiment, the two adaptive filters w_{12} and w_{21} are close to $\delta(n)$. One can observe that the signal u_l is strongly attenuated. For the output s_l , we see that the attenuation is compensated and that the original speech signal is restored. Similar behaviour has been observed with the closed-loop post-filter. In this critical situation (closely spaced microphones), we have evaluated the CD criteria for the two methods (see Fig.9).

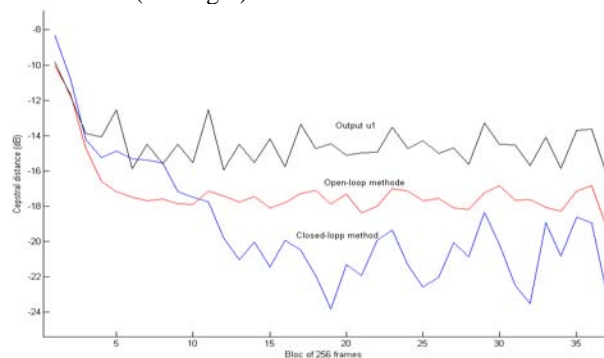


Fig. 9. Comparison of the CD of the two methods with closely spaced microphones

We have obtained -17.67 dB for the first method and -20.57 dB with the second one. We have confirmed the good behaviour of the second method through informal listening tests.

6. CONCLUSION

We have proposed two new frequency domain methods to compute the post-filters to compensate for the spectral distortion caused by the forward BSS structure. The first method is based on an open-loop amplitude equalization structure. The second one uses a frequency domain adaptive filter to estimate recursively the post-filter. In this second method, we have introduced a variable step size dependent on the SNR to cope with the various situations that can be encountered in practice. Hence, this new modification permits robust convergence of the adaptive equalizer even in critical situations. All the simulations have shown the good behaviour of the two proposed methods with superior performance of the second one. Note that even more striking improvement of the cepstral distance with the two methods has been obtained with very short impulse responses in the mixing model (10 points), which produce a very high spectral distortion at the output of the separation structure. Note that we have also tested successfully the two proposed methods with real signals recorded in a car. As mentioned in the conclusion of [4], we note that the amplification of the non-coherent noise by the post-filters (not discussed here) may be mitigated by the fact that the uncorrelated noise is generally lower in case of closely spaced microphones. Further work will deal with this problem.

7. REFERENCES

- [1] B. Widrow and al., "Adaptive noise cancelling: principles and applications", *Proceedings of the IEEE*, vol. 63, no.12, pp.1692-1716, Dec.1975.
- [2] 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, July 1989.
- [3] S. Van Gerven and D. Van Compernelle, "Signal separation by symmetric adaptive decorrelation: stability, convergence, and uniqueness," *IEEE Trans. Signal Proc.*, vol. 74 No. 3, pp.1602-1612, July 1995.
- [4] 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. ASSP*, vol.3, pp. 744-747, May 2006.
- [5] E. R. Ferrara, "Fast implementation of LMS adaptive filter," *IEEE Trans. ASSP*, vol. ASSP-28, pp. 474-475, Aug. 1980.
- [6] P. Scalart, and J. Vieira Filho, "Speech Enhancement Based on a Priori Signal to Noise Estimation," *IEEE Int. Conf. ASSP*, pp. 629-632, May 1996.
- [7] C. Plapous, C. Marro, P. Scalart, and L. Mauuary, "A Two-Step Noise Reduction Technique," *IEEE Int. Conf. ASSP*, vol. 1, pp. 289-292, 17-21 May 2004.