ANALYTIC COMPLEX WAVELET PACKETS FOR SPEECH ENHANCEMENT

Thomas Weickert, Claus Benjaminsen and Uwe Kiencke

Institute of Industrial Information Technology
Universität Karlsruhe (TH)
76187 Karlsruhe, Germany

ABSTRACT

In previous work, the authors found the lack of shift invariance of real wavelet packets very disadvantageous for speech enhancement in the case of periodic noise. Therefore, this paper investigates the positive properties of the Dual-Tree Complex Wavelet Transform (DTCWT). This transform is nearly shift invariant at moderate additional computational cost. However, the straightforward approach of extending the DTCWT to wavelet packets by decomposing the high pass coefficients as well led to non-analytic basis functions. Because analytic basis functions were required for the desired properties, a filter swapping scheme was developed to preserve analyticity. This analytic Complex Wavelet Packet Transform showed improved denoising performance for the application of speech enhancement and promises improvements for other applications like general filtering and signal analysis.

Index Terms— wavelet transforms, wavelet packets, complex wavelets, nonstationary noise, speech enhancement

1. INTRODUCTION

Speech enhancement aims at denoising a speech signal corrupted by noise and to improve the intelligibility. Important applications are communication in noisy environments or preprocessing for speech recognition systems. A powerful tool for denoising are wavelet packets (WP) which are a redundant time-frequency representation. Simple thresholding schemes are very effective for stationary white and colored noise and were applied to speech enhancement in many papers, for instance in [1].

However, most papers do only regard stationary noise. Hence, the probably biggest advantage of time-frequency representations is not leveraged: the analysis of nonstationary signals. In [2], the authors consider periodic noise, whose power density changes periodically over time, as a special class of nonstationary noise. This class is important to speech enhancement because of its common appearance in the presence of engines and machines.

During their work, the authors found that even small variations of the noise period lead to a relatively large decrease in filtering performance. This is due to the shift variance of the wavelet packet transform caused by the downsampling operations. The dual-tree complex wavelet transform (DTCWT) from Kingsbury et al. [3] [4] promises near shift-invariance and other desirable properties at small additional computational costs. This paper shows that a straightforward extension of the DTCWT to wavelet packets is not analytic and presents a filter swapping scheme in order to solve this issue.

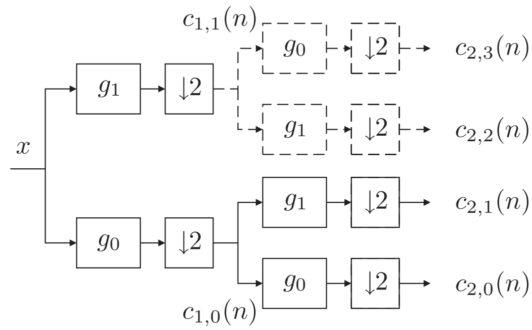
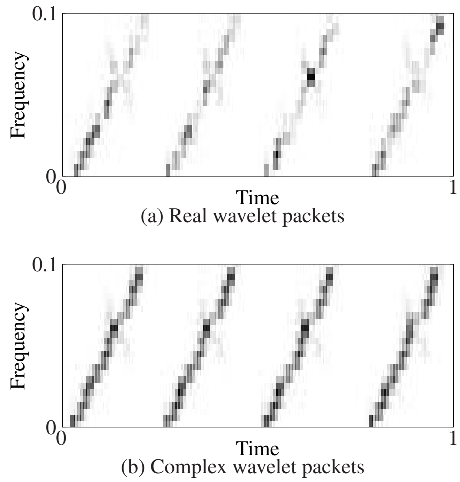


Fig. 1: Wavelet Packet filter bank of max. depth $D = 2$, g_0: low pass, g_1: high pass, solid line: wavelet transform (subset of a wavelet packet), dashed: additional decomposition for wavelet packets

This work is organized as follows: Section 2 gives a review of wavelet packets and wavelet packet denoising. The basics of the DTCWT are explained in section 3. Afterwards, section 4 describes the extension of the DTCWT to complex wavelet packets. The improvement of speech enhancement by complex wavelet packets is demonstrated in section 5. Finally, the conclusions are drawn in section 6.

2. WAVELET PACKET THRESHOLDING

The discrete wavelet transform can be computed efficiently by a Conjugate Quadrature Filter bank (CQF), see figure 1 solid lines only [5]. A discrete-time signal x is split up by high pass and low pass filters and downsampled. This is done recursively for the low pass coefficients. The extension to wavelet packets is straightforward. The high pass coefficients are split up as well. The coefficients can be organized in a



(a) Real wavelet packets

(b) Complex wavelet packets

Fig. 2: Chirps in time-frequency planes

binary tree and are addressed by

$$c_{d,b}(n) = WP(d,b,n)$$
$$\text{with} \quad 0 \le d \le D \quad, 0 \le b \le 2^d - 1 \quad, 0 \le n \le \frac{N}{2^d} - 1$$

where d denotes the depth in the tree, b is the node number in this depth, n is the coefficient index in the specific node, D is the maximum depth and N is the signal length. A wavelet packet is a redundant decomposition. A Best Basis can be chosen to represent the signal in few but large coefficients.

The established denoising technique in the wavelet packet domain is called “wavelet packet thresholding”. The WP coefficients of the noisy signal are thresholded by a nonlinear thresholding function. In the case of stationary white noise, the threshold can be constant for all coefficients because the energy of white noise is distributed equally across the coefficients of an orthonormal basis. For stationary colored noise, a frequency-dependent threshold can be applied. That is, each node (d, b) has a certain threshold which is constant for all coefficients of this node.

If the noise is nonstationary, the threshold shouldn’t be constant over time to get a good filtering result. In [2], the authors consider periodic noise as a special case of nonstationary noise which is very common in the presence of machines and engines. The time-frequency energy distribution of a noise period is adaptively estimated and used as thresholds of the denoising scheme.

However, the authors found that the shift variance of real wavelets due to the downsampling operations decreases the filtering performance if the period is not perfectly constant. Figure 2 shows time-frequency planes of chirp noise with period variations of 5 %. Real wavelet packets yield the coefficient energy distribution in figure 2a. Two undesired effects are revealed here. Firstly, the energy is not distributed equally over all frequencies of the chirp. Secondly, the energy distribution over frequency changes for every period. For instance, the first period has a maximum at low frequencies while the last two periods have their maxima at medium or high frequencies. This negative effect due to the real wavelet shift variance motivates the usage of a more shift robust representation. The following section will therefore summarize the theory of the dual-tree complex wavelet transform, which is nearly shift invariant.

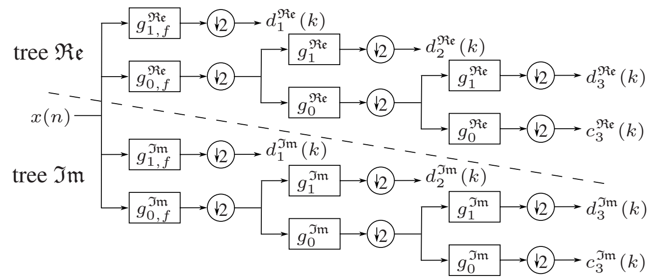


Fig. 3: Dual-Tree Complex Wavelet Transform

3. THE DUAL-TREE COMPLEX WAVELET TRANSFORM

Why are complex basis functions more shift invariant than real basis functions? Shift invariance is a well known property of the Fourier transform where a shift of the input signal only results in a phase change. The amplitude will stay unchanged. This is due to the analytic basis functions which lead to complex coefficients. Another advantage of analytic basis functions is, that their amplitude and therefore the amplitude of their coefficients is smoother than the coefficient amplitudes of real basis functions which oscillate naturally.

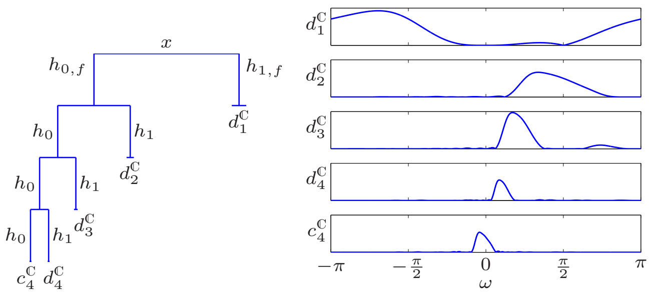
Research in applying analytic basis functions to the Discrete Wavelet Transform has led to the Dual-Tree Complex Wavelet Transform [3]. The real part of the wavelet coefficients $d_l^{\mathfrak{Re}}(k)$ and the imaginary part $d_l^{\mathfrak{Im}}(k)$ are calculated separately by two filter banks, see figure 3.

The complex coefficients can be calculated by

$$d_l^{\mathbb{C}}(k) = d_l^{\mathfrak{Re}}(k) + j\, d_l^{\mathfrak{Im}}(k) \quad .$$

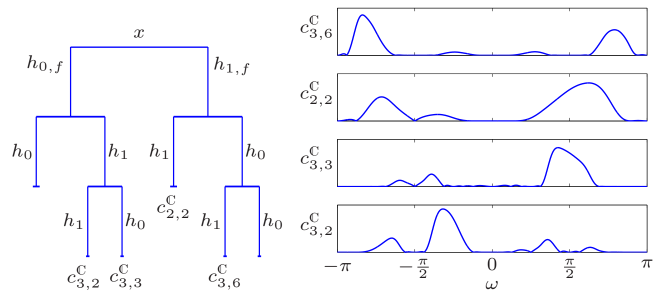
Since the complex coefficients have to be analytic, the basis functions of tree $\mathfrak{Im}$ have to be the Hilbert transforms of the basis functions of tree $\mathfrak{Re}$. This can be translated into design criteria for the synthesis low pass filters $h_0^{\mathfrak{Im}}$ and $h_1^{\mathfrak{Re}}$. The so called half-sample delay condition is given by

$$h_0^{\mathfrak{Im}}(n) = h_0^{\mathfrak{Re}}\left(n - \frac{1}{2}\right)$$
$$\Rightarrow \begin{cases} \left|H_0^{\mathfrak{Im}}\left(\mathrm{e}^{j\omega}\right)\right| = \left|H_0^{\mathfrak{Re}}\left(\mathrm{e}^{j\,\omega}\right)\right| \\ \Delta\varphi_{h_0} = \angle H_0^{\mathfrak{Im}}\left(\mathrm{e}^{j\omega}\right) - \angle H_0^{\mathfrak{Re}}\left(\mathrm{e}^{j\omega}\right) = -\frac{1}{2}\omega \end{cases} \tag{1}$$



(a) DTCWT Tree (b) Amplitude spectra of basis functions

Fig. 4: Analytic basis functions of the DTCWT



(a) Example Tree (b) Amplitude spectra of basis functions

Fig. 5: Non-analytic basis functions of straightforward approach

It is valid for all stages with $l > 1$. For the first stage however, the condition is

$$h_{0,f}^{\mathfrak{Im}}(n) = h_{0,f}^{\mathfrak{Re}}(n-1) \quad .$$

Therefore, any existing perfect reconstruction filters can be used for the first stage if they are shifted by one sample.

For later stages, the half-sample delay condition (1) cannot be fulfilled perfectly by FIR filters. Kingsbury developed q-shift filters [6] which approximate condition (1). These filters are also used for the CWP later. The basis functions corresponding to these filters are very close to being analytic (except for the lowest and highest frequency band), as figure 4 shows. The discrete basis functions were calculated by setting all coefficients to zero, except a single coefficient set to one, and performing the inverse transform with synthesis filters $h_0^{\mathfrak{Re}}$, $h_1^{\mathfrak{Re}}$, $h_0^{\mathfrak{Im}}$ and $h_1^{\mathfrak{Im}}$. The following section will show how this already powerful approach can be extended to analytic complex wavelet packets.

4. COMPLEX WAVELET PACKETS

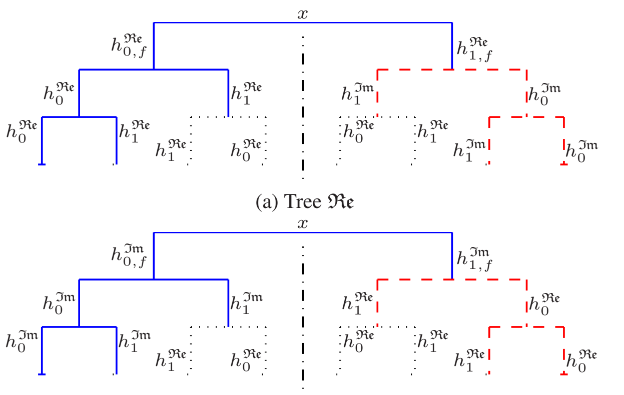
One would expect the extension from the DTCWT to CWP to be straightforward: Instead of only decomposing the low pass coefficients, the high pass coefficients should be decomposed as well in both trees. However, if this is done and the basis functions are calculated, figure 5 reveals that most of the basis functions are far from being analytic.

Examining how the basis functions of $d_4^{\mathbb{C}}$ in figure 4 become analytic by the inverse transform helps to understand this problem. According to the half-sample delay condition (1), the synthesis low pass filters give a phase difference of $\Delta\varphi_{h_0}(\omega) = -\frac{1}{2}\omega$. For the high pass filters follows a phase difference of $\Delta\varphi_{h_1}(\omega) = \frac{1}{2}(\omega - \pi)$ because the high pass filters can be calculated from the time-reversed and shifted low pass filters. Therefore, the first synthesis filter stage of $d_4^{\mathbb{C}}$, a high pass filter pair $h_1^{\mathfrak{Re}}$ / $h_1^{\mathfrak{Im}}$, gives the coefficients a phase difference with a slope of $\frac{1}{2}$. The further processing of the coefficients are upsampling operations, which double the slope of the phase difference to 1, and synthesis low pass filter pairs $h_0^{\mathfrak{Re}}$ / $h_0^{\mathfrak{Im}}$, which add a phase difference of $-\frac{1}{2}\omega$ thus compensating the positive slope introduced by the high pass filter partially. In the end after the top filter stage $h_{0,f}^{\mathfrak{Re}}$ / $h_{0,f}^{\mathfrak{Im}}$, the phase difference will be flat except a jump from $\pi/2$ to $-\pi/2$ at $\omega = 0$. This is exactly the desired phase difference for Hilbert transformed basis functions. This also explains why the basis functions of $c_4^{\mathbb{C}}$ and $d_1^{\mathbb{C}}$ cannot be analytic. If the coefficients are only processed by low pass filters (or high pass filters respectively), the phase difference added by the complete path can never be flat.

These considerations enable the creation of analytic complex wavelet packets. A complete wavelet packet of depth $D = 3$ is shown in figure 6. Firstly, the left half of the packet is regarded. The solid line filter stages are the known DTCWT as a subset of the complex wavelet packet. The complete filter bank paths from signal x down to the ends of the DTCWT add the correct phase difference (except for the highest and lowest frequency band of course). Thus, the dotted decomposition stages after the DTCWT stages need no additional phase difference. Identical filters can be used for both trees, that is the $h_0^{\mathfrak{Re}}$ / $h_1^{\mathfrak{Re}}$ or the $h_0^{\mathfrak{Im}}$ / $h_1^{\mathfrak{Im}}$ filters. In our implementation, the $\mathfrak{Re}$ filters are used.

If the right half of the wavelet packet is considered, the dashed filter stages are mirrored to the original DTCWT filter stages. The difference is the top level filter stage. In the right half-tree it is a high pass filter $h_{1,f}$ while in the left half it's a low pass filter $h_{0,f}$. The high pass filter adds a phase difference of positive slope while the low pass filter phase difference has a negative slope. Therefore, the lower filter stages in the right half (dashed) must add the opposite phase difference than their counterpart in the left half. The filter pairs $h_0^{\mathfrak{Re}}$ / $h_0^{\mathfrak{Im}}$ and $h_1^{\mathfrak{Re}}$ / $h_1^{\mathfrak{Im}}$ can be swapped in the right half in order to achieve this. The dotted filter stages following the dashed ones are identical to their counterparts in the left half. This filter swapping scheme applies to the analysis filters in the same way.

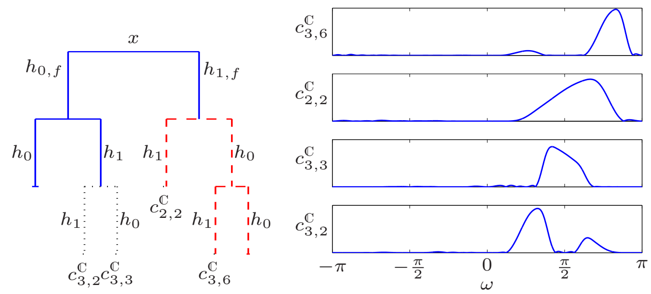
Considering the same example as in figure 5, figure 7 demonstrates that the basis functions of the proposed Complex Wavelet Packets are analytic. The analysis of chirp signals with shift variations is shown in figure 2b. In contrast to



(a) Tree $\mathfrak{Re}$

(b) Tree $\mathfrak{Im}$

Fig. 6: Complex Wavelet Packet Tree; solid: regular filter pairs ($\mathfrak{Re}$, $\mathfrak{Im}$), dashed: swapped filter pairs ($\mathfrak{Im}$, $\mathfrak{Re}$), dotted: identical filters ($\mathfrak{Re}$, $\mathfrak{Re}$)



(a) Example Tree (b) Amplitude spectra of basis functions

Fig. 7: Analytic CWP basis functions

the real wavelet packet analysis in figure 2a, the energy is distributed more equally over all frequencies and there is almost no influence from shifts.

5. RESULTS

This section shows the performance improvement of the denoising due to the usage of complex wavelet packets. The coefficient energy distribution of a period of noise was estimated from only two periods.

5.1. Chirp Noise

A series of chirps with shift variations like in figure 2 was generated and added to a speech signal. The SNR of the noisy signal was -3.96 dB. Denoising based on real wavelet packets yielded an improved SNR of 1.15 dB. Utilizing Complex Wavelet Packets improved the SNR further to 3.94 dB.

5.2. Engine Noise

Engine noise recorded from a car engine was used. The noisy SNR was -4.04 dB. It was improved by real wavelet packet thresholding to 2.38 dB. Complex Wavelet Packet denoising improved the SNR to 4.26 dB which is again an improvement compared to real wavelet packets.

6. CONCLUSIONS

In previous work [2], the authors found that the shift variance of real wavelet packets decreases the performance of a wavelet packet denoising scheme for periodic noise. Therefore, this paper concentrated on extending the dual-tree complex wavelet transform (DTCWT) to complex wavelet packets, thereby incorporating its nearly shift invariance as well as other positive properties [3]. It was found that the straightforward extension by decomposing the high pass coefficients of the DTCWT does not yield analytic basis functions, which is necessary to preserve the desired properties of the DTCWT. A filter swapping scheme was developed in order to solve this problem and the basis functions became nearly analytic. The simulation results showed an improved performance for computer generated chirp noise and real-world engine noise.

It should be noted that complex wavelet packets are not only useful in the specific periodic noise filtering method but are in general a better representation because they preserve the properties of the DTCWT: shift invariance, smooth coefficients and better directionality (in higher dimensions). They should yield better results in applications like compression, general filtering or signal analysis.

7. REFERENCES

[1] Hamid Sheikhzadeh and Hamid Reza Abutalebi, “An improved wavelet-based speech enhancement system,” in *EUROSPEECH 2001*, 2001, pp. 1855–1858.

[2] Thomas Weickert and Uwe Kiencke, “Adaptive estimation of periodic noise energy distributions for speech enhancement,” in *9th IFAC Workshop ALCOSP’07*, 2007.

[3] I. Selesnick, R. Baraniuk, and N. Kingsbury, “The Dual-Tree Complex Wavelet Transform,” *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, November 2005.

[4] N.G. Kingsbury, “The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters,” in *Proc. 8th IEEE DSP Workshop*, 1998.

[5] Stéphane Mallat, *A wavelet tour of signal processing*, Academic Press, 2. ed. edition, 1999.

[6] N. Kingsbury, “A dual-tree complex wavelet transform with improved orthogonality and symmetry properties,” in *Proc. International Conference on Image Processing, 2000*, 2000, vol. 2, pp. 375–378.