HARMONIC-BAND WAVELET COEFFICIENT MODELING
FOR PSEUDO-PERIODIC SOUND PROCESSING

Pietro Polotti, Gianpaolo Evangelista

Laboratoire de Communications Audiovisuelles (LCAV)
École Polytechnique Fédérale de Lausanne, Switzerland
`pietro.polotti@epfl.ch`   `gianpaolo.evangelista@epfl.ch`

ABSTRACT

In previous papers [1], [2] we introduced a model for pseudo-periodic sounds based on Wornell results [3] concerning the synthesis of $1/f$ noise by means of the Wavelet transform (WT). This method provided a good model for representing not only the harmonic part of real-life sounds but also the stochastic components. The latter are of fundamental importance from a perceptual point of view since they contain all the information related to the natural dynamic of musical timbres. In this paper we introduce a refinement of the method, making the spectral-model technique more flexible and the resynthesis coefficient model more accurate. In this way we obtain a powerful tool for sound processing and cross-synthesis.

1. INTRODUCTION

The main motivation of our $1/f$ pseudo-periodic model derives from the observation that the sidebands of the harmonics of a large class of voiced sound have an approximate $1/f$ behavior. These sidebands contain the stochastic components of the sound, i.e., the information concerning the "stochastic evolution" of the timbre and the added noises, as breath in wind instruments, so important in order to perceive a sound as a natural one. In [1, 2] we introduced a new set of wavelets, which we called the Harmonic Band Wavelets (HWBT). By means of the HBWT we were able to extend Wornell result about the synthesis of $1/f$-like processes by means of WT to the pseudo-periodic $1/f$-like case. This was realized by means of the demodulation-modulation scheme introduced in [2]. We showed that it is possible to synthesize pseudo-periodic $1/f$-like power spectra by employing white noise, with wavelet band dependent energy, as coefficients. The spectrum of the synthetic signal can be adapted to that of a real-life pseudo-periodic sound by extracting a limited set of parameters from the HBWT analysis coefficients to control the energies of the spectral sidebands.

Different levels of approximation can be achieved in the synthesis of noise components of pseudo-periodic sounds. A refinement of the technique consists in setting our method free from the strict constraints of the $1/f$ model in order to obtain a better approximation of the spectrum shape. This can be achieved by employing the Frequency Warped Wavelet Transform (FWWT), recently introduced by one of the authors [4, 5]. We obtain arbitrary segmentation of the frequency axis, i.e., of the wavelet analysis and synthesis bands. In this way we can better reproduce the deviation of real spectra with respect to the strict pseudo-periodic $1/f$-like model (see for example Fig. 1, where relevant non-harmonic peaks are present in the spectrum of a french horn).

Furthermore white noise resynthesis coefficients are only a rough approximation, which allows for very low coding rates but involves large distortion. In fact the HBWT analytical results show an important aspect, which the white noise approximation does not take into account: the existence of a small but non-zero correlation in the analysis coefficients. We model this correlation by means of AR filters driven by white noise. The filters are obtained by performing LPC analysis on the coefficients of the HWBT analysis of the signal.

In our method we model only the steady part of sound and we assume that transients have been extracted from the rest of the signal. A pitch detector can be successfully employed for automatic detection of the transient length. As soon as the pitch becomes sufficiently stable, the HBWT analysis starts.

This technique is a powerful approach to audio data compression. Compression ratios of about 1/40 can be achieved, while granting optimal sound quality and naturalness.

From a musical point of view straightforward applications of our method are, for instance, cross synthesis and morphing.
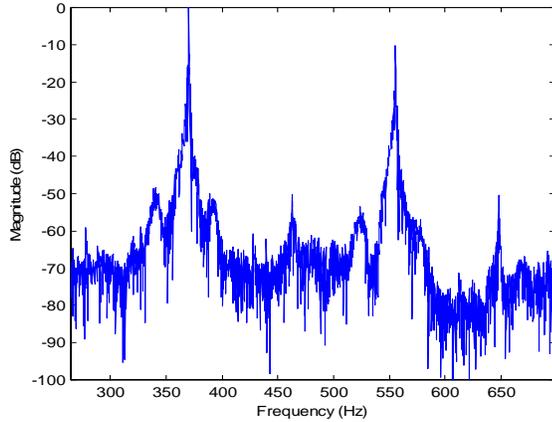

Fig. 1 *Two harmonics of the spectrum of a real life horn*

The paper is organized as follows. In the next two sections we briefly review the Frequency Warped Wavelets and the Harmonic-Band Wavelets, respectively, and we introduce the Frequency Warped HBWT. In section 4 we present the synthesis method and the coefficient modeling techniques. In section 5 we describe the sound processing potentialities of the method. In the last section we finally draw our conclusions.

2. WAVELETS AND FREQUENCY WARPED WAVELETS

Wavelet transforms are a mathematical tool performing multiresolution analysis. Their most appealing feature from a DSP point of view is related to the non-uniform octave band subdivision of the space-frequency or time-frequency spaces in image and sound processing respectively. The octave band subdivision as well as the principle of selfsimilarity, lying behind it, seems to be successful from a perceptual point of view for both our hearing and visual system.

What we are trying to do here is to model the spectra of pseudo-periodic signals. The harmonic-peaks of these spectra have an approximately $1/f$-like behavior, well fitting the power-2 "observation perspective" of the wavelet transform. This model can be improved if we make the power-2 law more flexible and adaptable to real-life deviations from the model itself. This can be obtained by introducing the frequency warped WT, i.e., a WT with an arbitrary non-uniform subdivision of the frequency axis replacing the octave-band subdivision [5].

We now introduce the mathematics. We know that, given an ordinary wavelet set orthonormal and complete in l^2, we can write any signal $s(l) \in l^2$ as:

$$s(l) = \sum_{n=1}^{N} \sum_{m} b_{n,m} \psi_{n,m}(l) + \sum_{m} a_{N,m} \phi_{N,m}(l), \quad (1)$$

where the symbols ψ and ϕ denote, respectively, a Wavelet sequence and a Scale sequence [6, 7]. The index n denotes the wavelet scale; m is the time shift. Equation (1) is a finite scale N expansion of $s(l)$ on the set of functions $\{\psi_{n,m}(l)\}$ and $\{\phi_{N,m}(l)\}$, where $n=1,\dots,N$ and $m \in \mathbf{Z}$. The wavelet expansion is computed by a cascade of filters g and h, implementing a two-channel critically sampled filter-bank [6, 7]. At each scale the high-pass filter g and the low-pass filter h implement the wavelet projection and the scale projection respectively. In (1) and in the rest of the paper a sum with unspecified boundaries denotes that the index runs from $-\infty$ to $+\infty$.

As mentioned above, in order to set the wavelet expansion free from the rigid power of 2 time-frequency plane subdivision, the frequency warped wavelets were introduced [4, 5]. The frequency warped wavelets $\widetilde{\psi}_{n,m}(l)$ and their corresponding frequency warped scale sequences $\widetilde{\phi}_{n,m}(l)$ obey, respectively, to the following recursive relation:

$$\widetilde{\psi}_{n,m}(l) = \sum_{k=0}^{\infty} g_{n,m}(k) \widetilde{\phi}_{n-1,k}(l) \quad (2)$$

and

$$\widetilde{\phi}_{n,m}(l) = \sum_{k=0}^{\infty} h_{n,m}(k) \widetilde{\phi}_{n-1,k}(l), \quad (3)$$

where the $g_{n,m}$ and the $h_{n,m}$ are some auxiliary sequences given by:

$$g_{n,m}(k) = \sum_{r=0}^{\infty} \lambda_{n,r}(k) g(r-2m) \quad (4)$$

and

$$h_{n,m}(k) = \sum_{r=0}^{\infty} \lambda_{n,r}(k) h(r-2m), \quad (5)$$

where the symbol $\lambda_{n,r}$ denotes a Laguerre sequence [4] of order r associated to the nth wavelet scale. The ordinary quadrature mirror filters g and h in this case play the role of coefficients of the Laguerre expansion of the functions $g_{n,m}$ and $h_{n,m}$ respectively.

The Frequency Warped Wavelets form orthonormal and complete sets. For any $s(l) \in l^2 (\mathbf{N} \cup \{0\})$ we can write:

$$s(l) = \sum_{n=1}^{N} \sum_{m} b_{n,m} \widetilde{\psi}_{n,m}(l) + \sum_{m} a_{N,m} \widetilde{\phi}_{N,m}(l). \qquad (6)$$

Frequency warping or frequency axis deformation, obtained by means of the Laguerre Transform, is controlled by the parameter d_n according to the following recurrence:

$$d_1 = \tan[(\pi - 2\omega_1)/4] \qquad (7)$$

and

$$d_n = \tan\left[\frac{\pi}{4} - \Omega_{n-1}(\omega_n)\right], \qquad (8)$$

where the ω_n are the arbitrary cut-off frequencies by which we subdivide the frequency axis, with $\omega_1 > \omega_2 > \ldots > \omega_n$ and the frequency mapping $\Omega_n(\omega)$ is given by [5]:

$$\Omega_n(\omega) = \theta_n(2\theta_{n-1}(\ldots 2\theta_2(\theta_1(\omega))\ldots)), \qquad (9)$$

where $\theta_k(\omega)$ is the negative phase response of a first-order allpass filter frequency response:

$$\theta_k(\omega) = \omega + 2\tan^{-1}\left(\frac{d_k \sin\omega}{1 - d_k \cos\omega}\right). \qquad (10)$$

3. HARMONIC-BAND WAVELETS (HBWT) AND FREQUENCY WARPED HWBT

In [2] we introduced a new wavelet set, i.e., the Harmonic-Band Wavelet Transform (HBWT). The HBWT are orthonormal and complete in l^2. They consist of a P-channel filter-bank based on the Discrete Cosimne transform (DCT) followed by a WT of each channel. Any signal $s(l) \in l^2$ can be expanded on a HBWT set:

$$s(l) = \sum_{p=1}^{P}\left(\sum_{n=1}^{N} \sum_{m} b_{n,m,p} \xi_{n,m,p}(l) + \sum_{m} a_{n,m,p} \zeta_{n,m,p}(l)\right), \qquad (11)$$

where the $\xi_{n,m,p}(l)$ and the $\zeta_{n,m,p}(l)$ represent a Discrete Harmonic-Band Wavelet and a Discrete Harmonic-Band Scale Function and are given by [8, 2]:

$$\xi_{n,m,p}(l) = \sum_{i} \psi_{n,m}(i) f_p(l - iP) \qquad (12)$$

and

$$\zeta_{n,m,p}(l) = \sum_{i} \phi_{n,m}(i) f_p(l - iP). \qquad (13)$$

The filters f implement the multichannel filterbank based on the DCT type IV ([2]).

P is the number of channels and p is the channel index. If $s(l)$ is a pseudo-periodic signal, P correspond to the pitch of $s(l)$.

In this paper we are interested in pseudo-periodic signals. The HBWT decomposition can be stated in a more intuitive way, if we separate its two levels: the P-channel filtering and the WT. In fact we can write $s(l)$ as:

$$s(l) = \sum_{p=1}^{P} \sum_{i} v_p(i) f_p(l - iP), \qquad (14)$$

where each $v_p(i)$ is the P-downsampled version of the pth band. If $s(l)$ is a pseudo-periodic signal we can "tune" P with the pitch of $s(l)$. In this case $v_p(i)$ represents one of the two sideband of the kth harmonic, with $k = \left\lfloor \dfrac{p+1}{2} \right\rfloor$.

The left sidebands correspond to odd p while the right sidebands corresponds to even p (see Fig. 2). The HBWT coefficients $b_{n,m,p}$ and $a_{n,m,p}$ are obtained by wavelet transforming the $v_p(i)$:

$$v_p(i) = \left(\sum_{n=1}^{N} \sum_{m} b_{n,m,p} \psi_{n,m}(i) + \sum_{m} a_{N,m,q} \phi_{N,m}(i)\right). \qquad (15)$$

In order to obtain the Frequency Warped HBWT we need simply to substitute ψ and ϕ by $\widetilde{\psi}$ and $\widetilde{\phi}$:

$$v_p(i) = \left(\sum_{n=1}^{N} \sum_{m} b_{n,m,p} \widetilde{\psi}_{n,m}(i) + \sum_{m} a_{N,m,q} \widetilde{\phi}_{N,m}(i)\right). \qquad (16)$$

The great advantage is that each subband of each sideband can be adjusted by an optimization procedure, in order to fit any real-life spectrum of the kind of Fig. 1. In Fig. 3 we show an example of how the frequency spectrum of Fig. 2 can be modified by means of frequency warping. The position and bandwidth of each subband can be independently set by means of properly chosen parameters $\{d_{n.p}\}$ for each p and n.
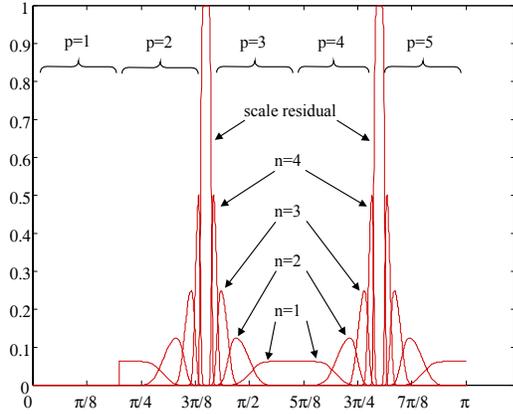
Fig. 2 *Frequency response of a filter-bank implementing ordinary HBWT: two harmonics.*
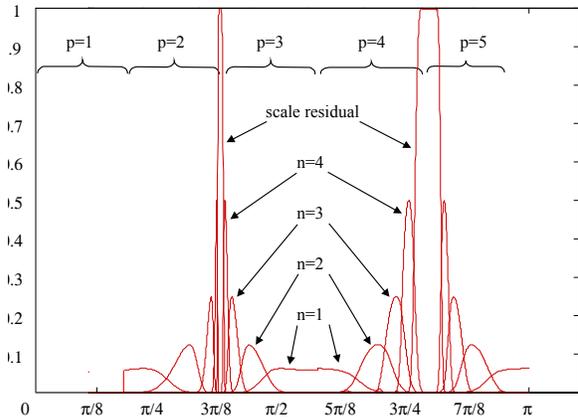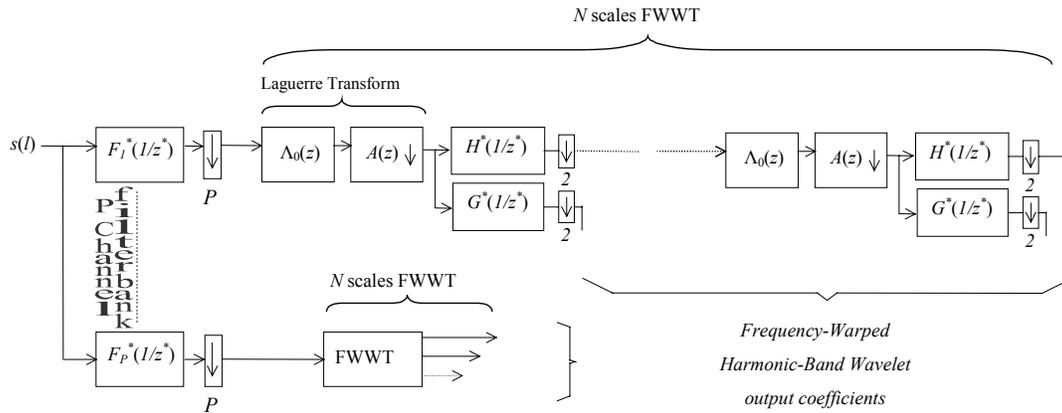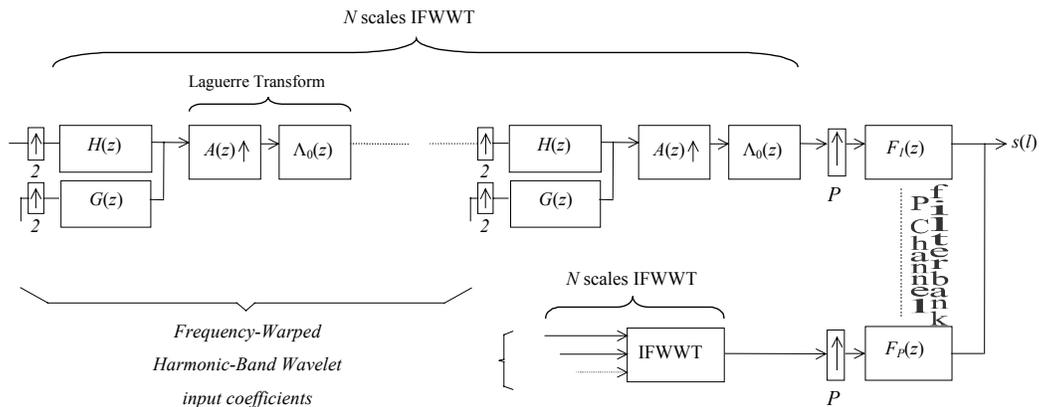
Fig. 3 *Frequency response of a filter-bank implementing Frequency-Warped HBWT: two harmonics.*

a) *Frequency-Warped HBWT Analysis Filter Bank*

b) *Frequency-Warped HBWT Synthesis Filter Bank*

Fig. 4 Analysis and Synthesis Frequency-Warped HBWT filter banks. F, G and H are the frequency responses of the filters f, g, h, respectively. Λ_0 and A are the frequency responses of the zero-order Laguerre sequence and of the dispersive delay lines implemented by a cascade of all-pass filters [5].

At the same time we obtain a finer tool for verifying the pseudo-periodic $1/f$-like model on data. In fact by means of the HBWT analysis we obtain for each sideband p a set of parameters corresponding to the energies of the subbands. Each parameter is a point of the hypothetical $1/f$-like spectrum of the sideband p. By subdividing the latter with a finer resolution we have more points at our disposal, by which we can test the validity of the pseudo-periodic $1/f$-like model. The experimental results confirm once more the already positive response obtained in [2].

In Fig. 4 we show the filer bank scheme, implementing the Frequency Warped HBWT.

4. RESYNTHESIS AND COEFFICIENTS MODELING

The main idea of the synthesis method we introduced in [1] and [2] was to "dress" the frequency response of Fig. 2 on the real-life spectrum for any given pseudo-periodic sound. This requires a parameter extraction, which is made by means of a HBWT analysis. More precisely we showed that the process

$$s(l) = \sum_{p=1}^{P}\sum_{n=1}^{N}\left(\sum_{m} 2^{\frac{n}{2}\gamma_p} v_{n,m,p}\,\xi_{n,m,p}(l) + \sum_{m} 2^{\frac{N}{2}\gamma_p} \mu_{n,m,p}\zeta_{n,m,p}(l) \right) \quad (17)$$

has time-averaged power spectrum approximately $1/f$ near each harmonic, where the $\xi_{n,m,p}(l)$ form a discrete Harmonic-Band Wavelet (*HBWT*) base and the $v_{n,m,p}$ and the $\mu_{n,m,p}$ are collections of white noise zero-mean coefficients with properly scaled energies. As already remarked, the index p in (17) represents the sidebands of the harmonics $q\dfrac{2\pi}{P}$, with $q = \left\lfloor \dfrac{p}{2} \right\rfloor$ and $q=1,\ldots,\left\lfloor \dfrac{P-1}{2} \right\rfloor$; n corresponds to the wavelet scale, m is the time shift and the γ_p are synthesis parameters derived from the HBWT analysis. The parameters γ_p shape the $1/f$ –like spectral behavior. The coefficients $\mu_{n,m,p}$ correspond to the deterministic-harmonic components of pseudo-periodic sounds. In the resynthesis of the harmonic components we do not employ the white noise coefficient model but the whole set of the HBWT analysis coefficients $\{a_{N,m,p}\}$. Their rate is low since each wavelet scale corresponds to a downsampling of order 2.

As a first refinement we set the method free from the rigid $1/f$-like pseudo-periodic model. In order to do this we replace the coefficients $2^{\frac{n}{2}\gamma_p}$ with a set of coefficients $\beta_{p,n}$ corresponding to the energies of the single subbands. Using four wavelet scales levels implies a growth of the number of parameters of approximately a factor 4, but we still obtain a very good coding rate.

Furthermore we corrected the first approximation, concerning the use of zero-correlation coefficients for the resynthesis. The non-zero correlation of the HBWT analysis coefficients is not negligible from an acoustical point of view. In order to reproduce it, we used an LPC synthesis of the HBWT coefficients. We obtained a set of AR filters, one for each p and n, and we used them to color each set of coefficient $\{v_{n,m,p}\}$. In other words we add a preliminary filtering of all the inputs of Fig. 4b. These AR filters are fed with uncorrelated noise with wavelet-scale-dependent variance.

A short-time version of the parameter extraction was also implemented. After computing the complete set of analysis HBWT coefficients, we update the parameters $\beta_{p,n}$ every 20 or less coefficients (the short-time window lengths depending upon the scale level).

Finally we specify once more that the model does not include the transient. An automatic detection of the transient length is made by means of a pitch detector. As soon as a stationary pitch is detected, we start HBW transforming the signal, while the transient is separated and preserved for the resynthesis as it is.

5. DIGITAL AUDIO APPLICATIONS

The HBWT analysis allows one to perform a partial reconstruction of the signal. We can separate the different subbands of all the harmonics (a single wavelet-band n of all the channels p) or one specific subband of one specific sideband, (fixed n and p) as well as any other arbitrary combination of subbands. This provides several possibilities in terms of sound processing results. More precisely we can define the nth noise subband (Nsb) as:

$$s_n(l) = \sum_{p=1}^{P}\sum_{m} \beta_{p,n} v_{n,m,p}\,\xi_{n,m,p}(l) \quad (18)$$

and the nth noise subband of the pth sideband (NSBsb) as:

$$s_{p,n}(l) = \sum_{m} \beta_{p,n} v_{n,m,p}\,\xi_{n,m,p}(l). \quad (19)$$

Also we can separate the noisy component of a single harmonic sideband. The pth noise sideband (NSB) is:

$$s_p(l) = \sum_{n=1}^{N}\sum_{m}\beta_{p,n}\nu_{n,m,p}\,\xi_{n,m,p}(l). \qquad (20)$$

In such a way timbre hybridization is straightforward: We can realize any "mixture" of subbands coming from the analysis of different instruments. A very simple example can be obtained combining the reconstructed harmonic component

$$s(l) = \sum_{p=1}^{P}\sum_{m}a_{n,m,p}\zeta_{n,m,p}(l) \qquad (21)$$

of one instrument with the noise sidebands $s_n(l)$ of another one. This can be successfully employed as a new cross synthesis technique. We obtained interesting results in combining the subbands of a horn, a trumpet, a bassoon, a clarinet and an oboe.

All the results of the last two sections are extendable to the frequency warped case.

6. CONCLUSIONS

In this paper we reviewed our synthesis technique based on the HBWT and we introduced many refinements to the method. A better spectral modeling can be obtained by introducing energy-parameters independently for each wavelet scale and by means of a Frequency Warping version of the HBWT. We improved the resynthesis coefficient model, introducing a pre-filtering of the white noise coefficients by means of AR filters, derived from the analysis.

The possibility of partial reconstruction and resynthesis of sounds gives interesting results. We can in fact employ the method as a tool for cross-synthesis and timbre hybridization.

Further developments of the method include first of all a pitch synchronous version. The starting point is the time varying case of modulated filter banks realized in [9]. This would give to the method a higher level of generality. The technique could then be applied to variable pitch sounds (including for instance, vibrato and monodic tunes).

7. REFERENCES

[1] P. Polotti, G. Evangelista, "Dynamic Models of Pseudo-Periodicity", *Proceedings of the 99 Digital Audio Effects Workshop*, pp. 147-150, Trondheim, Dec. 1999.

[2] P. Polotti, G. Evangelista, "Analysis and Synthesis of Pseudo-Periodic 1/*f*–like Noise by means of Wavelets with Applications to Digital Audio", to appear in *Applied Signal Processing. The International Journal of Analog and Digital Signal Processing*, Springer-Verlag, Berlin, special issue, Dec. 2000.

[3] G. W. Wornell, "Wavelet-Based Representations for the 1/*f* Family of Fractal Processes", *Proc. IEEE*, Vol. 81, No. 10, pp. 1428-1450, Oct. 1993.

[4] G. Evangelista, S. Cavaliere, "Frequency-Warped filter Banks and Wavelet Transforms: A Discrete Time Approach via Laguerre Expansion", *IEEE Transaction on Signal Processing*, Vol. 46, No 10, Oct. 1998.

[5] G. Evangelista, S. Cavaliere, "Discrete Frequency Warped Wavelets: Theory and Applications", *IEEE Transaction on Signal Processing*, Vol. 46, No 4, Apr. 1998.

[6] I. Daubechies, *Ten Lectures on Wavelets,* SIAM CBMS series, Apr. 1992.

[7] M. Vetterli, J. Kovacevic, *Wavelets and Subband Coding*. Prentice Hall PTR, 1995

[8] G. Evangelista, "Comb and Multiplexed Wavelet Transforms and Their Applications to Signal Processing", *IEEE Trans. on Signal Processing*, vol. 42, no. 2, pp. 292-303, Feb. 1994.

[9] G. D. T. Schuller, T. Karp, "Modulated Filter Banks with Arbitrary System Delay: Efficient Implementations and the Time-Varying Case.", *IEEE Transaction on Signal Processing*, Vol. 48, No 63, March 2000.

[10] G. Evangelista , P. Polotti "Analysis and Synthesis of Pseudo-Periodic 1/*f*-like Noise by means of Multiband Wavelets", *Proc. of XII CIM,* pp.35-38, Sept. 1998.

[11] P. Polotti, G. Evangelista, "Time-Spectral Modeling of Sounds by Means of Harmonic-Band Wavelets", *Proceedings of the ICMC 2000*, Aug. 2000.

[12] P. Polotti, G. Evangelista "Sound Modeling by means of Harmonic-Band Wavelets: New Results and Experiments", *Proc. of XIII CIM,* pp.43-46, Sept. 2000.