

TIME-SPECTRAL MODELING OF SOUNDS BY MEANS OF HARMONIC-BAND WAVELETS

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 spectrum of any voiced sound in speech and music shows harmonic peaks whose sidebands behave approximately as hyperbolic segments. These sidebands contain the frequency representation of the stochastic micro-fluctuations of the signal with respect to a pure periodic behavior. This approximately periodic behavior induces us to think about voiced sounds as pseudo-periodic signals.

Thanks to the particular time-frequency tiling, the Harmonic-Band Wavelet Transform provides an ideal tool to deal with such a spectral behavior from both the analysis and synthesis points of view. Their peculiarity is to reproduce the ordinary wavelet spectral behavior in a periodic way. These periodic spectra can be tuned and reshaped according to the pitch and spectrum of any pseudo-periodic signal. Our technique, i.e., it can reconstruct the stochastic component of pseudo-periodic signals, so important in order to perceive sounds as natural ones.

The model, combined with a deterministic synthesis of the harmonic part of the sound, also based on wavelets, forms a powerful synthesis method that preserves the naturalness of sounds with respect to sinusoidal models.

1 Introduction

In a previous paper [4] we introduced the $1/f$ -like pseudo-periodic model in order to provide a synthesis model for pseudo-periodic voiced sounds in music. In this paper we present some extensions of that model where we obtain a more accurate spectral representation and we take into account some time dependent components which can be better reproduced by means of shot noise models [7]. In the $1/f$ pseudo-periodic model we used white noise as synthesis coefficients and we approximated each

hyperbolic-like segment by means of only one parameter. This parameter controlled the spectral shape of the whole sideband. We now refine our technique by modeling separately each wavelet subband of each harmonic sideband. This provides a more detailed approximation of the spectrum. A further significant improvement is achieved considering the loose but not null correlation of the Harmonic-Band Wavelets analysis coefficients. Speech and music voiced sounds expansion coefficients are well approximated by an AR analysis and resynthesis model, employing white noise as excitation and reproducing the above-mentioned loose correlation. With these coefficients, the Harmonic Wavelets synthesis scheme provides a good model for reconstructing the sidebands of the harmonics by means of a still very restricted set of parameters.

The model can be improved by employing the Arbitrary Bandwidth Wavelet Transforms [10]. These transforms release our technique from the strict frequency domain subdivision by negative power of two of ordinary wavelets. Applying the arbitrary bandwidth wavelet to pseudo-periodic signals, we can realize higher resolution spectral modeling of each sideband.

Furthermore we have to deal with time-localized noisy events, such as blowing impulses in wind instruments. Shot-like noises [7] are time localized energy peaks. By a well-suited threshold-based combined pitch and peak detection is then possible to record the impulses occurrence times and to resynthesize them by means of a limited collection of properly overlapped elementary waveforms samples.

The paper is organized in the following way: in section 2 and 3 we provide a short review of the Harmonic-Band Wavelet transform (HWBT) and of the $1/f$ pseudo-periodic model, respectively. In section 4 we detail the analysis and synthesis techniques. In Section 5 we illustrate some experimental results. In section 6 we draw our conclusions.

2 Harmonic-Band Wavelets: a review

Harmonic-Band Wavelets are a generalization of the Multiplexed Wavelets (MWT) [3]. The MWT, the HBWT allows one to control independently the intensity and shape of each harmonic side-band of the spectrum, which are consolidated in single basis elements in the MWT. The computation of the transform consists in band-pass filtering and critically downsampling each sideband. The resulting signals are then separately wavelet transformed (see Fig. 1).

In [5] we introduced the set of functions

$$\left\{g_{q,r}(t)\right\}_{q=0,1,\dots,k \in \mathbf{Z}}, \quad \text{with}$$

$$g_{q,r}(t) = g_{q,0}(t - rT_P)$$

$$g_{q,0}(t) = \frac{1}{\sqrt{T_P}} \cos\left(\frac{2q+1}{2T_P} \pi t\right) \text{sinc}\left(\frac{t}{2T_P}\right), \quad (1)$$

which is easily shown to form an orthonormal and complete basis.

Similar sets are available for the discrete case. We chose the Type IV cosine modulated bases:

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

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

where the lowpass prototype impulse response $W(l)$ of length M satisfies some technical conditions [6].

The $\{g_{q,r}(l)\}_{q=0,1,\dots,P,k \in \mathbf{Z}}$ are the bandpass filters, which separate the harmonic band of a pseudo-periodic signal of period P .

The HBWT are then defined in the following way:

$$\xi_{n,m,q}(l) = \sum_r \psi_{n,m}(r) g_{q,r}(l) \quad (3)$$

where $\psi_{n,m}(r)$ are discrete-time ordinary wavelets [2].

3 1/f-like Pseudo-Periodic Model: a review

In this section we briefly review the results related to the 1/f-like pseudo-periodic model. The goal is to reproduce the casual but time-correlated micro-

fluctuations with respect to a strictly periodic behavior of voiced signals in speech and music. These fluctuations play a relevant role in the emulation of naturalness of these sounds. We proved that it is possible to extend Wornell's results in [1] to the 1/f-like pseudo-periodic case. In Theorem 3 Wornell proves that a process

$$x(t) = \sum_{n=-\infty}^{\infty} a_n(m) \psi_{n,m}(t),$$

where $\psi_{n,m}(t)$ form a orthonormal wavelet basis and the $a_n(m)$ are collections of white noise zero-mean coefficients with properly scaled energies, is nearly 1/f, i.e., its time-averaged power spectrum

$$\bar{S}_x(\omega) = \sigma^2 \sum_{n=-\infty}^{\infty} 2^n |\Psi(2^n \omega)|^2$$

satisfies the relations $\frac{\sigma_L^2}{|\omega|^{\gamma}} \leq \bar{S}_x(\omega) \leq \frac{\sigma_U^2}{|\omega|^{\gamma}}$ for some

$$0 < \sigma_L^2 \leq \sigma_U^2 < \infty.$$

Similarly we showed that a process

$$x(t) = \sum_{q=0}^{\infty} \sum_{n=1}^N \sum_{m=-\infty}^{\infty} a_q^n(m) \xi_{n,m,q}(t)$$

where the $\xi_{n,m,q}(t)$ form a HBWT base and the $a_q^n(m)$ are collections of white noise zero-mean coefficients with properly scaled energies, has time-averaged power spectrum

$$\bar{S}(\omega) = \frac{1}{T_P} \sum_{q=0}^{\infty} \sigma_q^2 |G_{q,0}(\omega)|^2 \left(\sum_{n=1}^N 2^{Nq} |\Psi_{n,0}(\omega T_P)|^2 + 2^{Nq} |\Phi_{N,0}(\omega T_P)|^2 \right)$$

where the $|G_{q,0}(\omega)|^2 |\Psi_{n,0}(\omega T_P)|^2 = |\Xi_{q,n,0}(\omega)|^2$ and $|G_{q,0}(\omega)|^2 |\Phi_{N,0}(\omega T_P)|^2 = |\Theta_{q,n,0}(\omega)|^2$ are the Fourier Transforms of the HBWT and the respective Harmonic Scale Functions. This spectrum is approximately 1/f near each harmonic $k \frac{2\pi}{T_P}$, with $k = \left\lfloor \frac{q+1}{2} \right\rfloor$, i.e.,

$$\frac{\sigma_{L,q}^2}{\left|\omega - k \frac{2\pi}{T_P}\right|^{\gamma_q}} \leq \bar{S}_N(\omega) \leq \frac{\sigma_{U,q}^2}{\left|\omega - k \frac{2\pi}{T_P}\right|^{\gamma_q}}$$

for some $0 < \sigma_{L,q}^2 \leq \sigma_{U,q}^2 < \infty$ (see Fig. 2b).

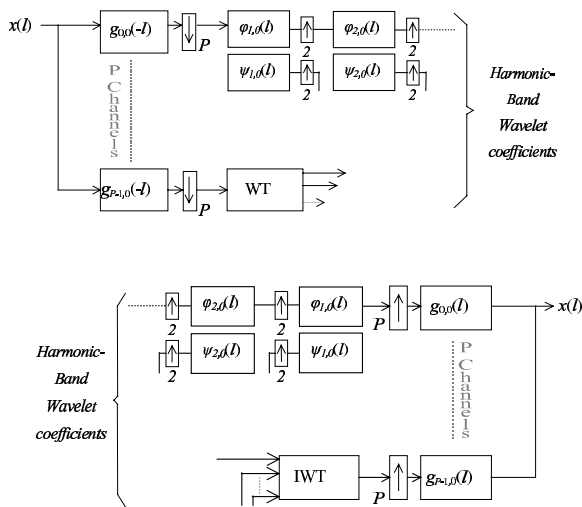


Fig. 1 HBWT analysis and synthesis scheme.

4 The analysis and synthesis model

In the analysis structure of the Discrete Harmonic-Band Wavelet Transform, the signal is sent to a P channel filter bank, where P is the discrete pitch of the analyzed signal. Each output corresponds to a single sideband of a harmonic. The resulting signals are critically downsampled and then Wavelet transformed. Signal reconstruction is achieved by separately inverse Wavelet transforming the Harmonic-Band Wavelet coefficients and passing these sequences through the inverse P channel filter bank.

The synthesis coefficients are produced according to the analysis results. The first step consists in measuring and recording the spectral energies of each wavelet filter output. Afterwards we perform an LPC analysis of each subband HBWT coefficients. We obtain a set of AR filters able to take into account the analysis coefficients loose correlation.

In the synthesis process we model the expansion coefficients by filtering white noise by means of these AR filters. Then the energy of each subband of each harmonic sideband is rescaled according to the spectral energies extracted from the analysis results. In practice, wavelet analysis is performed up to the 4th scale. The residual scale functions coefficients are left unchanged, in order to obtain a perfect reconstruction of the deterministic harmonic part of the sound. With respect to the simple $1/f$ -like model the amount of parameters increases. However we have a robust compression ratio (approximately 1/40)

providing on the other side a very high quality method for coding voiced audio signals.

In some cases, as in brass instruments, the noise of the first subband of the harmonics is not well representable by a white noise source model. Rather, a shot noise or elementary waveform (wfs) ([7], [8]) model seems to be more suited in order to simulate the stochastic part of the sound due to the breath and saliva gurgling, which is pretty significant in the case of brass instruments.

One can also achieve better spectral approximations by employing the recently introduced Arbitrary Band Wavelet Transforms [9], [10]. A significant improvement is obtained in the case of noisier sounds such as single reeds (oboe, bassoon) at a higher computational cost. One has to implement an iterated Laguerre transformation of the DWT channels in order to get rid of the power-of-2 subdivision constraint. In this way the subbands can be freely adapted to a spectrum shape not necessarily $1/f$ -like (see Fig. 2).

A similar frequency warping can be performed in order to deal with non-harmonic sounds. The P -band filter bank is warped to be adaptable to any distribution of the partials.

We apply the model only to the steady part of the sounds, while keeping the transient data.

A lot of improvement can be still realized in terms of computational optimization, adapting the technique to the particular sound that one wants to reproduce or compress. In our model we make the same amount of computation over all the frequency range. Actually not all the harmonics need an LPC analysis. Also, the higher harmonic can be successfully shaped by means of a much more restricted set of parameters, or can be represented by a higher wavelet scale level. In this way it is possible to reduce considerably the amount of resynthesis data.

5 Experimental results

We tested our method on a wide range of instrumental sounds. We used the synthesis bank to reproduce each HBWT subband over all the harmonic range. Then we tested the resynthesis technique on each subband, obtaining very good results from the acoustical point of view. A higher or lower level of approximation can be obtained according to the use of normal DWT banks or of Arbitrary Band Wavelets banks at the cost of a lower or higher computational complexity. In cases of sounds with variable dynamics we can adopt a short-time computation of the energies of the number HBWT expansion coefficients. In this case the number of parameters increases but still remains in a convenient range.

As mentioned in the above, in the case of brass instruments higher fidelity is obtained by adopting also a

elementary wfs resynthesis (see Fig. 3). We considered the cases of a trumpet and of a trombone; in both cases it is easy to select some elementary waveforms according to a good pitch detector. In fact, the waveform of the first subband essentially contains the part of the periodic breath impulses produced by the player.

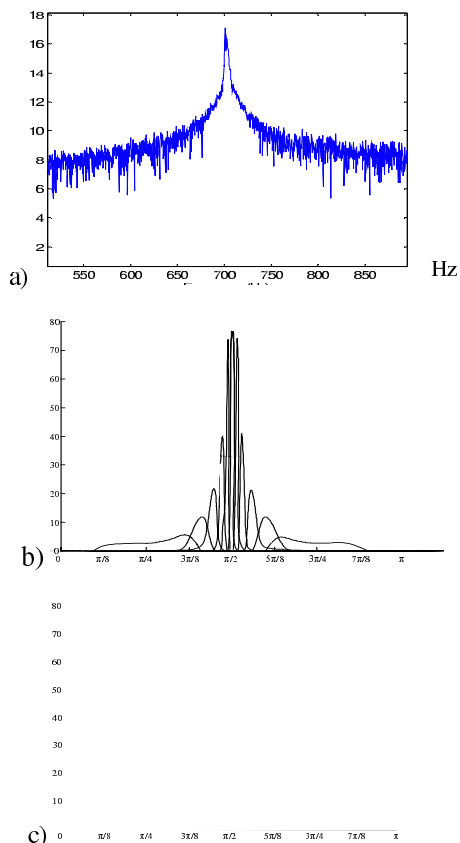


Fig. 2 a) Spectrum of one harmonic of a real clarinet
 b) Spectrum of one harmonic of HBWT.
 c) Spectrum of one harmonic of warped HBWT

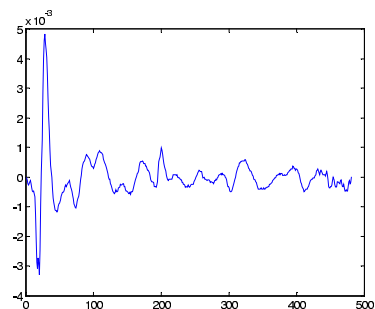


Fig. 3 Example of an elementary waveform extracted from the reconstructed first subband of a trombone with pitch $P=480$ samples. This is one of the wfs employed for the resynthesis.

Also in this case the quality of reproduction is very good compared to the little number of wfs needed to obtain it.

6 Conclusion

In this paper we presented some further developments of a recently introduced analysis and resynthesis technique based on a new family of multiwavelets, i.e., the HBWT. This method concerns voiced sounds in speech and music and shows many interesting features in the context of structured audio.

From the point of view of sound synthesis it solves the problem of sustained tones. Our method is able to render the natural dynamics of the timbre by a relatively restricted set of parameters. The longer is the sound one wants to reproduce the more efficient is our method from the computational point of view.

From an audio data compression point of view our technique guarantees very good rate values for a high quality resynthesis.

References

- [1] G. W. Womell, "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 Processing*, 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 Processing*, Springer-Verlag, Berlin, special issue, Dec. 2000.
- [7] C. d'Alessandro, G. Richard, "Random Wavelet Representation of Unvoiced Speech"
- [8] C. d'Alessandro, J.S. Lienard, "Decomposition of Speech Signal into Short-time Waveforms using Spectral Segmentation",
- [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.
- [10] G. Evangelista, S. Cavaliere, "Discrete Frequency Warped Wavelets: Theory and Applications", *IEEE Transaction on Signal Processing*, Vol. 46, No 4, Apr. 1998.