COMBINED WIENER AND COHERENCE FILTERING IN WAVELET DOMAIN FOR MICROPHONE ARRAY SPEECH ENHANCEMENT

Djamila Mahmoudi and Andrzej Drygajlo

Signal Processing Laboratory
Swiss Federal Institute of Technology at Lausanne
CH-1015 Lausanne, SWITZERLAND
e-mail:Djamila.Mahmoudi@epfl.ch

ABSTRACT

Wiener filter based postfiltering has shown its usefulness in microphone array speech enhancement systems. In our earlier work, we developed a postfilter in the wavelet domain where better performance has been obtained compared to the algorithms developed in the Fourier domain. Furthermore, considerable computational savings are provided thanks to the multi-resolution and multi-rate analysis. This contribution shows that the coherence function, calculated between the beamforming output signal and the reference microphone output signal using wavelet transform, provides a revelant and exploitable information for further noise suppression. Thus, a nonlinear coherence filtering and a linear Wiener filtering are combined in the wavelet transform domain to improve the performance of the Wiener filter based postfilter, especially during pauses. Evaluations of the new algorithm confirm that speech quality is indeed improved with significantly reduced distortions. Finally, the results of the objective measures are presented.

1. INTRODUCTION

Most of the hands-free voice communication systems are intended to work in real conditions where interfering sources like noise and reverberation effects are often present. This leads to an important decrease in performance of such systems. To provide a good quality of communication, many speech enhancement algorithms have been proposed. Microphone array speech enhancement has shown its superiority over the techniques based on one channel thanks to its independence from the nature of the interfering signals. Furthermore, microphone arrays are very suitable for situations involving a moving source since they can alter their direction of reception automatically.

One of the commonly used microphone array speech enhancement methods is conventional delay-and-sum beamforming with an additional postfiltering, as proposed in [5, 10, 9]. This method has proved its efficiency for noise reduction. However, important distortions are introduced by the postfilter, particularly because of the inaccurate speech and noise power spectral densities estimates. On the other hand, poor performance is noticed at low frequencies which exhibit a high spatial coherence between noises in the microphone output signals. This is due to the small spacing required between the microphones.

To overcome these problems, a number of improvements, including smoothing techniques for power spectral densities estimates, have been proposed [4]. We have proposed a postfilter using Wiener filter developed in the wavelet domain [7, 8]. Good performances are obtained with this algorithm thanks to the multi-resolution induced by the wavelet transform and represented by the sub-band filters of various lengths. In addition, computational load is reduced.

In this paper, we extend our approach by including a different processing based on a coherence function. A wavelet transform based coherence function is introduced to estimate the degree of similarity between two signals in the time-frequency domain. Using this function which is analogous to the FFT based coherence function, we develop a nonlinear filter to improve the noise suppression obtained with the Wiener filter alone. Unlike the FFT based coherence function calculated between two microphone output signals as in [1, 2], the proposed coherence function is calculated between the beamforming output signal and the reference microphone output signal in the wavelet domain.

2. PROBLEM STATEMENT

The idea of the proposed system consists in the decomposition of a non-symmetric logarithmic microphone array into sub-arrays and in the beamforming operation performed in sub-bands [6]. As detailed in [7], wavelet transform is very suitable for the proposed decomposition. Furthermore, the advantage of its use is the computational saving achieved by the multi-resolution multi-rate processing represented by the critical sub-sampling. The processing is applied in the time-frequency domain where both sub-array based beamforming operation and postfiltering are performed using wavelet transform (Fig. 1).

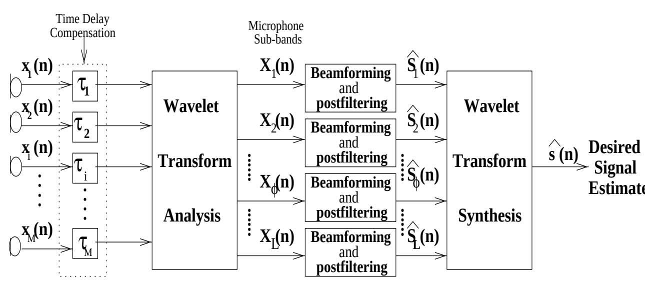

Figure 1: *Block diagram of the noise reduction technique.*

Let $x(n)$ be the noisy signal samples where $x(n) = s(n) +$

$n(n)$. The clean speech signal $s(n)$ and the additive noise $n(n)$ are assumed to be independent. The wavelet transform, represented by a transform matrix $\mathbf{A}$ is performed on the corrupted input vector, $\mathbf{x}$, yielding

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{x} = \mathbf{A} \cdot \mathbf{s} + \mathbf{A} \cdot \mathbf{n} = \mathbf{S} + \mathbf{N},$$

where $\mathbf{S}$, $\mathbf{N}$ and $\mathbf{X}$ are the wavelet transform coefficients of the desired speech, noise and corrupted vectors respectively.

The input vector to the noise reduction procedure, which consists of the wavelet transform coefficients of the time delay compensated microphone signals and which corresponds to the ϕth sub-band, $\phi = 1, \cdots, L$, is expressed as:

$$\mathbf{X}_\phi(n) = [X_{1,\phi}(n), \cdots, X_{i,\phi}(n), \cdots X_{M,\phi}(n)]^T, \qquad (1)$$

where M is the number of microphones. The components of this vector are summed and averaged forming $\bar{X}_\phi(n)$. Then, we pass $\bar{X}_\phi(n)$ through a postfilter to obtain a sub-band signal estimate $\hat{S}_\phi(n)$. Generally, an adaptive Wiener filter is used as postfilter to provide this estimate [10, 9].

3. WAVELET TRANSFORM BASED WIENER FILTER

In our earlier paper [7], we developed a Wiener filter H_w in the wavelet transform domain. We showed that it is provided by the same formula as in the Fourier domain. Under the assumption of the diagonality of the wavelet covariance matrices of the desired signal and noise, the scalar Wiener filter is expressed as:

$$H_w = \frac{E\{S^2\}}{E\{X^2\}}, \qquad (2)$$

where the quantities $E\{S^2\}$ and $E\{X^2\}$ are the wavelet power spectra of the desired and the noisy speech signals, respectively. The wavelet transform Wiener filter H_w applied in sub-band can be described by the following expression:

$$H_{w_\phi}(n) = \frac{\frac{2}{M \cdot M-1} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} X_{i,\phi}(n) \cdot X_{j,\phi}(n)}{\frac{1}{M} \sum_{i=1}^{M} X_{i,\phi}(n)^2}. \qquad (3)$$

4. WAVELET TRANSFORM BASED COHERENCE FILTER

The well-known coherence function is usually used to measure the similarity between two signals [3]. This function can also be calculated on the wavelet transform coefficients $X_1(n)$ and $X_2(n)$ of any two signals $x_1(n)$ and $x_2(n)$. For the purpose of this work, these two signals are modelled as follows:

$$\begin{cases} X_1(n) = \alpha S(n) + N_1(n) \\ X_2(n) = \alpha S(n) + N_2(n), \end{cases} \qquad (4)$$

where α is the attenuation factor. By analogy to the magnitude squared method, the wavelet coherence function takes the following form:

$$C_{X_1X_2,l}(n) = \rho_{X_1X_2,l}(n)^2 = \frac{\Gamma_{X_1X_2,l}(n)^2}{\Gamma_{X_1X_1,l}(n)\Gamma_{X_2X_2,l}(n)}, \qquad (5)$$

where $\Gamma_{X_1X_1,l}(n)$ and $\Gamma_{X_1X_2,l}(n)$ are the wavelet auto-power and cross-power spectra of the lth frame respectively. This quantities are calculated as follows:

$$\Gamma_{X_iX_j,l}(n) = \beta \cdot \Gamma_{X_iX_j,l-1}(n) + X_1(n)X_2(n) \quad i,j = 1,2 \quad (6)$$

where $\beta \in [0.1, 0.25]$ is a forgetting factor.

Let $G_{\phi,l}$ be the averaged coherence function over the whole sub-band. It provides one value per sub-band for each frame. If we consider the model (4) and if we assume that $N_1(n)$ and $N_2(n)$ are the uncorrelated noises that affect the clean signal $S(n)$, $G_{\phi,l}$ provides a significant information about the speech presence or absence. If $G_{\phi,l}$ is close to zero, it means that the lth frame contains only noise (pause) and can be removed using a nonlinear filter. Otherwise, the frame is kept untreated.

Unfortunately, a coherence function calculated between two microphone signals as done in [1] is far from being exploitable in noise suppression since noises at two microphones are strongly correlated because of the small microphone spacing.

The basic idea of the coherence based noise reduction algorithm proposed in this paper consists in exploiting the change in the additive noise structure as a result of the conventional beamforming. In other terms, if the time delay is compensated for all the microphone signals according to the desired source direction and the beamforming operation is performed, the noise components will also be added and averaged. However, they are still unphased since the interfering sources are coming from directions other than the one of the desired source. This results in a different noise in the beamforming signal compared to the initial noise at the microphone output signal while the same desired signal is present in these two signals under an ideal time delay compensation. Consequently, the model described by (4) is verified. Thus, the coherence function is calculated between the beamforming output signal $\bar{X}_\phi(n)$ and the microphone reference output signal $X_{1,\phi}(n)$. Fig. 2 presents the block diagram of the coherence based noise reduction procedure. We observed that the averaged coherence func-

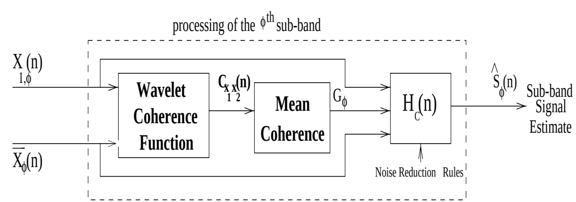

Figure 2: *Wavelet transform based coherence sub-band filter.*

tion $G_{\phi,l}$ for each frame generally exhibits a significant change of value when the speech energy is absent (see Fig. 3). This important property, that confirms our idea, leads to coherence coefficients which are small during pauses and close to one when the speech signal is present. Thus, the coherence function is exploited in a nonlinear postfilter H_C to enhance each sub-band beamformer output. The sub-band estimate of the desired signal using the coherence method is:

$$\hat{S}_\phi(n) = H_C(\bar{X}_\phi(n)). \qquad (7)$$

The enhancement is achieved using noise reduction rules which are derived by observing the averaged coherence function in the

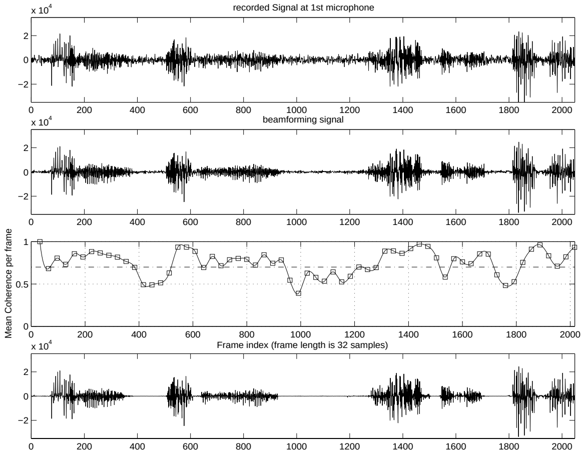

Figure 3: *Example of the wavelet coherence method in sub-bands. B=[500Hz,1000Hz]*

different sub-bands. Thus, the noise reduction rules are expressed by the following equations:

$$\hat{S}_\phi(n) = \begin{cases} \bar{X}_\phi(n) & \text{if} \quad G_{\phi,l} > T \\ G^{\gamma}_{\phi,l} \cdot \bar{X}_\phi(n) & \text{if} \quad G_{\phi,l} \le T \end{cases} \tag{8}$$

An appropriate threshold T is constant and chosen equal to 0.7 for all sub-bands. It results from the observations of the behaviour of $G_{\phi,l}$ and from listening judgements. γ is chosen equal to 25. Fig. 4 shows that noise is removed during pauses while in the speech parts, which have $G_{\phi,l}$ exceeding the threshold T and are not treated, some noise is still present. We note that the performance of the coherence based noise reduction procedure is poor in the energetic parts of the speech signal compared to the noise reduction algorithm based on Wiener filter (see Fig. 4).

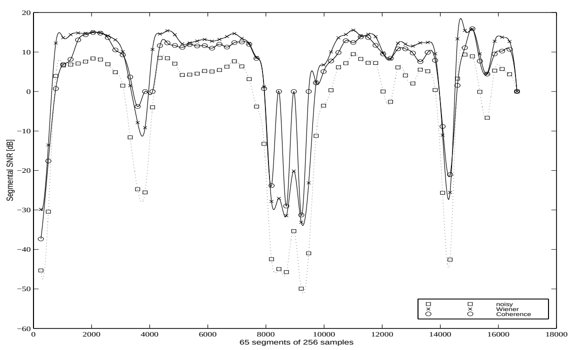

Figure 4: *Evaluation of the performance of coherence based noise reduction procedure in terms of SNR_{seg} (white noise).*

5. COMBINED WIENER AND COHERENCE FILTERING IN THE WAVELET DOMAIN

To increase the performance of the postfilter, we propose to combine the coherence nonlinear filter proposed in Sec. 5 with the Wiener filter presented in Sec. 3. The block diagram of the combined method is shown in Fig. 5. This procedure is represented by a nonlinear filter H which is defined as:

$$H = \begin{cases} H_W & \text{if} \quad G_{\phi,l} > T \\ H_C & \text{if} \quad G_{\phi,l} \le T \end{cases} \tag{9}$$

With this combined method, one can process both the coherent and noncoherent noise. The problem of poor performance at low frequencies encountered with the Wiener filter is overcome by the use of the coherence based noise reduction procedure.

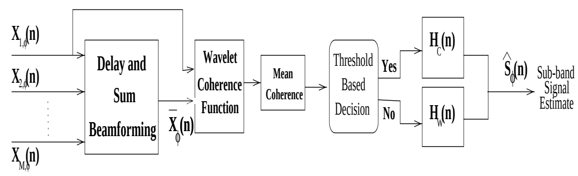

Figure 5: *Combined Wiener and coherence filters applied in wavelet transform domain for noise reduction technique.*

6. PERFORMANCE EVALUATION

The proposed algorithm was tested with a simulated non-symmetric logarithmic array with 6 microphones. The smallest spacing between the microphones was $d_{min} = 5cm$ and the sampling frequency for the speech enhancement part is 8kHz. In this work, different real noise signals with different SNRs are used and a 3rd order Daubechies' prototype filter is adopted for the wavelet transform. This order is found sufficient. For the performance evaluation, we have chosen two measures: the segmental SNR SNR_{seg} and the Log Area Ratio (LAR). The first measure permits to show the achieved noise reduction factor for the different parts of the speech signal while the second one exhibits a high correlation with the human auditory process. Computer simulation results and informal listening tests showed that the new algorithm yields a significant improvement of speech quality, especially in pauses where noise is well removed. The results of the new algorithm are shown in Figures 6 and 7. Furthermore, unlike the

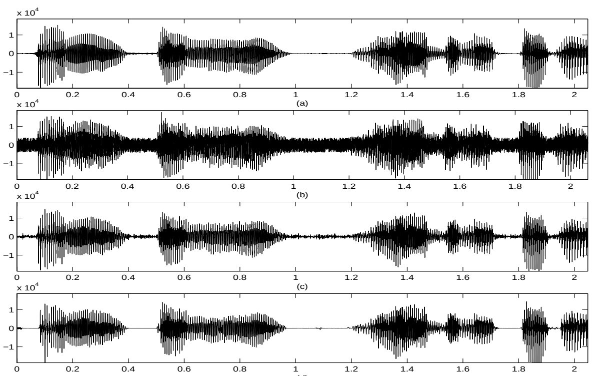

Figure 6: *Resulting speech signals from the noise reduction system (a) clean signal, (b) reference microphone output signal(SNR=4.5 [dB]), (c) wavelet transform based Wiener filter algorithm and (d) new algorithm (white noise, input SNR = 4.5 [dB]).*

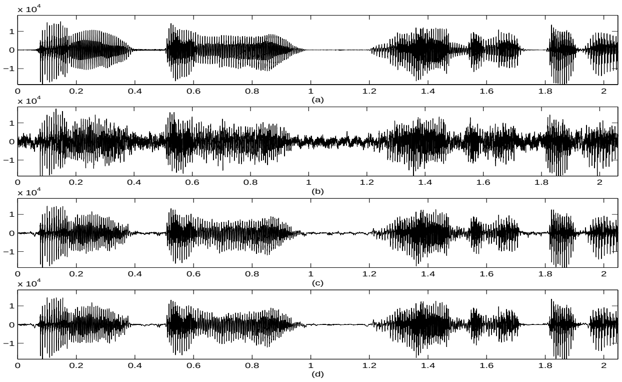

Figure 7: *Resulting speech signals from the noise reduction system (a) clean signal, (b) reference microphone output signal (SNR=6 [dB), (c) wavelet transform based Wiener filter and (d) new algorithm (real factory noise)*

already proposed noise reduction methods based on FFT coherence function [1], the resulting signal sounds natural and no discontinuities are noticed. This is confirmed by the listening tests and LAR improvements shown in Fig. 9. It should be noted that a smaller value of LAR corresponds to a better perceptual quality of the speech signal. Figures 8 and 9 give an idea about the

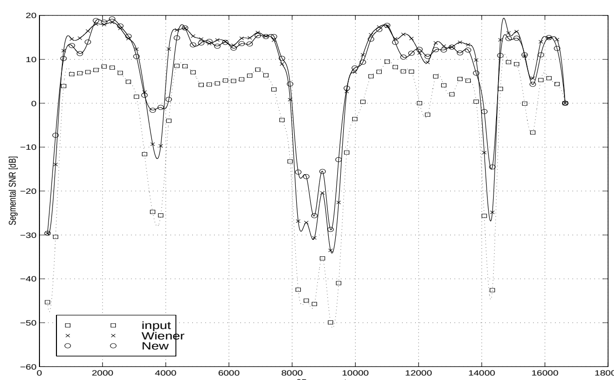

Figure 8: *Segmental SNR improvement of the proposed combined Wiener and coherence filters procedure (White noise).*

achieved noise reduction in terms of SNR_{seg} and LAR for the different parts of the speech signal. Here, the performances are compared between the Wiener filter alone and combined Wiener and coherence filters, both performed in the wavelet domain. The new algorithm provides better performances when the speech signal is absent.

7. CONCLUSIONS

In this paper, a new noise reduction procedure based on the coherence function and applied in the wavelet domain is introduced as a nonlinear postfilter. The coherence filter is combined with a previously proposed wavelet transform based Wiener filter in order to improve the noise reduction, especially in pauses and for the coherent components of the microphones signals. The new algorithm

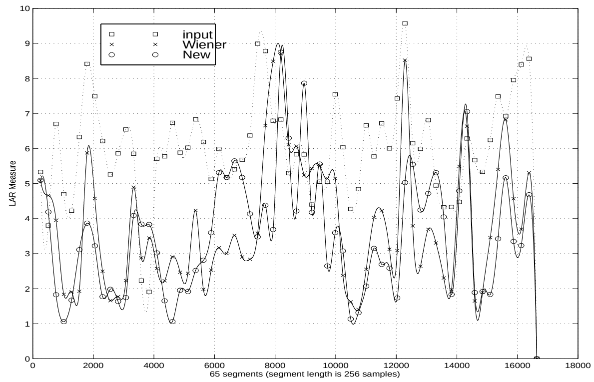

Figure 9: *LAR improvement of the proposed combined Wiener and coherence filter procedure (white noise).*

provides a significant noise suppression with less distortions. This is confirmed by the objective measures and the informal listening tests. The proposed method is thus suitable for an efficient speech enhancement system with a minimum computational complexity.

8. REFERENCES

[1] R. Le Bouquin and G. Faucon. Using the Coherence Function for Noise Reduction. *IEE Proceedings*, 139(3):276–280, June 1992.

[2] B. Le Bouquin-Jeannes, A. A. Azirani, and G. Faucon. Enhancement of Speech Degraded by Coherent and Incoherent Noise Using a Cross-Spectral Estimator. *IEEE Trans. on Speech and Audio Proc.*, 5(5):484–487, September 1997.

[3] G. C. Carter. Coherence and Time-Delay Estimation. *Proc. of IEEE*, 75:236–255, February 1987.

[4] S. Fisher and K. U. Simmer. Beamforming Microphone Arrays for Speech Acquisition in Noisy Environment. *Speech Communication*, 20(3-4):215–227, December 1996.

[5] Y. Kaneda and J. Ohga. Adaptive microphone array system for noise reduction. *IEEE Trans. on ASSP*, 34(6):1391–1400, 1986.

[6] D. Mahmoudi. Multiresolution Array Processing for Speech Source Tracking in Voice Communication Systems. In *Proc. of ICSPAT*, pages 346–350, October 1996.

[7] D. Mahmoudi. A Microphone Array for Speech Enhancement using Multiresolution Wavelet Transform. In *Proc. of Eurospeech'97, Rhodes, Greece*, pages 339–342, September 1997.

[8] D. Mahmoudi and A. Drygajlo. Multiresolution Microphone Array for Speech Source Acquisition and Tracking. In *Proc. of IWAENC'97, London*, pages 116–119, September 1997.

[9] K. U. Simmer and A. Wasiljeff. Adaptive Microphone Arrays for Noise Suppression in the Frequency Domain. In 2^{nd} *Cost 229 Workshop on Adap. Algo. in Com., France*, pages 185–194, Sep. 1992.

[10] R. Zelinski. A Micropohone Array with Adaptive Postfiltering for Noise Reduction in Reverberant Rooms. In *Proc. of ICASSP'88*, pages 2578–2581, 1988.