Novi Sad J. Math.
Vol. 41, No. 1, 2011, 11-20

WAVELETS IN A PROBLEM OF SIGNAL PROCESSING

Wendel Cleber Soares[1], Francisco Villarreal[2], Marco Aparecido Queiroz Duarte[3], Jozué Vieira Filho[4]

Abstract. In this work a new method is proposed for noise reduction in speech signals in the wavelet domain. The method for signal processing makes use of a transfer function, obtained as a polynomial combination of three processings, denominated operators. The proposed method has the objective of overcoming the deficiencies of the thresholding methods and the effective processing of speech corrupted by real noises. Using the method, two speech signals are processed, contaminated by white noise and colored noises. To verify the quality of the processed signals, two evaluation measures are used: signal to noise ratio (SNR) and perceptual evaluation of speech quality (PESQ).

AMS Mathematics Subject Classification (2010): 00A69, 00A05

Key words and phrases: discrete wavelet transform, noise reduction, polynomial combination

1. Introduction

In speech processing, one of the factors that has been attracted interest of the scientific community is how to reduce its present noise without deteriorating speech quality [11]. In general, speech can be contaminated by artificial or real noises. Given a clean signal, white or colored noise is added, resulting the noisy signal in time domain. In this sense, noise reduction is important in the most varied applications involving signal processing. In order to reduce noise, there are several methods, some of them using Fourier transform [12] and others using the discrete wavelet transform (DWT)[5, 3, 9]. The applications involving the DWT have increased in the last years due to the form that this transform acts on the signal that is being processed [7]. The most used noise reduction methods in the wavelet domain are the thresholding methods, because they yield good results for signals corrupted by white noise, but they are not so efficient when colored noise is considered. Colored noise is more common noise in real situations. In those methods, threshold is usually calculated in silence intervals, when speech is considered, and applied for the whole signal. The coefficients in

[1]Adamantinenses Integrated Universities - FAI SP, Brazil
[2]Depart. of Mathematics, FEIS, UNESP, SP, Brazil
[3]Depart. of Mathematics, UUC, UEMS, MS, Brazil,
e-mail: wendelsoares@aluno.feis.unesp.br
[4]Depart. Electric Engineering, FEIS, UNESP, SP, Brazil

the wavelet domain are compared to this threshold and those that have absolute value below it are eliminated or reduced, making a linear application of this threshold. However, such elimination frequently causes time and frequency discontinuities in the processed signal. Therefore, the form in which the threshold is computed can deteriorate voice segments of the processed signal, mainly in cases where the threshold depends strongly on the last window of the last silence interval. In this work, the wavelet transform is applied to speech, obtaining the signal in the wavelet domain, and the proposed methodology is applied to this signal in order to reduce or attenuate noise without thresholding. The noise reduction methodology proposed consists in the application of three processings, denominated as partial and independent operators. The application aims to reduce the present noise making the signal intelligible. The operators are independently generated, however the curve used for noise reduction is generated by a combination of them, resulting in a polynomial combination. Each operator acts according to its features in voice and silence segments of the speech, without a threshold usage. The three processings executions are condensed in a single polynomial function, a transfer function denoted by *fil*, and it acts as a filter in the speech processing. Thus, a filtering technique is introduced and its processing makes noise elimination or reduction of noise coefficients without thresholding the signal. The objective is reducing the noise and avoiding distortions the applying a polynomial combination in the wavelet domain. The quality of signals used for test is evaluated using two objective measures, SNR and PESQ.

2. Proposed Methodology

The processing proposed in this work depends on the noisy signal alone or on a combination of this with the background noise estimate. Each of these operators is individually generated, in order to extract different characteristics of the signal to be processed. Their polynomial combination generates a curve used as a filter for the noise present in the signal. In equation (1) the basic processing function is presented $f(s,r)$, where s represents the noisy signal and r is the noise present in the signal or even the same signal. The variable $d = d(s,r)$ depends on each processing type.

$$f(s,r) = \frac{|s| + |r|}{d(s,r)} \tag{1}$$

To explain the three processing applications, consider $S[n]$ in the wavelet domain, a finite length observation sequence (considered here N a power n of base 2: $N = 2^n$) of a clean signal $X[n]$ corrupted by addictive noise, artificial or real, $R[n]$, according to the following equation:

$$S[n] = X[n] + R[n] \tag{2}$$

The clean signal $X[n]$ and the noise $R[n]$ are independent random processes. The objective is to recover the clean signal $X[n]$ from $S[n]$. The three operators will be applied to the noisy signal $S[n]$, in the wavelet domain. The coefficients of $S[n]$ are given as follows.

$$S[n] = \{s_0, s_1, s_2, \ldots, s_{N-1}\} \tag{3}$$

Each processing applies different functions, here denominated operators. For the application of the first processing, an operator called *simple average operator* (SAO) is considered. The two others are *noise to signal ratio a priori* (SNRP) and *and noise to signal ratio a posteriori* (SNRPR). Executing the operators, in each case, gets the respective set $F[n]$ associated to it, presented in sections 2.1, 2.2 and 2.3, respectively. The elements of the referred set are coefficients computed according to the respective processing. The elements of each set $F[n]$ will be indicated by f_i:

$$F[n] = \{f_0, f_1, f_2, \ldots, f_{N-1}\} \tag{4}$$

Speech noise reduction in the wavelet domain, applying the three operators, in each case, multiplies the respective coefficients of the noisy signal by the corresponding coefficients of $F[n]$, according to the following equation.

$$Y_s[n] = \{y_0, y_1, y_2, \ldots, y_{N-1}\} \tag{5}$$

where $y_i = s_i.f_i$. The elements of $Y_S[n]$ are the coefficients resulting from the characterization of the noise reduction operation. After that, the inverse wavelet transform is applied to the coefficients $Y_S[n]$ to obtain the reconstructed signal $\hat{y}_S(n)$ in time domain. As a result, finally, from the above operations, a polynomial combination is made (equation (9)), using the operators mentioned above. After that, a sigmoidal adjustment is performed. In sections 2.1, 2.2 and 2.3, the operators executions are described, and in section 2.4, the polynomial combination and the transfer function *fil*, which makes the sigmoidal adjustment, are explained.

2.1. First Processing - SAO Operator

Let $S[n] = \{s_0, \ldots, s_{N-1}\}$ be a noisy signal, where N is a power n of base 2 ($N = 2^n$), in the wavelet domain. Noise reduction is performed using the simple average operator (SAO), whose execution is defined by the following equation,

$$f_i = \frac{|s_i| + |s_{i+1}|}{2} \tag{6}$$

for $i = 0, \ldots, N-2$ and $f_{N-1} = f_{N-2}$. For the reduction accomplishment, an associated signal $Y_S[n]$ is generated from $S[n]$, also in the wavelet domain, using the set $F[n] = \{f_0, \ldots, f_{N-1}\}$ (see equation (4)). This noise reduction is made through the simple multiplication of the respective elements s_i of $S[n]$ by the correspondents elements f_i of $F[n]$ (see equation (5)). After the noise reduction, the signal is reconstructed from $Y_S[n]$, using the inverse wavelet transform, obtaining $\hat{y}_S(n)$ in time domain. The operator SAO is efficient in noise reduction, because, although it reduces strongly the amplitude of the signal, the coefficients of the speech interval are enhanced. Thus, when multiplying s_i by f_i, the signal waveform is preserved, however, there is a strong amplitude reduction, and, of course, also a strong noise reduction justified by the drastic amplitude reduction. This strong reduction in the amplitude indicates the introduction of distortions in voice segments, consequently, the voice coefficients are also eliminated. However, when it is compared to the original signal, it is possible to verify that its waveform is preserved, which implies that significant coefficients in the wavelet domain are preserved, although with reduced magnitude, they are not eliminated, but attenuated. These coefficients keep the characteristics of the signal.

2.2. Second Processing - SNRP Operator

The processing of the signal to noise ratio a priori operator (SNRP) is similar to SAO. Considering a noisy signal $S[n] = \{s_0, \ldots, s_{N-1}\}$, this operator is defined as follows

$$f_i = \frac{|s_i|}{\alpha + |s_i|} \tag{7}$$

for $i = 0, \ldots, N-2$ and $f_{N-1} = f_{N-2}$, where α is a value between 0 and 1. To accomplish the reduction, the adopted procedure is the same as in the previous section, the reduction is performed by the multiplications of s_i by f_i . This operator also reduces the amplitude of the signal, being efficient in the noise reduction. Although a reduction in the amplitude exists, it is not so strong as in SAO. The SNRP operator, computed by equation (7), estimates the noise present in each speech coefficient, this is a characteristic of the methods based on the SNR a priori [2]. Therefore, when executing the multiplication of s_i by the corresponding f_i, there is distortion introduced in the voice segments. The SNRP operator also maintains the waveform of the signal, meaning that when applied to a signal in the wavelet domain the significant coefficients are kept or only attenuated, however this reduction is much smaller than in the previous operator, so that the distortions are also smaller.

2.3. Third Processing - SNRPR Operator

The signal to noise ratio a posteriori operator (SNRPR) is applied to a noise signal $S[n] = \{s_0, \ldots, s_{N-1}\}$, being $(N = 2^n)$, in the wavelet domain, using the same procedures of SAO and SNRP, and its execution is defined by

$$f_i = \frac{|s_i|}{\alpha + |sr_i|} \tag{8}$$

where $i = 0, \ldots, N-2$ and $f_{N-1} = f_{N-2}$, α is a value between 0 and 1, sr_i is a vector of the same length as s_i and each component represents the average noise in the correspondent components of the windows in the last silence interval. Noise reduction is performed as in SAO and SNRP. It is verified, although there is a strong noise reduction, that there is also a great distortion in voice segments. The SNRPR operator is computed practically as SNRP, however using the noise average of the last silence interval. This is justified principally when stationary noise is considered, because that kind of noise can cause abrupt changes in speech.

2.4. Processing via Polynomial Combination

The Polynomial Combination combines the three operators presented previously, in order to reduce noise and avoid distortion. The processing is the same for voice and silence segments. Two of the presented processings reduce noise significantly, naturally introducing strong distortion in voice intervals, because the noise reduction is also strong in those intervals. Thus, by minimizing the noise reduction, distortion will be minimized too, considering that noise spreads uniformly over the signal. As an alternative for avoiding this voice distortion a polynomial combination of the three operators presented before is performed. Considering a signal of $N = 2^n$ coefficients, in the wavelet domain, and the previous operators presented again, respectively, in equations (6), (7) and (8):

$$y_j = \tfrac{|s_j|+|s_{j+1}|}{2}$$

$$z_j = \tfrac{|s_j|}{\alpha+|s_j|}$$

$$w_j = \tfrac{|s_j|}{\alpha+|sr_j|}$$

the *polynomial* combination proposed is the function $F_{cp}[n] = \{f_1, \ldots, f_{N-1}\}$, whose elements f_j are obtained according as

$$f_j = y_j^3 + z_j^2 + w_j \tag{9}$$

This polynomial combination acts in the wavelet domain in the same way as the three operators previously considered. This processing application enlarges significantly the amplitude of the signal and, consequently, introduces strong distortions in voice segments. This problem is caused because wavelet coefficients, after being processed by equation (9), have the magnitude out of the interval $[-1, 1]$, used for MATLAB wav format. In order to avoid the amplitude enlargement, a sigmoidal amplitude adjustment is performed. The signal that is being considered is processed using the function defined in equation (10). The sigmoidal function has image in the interval $[-1, 1]$, thus, this sigmoidal adjustment keeps the amplitude of the signal in the original interval,

$$fil_j = \left| \frac{1 - e^{-\alpha f_j}}{1 + e^{-\alpha f_j}} \cdot \frac{1 - e^{\alpha f_j}}{1 + e^{\alpha f_j}} \right| \tag{10}$$

f_j is defined in (9). The objective of this adjustment is the combined use of the three processings presented in sections 2.1, 2.2 and 2.3, maintaining the amplitude of the original signal. Thus, the proposed method that uses a transfer function *fil* during the signal processing, acts simultaneously with the equation (9). The function *fil* is generated by the multiplication of two inverse sigmoidal related to the horizontal axis. This transfer function, *fil*, acts as a filter in speech processing. But, a filter separates the signal in high and low frequencies bands, through convolution, and it can occur in several domains. Here, the transfer function, *fil*, acts in the wavelet domain as the previous processes, only multiplying coefficients independently of the frequency band. This method is efficient in the noise reduction and makes a uniform reduction in the whole signal. The amplitude of the signal neither increases nor decreases. And, although there are differences among the two signals in the speech intervals, those differences are minimum and do not represent distortion for the processed signal [10].

3. Results, analysis and discussion

In this section, implementations results of the proposed method and evaluations of the processed signals are presented. In order to verify the quality of the processed signals, noise reduction and distortion levels are evaluated using signal to noise ratio (SNR) and PESQ (Perceptual Evaluation of Speech Quality), proposed by ITU (International Telecommunications Union) through recommendation P.862 in February, 2001 [1, 6, 8]. SNR is calculated according to equation (11), its computation is performed by the rating between a voice segment and a silence segment in each signal, according to [4]:

$$\mathrm{SNR} = 10\log_{10}\left[\frac{\sum_{i=0}^{N-1} x_i^2}{\sum_{i=0}^{N-1} r_i^2}\right] \tag{11}$$

where N is the number of samples in the chosen segment, $x(n) = \{x_i | i = 0, \ldots, N-1\}$ the samples in a voice interval and $r(n) = \{r_i | i = 0, \ldots, N-1\}$ the samples in a silence interval. SNR values presented in Tables 1, 2 and 3 are averages of the SNR values obtained by the application of equation (11) in three segments of the speech signal. In the SNR of the processed signal should be observed the following; if after the processing, it is very low comparing to the SNR of the clean signal, it means that there was little noise reduction. If, after processing, it goes much larger than SNR of the clean signal, it means that there was great noise reduction and, possibly, the signal should be enough distorted. Therefore, after processing, signals that have SNR close to the SNR of the original signal are more attractive, because they indicate that the noise reduction did not distort voice segments [11].

3.1. Signals and Results

In order to verify the efficiency of the method two speech signals were used, A and B, recorded by a male and a female, respectively. These signals are recommended by IEEE, obtained at the site `http://www.utdallas.edu/~loizou/`

`speech/noizeus`. The sentences are:
Signal A - WE FIND JOY IN THE SIMPLEST THINGS.
Signal B - SHE HAS A SMART WAY OF WEARING CLOTHES.
The signals A and B were sampled at 8 kHz, with a sampling rate of 16 bits. During the tests, files were used in it wav format, normalized in the interval [-1,1]. These speech segments were obtained using a 256-sample Hanning window with overlap of 50 % among the segments. Using a 256-samples window, in the wavelet domain, it is possible to decompose the signal up to 8 resolution levels. For computational implementation MATLAB was used and the wavelet function used in DWT was the Daubechies 10 (Daub10). Signals A and B were corrupted by two noise types: white Gaussian and colored (car noise). Two SNR levels were tested for each noise type, 5 dB and 10 dB. In order to verify the level of noise reduction in the processed signals, SNR was calculated for the clean signal, before contamination, for the noisy signal, of course it was 5 dB or 10 dB, and, finally, for the processed signal. Table 1 shows SNRs, in dB, for clean signals.

As an illustration, Figure 1 displays the waveforms of signal A clean and

	SNR
Signal A	45.922
Signal B	41.754

Table 1: SNR for the clean signals

processed by the proposed method, corrupted by white noise (10 dB), where n is the number of samples and AN is the numeric amplitude signal.

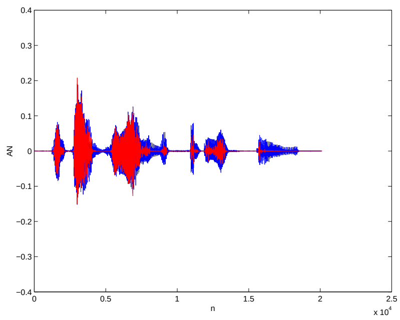

Figure 1: **Signal A (10 dB): Clean signal (blue) and processed signal by the proposed method (red)(white noise)**

Tables 2 and 3 show SNRs, in dB, for the processed signals corrupted as mentioned previously, using the proposed method, with SNRs of 5 dB and 10 dB, respectively. SNR is an objective measure used to measure noise in the signal.

	White noise	Colored noise
Signal A	19.851	10.880
Signal B	15.765	23.209

Table 2: SNR for the processed signals (5 dB)

	White noise	Colored noise
Signal A	46.635	33.528
Signal B	39.382	34.151

Table 3: SNR for the processed signals (10 dB)

SNR does not measure the intelligibility of a signal, but indicates its noise reduction level after processing. Thus, having three signals; reference (clean), noisy and processed, an effective noise reduction, should present close SNR values for clean and processed signals. When clean and processed signals have very different values of SNR, it means that: either there was a strong noise reduction and, consequently, distortion introduction or there was weak noise reduction and the intelligibility of the signal is still committed. The results presented in Table 3, mainly for white noise, show that the two signals had a significant level of noise reduction, meaning noise reduction without distorting the signal, because their SNR are close to the SNR of the clean signals. For colored noise, the variations of SNR of the processed signals were around 74 % to 82 % comparing to the SNR of the clean signals, respectively. The results already presented in Table 2 show that the two processed signals present SNR below 50 %, except for the signal B corrupted by colored noise which has SNR close to 50 %, but above, when compared to the clean signals. However, the SNR values obtained by the proposed method indicate noise reduction without introducing strong distortions in the processed signals. Table 5 presents PESQ results for the processed signals corrupted by white and colored noises with 5 dB and 10 dB SNRs, respectively. PESQ results for noisy signals, before processing, are presented in Table 4. The largest score that PESQ attributes to a signal is 4.5 and signals with PESQ scores starting from 3 are considered as good quality signals. PESQ scores for the processed signals A and B with 10 dB SNR, in Table 5, are greater than 3, except for colored noise corrupted signal B, but it is very close to 3, indicating that it is good for audition. PESQ scores for the processed signals A and B when SNR is 5 dB are all close to 3. Although below 3, when compared to the noisy signal, it can be verified that there was a significant improvement in their PESQ scores, reducing the noise without less distortion.

Observing Figure 1, it can be seen that the method produces a uniform noise reduction along the signal and, comparing processed and clean signals, there is no distortion introduction in the voice segments.

	White noise		Colored noise	
	5 dB	10 dB	5 dB	10 dB
Signal A	1,805	2,273	1,950	2,220
Signal B	1,617	2,021	1,848	2,310

Table 4: PESQ for noisy signals

	White noise		Colored noise	
	5 dB	10 dB	5 dB	10 dB
Signal A	2.842	3.315	2.715	3.054
Signal B	2.448	3.065	2.708	2.992

Table 5: PESQ for processed singnals

4. Conclusions

Observing the results showed in the last section, it is possible to conclude that the proposed method is efficient, because it performs noise reduction without introducing distortions and without changing the amplitude of the processed signal. The main objective of this method is the overcoming of the thresholding methods drawbacks and the effective processing of real noise corrupted signals. The main advantage of this method over the other noise reduction methods in the wavelet domain is the use of a transfer function which does not use threshold to reduce noise. This transfer function is not discontinuous and it does not introduce distortions in voice segments of the processed signal.

References

[1] Beerends, J. G., Hekstra, A. P., Rix, A. W., Hollier, M. P., Perceptual evaluation of speech quality (pesq) the new itu standard for end-to-end speech quality assessment part ii: psychoacoustic model. AES Journal, New York, Vol. 50(10) (2002), 765-778.

[2] Cohen, I., Speech Enhancement Using a Noncausal a Priori SNR estimator. IEEE Signal Processing Letters, Vol. 11(9) (2004), 725-728.

[3] Dae-Sung, K., Jung-Go, C., Geun-Taek, R., Mun-Seob, B., Hyeon-Deok, B., Noise Reduction Using Coefficients Smoothing Wavelet Transform. International Conference on Signal Processing Applications and Technology, ICSPAT-2002, San Jose California, September 2002, pp. 29-33.

[4] Deller, J. L., Proakis, J. G., Hansen, J. H., Discrete-Time Processing of Speech Signals. New York: Macmillam, 1993.

[5] Donoho, D. L., De-Noising by Soft-thresholding. IEEE Transactions on Information Theory, Vol. 41(3) (1995), 613-627.

[6] International Telecommunications Union, Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. Geneva: ITU, (ITU-T P.862), 2001.

[7] Rioul, O., Vetterli, M., Wavelets and Signal Processing, IEEE Signal Processing Magazine, Vol. 8(4) (1991), 14-38.

[8] Rix, A. W., Beerends, J. G., Hollier, M. P., Hekstra, A. P., PESQ - the new ITU standard for end-to-end speech quality assessment. In: AES CONVENTION, 109, 2000, Los Angeles. Aes Convention, Los Angeles: S.n., 18p., 2000.

[9] Sheikhzadeh, H., Abutalebi, H. R., An Improved Wavelet-Based Speech Enhancement System. Eurospeech 2001, September 2001, pp. 1855-1858.

[10] Soares, W. C., A non thresholding speech noise reduction method in the wavelet domain. Doctoral Thesis, FEIS-UNESP,Ilha Solteira-SP, 205p., 2009.

[11] Soares, W. C., Duarte, M. A. Q., Villarreal, F., Vieira Filho, J., Non thresholding noise reduction in the wavelet domain. Proceedings of the Brazilian Conference on Dynamics, Control and Their Applications, DINCON'2008, Presidente Prudente: SBMAC, 2008, pp. 961-967.

[12] Vieira Filho, J., Speech noise reduction in vehicle mobile radio systems. Doctoral Thesis, FEEC-DC-UNICAMP, Campinas, 1996.

Received by the editors February 5, 2010