MULTIRESOLUTION SINUSOIDAL/STOCHASTIC MODEL FOR VOICED-SOUNDS

Pietro Polotti, Gianpaolo Evangelista

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

ABSTRACT

The goal of this paper is to introduce a complete analysis/resynthesis method for the stationary part of voiced-sounds. The method is based on a new class of wavelets, the Harmonic-Band Wavelets (HBWT). Wavelets have been widely employed in signal processing [1, 2]. In the context of sound processing they provided very interesting results in their first harmonic version: the Pitch Synchronous Wavelets Transform (PSWT) [3]. We introduced the Harmonic-Band Wavelets in a previous edition of the DAFx [4]. The HBWT, with respect to the PSWT allows one to manipulate the analysis coefficients of each harmonic independently. Furthermore one is able to group the analysis coefficients according to a finer subdivision of the spectrum of each harmonic, due to the multiresolution analysis of the wavelets. This allows one to separate the deterministic components of voiced sounds, corresponding to the harmonic peaks, from the noisy/stochastic components. A first result was the development of a parametric representation of the HBWT analysis coefficients corresponding to the stochastic components [5, 7]. In this paper we present the results concerning a parametric representation of the HBWT analysis coefficients of the deterministic components. The method recalls the sinusoidal models, where one models time-varying amplitudes and time varying phases [8, 9]. This method provides a new interesting technique for sound synthesis and sound processing, integrating a parametric representation of both the deterministic and the stochastic components of sounds. At the same time it can be seen as a tool for a parametric representation of sound and data compression.

1. INTRODUCTION

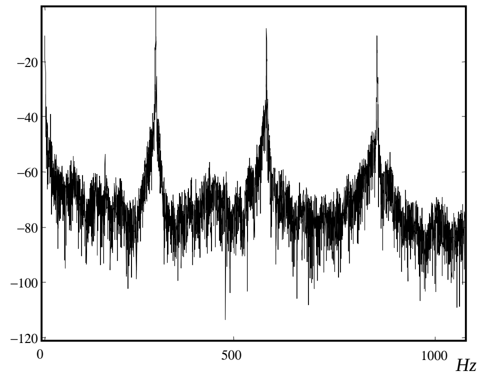
In a previous DAFx paper we defined the HBWT orthogonal and complete set of functions and the pseudo-periodic 1/*f*-like spectral model for voiced sounds [4]. This model allows one to deal with the deterministic and the stochastic components of sound separately. Subsequently we developed a method for the resynthesis of the stochastic components of voiced sounds based on a parametric representation of the analysis HBWT coefficients corresponding to these components [5, 7]. The resynthesis parameters control the time varying variance of white noise coefficients and their spectral shaping.

In this paper we present some new results concerning the parametric representation of the HBWT analysis coefficients corresponding to the deterministic components of voiced sounds.

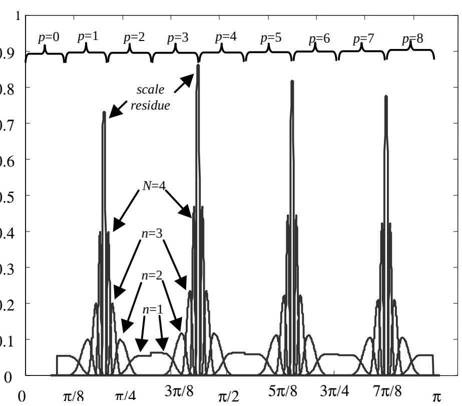
An intuitive preview of the method is given in figure 1, where we show the spectral partition performed by the HBWT analysis. The scale index *n* corresponds to the harmonic-band wavelet coefficients, while the "scale residue" corresponds to the harmonic-band scale coefficients. The first ones are related to the stochastic components of sound and the second ones to the deterministic components of sound. The new method we are going to introduce in section 3 is for the representation of the deterministic components, i.e., for the harmonic-band scale coefficients. Our model is reminiscent of sinusoidal models. However, being based on the MDCT transform, what we model is not the set of partials of the sound itself but a complex version of its harmonic-band scale analysis coefficients.

2. THE PSEUDO-PERIODIC 1/F-LIKE SPECTRAL MODEL. A REVIEW

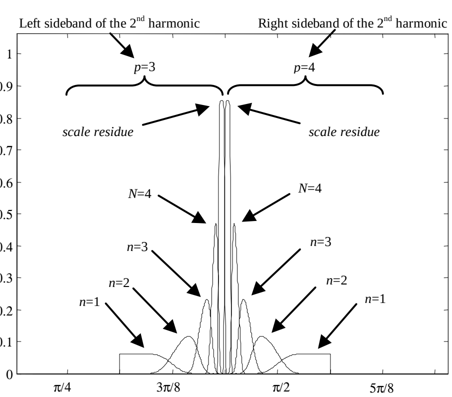
The main idea of the pseudo-periodic 1/*f*-like spectral model is to represent voiced sound spectra as bandshifted approximately 1/*f* spectral segments. The 1/*f*-like sidebands (right and left) of each harmonic take into account not only the harmonic peak but also the information relative to the noise contained in the spectrum band of the harmonic. This noise is due to both microfluctuations with respect to a pure harmonic behavior and to the noise of the physical excitation. The well-suited mathematical tool to manipulate this model is provided by the Harmonic-Band Wavelet Transform HBWT [4, 7].



a)



b)



c)

Figure 1. a) *Real-life violin spectrum. The first harmonics*.
b) *Magnitude Fourier transforms of the HBWT basis set.*
c) *Magnitude Fourier transforms of the HBWT subband decomposition of a single harmonic. Left and right sidebands.*

Figure 2 and figure 3 represents the implementation of a HBWT analysis and synthesis scheme, respectively. In figure 2 one can see first a *P*-channels filter bank, realized by means of the set of filters $g_{p,0}$. These filters implement a cosine-modulated filter-bank and are given by:

$$g_{p,r}(l) = g_{p,0}(l - rP) \;, \qquad p = 0,....,P-1; \quad r \in \mathbf{Z} \quad (1)$$

with

$$g_{p,0}(l) = W(l)\cos\left(\frac{2p+1}{2P}\left(l - P + \tfrac{1}{2}\right)\pi - (-1)^p \frac{\pi}{4}\right), \; (2)$$

where $W(l)$ is a $2P$-length window satisfying some technical constraints [7]. The number of channels *P* is tuned to the average pitch of the analyzed sound. The filters separate the *P* sidebands of the $\left\lfloor \frac{P}{2} \right\rfloor$ harmonics represented in the digital spectrum of the sound (see figure 1). Each sideband is then subdivided into *N* spectral subbands by means of a downsampling of order *P* and an ordinary wavelet analysis.

The *N* subbands are in a dyadic relationship. In this way we obtain a natural subdivision of the sideband with a higher frequency resolution in the range where the spectrum varies rapidly and a lower resolution where the spectrum tends to be flat (see Fig. 1a and b). This has a positive effect in terms of coding as well as in terms of perceptual significance.

In a following paper [6] we also presented a refined version of the model by introducing the Harmonic-Band Frequency-Warped Wavelet Transform (HBFWWT). The HBFWWT allows a more flexible subdivision of the sidebands. In this refinement the subbands are not any more related to each other by a dyadic law. One can individually select the bandwidth of each subband. This allows us to optimize in an arbitrary way the spectral subband subdivision of each sideband.

Figure 2 also shows the extraction of the parameters describing the behavior of the HBWT analysis coefficients. In previous papers [4, 7] we showed that the coefficients, coming from the output of the filters $k_{n,0}(l)$ of figure 2, can be efficiently modeled in terms of energy scaled and filtered white noise coefficients. This is the main result concerning the stochastic part of our model.

From the short-time energy analysis of the analysis coefficients we extract an amplitude envelope for each subband by means of a simple polynomial interpolation. The unitary variance white noise resynthesis coefficients are energy-scaled by means of these envelopes.

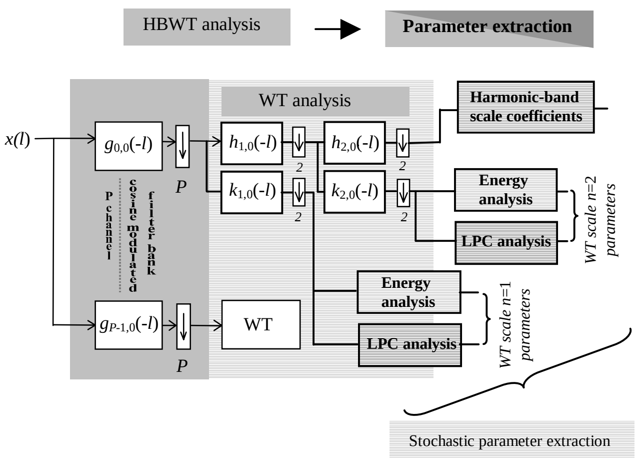


Figure 2. *Harmonic-Band Wavelet (HBWT) analysis scheme and resynthesis parameter estimation. The* $g_{p,0}(l)$ *represent the impulse responses of the filters implementing a MDCT. The* $h_{n,0}(l)$ *and* $k_{n,0}(l)$ *represent the impulse responses of pairs of QMF filters implementing a wavelet transform.*

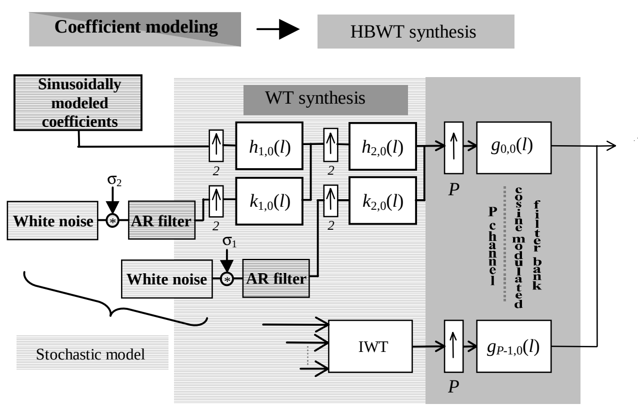


Figure 3 *Harmonic-Band Wavelet (HBWT) resynthesis scheme*

As a second step we model the little but not zero autocorrelation of the HBWT analysis coefficients. This autocorrelation is important from a perceptual point of view. In order to reproduce it we perform an LPC analysis of the HBWT analysis coefficients of each subband of all the harmonic sidebands (for each subband n of al the P channels). The resulting AR filters are employed to shape the spectra of the white noise resynthesis coefficients. The resynthesis coefficients are then given by energy scaled white noise samples, whose spectra are shaped by AR filters derived by LPC-analysis (see figure 3).

What we still need is to provide a model for the output of the last filter $h_{N,0}(l)$ for each channel p, i.e., to define a model for the HBWT analysis coefficients corresponding to the deterministic harmonic part of sounds.

3. SINUSOIDAL MODEL OF THE HARMONIC-BAND SCALE COEFFICIENTS

In order to model the Harmonic-Band (HB) scale coefficients we resort to a complexification of pairs of adjacent channels, corresponding to the scale residue of two sidebands of one harmonic (see fig. 1). In other words, we consider the set of coefficients:

$$A_r[m] = a_{N,2r-1}[m] + j a_{N,2r}[m] \qquad (3)$$

where the $a_{N,2r-1}[m]$ and the $a_{N,2r}[m]$ are the HB scale-N coefficients of the left and right sideband of the $r=(p+1)/2^{th}$ harmonic, respectively, i.e., the output of the $h_{N,0}(l)$ filters of the p^{th} and the $(p+1)^{th}$ channels, respectively, with $p=2r-1$ (see scheme of figure 2).

We write $A_r[m]$ in polar form:

$$A_r[m] = \left|A_r[m]\right| e^{j \arctan\left[a_{n,2r}[m] / a_{n,2r-1}[m] \right]} \qquad (4)$$

The experimental results (see figure 4) show that the HB scale coefficients form smooth and slowly oscillating curves. As a consequence we can efficiently approximate the amplitude $\left|A_r[m]\right|$ and the phase $\phi_r[m] = \arctan\left[a_{n,2r}[m] / a_{n,2r-1}[m] \right]$ by means of linear splines. The results in the case of a violin are shown in figure 5 and 6, respectively. We considered a two-level HBWT analysis of a violin sound of length 201064 and pitch P=150, corresponding to 344 HB scale coefficients. We employed splines of order 2 with 9 knots as interpolating functions for the amplitudes and with 11 knots for the phases. The amplitudes of the $A[m]$ are nothing but the time envelopes of the harmonic partial downsampled by a factor $P2^N$. The phases represent the slow quasi-sinusoidal variation of the coefficients due to the difference between the central frequencies of the harmonic scale filters and the corresponding harmonics themselves. The resynthesis coefficients modeled by means of spline-interpolation provide high quality results from an acoustical point of view. Differences between the original and the synthetic sounds are hardly perceivable.

The transients are perfectly reconstructed from the original analysis coefficients.

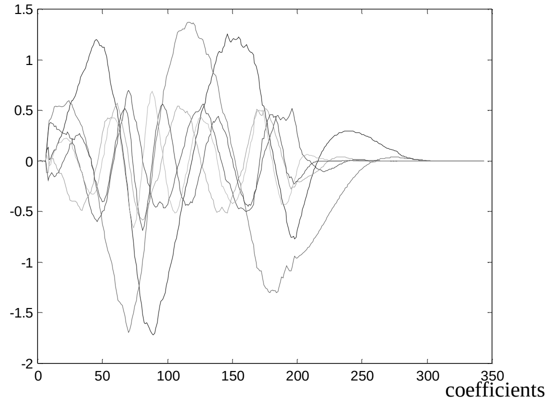


Figure 4. *HB scale analysis coefficients of a violin sound at 294 Hz, 3 harmonics, i.e., 6 analysis channels.*

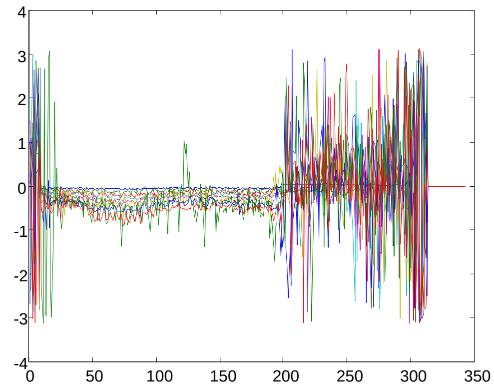


Figure 7 *Second derivative of the harmonic band coefficient phase*

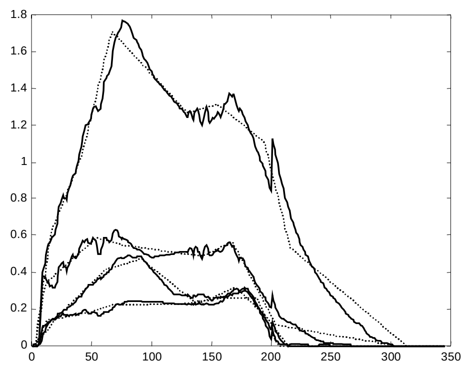


Figure 5. *First 4 harmonics complex coefficient amplitude interpolation*

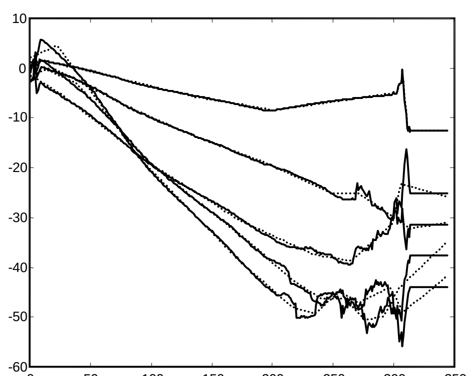


Figure 6 *First 4 harmonics complex coefficient phase interpolation*

A very interesting byproduct of the analysis of the harmonic-band scale coefficients is that the second derivatives of the $\phi_r[m]$ provide a very efficient transient detector (see figure 7). Where the sound is stationary, the absolute value of second derivative is less than 1. This becomes more than three times larger where the transients occur. We employed this result in order to define automatically the borders of the attack and decay transients.
Beside the case of the violin, very good experimental results were obtained with other musical instruments: a bassoon, a clarinet, a trumpet, an oboe and a cello. All of them gave equally satisfying results.
As further developments of our work we need a definition of a model for the transients. A good starting point could be the work of Verma and Meng [10], to be integrated with our wavelets techniques.
Also a pitch synchrounous version is already at work. The goal is to maintain a PR structure, being able to follow the slight pitch deviation or the vibrato of a single musical tone. This means to build a *P*-channel filterbank where one can change the number of channels *P* at each period.
More generally our model is still missing a sufficient flexibility in the design of the *P*-channel filterbank. In other words it lacks the possibility of subdividing the whole frequency range in an arbitrary way, in order to deal with non-harmonic or polyphonic sounds. This could be obtained by giving up the PR constraints of the whole system or by means of some generalized technique of frequency warping, at the expense of an increased computational complexity. We would then still exploit the wavelet spectral tiling in order to model the noise and the peaks of the partials. The latter approach is already at study by one of the authors.

4. CONCLUSIONS

We defined a method for sound synthesis, which allows us to control and reproduce with high fidelity the stationary part of real life voiced sounds by means of a restricted number of parameters.

This method is a sort of additive synthesis where one synthesizes and adds deterministic components and stochastic components separately.

This method can be seen as is a part of a wider system for a complete Structured Audio Representation. Different kinds of musical applications can be devised.

5. REFERENCES

[1] Mallat, S. 1997. *A Wavelet Tour of Signal Processing*. Academic Press.

[2] Rioul, O. and M. Vetterli. 1991. "Wavelets and signal processing", *IEEE Signal Processing Magazine*, Volume: 84, Pages: 14 –38

[3] Evangelista, G. 1993. “Pitch Synchronous Wavelet Representations of Speech and Music Signals,” *IEEE Trans. on Signal Processing*, special issue on Wavelets and Signal Processing, vol. 41, no.12, pp. 3313-3330.

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

[5] Polotti, P. and G. Evangelista. 2000. "Time-Spectral Modeling of Sounds by Means of Harmonic-Band Wavelets", *Proceedings of the ICMC 2000*, pp. 388-391, Berlin, Germany.

[6] Polotti, P. and G. Evangelista. 2000. "Harmonic-band wavelet coefficient modeling for pseudo-periodic sounds processing", *DAFx-00 Proceedings*, pp. 103-108, Verona, Italy.

[7] Polotti, P. and G. Evangelista. 2001. "Analysis and Synthesis of Pseudo-Periodic 1/*f*-like Noise by means of Wavelets with Applications to Digital Audio", *EURASIP Journal on Applied Signal Processing*, Hindawi Publishing Corporation, pp. 1-14.

[8] Serra, X. and J. O. Smith. 1990. "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition.", *Computer Music Journal* 14(4): pp. 14-24.

[9] Serra, X. "Musical sound modeling with sinusoids plus noise", in *Musical Signal Processing*, A. Piccialli, G. De Poli, C. Roads and S. T. Pope. Swets & Zeitlinger, Amsterdam, Holland.

[10] Verma T.S., T. H. Y. Meng, 2000, "Extending Spectral Modeling Synthesis with Transient Modeling Synthesis", *Computer Music Journal* 24(2), pp 47-59, Summer 2000.