Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 50(64), Fascicola 1, 2005

Speech enhancement using Enriched Diversity Wavelet Transform and Bishrink filter

Cedric Clolus[1]

Abstract - This paper presents a new denoising method for speech signals corrupted by additive noise. This method is based on the use of a special wavelet transform, called enriched diversity wavelet transform and of a special MAP filter, called composed bishrink. Some simulations are presented. The results obtained are compared with the results of other denoising methods.

Keywords: wavelets, denoising, soft thresholding, bishrink.

I. INTRODUCTION

A lot of applications in the domain of signal processing require voice activity detection: speech coding, echoes reduction in telephony domain for example. Some solutions for systems with a voice activity detector have been standardized, but these solutions lack of robustness. It is why other solutions for voice activity detection are searched. Those one try to integer in the structure of the voice activity detector methods of denoising speech signal. In this paper, we consider a method of denoising based on two new tools of signal processing, the enriched diversity wavelet transform and the Bishrink filter. Section II deals with some already existing methods of denoising speech signal. In Section III, the proposed method is presented. Section IV focuses on the results obtained by simulation of the preceding methods. The last section is dedicated to some concluding remarks.

II. DENOISING METHODS

In this section, we will focus on two main methods to denoise speech signals, the spectral subtraction and a method based on statistic properties of the signal.

1. Spectral Subtraction [1] is a method for restoration of the power spectrum or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum. The noise spectrum is usually estimated from the periods when the signal is absent and only the noise is present. For restoration of time-domain signals, an estimate of the instantaneous magnitude spectrum is combined with the phase of the noisy signal, and then transformed via an inverse discrete Fourier transform of the time domain. In terms of computational complexity, spectral subtraction is relatively inexpensive. The block diagram of this method is presented by the figure 1.

2. The second method for speech enhancement on which we will focus is the system presented by Y.Ephraim and D.Malah [2][8] based on the utilization of a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator (MMSE STSA). The model used is a statistical model which utilizes asymptotic statistical properties of the Fourier expansion coefficients and more specifically, the fact that the Fourier expansion coefficients of each process can be modeled as statistically independent Gaussian random variables. The mean of each coefficient is assumed to be zero, since the processes involved are assumed to be zero mean. The variance of each speech Fourier expansion coefficient is time-varying, due to speech nonstationarity. This Gaussian statistical model is motivated by the central limit theorem, as each Fourier expansion coefficient is a weighted sum (or integral) of random variables resulting from the process samples. The MMSE STSA estimator depends on the parameters of the statistical model it is based on. In the algorithm proposed in [2], these are the *a priori* SNR of each spectral component and the variance of each noise spectral component. The *a priori* SNR was found to be a key parameter of the STSA estimator. Their method consists in combining a “decision-directed” method for estimating the *a priori* SNR with the MMSE STSA estimator which takes into account the uncertainty of signal presence in the noisy observations.

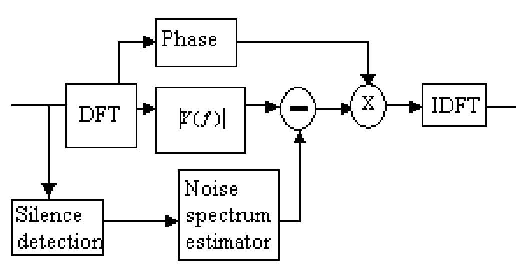

Fig.1. Block diagram of a spectral subtraction system

[1] ENSIETA - Brest, France

III. SPEECH ENHANCEMENT USING ENRICHED DIVERSITY WAVELET TRANSFORM AND BISHRINK FILTER

In recent years, the techniques that use multiscale and local transform-based algorithms have become popular in noise filtering applications. In particular the use of non-linear filters in the DCT domain was studied. In this section we proposed to present a denoising method based on the enriched diversity wavelet transform (EDWT) [5] and on the Bishrink filter [6] [7].The block diagram of this method is presented in figure 2.

For a given signal, in using different mother wavelets, different concentrations of energy are obtained. Thus, for a given input signal, there is a mother wavelet which maximizes the concentration of energy. The idea of the enriched diversity wavelet transform, proposed in [5], is obtaining a discrete wavelet transform which is less sensitive to the choice of the mother wavelet. The construction of this wavelet transform is based on the increase of the diversity. The parameters of the discrete wavelet transform are the mother wavelet and the number of iterations. Thus, the diversity can be increased in calculating for a same signal, *x(t)*, several discrete wavelet transforms. For each one, a different mother wavelet is used. So, the enriched diversity wavelet transform is obtained. It is a redundant discrete wavelet transform. This transform carries out association between the vector *x(t)* and a matrix *EDWT[t,l]*. Each column of this matrix represents one of the discrete wavelet transform of the signal *x(t)*. This transform can be inversed. For each columns of the matrix *EDWT[t,l]*, the corresponding inverse discrete wavelet transform is calculated. So, a new matrix is obtained. Each column of this matrix contains the signal *x(t)*. In calculating the mean of the columns of this matrix, the vector *x(t)* is obtained. The filtering in the domain of the enriched diversity wavelet transform is obtained by filtering each component of the transform.

In [6], Sendur and Selesnick proposed a wavelet-based denoising filter which considers the dependencies between the wavelet coefficients and their parents in detail. In fact, there are strong dependencies between neighbor coefficients such as between a cofficient, its parent (adjacent coarser scale location), and their siblings (adjacent spatial locations). From this observation, Sendur and Selesnick built four jointly non-Gaussian models to characterize the dependency between a coefficient and its parents, and derived the corresponding bivariate MAP estimators based on noisy wavelet coefficients in detail. From each model, a denoising method is proposed in taking into account the dependencies between the wavelet coefficients and their parents in detail.

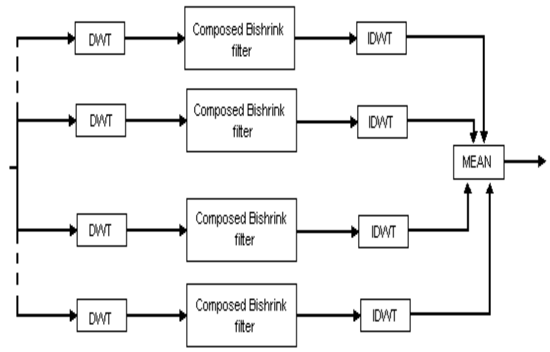

Fig.2. Block diagram of the proposed method

As is said previously and in [7], there is an important correlation between a wavelet coefficient at a given scale and the same coefficient situated in the same position at the next scale. This correlation can be exploited to construct adaptive filters acting at a given scale and using for the estimation of their parameters information obtained at the next scale. Using the parent and child wavelet coefficient of the input signal it is possible to estimate the child coefficients of the discrete wavelet transform of the useful part of the input signal, with the aid of a bishrink filter.

Concerning the model of the DWT of the useful component, in the case of our method's filter, named composed Bishrink filter, for the first iterations (for scales with a number of coefficients superior with 16), a Laplace distribution will be considered (like in the case of the bishrink filter). For the last iterations, a soft thresholding with a threshold equal to 3 σ_n, where σ_n is the noise variance (the rule of the 3 sigma), is used.

IV. SIMULATION RESULTS

This section will present the simulation done to test the new method and the comparison done with two others methods. The simulation have been carried out with two different noises, first a Gaussian white noise and secondly with a F16 noise, what is to say a noise recorded at the co-pilot's seat in a two-seat F-16, traveling at a speed of 500 knots, and at an altitude of 300-600 feet. These noises have been added to a clean speech signal (figure 3) with different Signal-to-Noise Ratios. The simulations have been carried out with four different denoising methods. The first two methods used for the simulation are the spectral subtraction proposed by Kamath [8] and the statistical method using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator (MMSE log-STSA) proposed by Ephraim and Malah [9]. The denoising has been also carried out with the method proposed in section III (EDWT). The final method used for the simulation is derived of the precedent method. Indeed, after the denoising of the noisy signal by the enriched diversity wavelet transform, a wrap detection is done on the denoised signal in order to obtain a voice

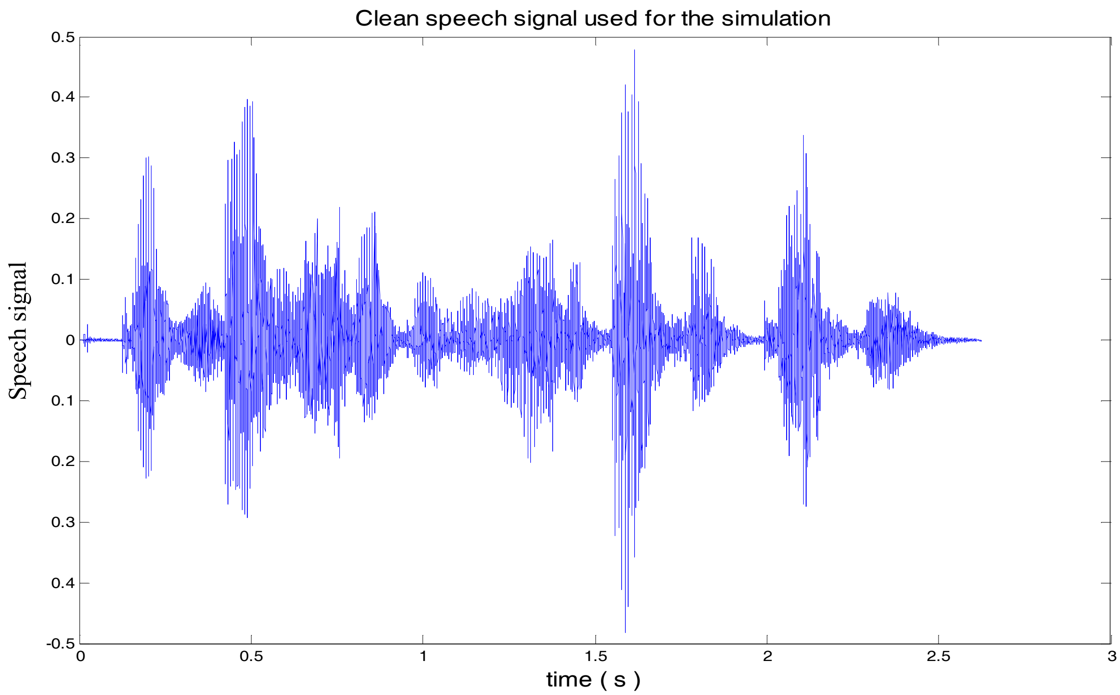

Fig.3. Clean speech signal used for the simulation

activity detection. The result is multiplied with the denoised signal to obtain the fourth result of the simulation (EDWT + VAD). The figures 4 and 5 represent the relation between the input SNR and the output SNR for the two different noises. So, with a signal corrupted by a Gaussian white noise, the proposed method obtained better results than the others, for any input SNR. However, for the F16 noise, the proposed method obtained best results only for an input SNR superior to 3 dB. Thus, the denoising by using Enriched Diversity Wavelet Transform and Bishrink filter is, in the most part of

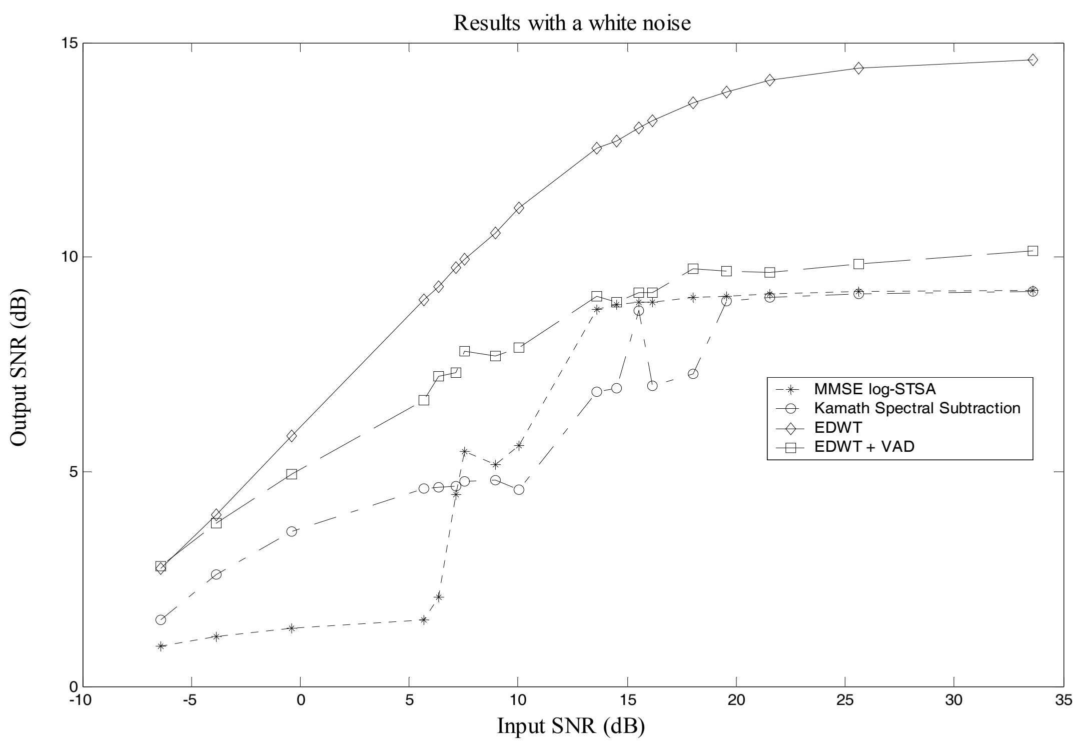

Fig.4. Results of simulation with a white noise

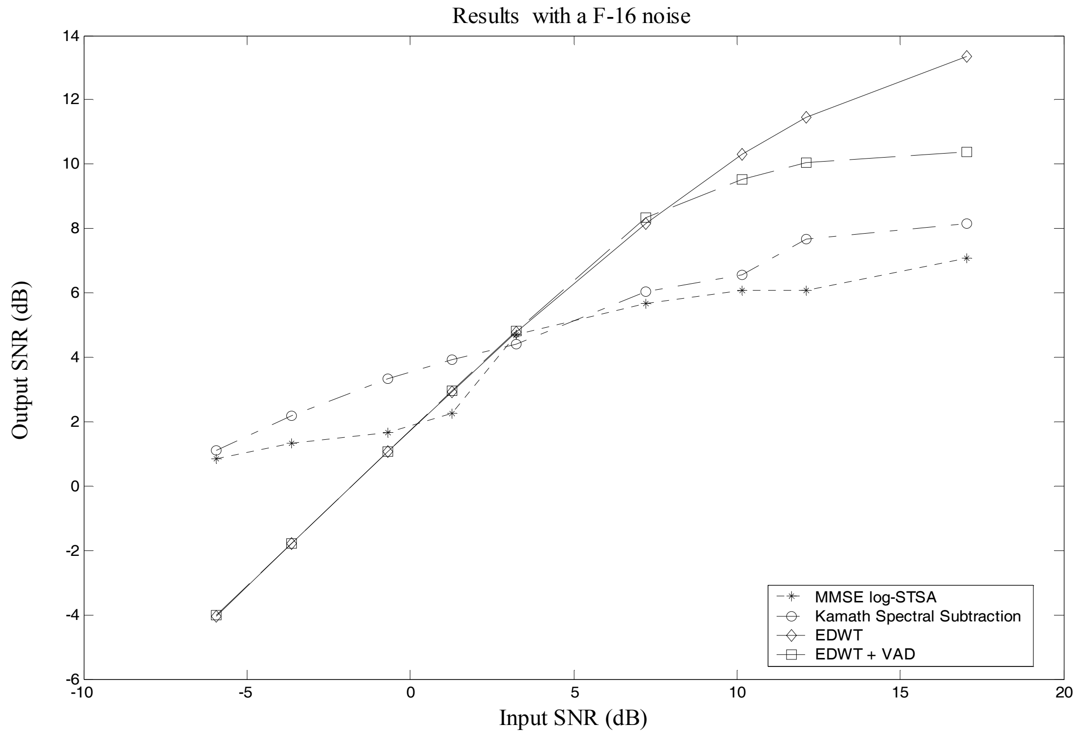

Fig.5. Results of the simulation with a F-16 noise

cases, more efficient than the others methods, but for some particular noise and for low input SNR, it obtains less good result.

V. CONCLUSION

In this paper is proposed a new denoising method based on the use of several discrete wavelet transforms and of the bishrink filter in the wavelet domain. This method takes into account also the statistics of the useful part of the input signal. That makes that this method to perform better in most cases than the denoising method based on spectral subtraction and some denoising method based on the statistical properties of the signal. Although the Bishrink filter was originally envisaged for denoising signals corrupted by white noise, it allows having good results with all kind of noise, even if the best results remain obtained with signals corrupted by white noise. A problem of the proposed method is its computational complexity more important than the computational complexity of the other methods used for the test presented in this article. Finally, the results obtained during the simulation showed that the proposed methods for denoising can be easily integrated in a voice activity detection system and allow obtaining good results.

REFERENCES

[1] S.V.Vaseghi, "Spectral Subtraction", *Advanced Digital Signal Processing and Noise Reduction*, Second edition, Chap. 11, 2000.

[2] Y.Ephraim and D.Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", *IEEE Trans. Acoust. Speech Signal Process.,* vol. ASSP-32, no.6, pp.1109-1121, Dec. 1984.

[3] Y.D.Cho, K.Al-Naimi, and A.Kondoz, "Improved Voice Activity Detection Based on a Smoothed Statistical Likehood Ratio", *CCSR, Proc. IEEE ICASSP,* May 2001.

[4] K. Al-Naimi, "A Robust Noise and Echo Canceller", *CCSR, Unis,* 2003.

[5] A.Isar, A.Quinquis, and M.Legris, "Le débruitage des images SAR", Journée d'Etudes de la SEE, Radar et environnement marin, Brest, 10 juin 2003.

[6] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency", IEEE Trans. on Signal Processing. 50(11): 2744-2756, November 2002.

[7] A.Quinquis, A.Isar, and D.Isar, "Denoising Over-Sampled Signals", *Scientifically Bulletin of "Politehnica" University, Electronics and Telecommunications series,* tom 49(63), 2004.

[8] Kamath, S. and Loizou, P. (2002). "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise, " Proceedings of ICASSP-2002, May 2002.

[9] Y. Ephraim and D. Malah, Speech enhancement using a minimum mean-square error log-spectral amplitude estimator, IEEE Trans. Acoust. Speech Signal Process. ASSP-33 (2), April 1985