ADDITIVE SYNTHESIS BASED ON THE CONTINUOUS WAVELET TRANSFORM: A SINUSOIDAL PLUS TRANSIENT MODEL

José R. Beltrán and Fernando Beltrán

Department of Electronic Engineering and Communications
University of Zaragoza, Spain
jrbelbla@unizar.es, beltran@unizar.es

ABSTRACT

In this paper a new algorithm to compute an additive synthesis model of a signal is presented. An analysis based on the Continuous Wavelet Transform (CWT) has been used to extract the time-varying amplitudes and phases of the model. A coarse to fine analysis increases the algorithm efficiency. The computation of the transient analysis is performed using the same algorithm developed for the sinusoidal analysis, setting the proper parameters. A sinusoidal plus transient schema is obtained. Typical sound transformations have been implemented to validate the obtained results.

1. INTRODUCTION

Additive Synthesis is a set of sound synthesis techniques based on the summation of elementary waveforms, called partials, to obtain more complex waveforms. Usually the process of partial extraction is based on the Short-Time Fourier Transform (STFT). The STFT maps the signal into a two-dimensional function of time and frequency. However, the size of the analysis window establishes a compromise in the resolution achievable in both domains [1].

The basic Spectral Modeling Synthesis (SMS) technique [2] models the sounds as the sum of sinusoids plus a residual. The sinusoidal components are extracted from the original signal by means of the Sort-Time Fourier Transform (STFT).

Wavelet Analysis [3] [4] can be considered as a windowing technique with variable-sized regions, which allows the use of long time intervals where more precise low frequency information is required, and shorter regions where the interest relies on high frequency information.

Mathematically, the Fourier Transform represents the process of the Fourier analysis:

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega t} dt \tag{1}$$

Therefore, the mathematical basis of the Fourier Transform are sine waves (the complex exponential can be broken down into real and imaginary sinusoidal components) of infinite duration.

Similarly, the Continuous Wavelet Transform (CWT) is defined as:

$$W_f(a,b) = \int_{-\infty}^{+\infty} f(t) \psi^*_{a,b}(t) dt \tag{2}$$

where * is the complex conjugate and $\psi_{a,b}(t)$ is the mother wavelet scaled by a factor a and dilated by a factor b:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right) \tag{3}$$

Hence, while the Fourier Transform consists of breaking up a signal into sine waves of various frequencies, the CWT consists of breaking up a signal into scaled and shifted versions of the wavelet basis $\psi_{a,b}(t)$.

If we define

$$\psi_a(x) = \frac{1}{\sqrt{a}} \psi \left(\frac{x}{a} \right) \tag{4}$$

equation 2 can be viewed as the inner product of the original signal $f(t)$ with the basis $\psi_a(t-b)$:

$$W_f(a,b) = \langle f, \psi_a(t-b) \rangle \tag{5}$$

or, equivalently, the wavelet transform is the convolution of the function f with the filter whose impulse response is $\tilde{\psi}_a(x)$

$$W_f(a,b) = f * \tilde{\psi}_a(b) \tag{6}$$

where

$$\tilde{\psi}_a(x) = \psi_a(-x) \tag{7}$$

The filter frequency response of the dilated mother wavelet $\psi_a(x)$ is:

$$\hat{\psi}_a(\omega) = \sqrt{a} \hat{\psi}(a\omega) \tag{8}$$

Taking the Fourier Transform on both sides of equation 6 the wavelet transform can be viewed as the filtering of the signal with a bandpass filter whose frequency response is $\hat{\psi}_a(\omega)$.

The real and imaginary parts of the wavelet transform can be obtained if a complex wavelet is used:

$$\begin{aligned} R_f(a,b) &= \Re\left(W_f(a,b)\right) \\ I_f(a,b) &= \Im\left(W_f(a,b)\right) \end{aligned} \tag{9}$$

Then, the modulus and phase of the complex wavelet transform are:

$$M_f(a,b) = \sqrt{R_f^2(a,b) + I_f^2(a,b)} \tag{10}$$

$$\Psi_f(a,b) = \arctan\left(\frac{I_f(a,b)}{R_f(a,b)} \right) \tag{11}$$

Equations 10 and 11 are the output of the Complex Continuous Wavelet Transform that is employed to perform the sound analysis in this work.

This paper is divided as follows. In section 2 we present the definition of the wavelet and the analysis parameters we use in the processing stage. In section 3 the sinusoidal analysis and additive resynthesis are presented. The transient analysis and some results are presented in section 4. In section 5 some typical transformations that can be implemented with the proposed model are described. Finally, the main conclusions are presented in section 6.

2. WAVELET DEFINITION

As proposed by Kronland-Martinet, Morlet and Grossmann [5] the analyzing wavelet we are going to use in this work is the complex generalization of the Morlet wavelet, given by the expression:

$$\psi(t) = C' e^{-\frac{t^2}{2}} \left(e^{j\omega_0 t} - e^{-\frac{\omega_0}{2}} \right) \tag{12}$$

This function requires small corrections to ensure that the formal conditions for an analyzing wavelet is satisfied [5]. However, if ω_0 is big enough for that the Fourier transform of Morlet's wavelet $\hat{\psi}(\omega)$ vanishes if $\omega < 0$, these corrections are numerically negligible. In practice taking $\omega_0 > 5$ is enough. In this case the Fourier transform of the complex Morlet's wavelet is:

$$\hat{\psi}(\omega) = C e^{-\frac{(\omega-\omega_0)^2}{2}} \tag{13}$$

C' and C are normalization constants in the time and frequency domain, respectively.

As it will be seen later, we will need to control the frequency resolution of the analysis. A parameter k, which determines the bandwidth of the mother wavelet has been included. So, the final expression of our wavelet in the frequency domain is:

$$\hat{\psi}(\omega) = C e^{-\frac{(\omega-\omega_0)^2}{2k}} \tag{14}$$

As proposed in [3] and [4] a dyadic set of scale factors is employed. This frequency division provides a logarithmic-resolution frequency axis. In our case, we want to be able to analyze each octave in a variable number of divisions D. The set of discrete scales is obtained by:

$$s_j = s_{min} 2^{\frac{j}{D}}, j = 1, \ldots, J \tag{15}$$

If $D = 1$ the spectrum is divided into J octaves. The minimum scale s_{min} is related to the maximum frequency that could be found in the analysis, f_{max}, and f_{max} is related to the sampling rate f_s by the Nyquist criterion, $f_{max} = f_s/2$.

Following equations 8 and 14, and changing the continuous scale factor a by the discrete one s, the scaled version of Morlet's wavelet can be expressed as:

$$\hat{\psi}(\omega) = C_s e^{-\frac{(s\omega-\omega_0)^2}{2k}} \tag{16}$$

From equation 16 it can be seen that the center frequency of the bandpass filter at scales s, located at the maximum of the exponential is:

$$\omega_c = \frac{\omega_0}{s} \tag{17}$$

Finally, we have to fit the filters bandwidth associated with the analyzing wavelet. These filters should be wide enough to cover the whole frequency axis. In figure 1, it is shown that this property is not properly carried out whatever the choice of the number of divisions by scale D and the resolution parameter k we give. In figure 1 we have chosen, arbitrarily, one division per scale, $D = 1$, and a resolution factor $k = 22$. When the filters are narrowed increasing k, it can be seen that some frequencies are not analyzed by the filter bank structure. So, divisions per scale and resolution are not totally independent.

To avoid this problem, a relationship between the number of divisions per scale D and the filters width k has been developed. If Q is defined as $Q = \omega_c/BW$, where BW is the filter bandwidth, it can be shown that:

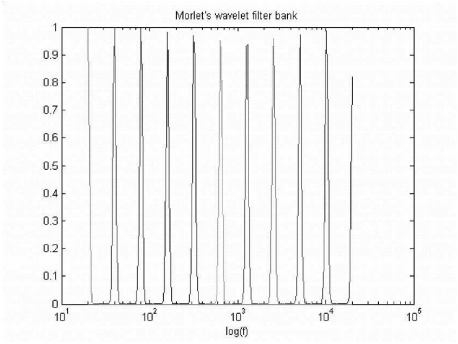


Figure 1: *Morlet's wavelet filter bank with D=1 and k=22. Some frequencies are not covered by the filter bank structure.*

$$k = \frac{1}{\ln(2)} \left(\frac{\omega_0}{2Q} \right)^2 \tag{18}$$

It is possible to obtain the expression relating Q and the filters center frequencies (or the corresponding scale) supposing that the spectrum is covered inside the -3 dB points of the bandpass filters. Then, using equation 17, we have:

$$Q = \frac{\omega_{c,j+1}}{\omega_{c,j+1} - \omega_{c,j}} = \frac{s_{j+1} - s_j}{s_j} \tag{19}$$

where $\omega_{c,j}$ is the center frequency of the bandpass filter at scale j and s_j is the j-th scale (see equation 15).

Equation 19 can be expressed in the particular case of $j = J$. A narrowing parameter q has been included. Then, Q can be expressed by:

$$Q = q \frac{s_J - s_{J-1}}{s_{J-1}} \tag{20}$$

Once we have chosen the number of divisions per scale D and the resolution parameter q, we can compute automatically Q and k using equation 20 and 18, respectively. In figure 2 we can see an example comparing the results of the filter banks obtained with $D = 1$ (figure 2(a)) and $D = 2$ (figure 2(b)), and $q = 1$ in both cases.

Using the expressions presented in this section we have all the necessary mathematical tools to afford the sinusoidal analysis presented in next sections.

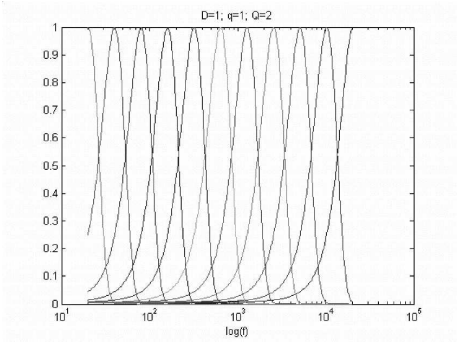
3. SINUSOIDAL ANALYSIS AND RESYNTHESIS

The block diagram of the sinusoidal analysis algorithm presented in this paper is shown in figure 3. It can be described in four steps.

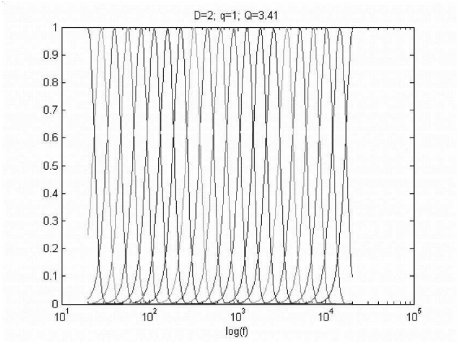
1. To compute the continuos wavelet transform of the input sound. The output of the transform is the CWT modulus and phase as described in equations 10 and 11.
2. To compute the scalegram. The scalegram is the sum over the time dimension of the modulus of the CWT coefficients for each scale.

$$S(a) = \int_{-\infty}^{+\infty} W_f(a,b) db \tag{21}$$

This is a representation of the energy contained in each scale of the analysis or, equivalently, each frequency band.



(a)



(b)

Figure 2: *(a) Filter bank structure of Morlet's wavelet with one division per scale and a resolution parameter q=1. The obtained Q is 2. (b) Filter bank structure of Morlet's wavelet with two divisions per scale and a resolution parameter q=2. The obtained Q is 3.41.*

To obtain a sinusoidal representation of the input signal we need to search the maximum energy spectral bands from the information given by the CWT modulus. A thresholding process, controlled by a user parameter, give us the scales where the most of the energy of the input signal is concentrated. This first and second steps are made at low resolution (for instance, $D = 1$, $q = 1$). This is a coarse analysis of the signal. In figure 4 we can see a plot of a recorded guitar note (figure 4(a)) and the output of the coarse analysis: scalegram (figure 4(b)) and modulus and phase of the CWT (figures 4(d) and 4(f), respectively).

3. To perform a finer analysis over the range of scales that contain some energy. We compute the CWT and the scalegram only at the marked scales. This computation is made over the input sound with high-resolution parameters (for example, $D = 8$ and $q = 1$). Figure 4 shows the scalegram (figure 4(c)), the modulus of the CWT (figure 4(e)) and the phase of the CWT (figure 4(g)) computed at high resolution over the signal in figure 4(a).
4. To search the maximum peaks in the scalegram inside the scale range that allow us to locate the constituent partials of the signal. Its temporal envelopes and the time-dependent phase describe the sound partials. The partials temporal envelopes are described in the modulus of the CWT at the scales corresponding to the obtained peaks. The partials envelopes obtained from the recorded guitar note are shown in figure 5(a). The unwrapped argument or phase of these coefficients provides the instantaneous phase of the partials and are presented in figure 5(b). The first derivative of the instantaneous phase determines the instantaneous frequency of the partials. With this analysis algorithm we have obtained a sinusoidal model of our signal.

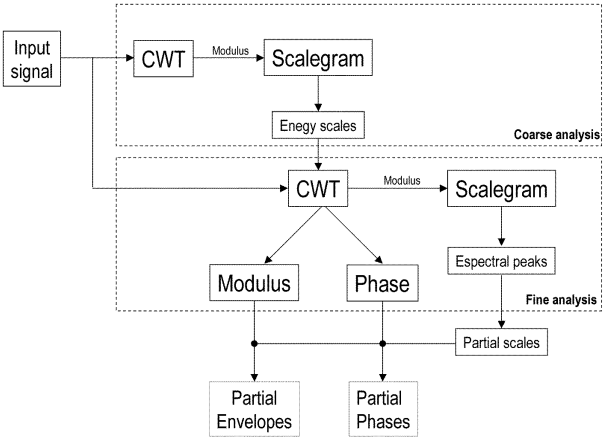


Figure 3: *Sinusoidal analysis algorithm block diagram.*

At the end of the process, having the envelopes and phases of the main partials, the resynthesis is performed using the additive synthesis expression:

$$\hat{s}_s(t) = \sum_{k=1}^{K} a_k(t) \cos \phi_k(t) \tag{22}$$

where $a_k(t)$ is the time depending envelope and $\phi_k(t)$ is the instantaneous phase of the k-th partial given by the continuous wavelet analysis.

The coarse-to-fine analysis presented here avoids the heavy computation of a fine analysis in the whole frequency range. In our approach, a full frequency analysis is only made at low resolution, so the computational effort is diminished. We are only dealing with sounds or phrases of one instrument. Musical programs that cover all the frequency range need a different processing approach and it is out of the scope of this paper.

4. TRANSIENT ANALYSIS AND RESULTS

As in the SMS framework [2] we are going to consider the residual signal. The residual signal is the difference between the original signal and the synthesized sinusoidal one. This residual contains information about transient and noise. We are going to model the residual by means of a conceptually identical schema that the one we have used in the sinusoidal analysis. The whole algorithm block diagram is presented in figure 6. We can observe that the difference between the synthesized sinusoidal and the original signal is processed with the same algorithm presented before. The algo-

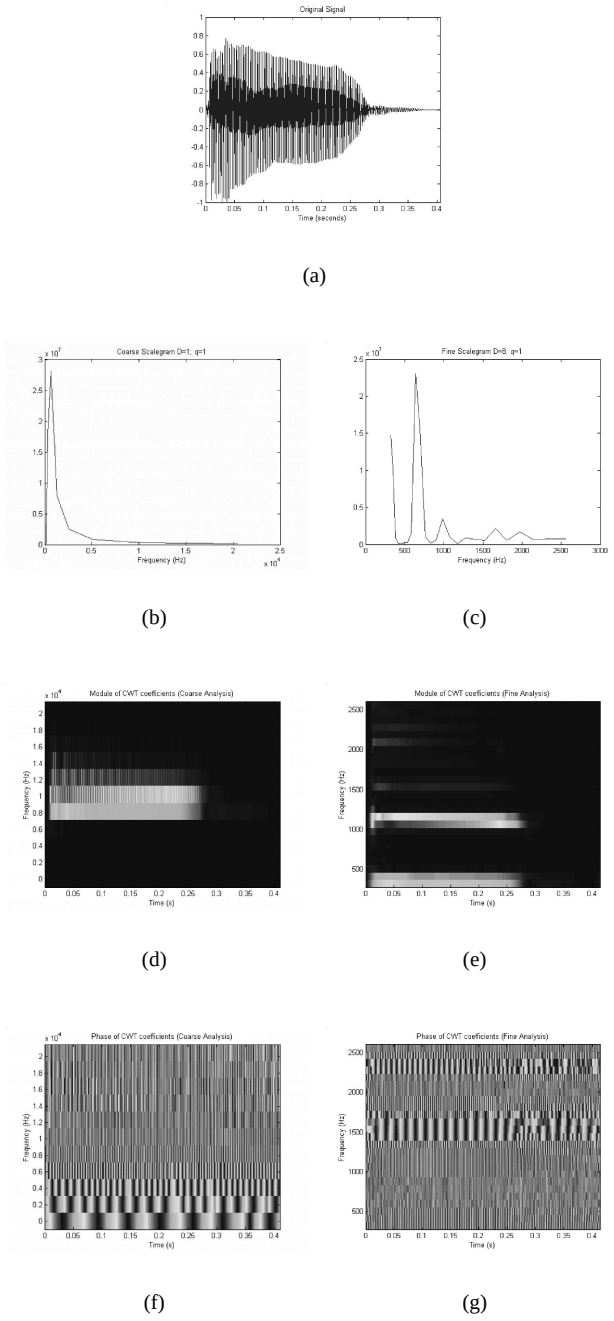


Figure 4: *(a) Recorded guitar note. (b) Scalegram at coarse resolution (D=1, q=1). (c) Scalegram at fine resolution (D=8, q=1). (d) Modulus of the CWT: coarse analysis. (e) Modulus of the CWT: fine analysis. (f) Phase of the CWT: coarse analysis CWT. (g) Phase of the CWT: fine analysis.*

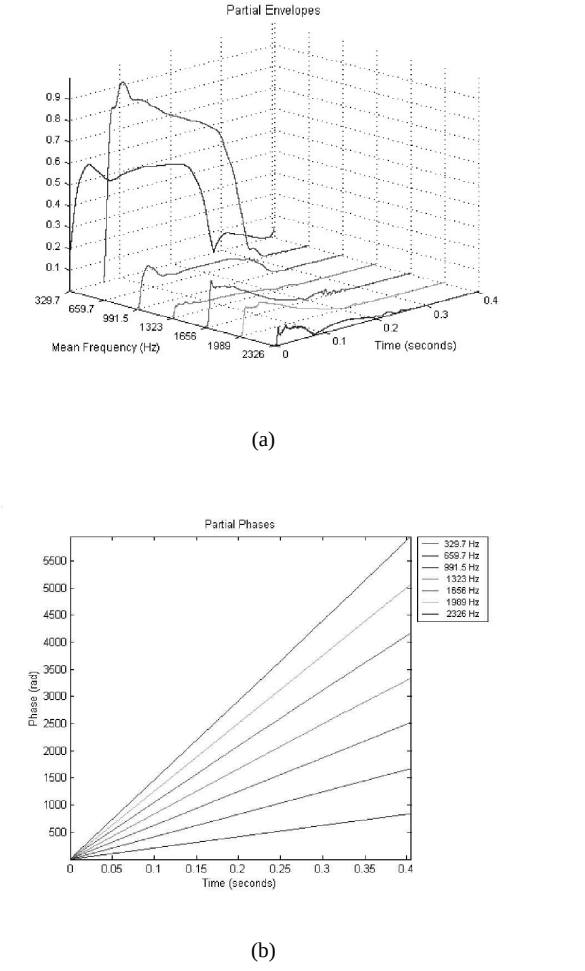


Figure 5: *(a) Partial envelopes obtained form peak detection on the fine analysis scalegram. (b) Unwrapped partial phases obtained from peak detection on the fine analysis scalegram.*

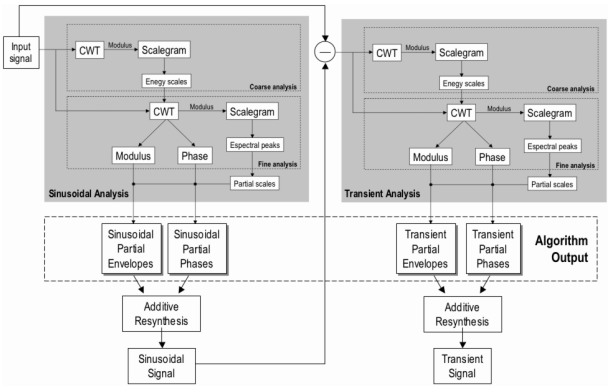


Figure 6: *Sinusoidal plus transient algorithm block diagram.*

rithm output are the envelopes and phases of the sinusoidal and the transient partials.

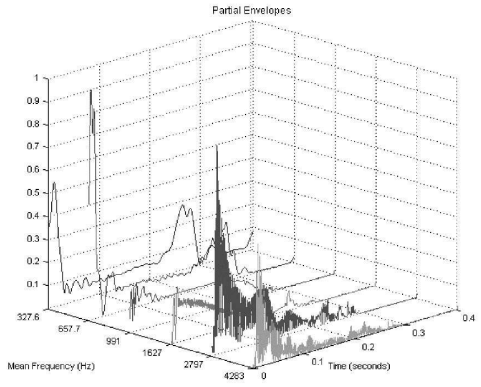
The low scales of the CWT contain a very fine estimation of the rapid variations of the envelopes and phases of the partials.The transient model is performed with wide-band filters. The algorithm parameters have been tuned selecting a few divisions per octave *D*

in equation 15, adjusting q (equation 20) to a value lesser to 1 (for example, $q = 0.5$).

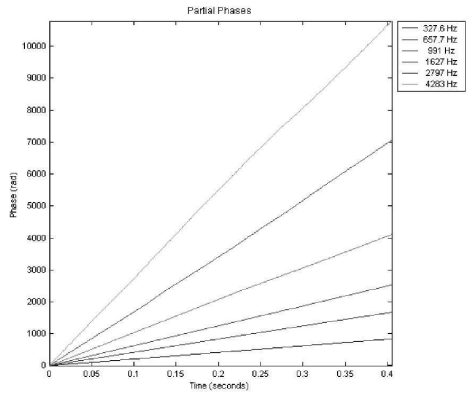
In this way, the residual is modeled by flexible partials envelopes and phases as in the sinusoidal analysis. The transient signal is synthesized by:

$$\hat{s}_t(t) = \sum_{l=1}^{L} a_l(t) \cos \phi_l(t) \tag{23}$$

In figure 7 we can see the envelopes and phases of the analyzed residual of signal in figure 4(a).



(a)



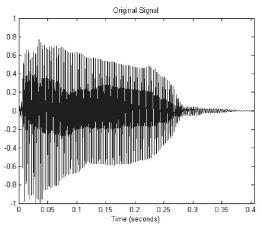
(b)

Figure 7: *(a) Partial envelopes of transient signal. (b) Unwrapped partial phases of transient signal.*

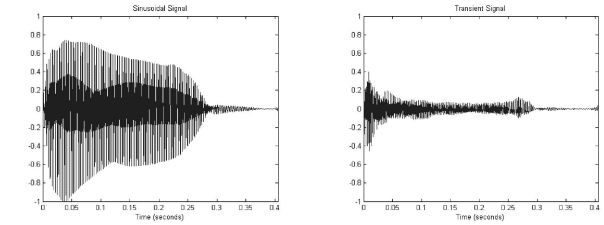
Finally, the synthetic signal is obtained by adding the synthesized sinusoidal and the synthesized transient:

$$\hat{s}(t) = \hat{s}_s(t) + \hat{s}_t(t) \tag{24}$$

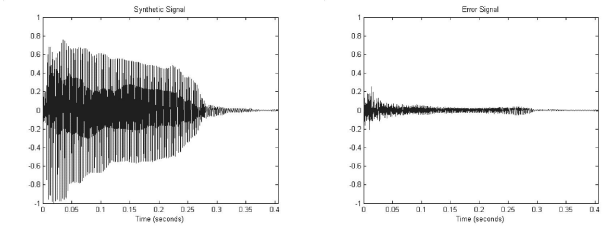
As a result we can see in figure 8 the recorded guitar note (figure 8(a)), the sinusoidal synthesis (figure 8(b)), the transient synthesis (figure 8(c)), the synthetic final signal (figure 8(d)) and the final total error (figure 8(e)). We can see the high quality of the reconstructed sound.



(a)



(b) (c)



(d) (e)

Figure 8: *(a) Original recorded guitar note. (b) Sinusoidal reconstruction. (c) Transient reconstruction. (d) Synthesized signal. (e) Final total error.*

5. TRANSFORMATIONS

At this point we have obtained a model that provides a satisfactory resynthesis of the original sound. Due to additive synthesis characteristics of the model (like in a classical phase vocoder [6]) we can easily implement the modifications of the model parameters that lead to some transformations musically interesting.

First, the note pitch can be estimated. The decomposition of the sound into constituent partials allows the determination of the fundamental frequency and thus, the pitch estimation.

5.1. Pitch shifting

The object of pitch-shifting is to alter the frequency content of a signal without affecting its time evolution [7]. We can define

an arbitrary pitch-scale time-varying function $\alpha(t)$ assumed to be slowly varying function of time. The resulting signal is obtained modifying the phases with the pitch-scale function while maintaining the same duration and amplitude of each partial. The pitch-shifted signal is expressed as:

$$s'(t) = \sum_{k=1}^{N} a_k(t) \cos\left(\alpha(t)\phi_k(t)\right) \tag{25}$$

For example, a vibrato execution could be implemented multiplying the instantaneous phases by a low-frequency sine function.

5.2. Time stretching

Time stretching consist on the modification of the time evolution of a signal without affecting its spectral content [7]. It should be defined a time mapping function $t \rightarrow t' = T(t)$. This mapping function is assumed to be a slowly varying function of time. We can define the slope of the time mapping function $\beta(t) = dT(t)/dt$, so the time stretched signal can be expressed as:

$$\begin{aligned} s'(t') &= \sum_{k=1}^{N} a_k(t') \cos\left(\beta(t)\phi_k(t')\right) \\ t' &= T(t) \end{aligned} \tag{26}$$

The amplitudes and phases of the partials can be stretched, but the phases should be multiplied by a correcting term $\beta(t)$ in order to maintain the pitch unaltered.

5.3. Cross synthesis and morphing

A typical transformation of a classical phase vocoder analysis is cross synthesis. Hybrid sounds taking the dynamic characteristics of a sound combined with the tonal attributes of another sound can be easily generated.

Finally, we can generate morphing or transition between two different sounds. Once the model parameters of both sounds have been associated, as in the generation of hybrid sounds, the envelopes and phases of the initial sound can be gradually modified until they are equal to the parameters of the final sound.

6. CONCLUSIONS

In this article we have presented the basic concepts involved in obtaining a musically meaningful signal representation based on the CWT. We have also discussed several ways to transform some features of the modeled sounds. The quality of the synthesized sounds is very high, maintaining the perceptual identity of the original sounds for most of the timbre families.

The main advantage of the developed technique is the very high flexibility of the model similar to a phase vocoder. Firstly, the separation of the stationary and transient components of the signal is performed. Secondly, each of these components is modeled as the sum of a set of partials, which allows the modification of the harmonic structure of the sounds. Finally, the dynamic and tonal features of each partial are isolated by the extraction of its temporal envelopes and instantaneous phases. This flexibility leads to a wide range of musically useful transformations.

An interesting refinement in order to improve this work will be the inclusion of a stochastic analysis to model the noisy background present in some sounds.

7. ACKNOWLEDGEMENTS

This work has been supported by the University of Zaragoza project UZ2002-TEC-01.

8. REFERENCES

[1] C. Roads, *The computer Music Tutorial*, M.I.T. Press, 1996.

[2] X. Serra and J. O. Smith, "Spectral modelling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, 1990.

[3] S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Tran. on Patt. Anal. and Machine Intell.*, vol. 11, no. 7, pp. 674–693, 1989.

[4] I. Daubechies, *Ten Lectures on wavelets*, vol. 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, 1992.

[5] R. Kronland-Martinet, J. Morlet, and A. Grossmann, "Analysis of sound patterns through wavelet transforms," *Int. J. of Patt. Recog. and Artif. Intell*, vol. 1, no. 2, pp. 272–302, 1987.

[6] T. F. Quatieri and R. J. McAulay, "Audio Signal Processing based on sinusoidal analysis/synthesis," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. Kluwer Academic Publishers, 1998.

[7] J. Laroche, "Time and pitch Scale Modification of Audio Signals," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. Kluwer Academic Publishers, 1998.