Sound Modeling by means of H rmonic-Band Wavelets: New Results and Experiments

Pietro Polotti, Gianpaolo Evangelista

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

Abstract

Musical signals contain both deterministic and stochastic components. The deterministic part provides the pitch and the global timbre of a sound; it is in a sense the fundamental structure of the sound. The stochastic part contains the "life of a sound", that is all the micro-fluctuations with respect to an electronic-like/non-evolving sound and noises due to the physical excitation system. These micro-fluctuations with respect to a pure harmonic behavior can be reconstructed from the power spectrum. A well-suited analysis and resynthesis tool of voiced sound spectra was introduced in [5] and [6], i.e., the Harmonic-Band Wavelet Transforms (HBWT).

The most attractive feature of the HBWT model is that resynthesis coefficients can be substituted in first approximation by white noise with proper scale-dependent energies [4]. At a more refined level one must take into account the little but non-zero correlation of the HBWT analysis coefficients of voiced sounds. This requires a pre-filtering of the resynthesis coefficients by means of AR filters. Furthermore an elementary waveform (wfs) model is employed for modeling the physical excitation system.

This method can be seen both as a musical tool for sound synthesis able to provide synthetic sounds with a natural timbre dynamic and as a compression technique.

1. Introduction

One of the most challenging aspects of sound analysis and representation is the definition of a good model for the noisy part of sounds. In other words we need a good representation of those components of sound whose spectra lie out of the frequency support of the partials.

In [4] we proposed a model for the particular case of voiced sounds, i.e., sounds with a harmonic spectrum, based on the Harmonic-Band Wavelet Transforms (HBWT). Thanks to the mathematical properties of the HBWT, the synthesis of signals with pseudo-periodic $1/f$ –like power spectra is straightforward and these spectra are very good approximations of those of real-life voiced sounds. In that model the only thing we needed was to control the energies of white noise coefficients, according to very few parameters derived from the analysis of real sounds.

In a more detailed perspective, the HWBT analysis reveals the existence of a little but not zero correlation between the coefficients. An AR analysis and resynthesis model, employing white noise as AR filters excitation and reproducing the above-mentioned loose correlation can substitute the trivial white noise coefficient model. As a further refinement of the technique we take into account scale-dependent time evolution of the resynthesis parameters. Also, our model cannot reproduce the excitation system of some instruments. In these cases we consider a second type of noise model, based on elementary waveforms (wfs) techniques [8].

The model does not fit the harmonic components. We preserve the restricted set of analysis wavelet coefficients corresponding to the narrow bands of the harmonics in order to have perfect reconstruction data for the deterministic part, i.e., the harmonics amplitude and their time envelopes. The attacks and decays are preserved as well.

The experimental results concerning all these refinements of the HBWT analysis and synthesis technique are the subject of this paper. In section 2 we make a short review of ordinary wavelets and Harmonic-Band Wavelets. In section 3 we review the pseudo-periodic $1/f$ model as developed in [4]. In section 4 and 5 we present the new developments of the synthesis method from the methodological and experimental points of view respectively. In section 6 we draw our conclusions.

2. Wavelets and Harmonic-Band Wavelets: a review

The wavelet transform provides a graded time-frequency representation of digital signals. In the particular case of audio signals, wavelets perform a time and frequency domain subdivision imitating the human perceptive system, i.e. a logarithmic tiling of the time-frequency plane (see Fig. 1-a). In other words we have a more detailed information in the low frequency area but a coarser sampling rate while a higher sampling rate but a coarser frequency resolution in the high frequency area.

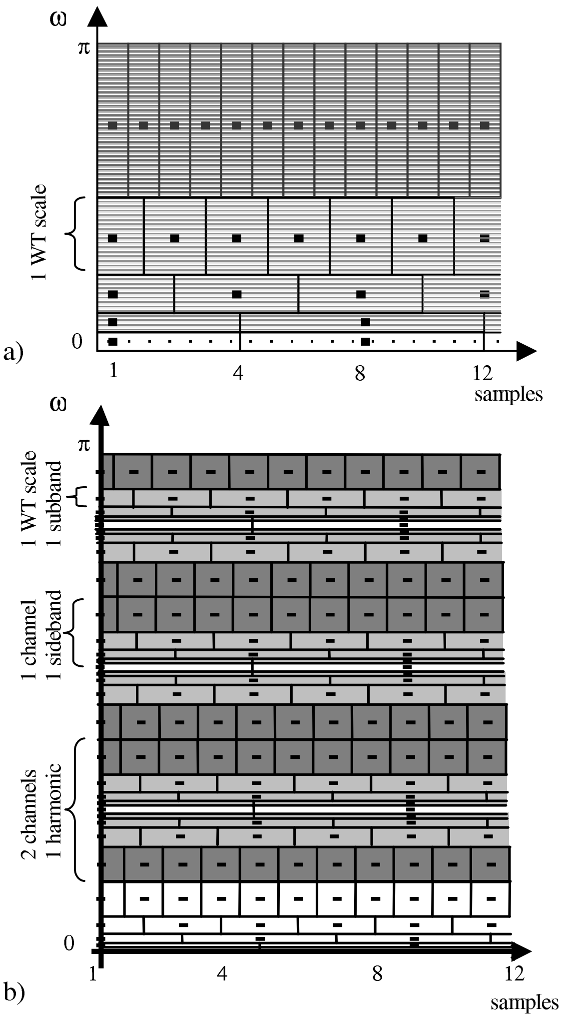

Fig. 1 Time-Fequency Plane Tesselation:
a) Ordinary Wavelets. Each dot corresponds to a wavelet coefficient. At each scale the sampling rate is divided by two.
b) HBWT. The tesselation is a frequency periodized version of a. The number of periods corresponds to the number of BP filters (that is, the number of channels) trapping each sideband of the harmonics.

The HBWT realizes a periodic version of the frequency domain subdivision of ordinary wavelets (see Fig. 1-b). This is obtained by means of the modulation and demodulation scheme described in [6]. With respect to ordinary wavelets, the HBWT provides a much more meaningful representation of voiced sounds. We can tune the frequency domain subdivision to the pitch of any given voiced sound, changing the number of channels of the HBWT filter bank (see [4]). In the ordinary wavelet representation the higher scales (corresponding to the low frequencies) represent the slow changes of a signal with respect to the "average" of the signal (0 in the case of an audio signal), i.e., with respect to a constant. On the other hand, lower scales (high frequencies) represent the changes with respect to the local mean at different rates. In the HBWT representation of voiced sounds the "local mean" is the average period, while the lower scales (the bands away from the harmonics) contain the information concerning the fluctuations with respect to the average period at different rates. In this way we are able to separate the harmonic part of voiced sounds from the different noisy components containing the "dynamics" of the sound.

3. The Pseudo-Periodic $1/f$ –like Model

At a previous CIM meeting we presented a paper about the pseudo-periodic $1/f$ -like noise analysis and synthesis [4]. The starting point was the experimental evidence, revealing an approximate pseudo-periodic $1/f$ behavior of the spectra of voiced sounds in music. The leading idea was to adapt the spectrum of a synthesized pseudo-periodic $1/f$ -like signal to that of a real life sound (see Fig. 2 and Fig. 3). The synthesis process is controlled by means of a very restricted set of parameters, defining the $1/f$ shape of each harmonic of the synthetic spectrum. The analysis tool necessary for the extraction of the resynthesis parameters as well as the synthesis tool is provided by a HBWT filter bank and its inverse, respectively. This model has the advantage to be extremely concise. The lower limit of one parameter per channel of the analysis and synthesis scheme introduced in [4] is an extremely good result from the point of view of data compression. Actually some refinements are necessary in order to reach a good quality in sound reproduction at the cost of an increase of the number of parameters.

4. The new Analysis and Synthesis Method

The new method we are going to describe in detail is a refinement of the pseudo-periodic $1/f$ -like spectral model [4]. The whole model is limited to the steady part of sound and is particularly well suited for long, sustained sounds. In our analysis and synthesis scheme the transients, that is the sound attack and decay, are just reconstructed from the complete set of HBWT coefficients resulting from their HBWT decomposition. A pitch detector based on the estimation of the autocorrelation is adopted in order to define the extension of the transient. Only the portion of sound where a steady pitch is detected is processed by means of the HBWT filter banks.

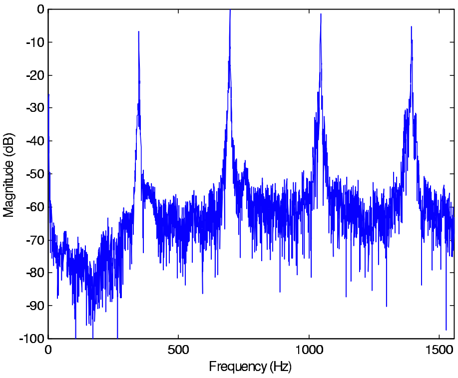

Fig. 2 Spectrum of a real-life trumpet

We subdivide the remaining portion of the time-frequency domain in different parts The first part corresponds to the deterministic components of the sound (see Fig. 1-b the white subbands). The analysis HBWT coefficients are preserved in order to obtain a perfect reconstruction of the harmonics and their time-envelopes. The spectrum portion close to the harmonics contains the micro-fluctuations with respect to pure periodicity (see Fig. 1-b the light gray subbands). Its behavior is approximately $1/f$. We estimate independent parameters controlling the energy of each HBWT subband separately. This allows us to find a better approximation of the spectrum. An even better spectral "design" can be obtained by means of a convenient and flexible analysis and resynthesis tool, recently introduced by one of the authors, i.e., the Arbitrary Bandwidth Wavelet Transforms [9]. These new wavelet transforms allow arbitrarily detailed, signal adaptable spectra subdivisions. This is necessary when the spectrum shape is not $1/f$ –like. We employ a mean square error criterion in order to find the more suitable piecewise approximation of the spectrum. Furthermore, an LPC analysis is applied to the HBWT analysis coefficients. The AR filters so obtained are used to color the white noise used as input to the resynthesis filter bank, in order to reproduce the time-correlation in the subbands.
The third spectrum portion includes the first subbands of the HBWT decomposition (see Fig. 1-b the dark gray subbands). According to the analysis results is possible to see how these subbands, which lie far away from the harmonics, contain the most significant and non-masked information concerning the additional noise due to the excitation systems.

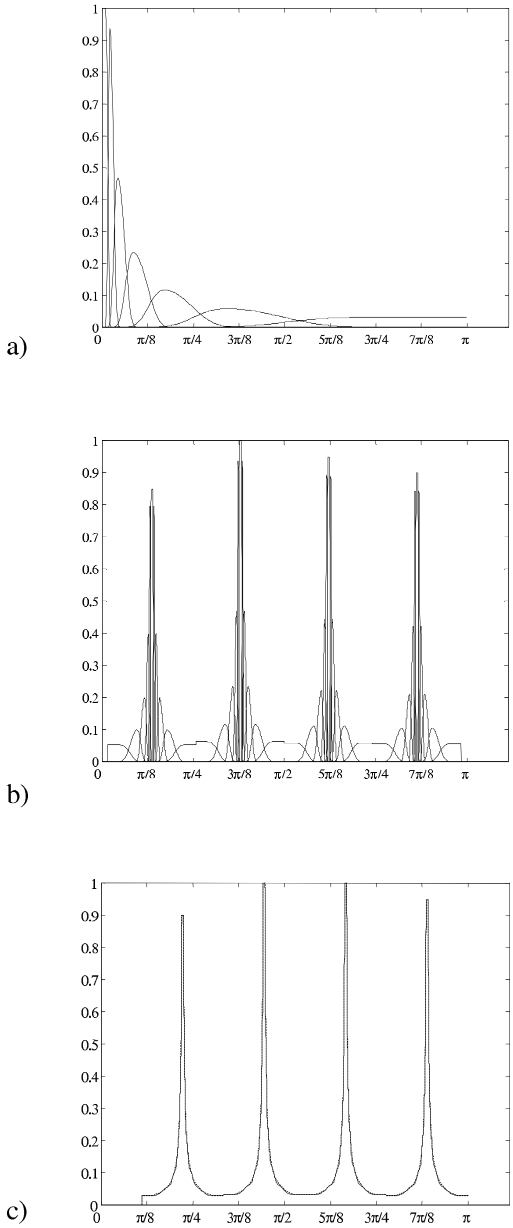

Fig. 3 1/f spectral model:

a) Ordinary Wavelets. The wavelet subbands with proper energies reproduce a $1/f$ spectral behavior.

b) HBWT. By means of a frequency periodized version of a) we are able to reproduce a pseudo-periodic $1/f$ -like spectral behavior. Each sideband has a $1/f$ behavior and is subdivided in wavelet subbands.

c) The resulting synthetic spectrum. This spectrum can be adapted to the spectrum of the sound we want to resynthesize. In this way both the harmonic components and the noisy components of sounds are reproduced.

In these frequency subdomains we look for the samples of elementary waveforms (wfs) which occur during the instrument continuos excitation. These noises include, for instance, saliva gurgling in wind instruments or bow noise in string instruments. These wfs are juxtaposed according to a statistical sampling and amplitude scaled according to the signal analysis itself. A more refined method should include the construction of a sort of codebook, i.e., a sample-case of elementary waveforms for each musical instrument.

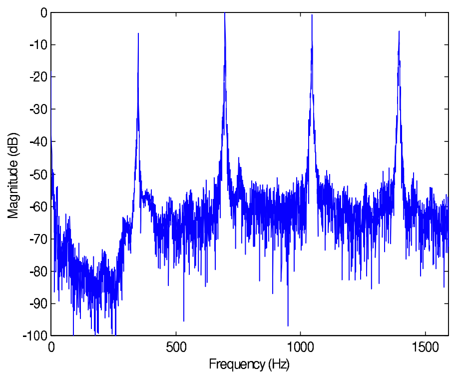

Fig. 4 Spectrum of a synthetic trumpet.

5. Experimental Results

The experimental results change significantly according to the instrument we analyze and resynthesize. The first step is the choice of the wavelet scale at which we stop the analysis. This defines which ratio of the spectrum we resynthesize by means of noisy coefficients and which ratio is perfectly reconstructed preserving the analysis coefficients. Normally 4 or 5 scales are the maximum values admitted in order to preserve the main time characteristic of the sound as the harmonics and their time envelope. The second step is to define the extension of the transients; the length of the attack and of the decay varies a lot according to the instrument, the pitch and the stabilization speed of the sound. Our algorithm gives good results for all the sounds we considered: a flute, an oboe, a clarinet, a bassoon, a trumpet, a french horn and a trombone. The AR filters employed are of the 10^{th} order for the second subband. The order diminishes with the order of the subbands. We performed a perfect reconstruction of each subband separately in order to compare them with the synthetic ones. The results from an acoustical point of view are very good.

A short-time version was also implemented, applying very short rectangular windows (20 coefficients) to the analysis coefficients. In this way we preserve the energy time envelopes of the subbands.

Finally we considered a dozen of elementary waveforms, extracted from the perfect reconstructed first subband, in order to synthesize the first subband. We mounted them in random order with amplitude following the short time energy analysis results.

Conclusions

We introduced an articulated method to deal with the different components of voiced sounds, focusing our attention on their noisy components, so important to maintain a real-life "color" in sounds. We think that our method provides a convincing noise model for voiced sounds in music, with a solid mathematical background [6] and very good experimental results.

The main improvement of the method, on which we will work in the immediate future, will be a pitch synchronous version, i.e., a time varying version of the filter banks. This would free the method from the limitations of a fixed number of channels, which restrict the set of sounds that we can analyze to those with a very well defined and stable pitch.

References

[1] 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.

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

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

[4] G. Evangelista, P. Polotti, “Analysis and synthesis of pseudoperiodic 1/*f*-like noise by means of multiband wavelets”, Proc. of XII CIM, pp.35-38, Sept. 1998.

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

[6] 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 Proc., Springer-Verlag, Berlin, special issue, Dec. 2000.

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

[8] C. d'Alessandro, G. Richard, "Random Wavelet Representation of Unvoiced Speech"

[9] 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.