SPATIAL COHERENCE EXPLOITATION WHICH YIELDS NON-STATIONARY NOISE REDUCTION IN SUBBAND DOMAIN

R. Atay*, E. Mandridake, D. Bastard and M. Najim
Equipe Signal et Image ENSERB and PRC - GDR ISIS
BP 99 - 33402 Talence Cedex - France
limby@goelette.tsi.u-bordeaux.fr najim@goelette.tsi.u-bordeaux.fr
*Faculté des Sciences, Département de Mathématiques et Informatique Meknès, Morrocco
rachid@goelette.tsi.u-bordeaux.fr

ABSTRACT

This paper deals with the enhancement of speech corrupted by real additive noises in a car when two observations are available. As far as we know, no enhancement system was capable of improving both the quality and the intelligibility of the noisy signals. We propose an enhancement method using thresholding, segmentation and filtering in subband domain. The main idea is to expand in subband signals the two observation speech signal and to exploit the spatial coherence of sound from the subband expansion. Noise reduction is conducted using two methods according to the degree of correlation of the subbands of the two observations. The proposed noise reduction approach is applied to realistic situations like speech signals, which are corrupted by non-stationary noises of diverse origins.

1 INTRODUCTION

This work addresses the problem of noise reduction in the speech signals recorded in a car. Since the car is a strongly noisy environment and the signal to noise ratio may be as low as 0 dB perhaps even negative, it is important to find a pertinent criterion to distinguish the useful signal from the disturbing noises.

In this application, we have not any reference concerning the noises. We have only two signals recorded from two microphones, which are put in the sun visor. Therefore, the two available observations contain noises and speechs in the same time. Faucon and al [5] [6] proposed several methods to resolve this problem when one or two observation are available, using Short Time FFT, spectral subtraction and coherence function. In this work, in view of the non-stationary nature of the corrupter noises, we propose to decompose first speech signals into subband signals. This subband decomposition yields subband speech signal components corrupted by the noise component in this subband. If the subband larger is narrow enough, noise component can be considered stationary so traditional noise reduction method is applicable in each subband. The other part of this work is the exploitation of the noise power subband indicator. In fact, there are two microphones, it is interesting to look at the correlation or the non-correlation between the two microphone subband signals. In a sense, having multiple observations brings more robustness for the noise reduction system in the non-stationary case in such way that it exploits the spatial coherence of sound fields better.

The association of the subband decomposition and the coherence function gives better results than using Short Time FFT and coherence in the case of uncorrelated noises. However when the two observations are strongly correlated, it is the case when the two microphones are close to each other, noise reduction using only the coherence function can induce an important alteration of speech. Whatever the degree of filtering, we can not attenuate the noise without degraded the intelligibility. So, we propose the division of the subbands into two sets. The first one is the set of subbands where the value of the coherence becomes significative to perform segmentation and filtering. The second one is the set of subbands where the two observations are too correlated one each other. In this case, we prefer using a method based on learning the power spectra of the noise in silence parts of the signal such as spectral substration. The remainder of the paper is organized as follows. The second section addresses a brief review of the subband decomposition. We present the analyzer/synthesizer filter bank structure, the polyphase implementation of the filters and the synthesis of the prototype filter. In the third section, we discuss the properties of the coherence function obtained from the subband coefficients of the two noisy signals. The behavior of this function is very helpful for determining the thresholds and the set of subbands where we can apply the algorithms presented in [8] for thresholding, segmentation and filtering. The second method presented here is the well-known spectral subtraction, which uses only one observation and needs a learning of noise characteristics in silence parts of the signal. So we combine this method with the segmentation algorithm based on the coherence issued from the first set of subbands and apply it to the second set of subbands. The subband speech enhancement scheme is presented in the fourth section. Finally, we give a comparison of the results issued from the subband and full-band methods. Simulations are conducted on real noisy speech signals recorded in a car.

2 SUBBAND DECOMPOSITION

FIR filter banks are used in a wide range of speech, image and other applications. The two familiar approaches to M-channel system design are the perfect reconstruction banks and pseudo QMF banks [1]. Unlike in speech coding, perfect reconstruction properties is not needed in our application because the enhancement algorithm modifies the subband coefficients and the output enhanced signal is very different than the input noisy one. So, we focus on the analysis aspects than perfect reconstruction aspects.

The subband scheme consists on a filter bank with one common input $x(n)$. This input is split by an analysis filter bank $(H_0(z)...H_{M-1}(z))$ and subsampled to yield the M system subband signals $(X_0(z)...X_{M-1}(z))$.

For an efficient implementation of this system, a polyphase implementation is adopted [2]:

$$\begin{aligned} H(z) \quad = \quad & \sum_{n=-\infty}^{+\infty} h\left(nM\right) z^{-nM} \\ & + z^{-1} \sum_{n=-\infty}^{+\infty} h\left(nM+1\right) z^{-nM} \\ & \cdot \\ & \cdot \\ & \cdot \\ & + z^{-(M-1)} \sum_{n=-\infty}^{+\infty} h\left(nM+M-1\right) z^{-nM} \end{aligned}$$

The FIR filter banks $H_i(z)$ and $F_i(z)$can be obtained by the modulation of a linear phase prototype [2]. In this method, only the prototype filter $P_0(z)$ (see below) is designed.

$$H_k\left(z\right) = H_0\left(zW_M^k\right) \quad \text{where} \quad W_M = e^{-j2\pi/M}$$

The frequency responses $H_k(e^{j\omega})$ are the uniformly shifted versions of the prototype. We can calculate the real coefficients of the analysis and reconstruction filter banks by using the cosine modulation:

$$h_k(n) = 2p_0\left(n\right)\cos\left(\frac{\pi}{M}\left(k+0.5\right)\left(n-\frac{N}{2}\right)+\theta_k\right)$$

$$f_k(n) = 2p_0\left(n\right)\cos\left(\frac{\pi}{M}\left(k+0.5\right)\left(n-\frac{N}{2}\right)-\theta_k\right)$$

where $p_0(n)$ is the impulse response of the prototype.

3 NOISE REDUCTION IN SUBBAND DOMAIN

The noise reduction system proposed here exploit two techniques. The first one is applicable when two input signals are available for processing and the second technique is applicable to the case where there is only one degraded input. The first technique uses the coherence function for segmentation, thresholding and filtering operations. The goal of the coherence is to cut the non-correlated part and to keep the correlated part of the signal. In the case of decorrelated or slightly correlated noises, this function is a pertinent criterion to perform these operations. The correlation between the noises depends principally on the distance between the microphones [5]. In subbands domain, it is known [6] that the coherence becomes not a good criterion along all the subbands. So, we propose to use this technique for segmentation and filtering only on the subbands where the noises are not too correlated. Otherwise, we exploit the segmentation issued from these subbands and combine it with the spectral subtraction on the subbands where the noises issued from the two observations are correlated.

The magnitude of the coherence function varies between 0 and 1 and gives for each subband, the percentage of signal energy coming from the two correlated sources. This value is close to one if we use only the speech for computing it. In opposition, the magnitude of the coherence function is close to zero when we use only the noises as input signals. In subbands, this behavior becomes not valid for all the subbands. The following example illustrates the behavior of the coherence function of subband signals in one subband, i.e. frequency bands of 250 Hz. Figures 3.1 and 3.2 represent the noisy speech inputs obtained from the two microphones. Figure 3.3 is the plot of the value of the coherence calculated up to 0.25 kHz of the two inputs. We can notice that this value is close to one in both silence + noise and speech + noise sequences. In the same way, in figure 3.4, coherence calculated between 2 and 2.25 kHz., the value of the coherence has a too bad behavior to be used for cleaning the speech. In the other hand, in figure 3.5, coherence calculated between 0.25 and 0.5 kHz, we can observe that the coherence has the needed behavior for processing the enhancement. This subband reflects, in general, the main speech activity.

So, the crucial step of this method is the choice of the subbands where we can exploit the coherence properties. This step must be achieved using a long training sequence of noisy speech.

In the next section, we describe the procedure, which for each subband, determines the appropriate method to apply according to the degree of coherence of the subband signals.

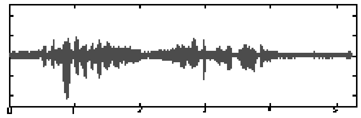


Figure 3.1 microphone 1

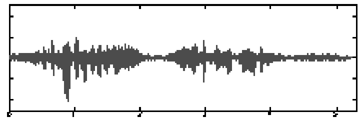


Figure 3.2 microphone 2

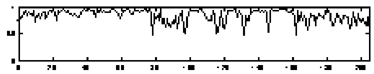


Figure 3.3 coherence function for the first Subband 0-0.25 kHz

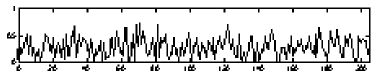


Figure 3.4 coherence function for the subband 2-2.25 kHz

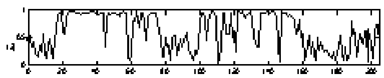


Figure 3.5 coherence function for the second subband 0.25-0.5 kHz

4 SUBBAND SPEECH ENHANCEMENT SCHEME

The system splits each signal into M subbands. For each subband we calculate the value of the coherence [4] with the following expression:

$$\gamma_{x_1 x_2}(f,n) = \frac{G_{x_1 x_2}(f,n)}{\sqrt{G_{x_1 x_1}(f,n)\, G_{x_2 x_2}(f,n)}}$$

where f is the subband index and n is the index of the current block of input data. x_1 and $x_2$are the input signals and can be expressed as follows under the assumption of speech and noise independences :

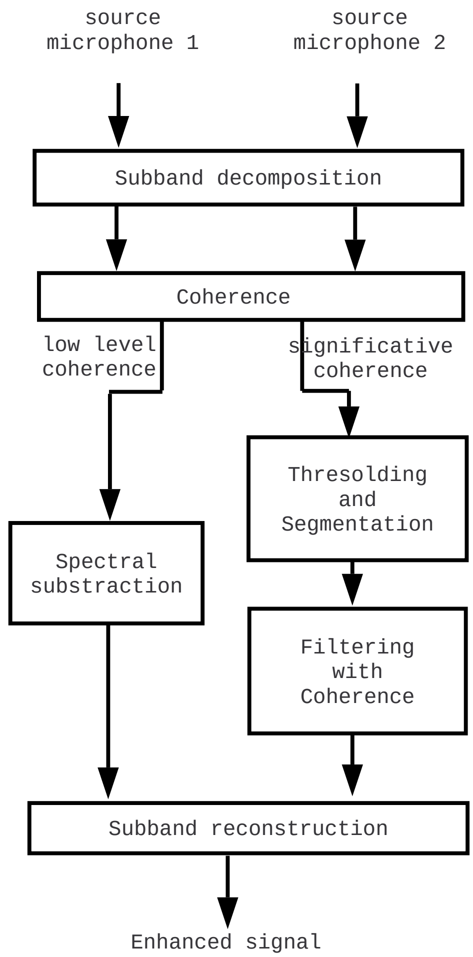
$$x_i(t) = s_i(t) + b_i(t)$$

$s_i(t)$ is the speech and $b_i(t)$ is the additive car noise recorder by the microphone i.

$G_{x_i x_j}(f,n)$ is the power spectra density calculated from the subbands coefficients vectors $X_i(f,n)$ and $X_j(f,n)$:

$$G_{x_i x_j}(f,n) = E\left[X_i(f,n) X_j(f,n)\right]$$

The second step of this method is the partition of the subbands into two sets. The first one is the set when the coherence is significative (i.e. when the noises are not too correlated) and the second one is the remainder of subbands. We use the coherence issued from the first set of subbands to detect speech and pause sequences in the current block with the algorithm proposed in [5]. After this segmentation, we use spectral subtraction to clean the second set of subbands and use the algorithms presented in [8] to filter the first set of subbands. Finally, we use the synthesizer scheme to reconstruct the enhanced signal.



5 RESULTS AND CONCLUDING REMARKS

Simulations were carried out with various real noisy speech signals, which are recorded in car. We use signals from the Matra Communications Database. Informal listening tests, which are the best estimation of efficiency of the proposed method, indicate a distinctly improving of the intelligibility and quality of the speech signals. We note an appreciably attenuation of the noises. In fact, the combination of the segmentation issued from the coherence function and the spectral subtraction allows a total elimination of the noise in the non-speech parts of the signal. So there is no musical noise at the output. In speech parts of the signal, the listening tests indicates that the quality of the output remains good even in the case of the low SNR inputs signals.

In our simulations, we use a 16-subband filter bank obtained from a 40 coefficients low pass filter prototype. The two noisy original observations are depicted in figures 3.1 and 3.2. Figure 5.1 is the spectrogram of the noisy free original signal (this signal is not used in our experiments. We present it only to illustrate the efficiency of our method) and figure 5.2 represents a same part of the traffic noisy signal (Input SNR = 5.09 dB). The following table summarizes the comparison between the proposed composite subband methods scheme and full-band methods.

Method	SNR	Listening
m_1 figure 5.3	9.3	no musical noise speech distorsion
m_2 figure *5.4*	9.7	musical noise speech distorsion
m_3 figure 5.5	10.5	no musical noise low speech distorsion

m_1: full band spectral subtraction[5]
m_2 : full band coherence [8]
m_3: subband speech enhancement scheme

Acknowledgment :
Mr. Atay would like to thank the French "Ministère des affaires Etrangères" for his support to the project A. I. 95 / 0841

References

[1] R. D. Koilpillai, P.P. Vaidyanathan, " Cosine-Modulated FIR Filter Banks Satisfying Perfect Reconstruction ". IEEE Trans. Signal Processing, vol. 40, pp. 770-783, Apr, 1992.

[2] P. P. Vaidyanathan, Multirate Systems and Filter banks. Englwood Cliffs, NJ: Prentice-Hall.

[3] H. Xu, W. Lu, A. Antoniou, " Efficient Iterative design Method for Cosine-Modulated QMF Banks " IEEE Trans. Signal Processing, vol. 44, pp. 1657-1668, July, 1996.

[4] G. C. Carter, "Coherence and Time Delay Estimation". Proc. of the IEEE, Vol. N°2, pp. 236-255, February 1987.

[5] R. Le Bouquin, "Enhancement of noisy speech signals: Application to radio mobile communications". IEEE Speech Com. Vol. 18, N°1, January 1996

[6] R. Le Bouquin, G. Faucon, "Speech Enhancement Using Sub-Band Decomposition and Comparison with Full-Band Techniques". Signal Processing VII: Theories and Applications. 1994 E.A.S.P.

[7] P. Pollak, P. Sovka and J. Uhlir, "Noise Suppression System for a Car".Eurospeech '93. PP. 1073-1076. Berlin, Sep. 1993

[8] F. Ehrmann, R. Le Bouquin, and G. Faucon, "Optimisation of a Two-Sensor Noise Reduction Technique". IEEE Signal Processing Letters, Vol. 2, N°6, pp. 108-110, June 1995.

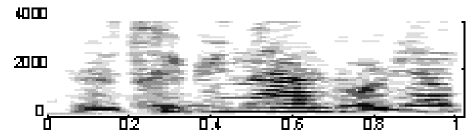


Figure 5.1: Original signal

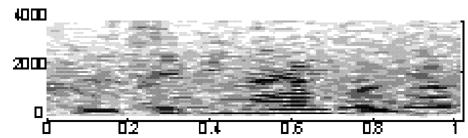


Figure 5.2: Noisy signal

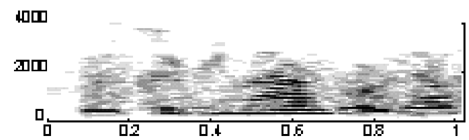


Figure 5.3: full band spectral subtraction

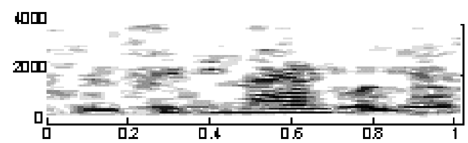


Figure 5.4: full band coherence

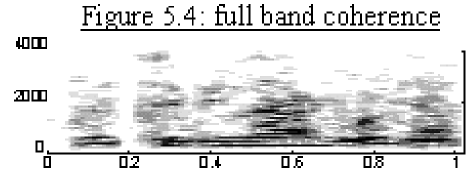


Figure 5.5: subband spectral subtraction and coherence