Dual-Tree Complex Wavelet Transform in the problem of speech enhancement

Caio C. E. de Abreu
Department of Electrical Engineering, FEIS - UNESP
15385-000, Ilha Solteira, SP
E-mail: caioenside@aluno.feis.unesp.br

Marco A. Q. Duarte
Department of Mathematics - UEMS
79540-000, Cassilândia, MS
E-mail: marco@uems.br

Francisco Villarreal
Department of Mathematics, FEIS – UNESP
15385-000, Ilha Solteira, SP
E-mail: villa@mat.feis.unesp.br

Abstract: The dual-tree complex wavelet transform is considered a relatively recent improvement for the discrete wavelet transform. In this paper, the applicability of such transform in the problem of speech enhancement is evaluated. For this purpose, a nonthreshold scheme is proposed. Two filters, one for the imaginary part and other for the real part of complex wavelet coefficients were designed. These two real filters were then averaged to obtain the final filter. A uniform noise reduction was performed for all wavelet scales. Simulation results show that, together, nonthresholding scheme and dual-tree complex wavelet transform obtained consistently results.

keywords: Speech enhancement, dual-tree complex wavelet transform, complex wavelet denoising.

1   Introduction

Important applications of speech enhancement are communication in noisy environments and automatic speech recognition. The problem of speech enhancement aims at denoising a speech signal corrupted by noise and to improve its intelligibility. In this sense, a powerful tool for denoising is the discrete wavelet transform (DWT), which provides a time-frequency analysis for the processing signal [2].

For the computation of the DWT it is required a digital filter bank in a tree structure [7]. The coefficients of the low-pass and high-pass filters, used in the decomposition process, are generated from wavelet filters chosen a priori. For each wavelet decomposition level, approximation and detail coefficients are acquired. In the transition from one wavelet level to another, the decomposition is applied again, however, only on the approximation coefficients. The DWT provides an efficient representation for many types of signals that appears often in practice, mainly for signals containing singularities (jumps and spikes), but it has some shortcomings [9]. The main one, for 1-D signals, it is shift variance, i.e., a small shift of the signal greatly perturbs the wavelet coefficients oscillation pattern around singularities (see Figure 1). In this sense, algorithms must be projected to be able of coping with the wide range of possible wavelet coefficient patterns caused by shifted singularities [9]. Fortunately, this problem is overcome by the complex wavelet transform. The dual-tree complex wavelet

transform (dual-tree $\mathbb{C}$WT) proposed by Kingsbury et al. [9],[5] is near shift-invariance and also provides other desirable properties at small additional computational costs [13]. This paper shows that the dual-tree $\mathbb{C}$WT is very promising for speech enhancement and provides good results when combined with a nonthreshold scheme.
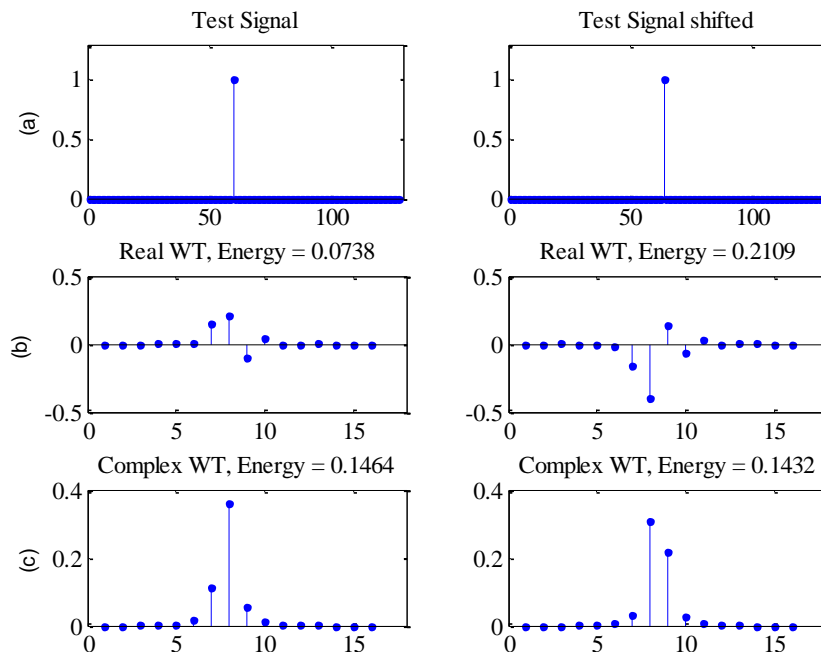

Figure 1. For two impulse signals $x(n) = \delta(n-60)$ e $x(n) = \delta(n-65)$ (a), follows the wavelet coefficients at a fixed scale j (b) and (c). (b) real wavelet coefficients computed using a conventional DWT, with Daubechies length 14 filters. (c) magnitude of the complex coefficients computed using a dual-tree $\mathbb{C}$WT , with length-14 filters from [6].

The earliest methods of speech enhancement in the wavelet domain were based on thresholding. Originally proposed by [3], wavelet thresholding considered as noisy coefficients those ones with absolute value under a certain value and then those coefficients were eliminated or modified. The thresholding schemes were applied to speech enhancement in many papers, for instance in [8]. In spite of the presence good results, these methods have some shortcomings and were overcome by nonthreshold methods, that act on the corrupted signal according to its own content [10]. Nonthreshold methods perform a uniform noise reduction throughout the signal, avoiding inconvenient sounds generated by searching the best threshold, being more enjoyable to the listener [11].

The remainder of the paper is organized as follows: In section 2, the Dual-Tree complex wavelet transform is presented. A nonthreshold scheme is presented and the performance of the dual-tree $\mathbb{C}$WT in the problem of speech enhancement is evaluated in section 3. In section 4, objective analyzes are presented. Finally, concluding remarks are given in section 5.

2   Dual-Tree complex wavelet transform

The dual-tree $\mathbb{C}$WT was introduced by Kingsbury [4] in 1998 as an enhancement to the DWT: it is nearly shift-invariant and directionally selective in two and higher dimensions [9].

Complex wavelet design takes into account some desirable characteristics of the Fourier transform [9]. Note that the Fourier transform is based on complex-valued oscillating sinusoids as in (1).

$$e^{j\Omega t} = \cos(\Omega t) + j \ \text{sen}(\Omega t), \qquad j = \sqrt{-1} \ . \tag{1}$$

The real and imaginary parts are 90° out of phase with each other, thus forming a Hilbert transform pair. According to [9], together they constitute an analytic signal. Furthermore, the magnitude of the Fourier transform is perfectly shift-invariant.

Inspired by the Fourier transform, the dual-tree $\mathbb{C}WT$ is an effective approach for implementing an analytic wavelet transform. The complex coefficients generated by the dual-tree $\mathbb{C}WT$ can be represented by:

$$d_c(j,n) = d_r(j,n) + j \ d_i(j,n). \tag{2}$$

The real part $d_r(j,n)$ and the imaginary part $d_i(j,n)$ are calculed individually by two filters banks, see Figure 2. In other words, the dual-tree $\mathbb{C}WT$ employs two discrete wavelet transforms, where the upper tree produces the real part and the lower tree produces the imaginary part of the complex coefficients. Both sets of filters are designed so that the overall transform is approximately analytic and near shift invariant [9].
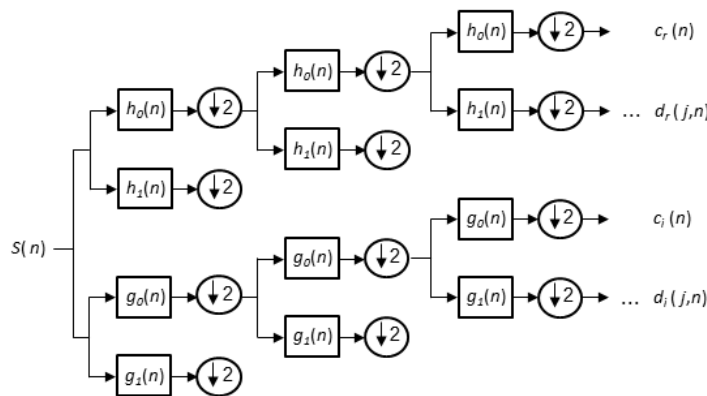

Figure 2. Filter bank analysis for the dual-tree $\mathbb{C}WT$ [9]. $h_0(n)$ and $g_0(n)$ represent low-pass filters whereas $h_1(n)$ and $g_1(n)$ represent high-pass filters. $c_r(n)$ and $c_i(n)$ are approximation coefficients. Similarly, $d_r(j,n)$ and $d_i(j,n)$ are details coefficients at a fixed scale j. $S(n)$ is the signal to be decomposed.

Approximation coefficients are obtained from real low-pass subband at the final level of the forward transform. In the case of details coefficients, they are obtained from complex high-pass subband at each level of the forward transform. The frequency content analysis is controlled by the scale factor j. All results in this paper were obtained by using near-symmetric (13,19)-tap filters at first level and q-shift (14,14)-tap filters at level $j > 1$ [9].

In order to invert the dual-tree $\mathbb{C}WT$, the inverse of each of the two real DWTs are used. After obtaining these two signals, they are averaged to obtain the inverse transform [9].

3   Dual-Tree $\mathbb{C}WT$ in the problem of speech enhancement

In order to evaluate the performance of the dual-tree $\mathbb{C}WT$ in the problem of speech enhancement, a nonthreshold scheme is proposed. Consider a noisy speech signal of length $n = 2^N$ to be processed:

$$y[n] = \left\{ y_{2k}, y_{2k+1} : 0 \le k \le 2^{N-1} - 1 \right\}. $$

Then,

$$y[n] = \left\{ y_0, y_1, \dots, y_{2k}, y_{2k+1}, \dots, y_{2^N-2}, y_{2^N-1} \right\} = \left\{ y_0, y_1, \dots, y_{2k}, y_{2k+1}, \dots, y_{n-2}, y_{n-1} \right\}.$$

First, apply the dual-tree $\mathbb{C}$WT to $y[n]$. The speech signal in the wavelet domain is estimated by difference equations used in the numerical solution of differential equations [12]. Equations (3)-(6) show the centered and advanced differences of even order.

$$c_i^2 = y_{i+1} - 2y + y_{i-1}, \tag{3}$$

$$c_i^4 = y_{i+2} - 4y_{i+1} + 6y_i - 4y_{i-1} + y_{i-2}, \tag{4}$$

$$a_i^2 = y_i - 2y_{i+1} + y_{i+2}, \tag{5}$$

$$a_i^4 = y_i - 4y_{i+1} + 6y_{i+2} - 4y_{i+3} + y_{i+4}. \tag{6}$$

After applying the differences, four vectors are obtained. The next step is the application of the a priori signal/noise ratio operator (SNRP), proposed in [11], on the four vectors generated by difference equations. This operator, responsible by the noise estimation, is defined in equation (7).

$$g(v_i) = \frac{|v_i|}{\alpha + |v_i|}, \quad \text{to } i = 0, \dots, n-2 \text{ and } \alpha \in]0,1[, \tag{7}$$

where $V[n] = \{v_0, \dots, v_{n-1}\}$ is any vector of finite difference.

The output signal is estimated by a linear combination of the SNRP operator when applied to all differences. This combination generates a vector $s[n] = \{s_0, \dots, s_{n-1}\}$, whose elements s_i are obtained from (8):

$$s_i = g(c_i^2) + g(a_i^2) + g(c_i^4) + g(a_i^4). \tag{8}$$

Finally, the filter coefficients $F[n] = \left\{ f_0, f_1, \dots, f_{n-1} \right\}$ must be obtained via sigmoidal adjustment as in (9):

$$f_i = \left| \frac{1 - e^{-\sigma s_i}}{1 + e^{-\sigma s_i}} \quad \frac{1 - e^{\sigma s_i}}{1 + e^{\sigma s_i}} \right|. \tag{9}$$

Proposed by [11], the sigmoidal adjustment in (9) normalizes the filter amplitude into [0,1]. The parameter σ controls the sigmoid slope and depends only of the noise profile:

$$\sigma = \beta \, \tau, \tag{10}$$

β is the standard deviation of the signal and τ is a value based on the noise power.

This step is very important because it keeps the original amplitude of the speech signal. The denoising process is performed through the product between the filter coefficients f_i and the noise speech coefficients y_i, in the wavelet domain. The enhanced speech signal is obtained by the inverse dual-tree $\mathbb{C}$WT.

The method consists in applying the nonthreshold scheme on the approximation and details coefficients for all the scales j, $j = 1,2, \dots, 6$. It is proposed to design two filters to the real and imaginary parts in (2), respectively. These two real filters are then averaged in order to obtain the final filter. The enhanced signal is obtained by the product between the real coefficients f_i and the complex coefficients $d_c(j,n)$ for all the scales. Note that the proposed method does not alter the phase of the noisy signal.

In order to evaluate a subjective observation of the enhanced speech by the proposed method, waveforms and spectrograms of the clean speech, noisy speech, and enhanced speech are presented in Figure 3. A speech signal in the male voice was enhanced, after being corrupted by car noise, generating a signal with 5 dB SNR. It can be noted that there was an improvement of 16.26 dB in the signal. Furthermore, it can be noted by the spectrograms that

the nonthreshold scheme together with dual-tree ℂWT performs a strong noise reduction and a low loss of speech information.

Objective analyzes are the subject of the next section, where, apart from the car noise, it was also used the cafeteria noise.
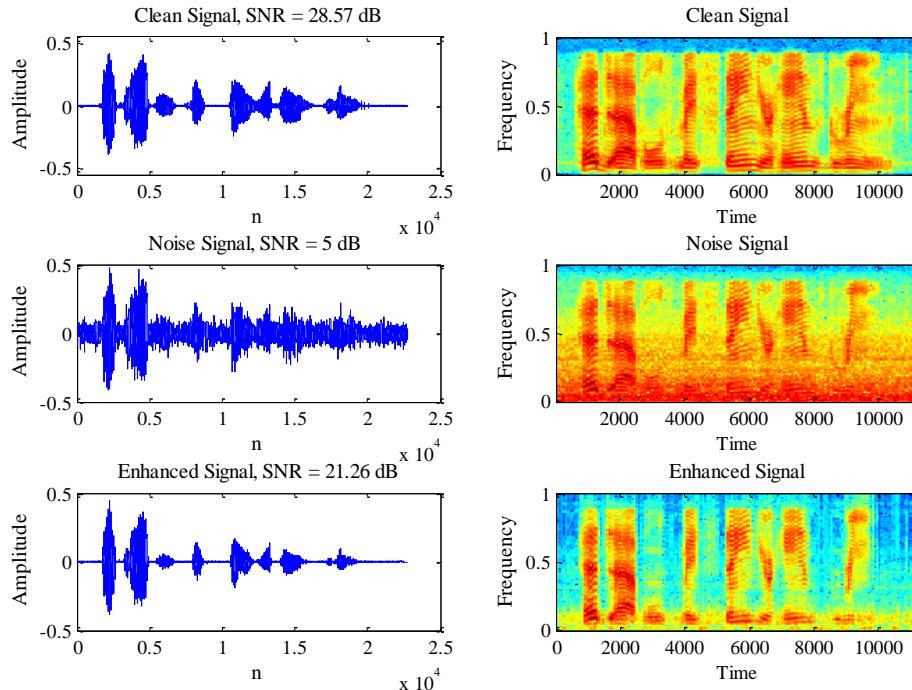

Figure 3. Waveform of clean speech, noise speech and enhanced speech. On the right, their respective spectrograms.

4   Objective analyzes

In order to check the performance of the nonthresholding method based on dual-tree ℂWT, four signals corrupted by car and cafeteria noise, divided into male and female voice, were used. Among them, there was a Portuguese male voice signal and the others were in English. All the signals used in the experiments are suggested by Test Signals for Telecommunication Systems (ITU-T).

Two quality objective measures were used, the Signal to Noise Ratio (SNR) and the Perceptual Evaluation of Speech Quality (PESQ). The assessment made by PESQ is based on psychoacoustic characteristics of the human ear. The PESQ score ranges between 1 and 5, being considered a signal of good quality the one with score equal or greater than 3 [1].

Table 1 shows the values of SNR for the clean signals and Table 2 shows the enhancement for the same signals, after being corrupted, generating signals with 5 dB and 10 dB SNR. Table 3 shows the PESQ scores to enhanced signals.

Clean signals	SNR (dB)
Male 1	32.48
Male 2	34.22
Female 1	37.20
Female 2	28.53

Table 1: SNR's of clean signals.

SNR evaluation				
	5 dB		10 dB	
	Car noise	Cafeteria noise	Car noise	Cafeteria noise
Male 1	23.76	16.90	38.56	26.62
Male 2*	23.67	13.82	38.76	26.51
Female 1	22.26	27.44	38.62	34.99
Female 2	26.31	21.66	37.39	27.77

Table 2: SNR of speech enhancement by the proposed method from four speakers
with 5 dB and 10 dB SNRs.
(*) Portuguese.

PESQ evaluation				
	5 dB		10 dB	
	Car noise	Cafeteria noise	Car noise	Cafeteria noise
Male 1	3.33	3.07	3.78	3.56
Male 2*	3.26	2.85	3.73	3.43
Female 1	2.81	2.75	3.05	2.88
Female 2	3.08	2.82	3.52	3.20

Table 3: PESQ score to speech enhanced by the proposed method from four speakers
with 5 dB and 10 dB SNRs.
(*) Portuguese.

The PESQ score measures the intelligibility of the processed signal. In this sense, analyzing the PESQ scores in the Table 3, it is observed that there was no strong distortion in the enhanced speech. Most of the enhanced signals obtained scores higher than 3. The others obtained scores very close to 3.

5   Conclusion

In this paper, a nonthresholding method based on dual-tree $\mathbb{C}$WT has been presented. The objective was to verify the use of dual-tree complex wavelet transform in the problem of speech enhancement.

The dual-tree $\mathbb{C}$WT is nearly shift-invariant and directionally selective in two and higher dimensions, being considered a relatively recent improvement for the DWT.

Nonthresholding scheme performs a uniform noise reduction throughout the signal, avoiding inconvenient sounds, generated by the choice of the best threshold in thesholding methods.

Simulation results show that together, nonthresholding scheme and dual-tree $\mathbb{C}$WT, obtained consistent results. The evaluation of SNR shows that the proposed method performs, in average, an improvement of 23.66 dB for the car noise and 16.96 dB for the cafeteria noise. In the case of PESQ evaluation, most enhanced signals had obtained scores higher or very close to 3. This shows that there were not serious distortions in the enhanced speech.

For future works, it is intended to develop a speech enhancement method based on the complex magnitude of the dual-tree $\mathbb{C}$WT. Furthermore, the possibility of a phase change will be studied. As Figure 3 shows, the advantages of the nearly shift invariant property are fully achieved by magnitude of the complex wavelets coefficients.

References

[1] J.G. Beerends; A. P. Hekstra; A. W. Rix; M. P. Hollier. 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, n.10, 765-778, (2002).

[2] I. Daubechies, "Ten Lectures on Wavelets", SIAM, (1992).

[3] D. L. Donoho, De-noising by soft-thresholding, *IEEE Transactions on Information Theory*, vol. 41, n. 3, 613-627, (1995).

[4] N.G. Kingsbury, The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters, in "*Proceedings 8th IEEE DSP Workshop*", Utah, paper no. 86, Aug. 9–12, (1998).

[5] N.G. Kingsbury, Complex wavelets for shift invariant analysis and filtering of signals, *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, 234–253, (May 2001).

[6] N.G. Kingsbury, Design of q-shift complex wavelets for image processing using frequency domain energy minimization, in Proc. "*Proceedings of the IEEE International Conference on Image Processing*", Barcelona, vol. 1, 1013–1016, (Sept. 2003).

[7] S. Mallat, A theory for multiresolution representation signal decomposition: the wavelet representation, *IEEE Transaction Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, 674-693, (Jul. 1989).

[8] T. F. Sanam; C. Shahnaz, Enhancement of noisy speech based on a custom thresholding function with a statistically determined threshold, *International Journal of Speech Technology*, vol. 15, Issue 4 , 463-475, (Dec. 2012).

[9] I. W Selesnick, R.G Baraniuk, and N. G Kingsbury. The dual-tree complex wavelet transform—A coherent framework for multiscale signal and image processing, *IEEE Signal Process. Mag.*, vol. 22, no. 6, 123–151, (Nov. 2005).

[10] W. C. Soares, M. A. Q. Duarte, F. Villarreal, J. Vieira Filho, Análise de métodos de redução de ruído por limiar no domínio wavelet, *TEMA: Tendências em Matemática Aplicada e Computacional*, São Carlos, v.9, n.3, 471-480, (2008).

[11] W. C. Soares, F. Villarreal, M. A. Q. Duarte, J. Vieira Filho, Wavelets in a Problem of Signal Processing, *Novi Sad Journal of Mathematics*, v.41, no.1, 11-21, (2011).

[12] D. Sperandio; J. T. Mendes; L. H. M. Silva, "Cálculo Numérico – Características Matemáticas e Computacionais dos Métodos Numéricos", Prentice Hall, São Paulo, (2003).

[13] T. Weickert; C. Benjaminsen; U. Kiencke, Analytic complex wavelet packets for speech enhancement, *IEE International Conference on Acoustic, Speech, and Signal Processing,* (ICASSP 2008), vol. 1, 3269-3272, (2008).