Research Journal of Applied Sciences 2 (2): 119-123, 2007

Simulation and Comparison of Noise Concellation Techniques in Speech Processing

Talbi Mourad
Signal Processing Laboratory, Science Faculty of Tunis, 1060 Tunis

Abstract: In this study we will develop a speech denoising interface which will be used for recognition, synthesis and coding applications. This interface developed under Matlab, uses wavelets transform. The results of this method are applied on several kind of noises then they will be compared with the spectral subtraction method.

Key words: Speech processing, denoising, wavelets, spectral subtraction, SNR, matlab

INTRODUCTION

Actually, there are numerous denoising techniques used in speech processing. Most of them include hypotheses on the original signal, as well as SNR ratio and distortion (Whitmal *et al.*, 1996). However these techniques do not cover all the explicit speech models. Each of them is associated with a particular type of distortion while maximizing noise-reduction effects. To rich this objective, we have to create a compromise between denoising and speech signal distortions by using a short-time spectral representation. Other techniques apply the new decomposing concept on an adapted basis of short time waves, called atoms or wavelets.

Short-time spectral attenuation methods: The best description of the functioning of short-time spectral attenuation methods is outlined by Mc Aulay and Mallpass (1980). Lets consider a noisy data:

$$y(n) = x(n) + b(n) \tag{1}$$

where x(n) designs the clean signal and b(n) the noise, the denoising principle using spectral attenuation, is provided in Fig. 1.

Noise attenuation is performed through a suppression law and a previous noise estimation. This suppression rule satisfies the following short-time spectral attenuation:

$$\begin{Bmatrix} G(w) = 1 \text{if } \hat{P}_x(w) >> \hat{P}_b(w) \\ G(w) = 0 \text{if } \hat{P}_x(w) << \hat{P}_b(w) \end{Bmatrix} \tag{2}$$

where $\hat{P}_x(w)$ and $\hat{P}_b(w)$ are respectively power spectral densities of clean signal x and noise b and G is the added gain to each value of the STFT. Generally, it is useful to include a supplementary procedure allowing to detect the signal absence during the processing.

Spectral subtraction: The spectral subtraction noise cancellation introduced by Boll (1979), is based on an estimation of the short-time spectrum magnitude of the original signal, taking into account the human auditory perception and the phase information. The Fig. 2 illustrates the principle of the spectral subtraction.

FT is the classical Fourier Transform and FT^{-1} is its inverse. α is a compression factor. $P_y(w)$ was obtained by suppressing the phase information which is regenerated during the estimation of the enhanced speech $\hat{s}(t)$, this is obtained after subtracting the estimate of power spectral density of the noise $\hat{P}_b(w)$ from $P_y(w)$.

Wavelets methods: The denoising technique using wavelets transform consists in applying the concept of decomposition in a wavelets adaptive base. This method has proved its efficiency in enhancement of signals corrupted by an additive noise (Chen, 1995).

Thresholding denoising method: Donoho and Johnstone, (1994) had developed a signal denoising method, which uses the wavelets coefficients contraction technique. It is described by the following algorithm: (Fig. 3 and 4).

In this study, two kinds of thresholding are used, a soft and hard thresholding.

There are several kinds of thresholds; constant or variable thresholds. References (Donoho and Johnstone, 1994, 1995) propose a value of global threshold,

$$\lambda = \sigma\sqrt{2\log n} \tag{3}$$

where σ is the noise variance.

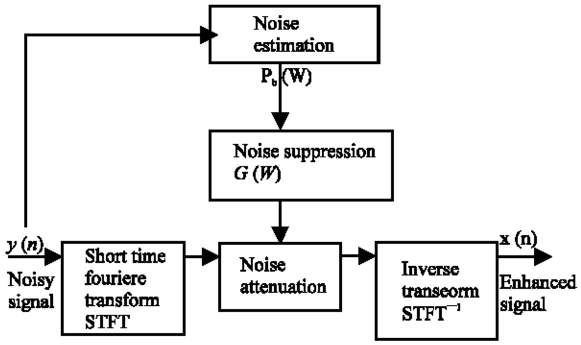

Fig. 1: Principle of noise reduction by spectral attenuation

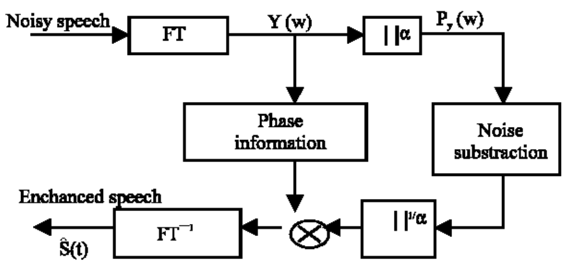

Fig. 2: Spectral subtraction principle

Denoising method with best bases: The linear signals decompositions are associated with paving of time-frequency plane by Heisenberg boxes having a fixed minimal value surface. Instead of considering a single basis, we can use a dictionary of orthonormals bases or consider a redundant family of time-frequency atoms and search among all, the best representation of the signal. From best orthonormal basis (Wickerhouser, 1994; Coifman and Wickerhouser, 1992), we define a cost function that expresses approximately the good adaptation of the orthonormal base. The obtained time-frequency representation is also a paving. Saito (1994) and Pesquet *et al.* (1998) had suggested a promising denoising method of signals corrupted by an additive white noise. This technique using wavelets packets and minimal description length criterion MDL (Rissanen, 1989), was improved later by Whitmal; Saito and Cofiman (1995), for enhancement of speech signal corrupted by correlated noise. This algorithm proposes to use the local discrimination bases LDB developed.

The used method: In the experimental part, we will compare the two denoising methods by spectral subtraction and thresholding. This comparison is based on two judgment criteria. An objective judgment based on

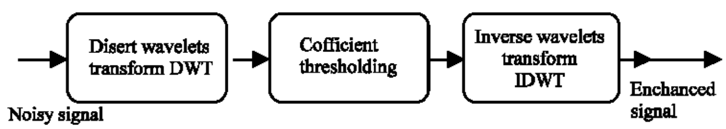

Fig. 3: Thresholding denoising principle

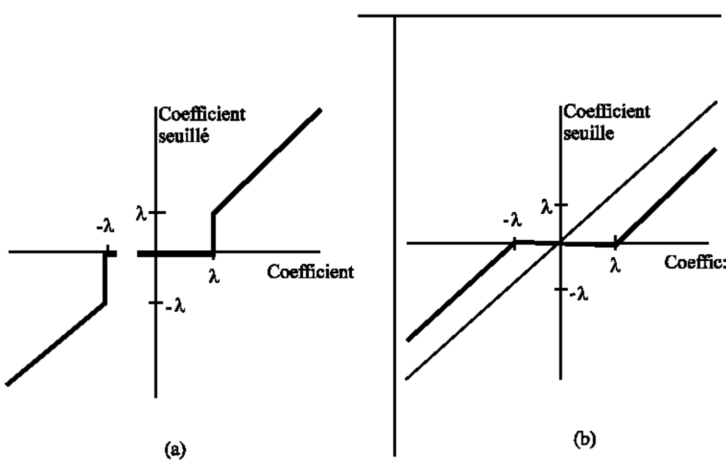

Fig. 4: (a) Hard thresholding. (b) Soft thresholding

the calculation of Signal to Noise Ratio SNR (in dB) and a subjective one by making a listening test and calculating the recognized words percentage from the total number of the words.

RESULTS AND DISCUSSION

We have used a speech database constituted by twenty sentences and 125 words pronounced in Arabic language by two speakers (male and female). All of these words and sentences are corrupted by two types of noise, large and narrow bands(a white noise and Volvo noise) with different values of the signal to noise ratio before denoising SNRi (25, 20, 15, 10, 5, 0 and -5dB). The obtained values of the signal to noise ratio after denoising SNRf are reported in the Table 1-4. Figure 5 represents a noisy speech sound sound with a car noise characterized by a lower SNR (-5dB).

SNR after enhancement: Table 1 to 4 show that soft thresholding denoising method improves the signal to noise ratio (SNRf≥SNRi) when SNRi takes the values 10dB, 5dB, 0dB and -5dB while spectral subtraction denoising method improves the RSB for all values taken by SNRi. In case of white noise, the wavelets denoising method is more efficient, however for volvo noises, the spectral subtraction seems to be more suitable and reliable.

Table 1: Male voice corrupted by a volvo (car) noise

RSBi SNRi	RSBf Denoising by wavelets transform Hard thresholding	Soft thresholding	Denoising by spectral subtraction
25dB	15.054	15.744	34.691
20dB	15.025	15.211	34.262
15dB	14.943	15.152	33.189
10dB	14.711	14.962	31.728
5dB	14.085	14.413	29.684
0dB	12.54	13.046	27.103
-5dB	9.699	10.646	24.135

Table 2: Male voice with white noise

SNRi	SNRf Denoising by wavelets transform Hard thresholding	Soft thresholding	Denoising by spectral subtraction
25dB	13.396	13.464	32.227
20dB	13.281	13.348	31.485
15dB	12.935	12.998	30.275
10dB	12.089	12.05	28.026
5dB	9.915	9.954	23.941
0dB	5.962	6.507	20.636
-5dB	0.459	2.145	14.415

Table 3: female voice corrupted by a volvo noise

SNRi	SNRf Denoising by wavelets transform Hard thresholding	Soft thresholding	Denoising by spectral subtraction
25dB	13.435	13.838	32.382
20dB	13.407	13.536	31.795
15dB	15.31	13.493	30.832
10dB	13.092	13.306	29.312
5dB	12.455	12.772	26.915
0dB	10.647	11.409	23.795
-5dB	8.084	8.759	20.661

Table 4: Female voice corrupted by a white noise

SNRi	SNRf Denoising by wavelets transform Hard thresholding	Soft thresholding	Denoising by spectral subtraction
25dB	14.786	14.923	34.497
20dB	14.682	14.823	34.089
15dB	14.325	14.466	33.187
10dB	13.385	13.501	31.82
5dB	11.314	11.382	30.237
0dB	7.875	7.919	27.604
-5dB	3.429	1.317	24.009

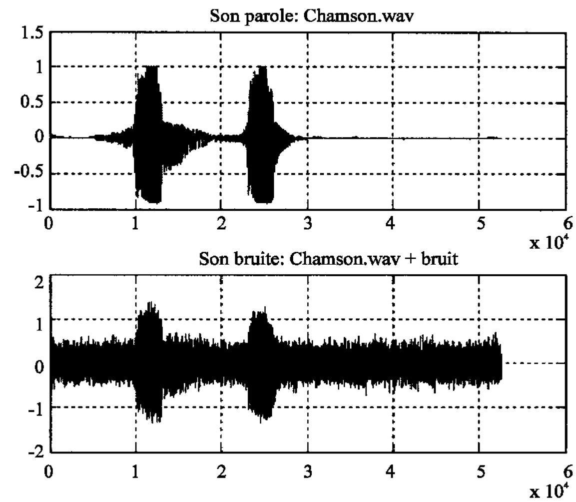

Fig. 5: Temporal waveform of a noisy male voice dharaba with SNR = -5dB

Listen test: We illustrated in Fig. 6a-c the percentage of recognized words versus the SNR ratio for three methods: Spectral subtraction, wavelets hard thresholding and soft thresholding related to a male voice corrupted with white and car noise.

The recognition percentage decreases considerably when the SNR moves from 25dB to -5dB. In this case, the results show that the thresholding denoising method is better than the second method when SNRi equals -5dB. For a of white noise, we have the same thing but instead of one value of SNRi, we have all values between -5dB and 0dB.

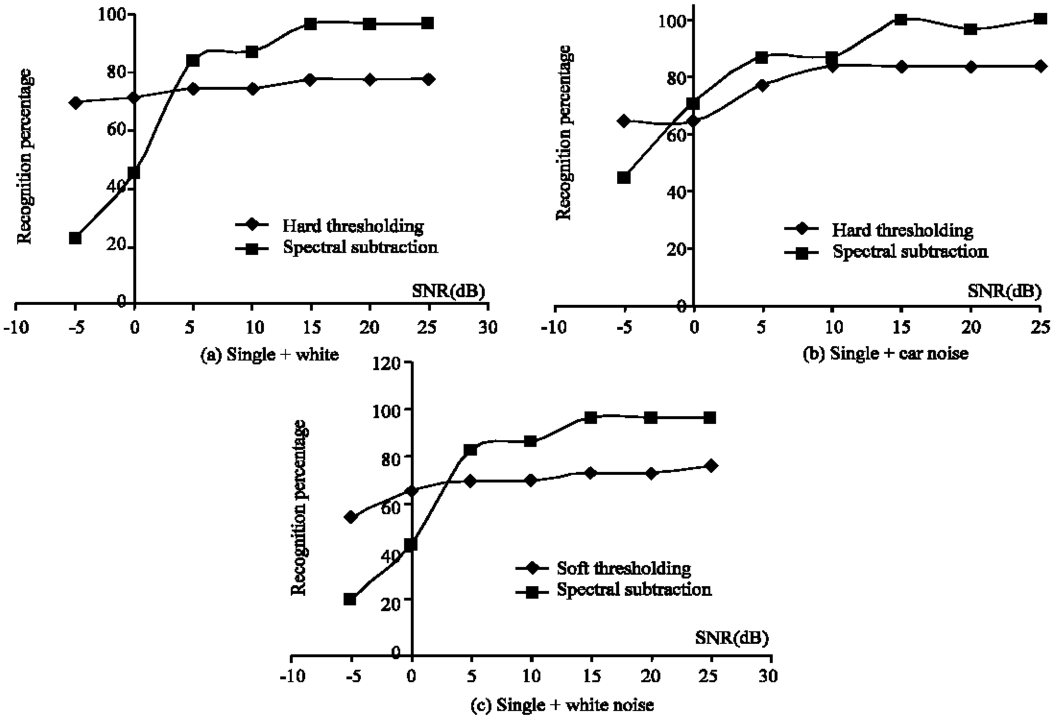

Fig. 6: Recognition ratio vs SNR with several noises

Fig. 7: Matlab speech processing menu

Enhanced signal representation: We have developed under Matlab a new speech processing interface using a noise cancellation procedure illustrated by Fig. 7. The pitch period is computing by the cepstral method, yet the formant frequencies are deduced from LPC spectra. The example illustrates the spectral parameters (speech waveform, zero-crossing, LPC spectra, pitch, spectrogram) of a voiced female sound sampled at 11025 Hz.

Figure 8 and 9 represent the temporal evolution (under Matlab) of noisy and enhanced speech signals using two wavelets thersholding methods (hard and soft).

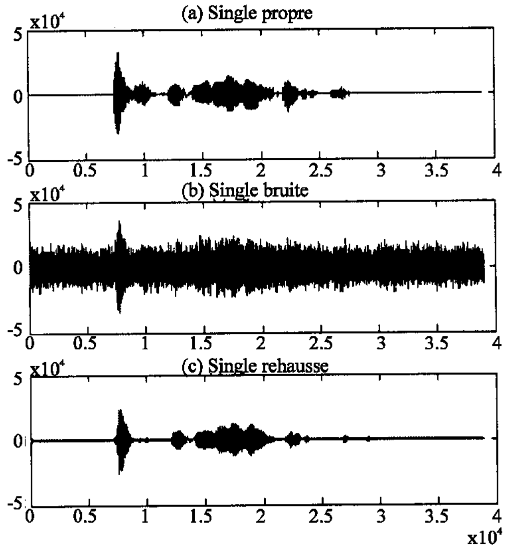

Fig. 8: Hard thresholding denoising method (SNR = -5dB)

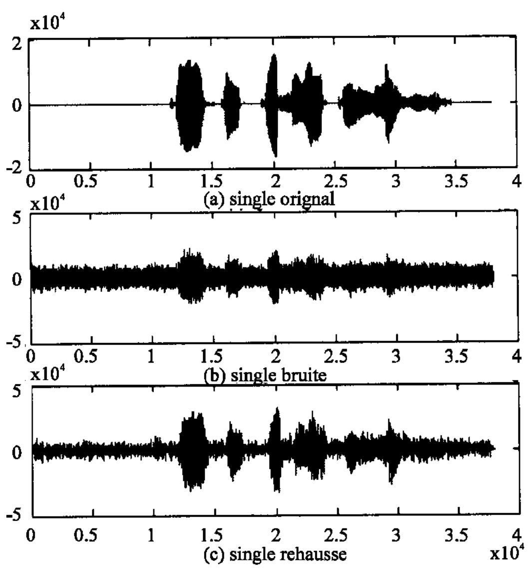

Fig. 9: Soft thresholding denoising method (SNR = -3dB)

We can observe that even for high noisy speech signals (SNR<0 dB) we succeeded to reconstitute a good quality of the enhanced signal especially with the hard thersholding wavelet method.

CONCLUSION

In this study, we have developed under Matlab a denoising speechlab program using wavelets method. The experiments are conducted by several speakers under noisy environment. The obtained results of enhanced signals, SNR ratios and listening tests show that denoising spectral subtraction and wavelets thresholding improve the SNR ratio and the speech intelligibility. The first method seems to be suitable for moderated noisy signals with SNR >0db, yet the second technique is useful for signals characterized by an SNR between 0dB and -10dB.

REFERENCES

Boll, S.F., 1979. Suppression of Acoustic Noise in speech using Spectral Substruction. IEEE Trans. on ASSP, Vol. 27.

Chen, S.S., 1995. Basis Pursuit, Ph.d Thesis, Standford University.

Coifman, R.R. and M.V. Wickerhauser, 1992. Entropy based Algorithm for best basis Selection. IEEE Transactions on Information theory, 38: 713-718.

Donoho, D., I.M. Johnstone, G. Kerkyacharian and D. Picard, 1995. Wavelet Shrinkage: Asymptotia, J. Royel Stat. Soc., Serie B, 57, pp: 3019-3069.

Donoho, D. and I.M. Johnstone, 1994. Ideal spatial Adaptation via Wavelet Shrinkage. Biometrika, 41: 425-455.

McAulay, R.J. and M.L. Mallpass, 1980. Speech enhancement using soft noise suppression filter. IEEE Transactions on Acoustics, speech and Signal Processing, 28: 137-145.

Pesquet, J.C., H. Krim, H. Carfatan and J.G. Proakis, 1993. Estimation of noisy signals using time invariant wavelet packets. 27th Asilomar Conf. Sig. Syst. Comp., pp: 635-640.

Rissanen, J., 1998. Stochastic Complexity in Statistical Inquiry. Singapore, World Scientific Publishning.

Saito, N. and R.R. Coifman, 1995. Local Discriminant Bases and their Applications. J. Math. Imag Vision, 5: 337-358.

Saito, N., 1994. Simultaneous noise suppression and signal compression using libary of orthogonal bases. Wavelets in Geophysics, Foufoula Georgiou, E. and P. Kumar, Eds, Academic Press.

Whitmal, N. , J.C. Rutledge and J. Cohen, 1996. Reducing correlated noise in digital hearing. IEEE Med. Biol. Magazine, 15: 88-96.

Whitmal, N., J.C. Rutledge and J. Cohen. Wavelet based noise reduction for digital hearing aids: Objective evaluation. TechReport (IEEE Eng. Med. Biol. Magazine), Departement of Electrical and Computer Engineering, Northwestern University.

Wickerhauser, M.V., 1994. Adapted Wavelet Analysis from Theory to Software. IEEE Press, A.K. Peters Wellesley Massachusetts.