Audio Engineering Society

Convention Paper 7416

Presented at the 124th Convention
2008 May 17–20 Amsterdam, The Netherlands

A Channel Vocoder using Wavelet Packets on a Reconfigurable Device

César Daniel Salvador Castañeda[1]

[1] Pontifical Catholic University of Peru, ECOS Research Group, Lima, Lima 32, Peru
salvador.cd@pucp.edu.pe

ABSTRACT

A channel vocoder using wavelet packets for computer music applications is proposed. The input audio signals are a modulating voice and a carrier melody. The wavelet packets channel vocoder transforms windowed frames of both signals to a subband domain, mixes the melody with the voice envelope, and transforms back the result to the original domain. Design is performed with Matlab/Simulink tools and real time implementations with Pure Data and Virtex II Pro FPGA board. Appropriate choices of frame length, wavelets, decomposition levels and envelope detector filter are proposed to achieve good quality sound effects. Finally, guidelines to improve transmission and compression rates in a future work are suggested.

1. INTRODUCTION

A new channel vocoder algorithm for computer music applications is introduced in this document. The novelty is the use of dyadic symmetric filter banks in the analysis and synthesis stages. Here, dyadic refers to two channel and symmetric to wavelet packets transform. Considering that filters are derived from wavelet functions, the proposed scheme corresponds with a perfect reconstruction filter bank, which transforms signals between time domain and subband domain. Thus, main processing is performed in subband domain.

The organization of this paper is as follows. Section 2 explains the classic channel vocoder algorithm from a subband coding point of view, setting the relation with two channel filter banks. Section 3 explains the wavelet packets transform and its realization with two channel filter banks. The proposed wavelet packets channel vocoder algorithm is presented in section 4, which is followed by a description of software and hardware real time implementations using Pure Data and a Virtex-II Pro Xilinx board in section 5. Finally, results in wavelet packets and time-frequency domains are reported in section 6 and concluding remarks based on results appear in section 7.

2. THE CHANNEL VOCODER

The Channel vocoder is used to achieve bandwidth compression in voice encoding. The inputs are a modulating signal, which is choice to be voice, and a carrier signal, which can be melody or noise. Voice and melody are typically subband encoded through a filter bank, then a rectifier followed by a low-pass filter is applied to each subband component of voice in order to estimate its energy in that particular bandwith. After mixing the corresponding components of both signals in each subband, a synthesized sound with almost the same frequency components as the carrier is obtained. Given that the synthesized sound has similar envelope as to the modulating part in time domain, the resulting sound is music-like human speech.

Dudley [1] introduced the channel vocoder, Gold and Rader [2] realized it with analog circuits, Rabiner and Huang [3] implemented a digital version using FFT, and Park [4] replaced FFT with wavelet transform using complex Morlet. The proposed WPCV, instead of assigning larger bandwidth to lower frequencies as in the wavelet case, it covers the whole spectrum with regularly distributed bands using wavelet packets.

A basic example on four subband decomposition with a four channel filter bank is shown in figure 1, where frequency scale appears normalized, with 1 corresponding with half the sampling frequency at the input of the corresponding block. That scheme corresponds with a filter bank based channel vocoder, since it is composed of the decimation filters of factor four $(\downarrow 4)H_0$, $(\downarrow 4)H_1$, $(\downarrow 4)H_2$ and $(\downarrow 4)H_3$, which cover the whole bandwith and set the context for subband mixing.

On the other hand, as shown in figure 2, decomposition into four subbands can also be done by applying three times a new two channel filter bank, which is composed of the decimation filters of factor two $(\downarrow 2)H_0$ and $(\downarrow 2)H_1$. The same reasoning can be extended to L levels of decomposition, say 2^L subband decomposition. Indeed, a 2^L subband decomposition requires a 2^L channel filter bank in a filter bank channel vocoder scheme, while the same operation requires $1+2+\ldots+2^L$ iterations of a two channel filter bank in the wavelet packet channel vocoder scheme.

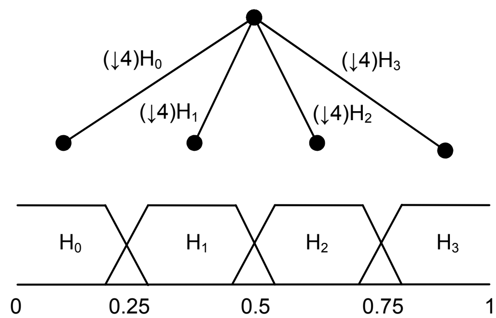


Figure 1: Four subband analysis in a filter bank channel vocoder scheme.

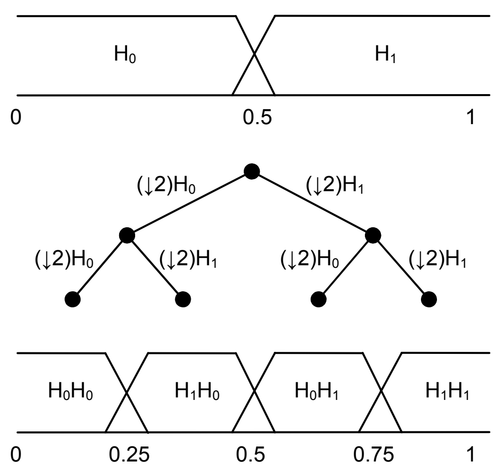


Figure 2: Four subband analysis in the wavelet packet channel vocoder scheme.

3. WAVELET PACKETS AND FILTER BANKS

The Wavelet Packets Transform (WPT) decomposes a broadband signal into its components with smaller bandwidths and slower sample rates. It uses a series of two channel analysis filter banks with filters H_0 and H_1 to repeatedly divide the input frequency range. The original signal can be reconstructed with the Inverse Wavelet Packets Transform (IWPT), which uses the dual two channel synthesis filter bank with filters $\hat{H}_0$ and $\hat{H}_1$ [6]. Figure 3 shows a three level decomposition, here WPT divides a signal u into its four subband components $u^{(00)}$, $u^{(10)}$, $u^{(01)}$ and $u^{(11)}$, whose samples can be multiplexed to allow the use of one single channel bus.

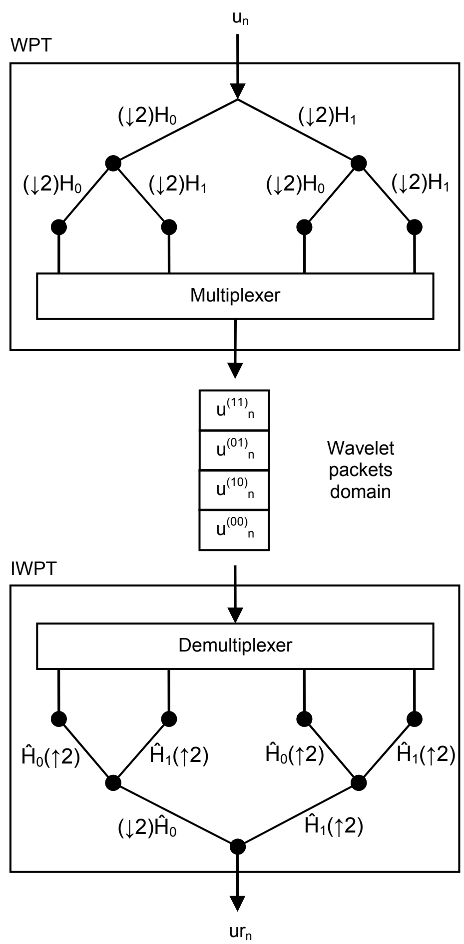


Figure 3: Two level analysis (top) and synthesis (bottom) of a signal using wavelet packets basis.

4. THE WAVELET PACKETS CHANNEL VOCODER

The wavelet packets channel vocoder (WPCV) algorithm uses wavelet packets to allow processing in regularly distributed bands. As shown in figure 4, this is achieved decomposing windowed frames of voice and melody in their subband components using WPT blocks. Thus, voice energy for each particular bandwith can be estimated through an envelope detector, which is composed of a rectifier followed by a low-pass filter. Components of signals are multiplied and the synthesized sound is reconstructed from the subband products using IWPT.

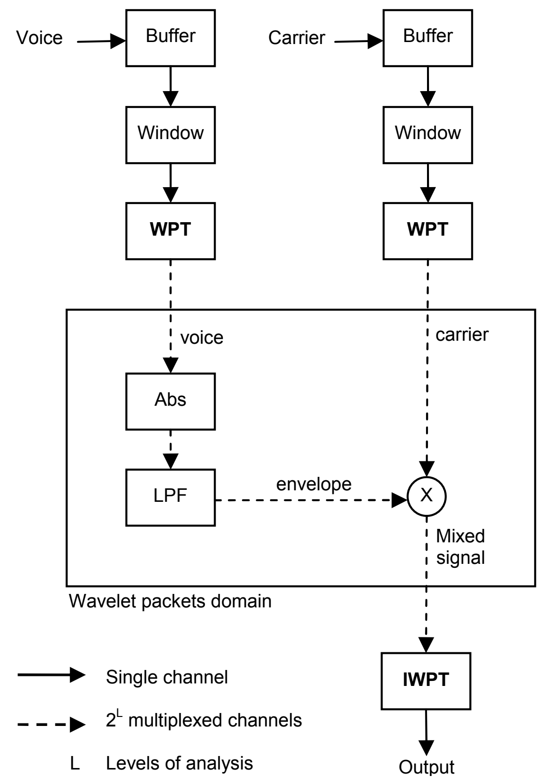


Figure 4: The wavelet packet channel vocoder.

Considering that an L level decomposition of a signal of length N gives 2^L subband signals with length $N/2^L$, then frame length multiples of 2^L must be chosen. Nevertheless, in order to avoid feedback effects in the synthesized sound, the WPCV uses frames of 2^L samples. Up to here, windowing can also be applied to frames for smoothing effects.

Since filter banks use a perfect reconstruction scheme then H_0, H_1, $\hat{H}_0$ and $\hat{H}_1$ are quadrature mirror filters. Hence, given H_0, filters H_1, $\hat{H}_0$ and $\hat{H}_1$ can be easily obtained using flip and alternating sign operations ([5], [6]). The number of decomposition level is 7 and the analysis and synthesis filters are derived from wavelet basis like Daubechies and Meyer. The same wavelet basis is used for both the carrier and modulating parts. Two sets of H_0 and H_1 filters for a two channel analysis filter bank appear in figure 5 and 6, respectively 32^{nd} order Daubechies wavelet filters and discrete Meyer wavelet filters.

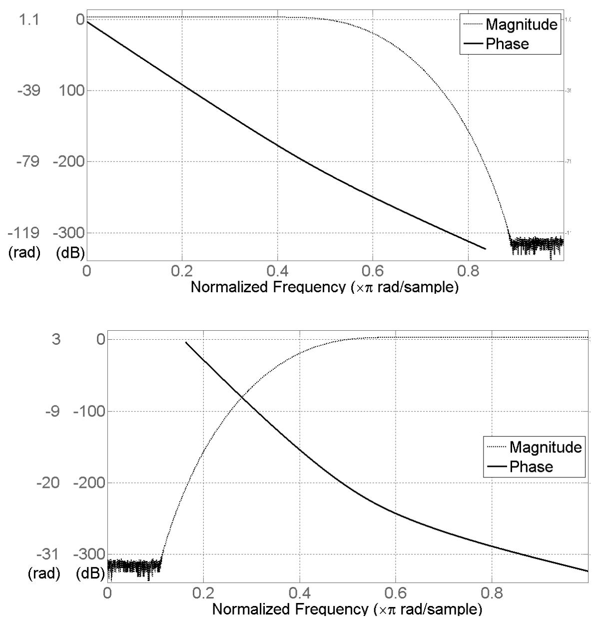


Figure 5: Frequency responses of low-pass (top) and high-pass (bottom) analysis filters corresponding to 32nd order Daubechies wavelet.

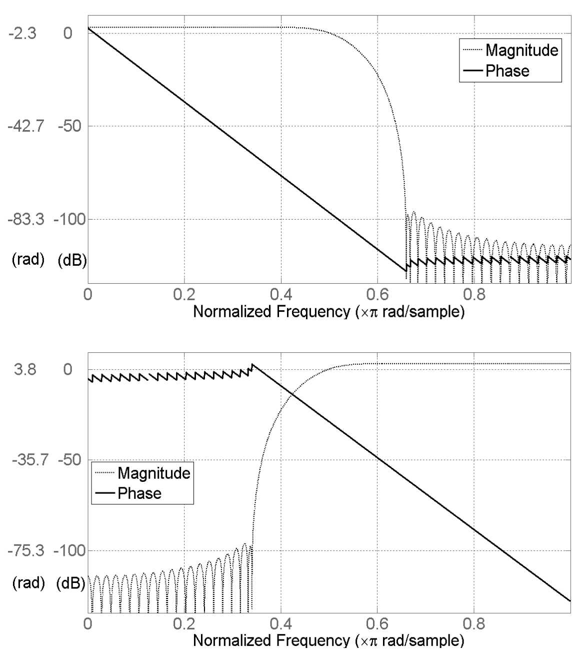


Figure 6: Frequency responses of low-pass (top) and high-pass (bottom) analysis filters corresponding to discrete Meyer wavelet.

The filter used for envelope detection is a low-pass FIR filter of order 16, which stop and pass frequencies are respectively 0.1 and 0.5. The order value was chosen in order to reduce complexity. The pass band takes into account the wavelet packets representation of the voice envelope shown in figures 9 and 10. This filter was designed using least 128th norm with density factor of 100. Least pth-norm method use density factor to control the density of the frequency grid over which the design method optimization evaluates the filter response function. The number of equally spaced points in the grid is density factor times (filter order + 1), in this case, 12800.

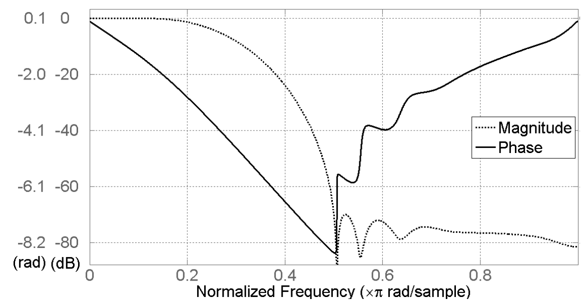


Figure 7: Frequency response of the low-pass FIR least 128th filter used for envelope detection.

When the multiplication in wavelet packets domain is being carried out, the voice energy in a particular bandwidth is multiplied by the corresponding frequency components in the carrier signal. This implies the insertion of energy of the voice into the carrier signal. Finally, the output is reconstructed from the subband products using IWPT.

5. IMPLEMENTATION

The use of wavelet packets simplifies the algorithm due to the use of dyadic filter banks. Thus, makes it suitable for hardware implementations. With the aim of avoiding convolution products that will be discarded by downsample operators in WPT and multiplications by zero due to upsample operations in IWPT, the two channel analysis and synthesis filter banks are implemented using a polyphase factorization [6]. A direct realization of the structure shown in figure 3 requires the system to run at the input data rate. This requirement can also be relaxed by taking advantage of the polyphase factorization [7].

Matlab/Simulink has been used for the design of the WPCV shown in figure 4, which uses the available dyadic filter banks built-in functions. System Generator has been used for the FPGA implementation of the proposed WPCV, which uses FIR filter cores to implement the WPT and IWPT architectures shown in figure 8. These allow for a feedback path to implement multiple stage decomposition trees. Additional delay units are used to synchronize control signals in the pipeline. The AC 97 audio codec integrated in Virtex II Pro FPGA board was also configured using EDK for real time processing.

A summary of resource consumption and performance results for the FPGA WPCV with seven levels WPT is reported in table 1. In all cases, the WPT block implements the discrete Meyer wavelet filters (102 taps fillers). The format column refers to input and output data and filter coefficients, where Fix-Nb-Bp means signed fixed point with Nb bits and bynary point Bp. For comparison, a software based floating point precision WPT and IWPT were implemented for Pure Data using the wavelet transform patch described in [8]. The maximum error obtained for each data format is reported in table 2.

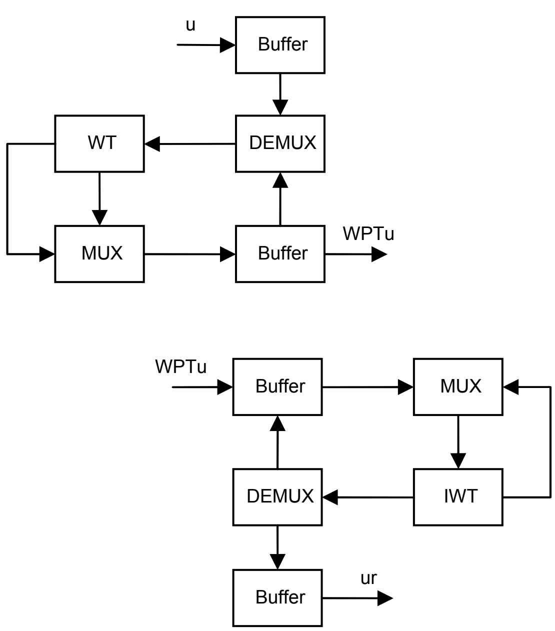


Figure 8: Block diagram of the FPGA WPT (top) and IWPT (bottom) with $1+2+\ldots+2^L$ iterations of a two channel filter bank (WT).

Format	Fix-16-14	Fix-24-22	Fix-32-30
Slices	4984	11328	15248
FFs	4240	6130	8098
BRAMs	0	0	0
LUTs	6568	13160	18444
Emb. Multipliers	24	72	72

Table 1 Summary of resource consumption and performance results.

Format	Fix-16-14	Fix-24-22	Fix-32-30
$\|\|e\|\|_\infty$	$2.7e10^{-5}$	$9.8e10^{-6}$	$5.5e10^{-8}$

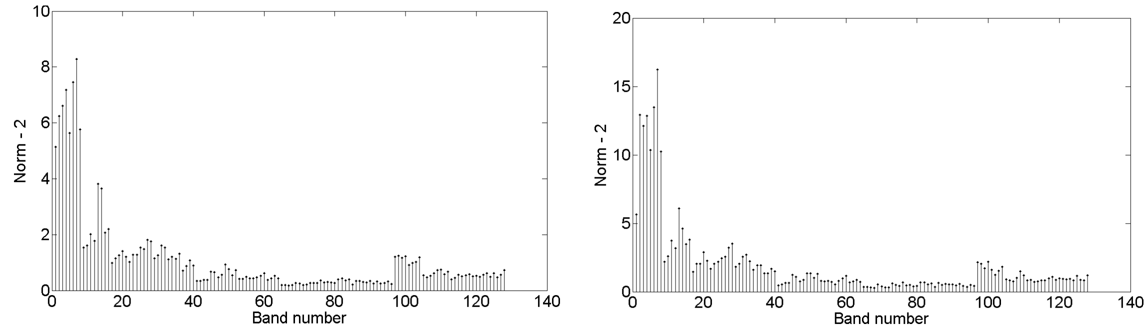
Table 2 Hardware software comparison.

6. RESULTS

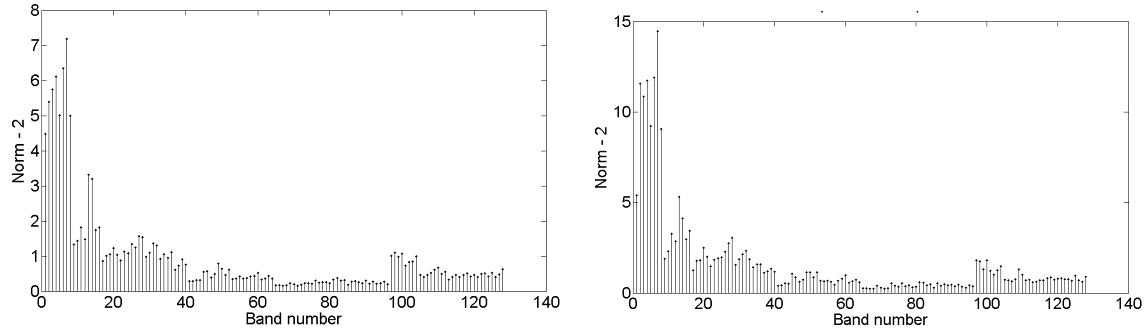
Results of internal signals in wavelet packets domain and external signals in time-frequency domain are reported in this section. The input signals are the same ones used in [4]. The modulating signal is composed of two loops of the speech fragment 'testing one two three', and the carrier signal is composed of two loops of an organ melody with two differentiated successive chords. Both signals are 16 bit PCM with sampling frequency of 22050 Hz.

Seven levels of analysis with WPT corresponds to 128 independent subbands each one with 172.25 Hz of bandwith. Figure 9 and figure 10 show the internal signals in wavelet packet domain, respectively for 32nd order Daubechies wavelet filters and discrete Meyer wavelet filters. From the L^2 and L^∞ norms figures it is clear that voice and its envelope has almost the same energy and maximum values in the lower subbands. Thus, energy as well as maximum values corresponding to each subband are better transmited approximately up to the 20th subband.

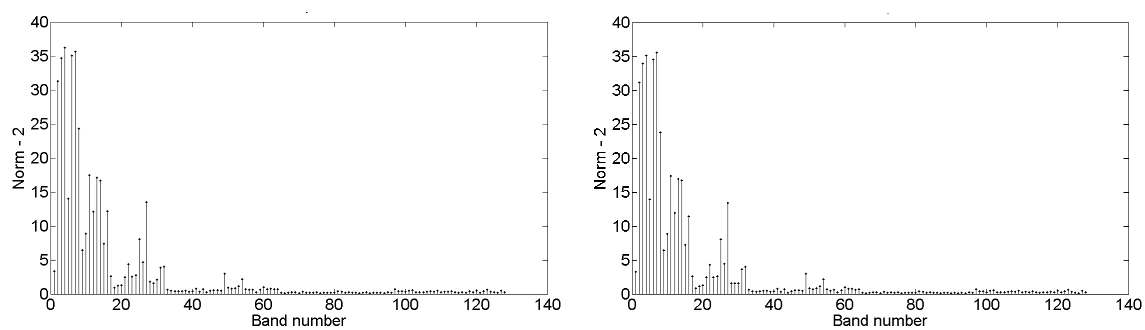
From the magnitude of the spectrogram of external signals in figure 11, it is clear that WPCV produces insertion of energy of the voice into the carrier signal. On the other hand, from the phase information of the spectrogram of the external signals, which appears unwrapped and rectified to center the attention on linearity, it is clear that discrete Meyer wavelet filter produces better results due to its linearity in phase, which can be seen in figure 6.



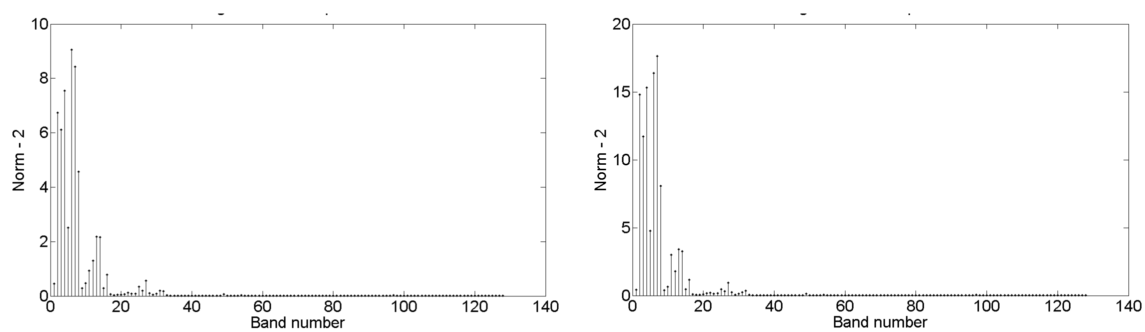
a) voice



b) voice envelope

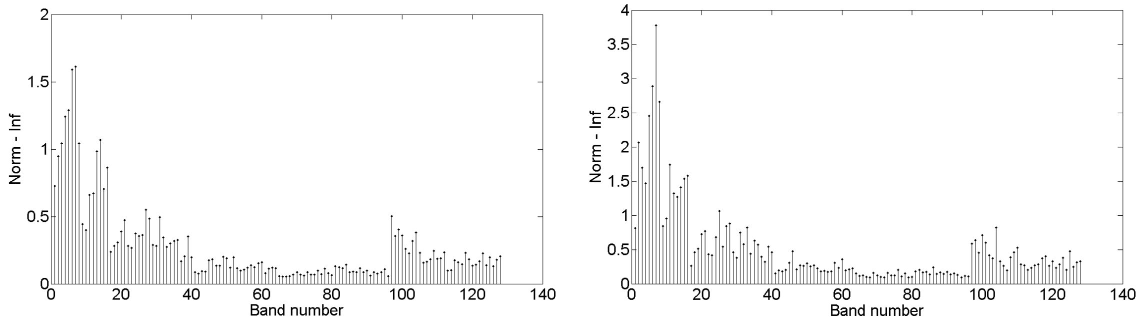


c) carrier

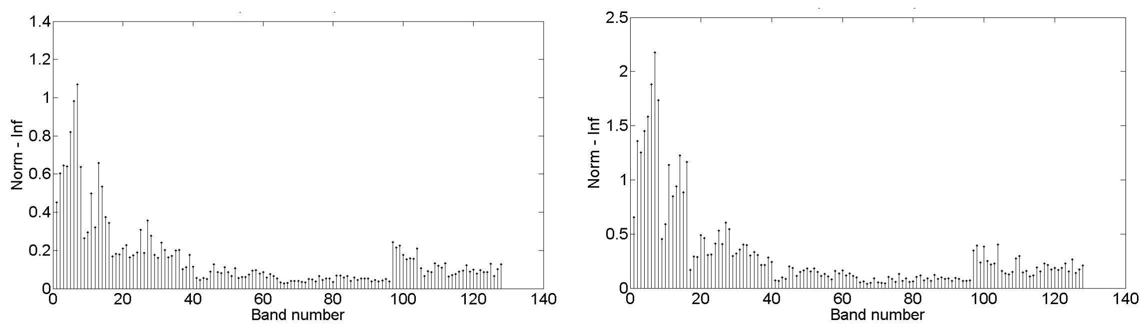


d) mixed signal

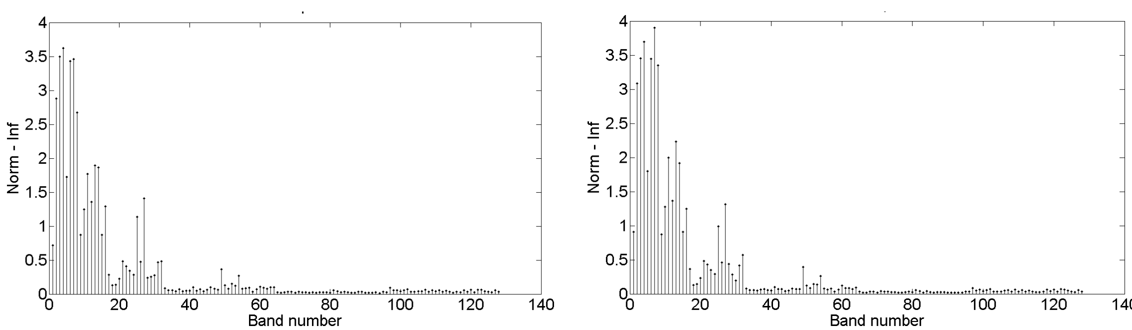
Figure 9: L^2-norm of signals in wavelet packets domain for 32^{nd} order Daubechies wavelet filters (right) and discrete Meyer wavelet filters (left).



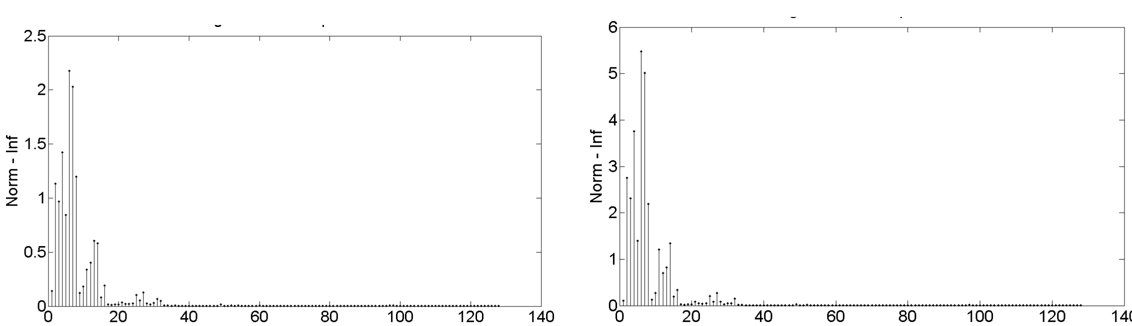
a) voice



b) voice envelope

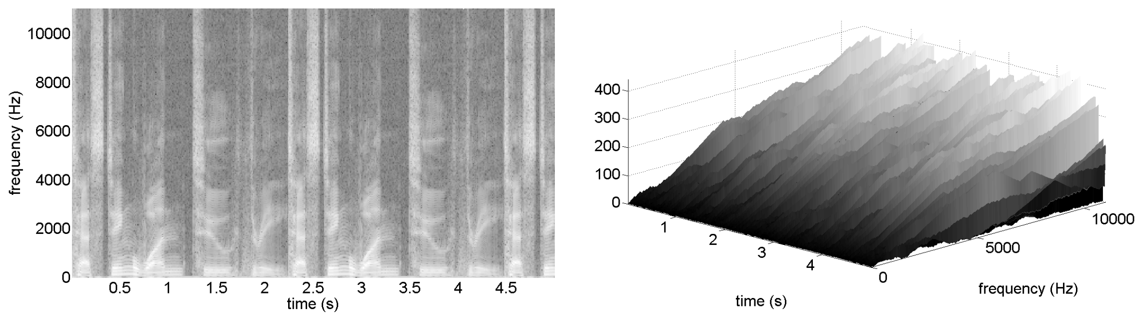


c) carrier

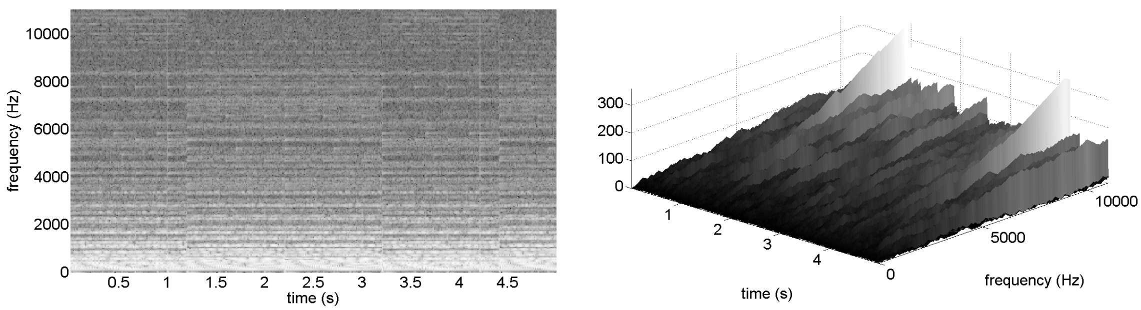


d) mixed signal

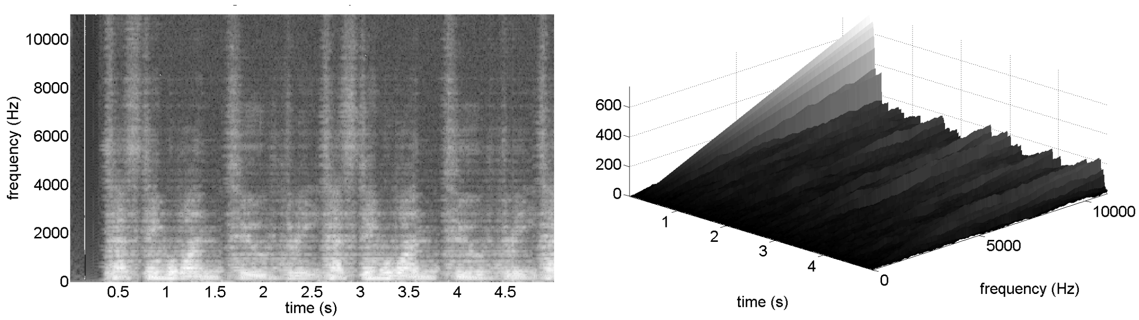
Figure 10: L^{∞}-norm of signals in wavelet packets domain for 32nd order Daubechies wavelet filters (right) and discrete Meyer wavelet filters (left).



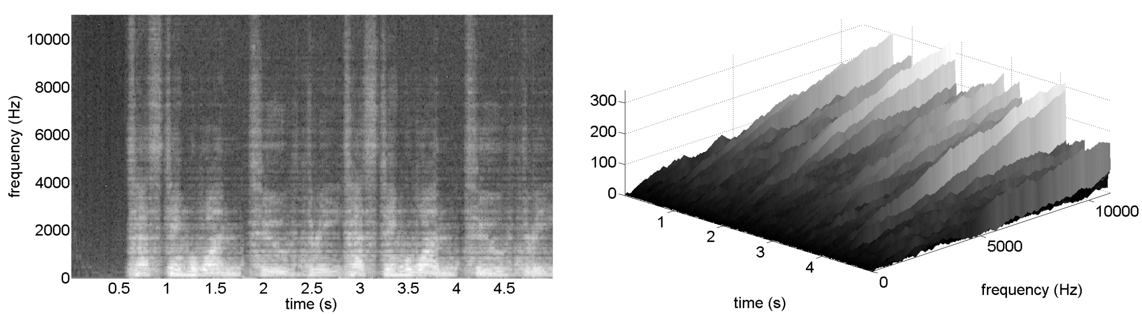
a) voice



b) carrier



c) output for 32nd order Daubechies wavelet filter



d) output for discrete Meyer wavelet filter

Figure 11: Time-frequency domain representation of input and output signals for the wavelet packets channel vocoder. Magnitude (left) and absolute value of unwrapped phase (right) of the spectrogram.

7. CONCLUSIONS

A channel vocoder using wavelet packets for computer music applications was proposed. Design has been realized with perfect reconstruction filter banks. A comparison between software and hardware real time implementations has been performed using Pure Data and a Virtex II Pro FPGA board.

More efficient results are produced using wavelet packets transform, given that wavelet basis considers time and frequency domain at the same time, being more suitable than Fourier basis. The reported results for 32nd order Daubechies and discrete Meyer wavelet filters, with mixing in the seventh level of analysis, showed that phase response of wavelet filters directly affects the synthesized sound.

In a future work two problems should be addressed. First, adapting the stop and pass frequencies of envelope detection low-pass filter using WPT information of voice envelope. Second, improving compression and transmission rates using wavelet packets transforms that map integers to integers. According to [9] results in integer arithmetic are possible using lattice factorization instead of a polyphase one in the implementation of two channel filter banks. Lattice factorization would directly improve the transmission rate by a factor of two, and integer arithmetic would allow the addition of coding and decoding stages in the transform domain.

8. ACKNOWLEDGEMENTS

The author would like to thank people from ECOS Research Group and DSP Laboratory at Pontifical Catholic University of Peru for their comments and technical support throughout this work. He wishes to thank Prof. William A. Sethares from University of Wisconsin–Madison for his idea on revisiting the channel vocoder from a multiresolution point of view. He also would like to thank Alonso Vera from University of New Mexico for his guidelines on FPGA wavelet transform.

9. REFERENCES

[1] H. Dudley, “Remaking speech,” Journal of the Acoustical Society of America, vol. 11, pp. 169-177 (1939 October)

[2] B. Gold, C. Rader, “The Channel Vocoder,” IEEE Transactions on Audio and Electroacoustics, vol. au-15, no. 4 (1967 December)

[3] L. Rabiner, B. Juang, Fundamentals of Speech Recognition, Prentice Hall, New York (1993)

[4] B. Park, Channel Vocoder using Multiresolution Method: Speech Synthesis, Project Report, University of Wisconsin–Madison (2005)

[5] S. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation”, IEEE Transactions on Pattern Recognition and Machine Intelligence, vol. 11, no. 7, pp. 674-693 (1989 July)

[6] T. Nguyen y G. Strang, Wavelets and Filter Banks, Wellesley-Cambridge Press, Wellesley (1996)

[7] U. Meyer-Baese, A. Vera, A. Meyer-Baese, M. Pattichis, and R. Perry, “Discrete wavelet transform fpga design using matlab/simulink,” in Independent Component Analyses, Wavelets, Unsupervised Smart Sensors, and Neural Networks IV, Proceedings of the SPIE, vol. 6247 (2006)

[8] R. Poblete, Manipulation of Audio in the Wavelet Domain processing a Wavelet Stream using PD, work for Toningenieur Diplomarbeit, Institut für Elektronische Musik, Graz, (2006)

[9] R. Calderbank, I. Daubechies, W. Sweldens,B. Yeo, “Wavelet transforms that map integers to integers,” Journal of Applied and Computational Harmonic Analysis, vol. 5, no. 3, pp. 332-369, (1998 July).