2 Application of Time-Frequency and Time-Scale Methods (Wavelet Transforms) to the Analysis, Synthesis, and Transformation of Natural Sounds

Richard Kronland-Martinet and Alex Grossmann

One of the critical problems in the use of digital synthesis of sound, whether in real time or deffered time, is the establishment of a correspondence between synthetic and natural sounds. Since 1957 the digital synthesis of sound has been successfully used in the domains of speech and music; it became clear, however, that one of the main problems would be to relate it with natural sound. The first synthetic sounds were obtained with calculations of samples given by simple mathematical models without direct reference to real sounds. However, it is musically interesting to be able to synthesize sounds that imitate or refer to natural sounds. These sounds would have the richness and distinctiveness of natural sounds and could still be manipulated as is done in all systems of synthesis.

To approach such a problem, it is necessary to bring two complementary aspects into play: the *analysis* and the *resynthesis* of signals (natural sound waves). This leads to the setting up of relationships between the physical (or psychoacoustic) parameters extracted from the analysis and the parameters of synthesis corresponding to a mathematical algorithm.

The analysis aspect should take into account significant parameters such as frequency, time envelopes, microvariations (accidental noise, random or regular modulations), and the distribution of partials. But it should also encompass *data reduction* resulting from the characteristics of auditory perception. For instance, the subjective notion of timbre cannot be always modeled with the help of the Fourier transform of the signal only. In the case of instruments with formants, it is useful to separate the contributions of the *resonances* of the physical system (the modes of the piano, resonant modes of the voice, and so on) from the effects of the system of *excitation* (the struck or plucked string, vibrating vocal chords, and the like).

The synthesis aspect consists of the creation of digitally and musically efficient algorithms. However, the parameters that determine the production of sound only rarely come from existing natural sounds. There are, however, cases where synthesis gives convincing results (for example, with trumpet or voice). Some methods, such as additive or subtractive synthesis, frequency modulation, or waveshaping can give adequate results, particularly if they are completed by adding microvariations or combined with other methods.

The work we present is an approach to the problem of extracting parameters for synthesis and sound modification based on the use of analysis and synthesis methods that combine time and frequency information. In this framework we deal with digital synthesis and analysis of signals by parametric and nonparametric methods. We stress in particular the exploration of acoustic applications of a new method of signal decomposition, the *wavelet transform.* This method, which is strictly speaking "time-scale" rather than "time-frequency," has turned out to be very fruitful, especially in the area of analysis-synthesis.

When one speaks of analysis, one thinks in general about a mathematical representation—as faithful as possible—of a physical phenomenon described in mathematical form. The parameters appearing in the representation must consequently be related in a straightforward way to physical parameters that represent the real world. In the case of audible signals, the "real world" is not limited to the phenomena of sound production and propagation but includes also a biological captor of the greatest importance: our ear. Although many studies in psychoacoustics have contributed to our understanding of the auditory system, it is nevertheless true that the only criterion for deciding on the auditory relevance of a physical parameter is still the ear. Using this criterion, Jean-Claude Risset has developed a powerful technique, *analysis by synthesis,* which consists of refining and characterizing the parameters of a method of synthesis on the basis of its psychoacoustical effect. These results have been quite conclusive, especially in the analysis of the timbre of the trumpet, pointing out the importance of the temporal aspect associated with the evolution of the spectral components.

Methods of digital synthesis attempt to simulate the sometimes rapid evolutions of sound through the manipulation of parameters that should—if possible—have psychoacoustical relevance. Relating those parameters to parameters coming out of analysis requires time-frequency analysis methods, so that the time evolutions of spectral properties of the analyzed signal can be described.

Time-frequency methods can be divided into two types: parametric methods and nonparametric methods.

Methods of the first kind consist of the determination of parameters in a specific model of sound production. Therefore they require some a priori knowledge about the signal being analyzed. In this chapter we are mostly concerned with "blind" analysis and so with nonparametric methods, supplemented when necessary by parametric methods after a "precharacterization" phase. However, in some situations (signals with formants)

parametric methods are very useful. This will be illustrated by the example of cross-synthesis of two natural sounds. In cross-synthesis the characteristics of one source sound are used to drive a system whose response is based on another source sound.

An Example of Parametric Analysis: Linear Prediction

Generally speaking, digital analysis and synthesis of natural sound signals requires the determination of a certain number of physical parameters. It is important to be able to isolate the ones that are most significant or relevant from an auditory point of view. When analyzing a natural sound, one is often led to distinguish between the contribution of the source of excitation and the contribution of the mechanical resonance of the system, which can be modeled by means of a filter. Both aspects are almost always present, but there are situations where the separation between excitation and resonance is not clear-cut. This is the case, for instance, in some percussive systems. Nevertheless in many cases one hopes to find a model of resynthesis by designing a digital filter, starting from a knowledge of the physics of the system that produces or analyzes the sound (figure 2.1).

If $s(t)$ is a real signal (the natural sound wave), then this model is described by a convolution,

$$s(t) = (e * r)(t) = \int e(u)r(t - u)\,du,$$

where $e(t)$ is the excitation or "source" signal, and $r(t)$ is the impulse response of the filter, taking into account propagation within the structure.

This model is then fully characterized by its *impulse response* $r(t)$.

In most cases the impulse response may be considered as fixed, for example, in "rigid" instruments. However, there exist signals that are produced by varying systems. Such signals give rise to time-varying filters (speech).

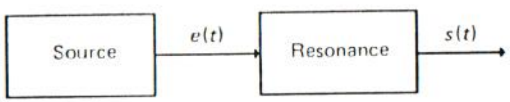


Figure 2.1
Schematic view of physical model synthesis. A source component generates an excitation signal $e(t)$ that is processed by a resonator to create the output signal $s(t)$.

In what follows we discuss this latter kind of signal and show how to apply automatic techniques for identification of the characteristics of time-varying filters and for resynthesis of signals.

We limit ourselves here to the case of filters that vary slowly, so that their frequency response can be considered stable over a time interval T. For the human voice T is usually between 10 and 20 ms. This is an experimental estimate, generally accepted in speech processing (Rabiner et al. 1971, Flanagan et al. 1970).

At this point let us also note that "resonance" and "excitation" can originate from different sources, so that we could perform cross-synthesis.

Analysis by Linear Prediction

The linear prediction technique was originally developed for the purposes of analysis and synthesis of speech (Atal and Hananer 1971). It consists of representing the system in terms of a recursive filter of order p with multiple resonances (figure 2.2). The transfer function is

$$R(z) = \frac{S(z)}{E(z)} = \frac{1}{1 - \sum_{k=0}^{p} a_k z^{-k}}$$

Here $E(z)$ is the z transform of the input signal e. For a voice it can be expressed as $e =$ pulse train (voiced sounds such as vowels) or $e =$ gaussian white noise (nonvoiced sounds such as consonants).

The analysis of a signal then consists of the determination of the coefficients a_k, which can be performed as follows (let us suppose that $e(t)$ is a white gaussian noise):

The system is characterized by

$$s(n) = \sum_{k=1}^{P} a_k \cdot s(n-k) + e(n).$$

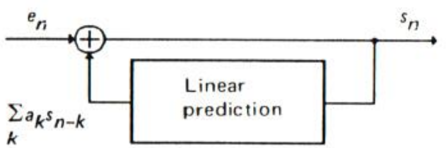


Figure 2.2
Schematic view of linear predictive analysis by means of a recursive filter

Let $\tilde{s}(n)$ be the estimated value of $s(n)$. We write $\tilde{s}(n)$ in the form

$$\tilde{s}(n) = \sum \tilde{a}_k \cdot s(n-k)$$

The prediction error is

$$C(n) = s(n) - \tilde{s}(n) = s(n) - \sum \tilde{a}_k \cdot s(n-k).$$

The $\tilde{a}_k$ are chosen so as to minimize the mean square error:

$$\langle C(n)^2 \rangle = \sum_n \left(s(n) - \sum_k \tilde{a}_k \cdot s(n-k) \right)^2. \tag{1}$$

Differentiating equation 1 with respect to $\tilde{a}_j$, $j = 1 \ldots p$, and setting the derivatives equal to zero, we obtain a system of p equations with p unknowns:

$$\sum_n \tilde{a}_k \sum_n s(n-k) \cdot s(n-j) = \sum_n s(n) \cdot s(n-j) \qquad j = 1 \ldots p. \tag{2}$$

We can write equation 2 in matrix form:

$$\Phi[a] = \psi.$$

Here Φ is the autocorrelation matrix of the signal,

$$\Phi_{i,j} = \sum s(n-i) \cdot s(n-j),$$

ψ is the autocorrelation vector of the signal,

$$\psi_j = \Phi_{i,j}$$

and $[a]$ is the vector of components $\tilde{a}_k$.

In practice one has to calculate the matrix Φ for every time interval compatible with the evolution of the signal and then solve the corresponding system of equations so as to obtain the prediction coefficients that characterize the resonance aspects of the system (Makhoul 1975, Kronland-Martinet 1988a, 1988b).

Subtractive Resynthesis by Predictive Filtering

Having identified the discrete evolution of the filter that serves as model for the instrument, we can perform the resynthesis by

$$s(n) = \sum a_k s_{n-k} + e(n)$$

The excitatory signal e can be obtained from the signal $s(t)$ either by inverse filtering (Moorer 1979) or else—if the mechanism of sound production is simple enough—by mathematical modeling (such as a simplified plucked string).

It is important to also notice the possibility of modeling $e(t)$ with the help of parameters of global synthesis (frequency modulation, amplitude modulation, waveshaping). This procedure yields—in addition to a significant reduction in the volume of data—a possibility of simplified control of the time-dependence of the signal.

Next we examine the possibility of using an excitation signal different from the one that corresponds to the instrument under consideration or its model. We shall now see an application of this possibility, which generalizes the notion of subtractive synthesis and opens up broad prospects for the synthesis of musical sound.

Transformation of Sound: Cross-Synthesis

The cross-synthesis of two sound signals is the construction of a hybrid signal that combines the resonant and the excitation aspects of two distinct sounds (Risset and Wessel 1982). We give here an example of cross-synthesis between voice and cymbal. This example was obtained in collaboration with the Groupe de Musique Expérimentale de Marseille (GMEM). Of course this is just one example out of various possible hybrids (voice + cello, voice + voice, and so forth).

The figures that follow are the spectral representations of, respectively, the voice signal (figure 2.3), the cymbal signal (figure 2.4), and the result of cross-synthesis (figure 2.5).

Here the cross-synthesized signal has been obtained by using the unfiltered cymbal signal as the excitation. This is justified by the fact that such cymbals do not present pronounced resonances.

We can see that resulting signal combines the excitation information of the cymbal (the distribution of partials) and the resonances of the voice (the spectral envelope). The resulting auditory effect is a "talking cymbal."

Parametric analysis obviously has the advantage of considerably reducing the amount of data necessary for the description of sound. Parametric analysis, however, is valid only if the parametric model corresponds to the actual sound generation mechanism. If this is not the case, the results obtained cannot be exploited and, in general, give rise to unstable filters. We shall not discuss here this delicate problem.

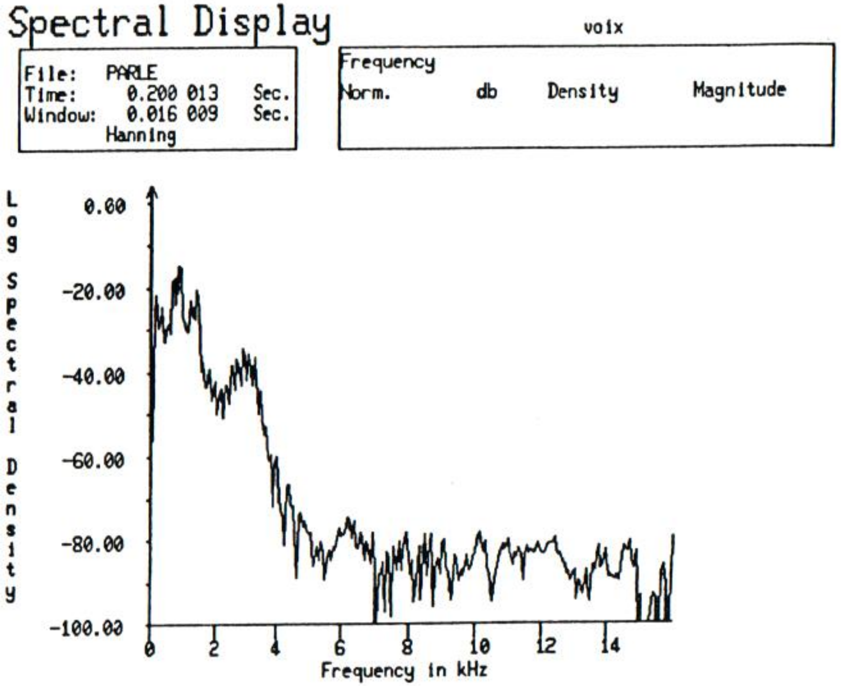


Figure 2.3
Spectral representation of a voice signal

Nonparametric Analysis

Nonparametric analysis methods often use the concept of *local spectrum*, where the signal is decomposed into a sum of complex exponentials (as in the short-time Fourier transform), weighted by a function that plays the role of a time localization window. This procedure is well suited for the identification of quasi-periodic structures in the signal; however, it is not appropriate for the study of rapid transients because events of short duration are delocalized, and their energy is distributed over a region determined by the width of the window.

Taking into account a simplified model of the auditory system, which consists of the representation of the receivers (ears) as a bank of filters with constant $\delta f/f$ for frequencies higher than about 800 Hz, we are tempted to look for a technique of analysis that stresses the time aspect at high

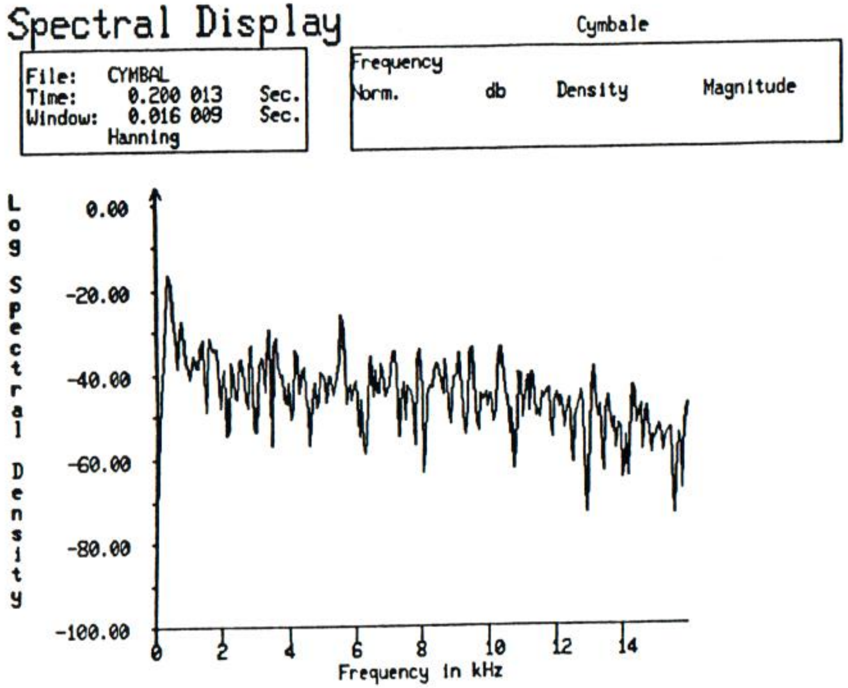


Figure 2.4
Spectral representation of a cymbal signal

frequencies and the frequency aspect at low frequencies. These requirements are satisfied by a new technique of analysis and synthesis, namely, the wavelet transform.

The wavelet transform is a linear transformation that consists of decomposing an arbitrary signal into elementary contributions. Those contributions called wavelets are generated by dilation and translation from a *mother function* called the *analyzing wavelet*.

One of the advantages of such a transform is the fact that one can choose (and consequently adapt) the analyzing wavelet depending on the information in the signal that one wants to point out. The choice of analyzing wavelet is subject only to some mathematical admissibility conditions that are not very restrictive in practice. The wavelet transform associates to a real signal (a real-valued function of time), a function depending on two variables (time and scale). Hence it gives rise to a two-dimensional represen-

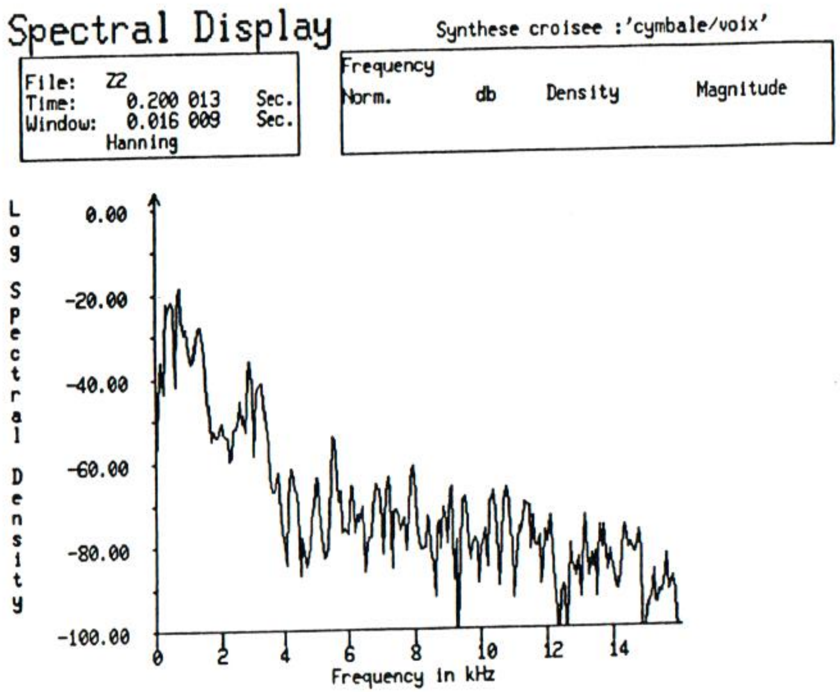


Figure 2.5
Spectral representation of cross-synthesis with a cymbal and a voice

tation (picture) that contains all the information carried by the signal. Because the transform is invertible, one can interpret these pictures as acoustic signatures of the sound. Our experience has indicated that it is useful to assume that the analyzing wavelet is "progressive," that is, that its Fourier components for negative frequencies vanish. Under these conditions the wavelet transform is complex valued. It can be suitably described by its modulus and its phase.

This double representation is interesting on several accounts, in particular because it establishes a relation between the result of the wavelet transform and physical parameters associated with the signal. Indeed the conservation of energy by wavelet transforms makes possible the interpretation of the square of the modulus of the transform as an energy density localized in regions of the time-scale half-plane that depend on the choice of the analyzing wavelet. Furthermore the phase of the wavelet transform can be used to define an *instantaneous frequency* at a given scale. Finally, the wavelet transform can be used to localize discontinuities and to extract modulation laws.

Before applying this method to the analysis, synthesis, and modification of real sounds, we describe the corresponding mathematical formalism. It is convenient to discuss in parallel wavelet transforms and the well-known short-time Fourier transform used in the phase vocoder because the two methods have many points in common as well as significant differences.

Elementary Functions: Grains

The underlying idea of any linear transformation associating a two-dimensional representation to a signal is the decomposition of this signal in terms of elementary grains, such that their sum, multiplied by appropriate coefficients, allows the reconstruction of the original signal (Gabor 1946). Such an approach is clearly useful if every coefficient has a physical or, even better, psychoacoustical interpretation. This interpretation depends strongly on the choice of the basic grains used and in particular on their behavior under variations of the parameters that define the representation.

To obtain decompositions of the kind that we mention, we have to define a family of elementary functions, preferably well localized with respect to the two variables of the representation.

In the case of wavelet transform, the elementary functions are copies of a mother function $g(t)$ (figure 2.6) that have been translated and dilated (or contracted, that is, subjected to a change of scale):

$$g_{\tau,a}(t) = \frac{1}{\sqrt{a}} g\left(\frac{t-\tau}{a}\right). \tag{3}$$

Here τ is the parameter of translation in time, and a is the rescaling (or dilation) parameter. τ is any real number, and a is necessarily positive. The function $g(t)$ is called the *analyzing wavelet*.

The localization in time of function 3 is its effective support region. The frequency localization of this grain is given by the Fourier transform of $g_{\tau,a}(t)$:

$$\hat{g}_{\tau,a}(\omega) = \sqrt{a}\hat{g}(a\omega)e^{i\omega\tau}, \tag{4}$$

where $\hat{g}$ is the Fourier transform of g.

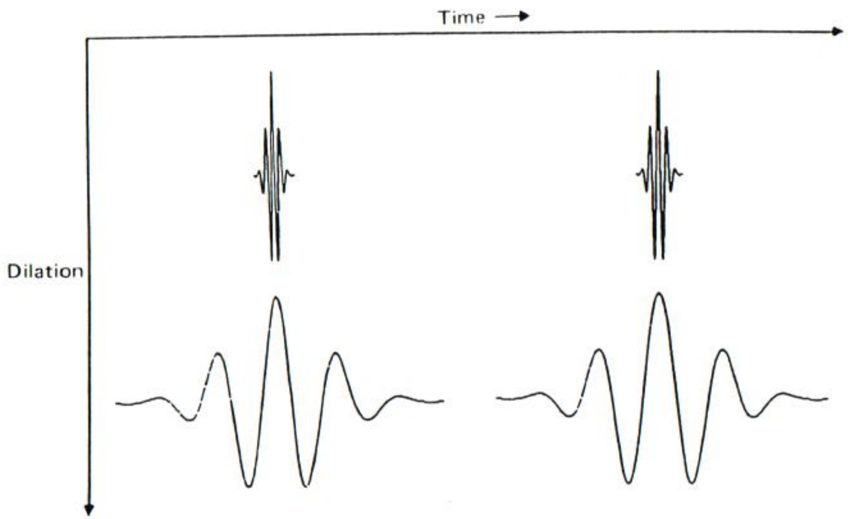


Figure 2.6
Elementary wavelets used for the wavelet transform in the time-dilation domain. The number of cycles in the wavelet does not change.

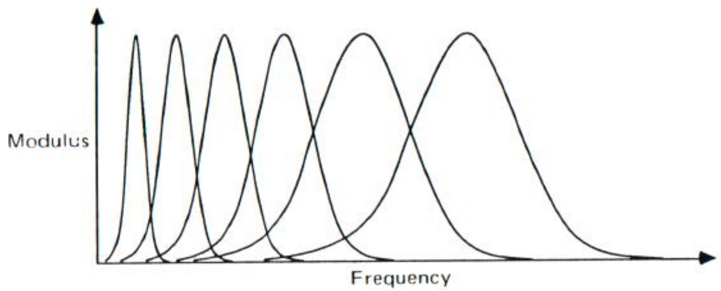


Figure 2.7
This family of curves represents the Fourier transform of the wavelet taken for various values of the scale parameter. The ratio $\Delta f/f$ does not change.

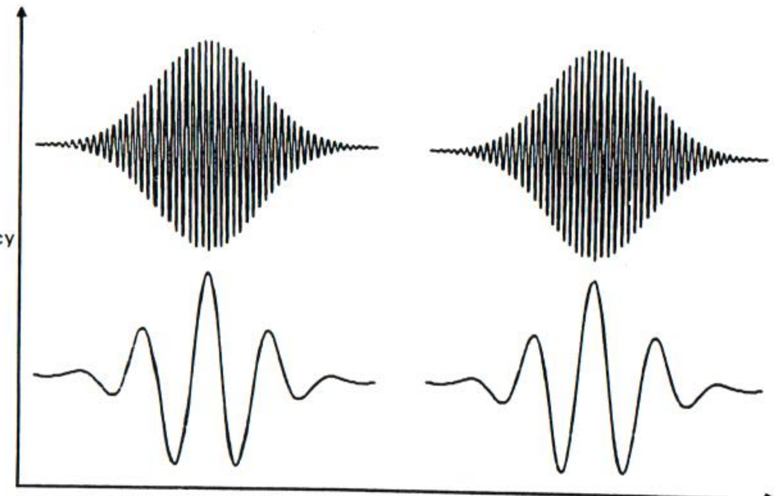


Figure 2.8
Elementary wavelets used for the Gabor expansion in the time-frequency domain

This localization depends on the parameter a (figure 2.7). The resulting decomposition will consequently be at $\Delta\omega/\omega$ = constant. Consequently wavelets can be interpreted as impulse responses of constant-Q filters.

The short-time Fourier transform uses elementary grains that are translations in time and in frequency of a mother function $h(t)$ (figure 2.8):

$$h_{\tau,\omega_0}(t) = h(t-\tau)e^{i\omega_0 t}, \tag{5}$$

where τ is the time translation parameter, ω_0 is the frequency translation parameter, and $h(t)$ is the window that ensures the time and frequency localization of equation 5.

In this case the time localization is given by the effective region of $h(t)$, and the frequency localization is given by the Fourier transform $\hat{h}(\omega)$ of the window. One has

$$\hat{h}_{\tau,\omega_0}(\omega) = \hat{h}(\omega-\omega_0)e^{i\omega\tau}. \tag{6}$$

An important aspect of this construction is that the localization width of equation 6 does not depend on ω_0 (figure 2.9). The resulting decomposition is consequently of $\Delta\omega$ = constant type. The elementary grains can be interpreted as impulse responses of filters of constant width.

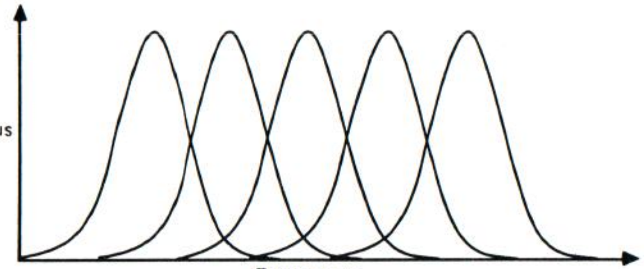


Figure 2.9
This family of curves represents the Fourier transform of the elementary wavelet used for the Gabor expansion. They are obtained from each other by frequency shift. The width Δf is constant.

Decompositions of an Arbitrary Signal

The time-scale and time-frequency transforms are obtained by decomposition of a signal of finite energy into *elementary grains*. The coefficients of this decomposition can be calculated in general as scalar products (correlation) between the signal and the grains, and represent the "weight" associated with each one of them. This is not immediately obvious because the grains are not mutually orthogonal, but is nevertheless true (Grossmann and Morlet 1985).

Consequently the wavelet transform is given by

$$S(\tau,a) = \langle g_{\tau,a} | s \rangle = \int \bar{g}_{\tau,a}(t)s(t)\,dt = \frac{1}{\sqrt{a}}\int \bar{g}\left(\frac{t-\tau}{a}\right)s(t)\,dt, \tag{7}$$

where the bar denotes complex conjugation.

In view of the Parseval relation, an expression equivalent to equation 7 and written in terms of Fourier transform is given by

$$S(\tau,a) = \langle g_{\tau,a} | s \rangle = \langle \hat{g}_{\tau,a} | \hat{s} \rangle = \int \bar{\hat{g}}_{\tau,a}(\omega)\hat{s}(\omega)\,d\omega$$

$$= \sqrt{a}\int \bar{\hat{g}}(a\omega)\hat{s}(\omega)e^{i\omega\tau}\,d\omega. \tag{8}$$

Similarly the short-time Fourier transform is given by

$$T(\tau,\omega_0) = \langle h_{\tau,\omega_0}|s\rangle = \int \bar{h}_{\tau,\omega_0}(t)s(t)\,dt = \int e^{i\omega_0 t}\bar{h}(t-\tau)s(t)\,dt, \tag{9}$$

$$T(\tau,\omega_0) = \langle \hat{h}_{\tau,\omega_0}|\hat{s}\rangle = \int \bar{\hat{h}}_{\tau,\omega_0}(\omega)\hat{s}(\omega)\,d\omega. \tag{10}$$

Let us define a voice as the restriction of the wavelet transform (or, respectively, the short-time Fourier transform) to a fixed value of the scale parameter a (respectively, the frequency ω_0). In other words a voice is the output of a filter with impulse response given by the corresponding grain. The transform so defined can be represented by a *continuous filter bank* with constant Q (wavelet) or with constant width (as in the short-time Fourier transform).

Inversion: Admissibility Conditions

Because the transforms just shown are linear, we may ask for an inversion (or resynthesis) formula by "resummation" of elementary grains weighted by the coefficients (granular synthesis).

In the case of the wavelet transform, we obtain (for progressive wavelets, see "Analysis of Signals: Graphical Representations")

$$s(t) = \frac{1}{c_g}\int S(\tau,a)g_{\tau,a}(t)\frac{d\tau\,da}{a^2}, \tag{11}$$

where c_g is a constant that depends on the choice of the analyzing wavelet:

$$c_g = 2\pi\int \frac{|\hat{g}(\omega)|^2}{\omega}\,d\omega. \tag{12}$$

Looking at equation 11, we notice that inversion is possible only if c_g exists, that is, if the integral in equation 12 converges. This condition that an analyzing wavelet has to satisfy is called the *admissibility condition*. In practice this condition reduces to

$$-E_g = \int |g(t)|^2\,dt < +\infty \quad (g \text{ has finite energy}),$$

$$-\hat{g}(0) = \int g(t)\,dt = 0 \qquad (g \text{ has zero mean value, that is, no DC bias}).$$

There exist other formulas for the inversion of wavelet transforms. One of them is useful because of its simplicity and consists of a reconstruction

of the signal by appropriate summation of coefficients at a fixed time:

$$s(t) = k_g\int S(t,a)\frac{da}{a^{3/2}}, \tag{11'}$$

where k_g is a constant that depends on the choice of the wavelet g.

In the case of the short-time Fourier transform, we obtain

$$s(t) = \frac{1}{c}\int T(\tau,\omega_0)h_{\tau,\omega_0}(t)\,d\tau\,d\omega_0, \tag{13}$$

where $c = 1/2\pi$ is now independent of the choice of $h(t)$. However, $h(t)$ has to be of finite energy.

$$E_h = \int |h(t)|^2\,dt < +\infty \tag{14}$$

Properties of the Wavelet Transform

Linearity The wavelet transform and the short-time Fourier transform are linear. This property is very useful. It means that the transform of the sum of signals is the sum of their transforms, which is convenient for the analysis of polyphonic signals. It should be remarked that the well-known Wigner Ville time-frequency representation is not linear but rather bilinear.

Conservation of Energy The total energy of the signal can be expressed in term of the values of the transforms by

$$E_s = \int |s(t)|^2\,dt = \frac{1}{c_g}\int |S(\tau,a)|^2\frac{d\tau\,da}{a^2}$$
$$= \frac{1}{2\pi E_h}\int |T(\tau,\omega_0)|^2\,d\tau\,d\omega_0, \tag{15}$$

where $E_h = \int |h(t)|^2\,dt$.

These expressions allow us to interpret the square of the modulus of the transforms as a density of energy distributed over the domains of representation. In the case of the short-time Fourier transform, this representation is well known; it is the *spectrogram*.

Constraints on the Coefficients The elementary grains are, in general, not orthogonal:

$$\langle g_{\tau,a}|g_{\tau',a'}\rangle = \int \bar{g}_{\tau,a}(t)g_{\tau',a'}(t)\,dt \neq 0,$$

$$\langle h_{\tau,\omega_0}|h_{\tau',\omega_0'}\rangle = \int \bar{h}_{\tau,\omega_0}(t)h_{\tau',\omega_0'}(t)\,dt \neq 0. \tag{16}$$

The coefficients of the transform are constrained to satisfy

$$S(\tau,a) = \int p_g(\tau,a;\tau',a')S(\tau',a')\frac{da'\,d\tau'}{a'^2}, \tag{17}$$

where $p_g(\tau,a;\tau',a') = (1/c_g)\langle g_{\tau,a}|g_{\tau',a'}\rangle$.

For the coefficients of the short-time Fourier transform, one has

$$T(\tau,\omega) = \int p_h(\tau,\omega;\tau',\omega')T(\tau',\omega')\,d\omega'\,d\tau', \tag{18}$$

where $p_h(\tau,\omega;\tau',\omega') = (1/2\pi E_h)\langle h_{\tau,\omega}|h_{\tau',\omega'}\rangle$.

The relations in equations 17 and 18 are important in the interpretation of the transform. One can show that it is possible to find a grid such that the analysis coefficients at the points of the grid allow an arbitrarily accurate reconstruction of the coefficients on the representation plane.

For wavelet transform the points of such a grid are of the form shown in figure 2.10A, that is, $\{a_0{}^j; k\tau_0 a_0{}^j\}$, where j and k are integers, and a_0 and τ_0 are real numbers that depend on the choice of the wavelet (Daubechies 1988).

For the short-time Fourier transform, the grid is of the form shown in figure 2.10B, that is, $\{m\tau_0; n\omega_0\}$, where m and n are integers, and ω_0 and τ_0 are real numbers that depend on the choice of the window $h(t)$.

Under certain conditions imposed on the grains, it is possible to construct orthonormal bases corresponding to some of the grids just defined. In the wavelet case the grids are dyadic ($a_0 = 2$). For details, see Meyer (1989).

The relations in equations 17 and 18 show that an arbitrary function of two variables need not be the transform of a signal. This may seem restrictive, if one wants to perform a synthesis starting from a picture that has been arbitrarily constructed or obtained by deformation of an existing transform. However, it can be shown that the inversion formula (equation 11) automatically projects into the appropriate subspace of functions that are wavelet transforms of signals.

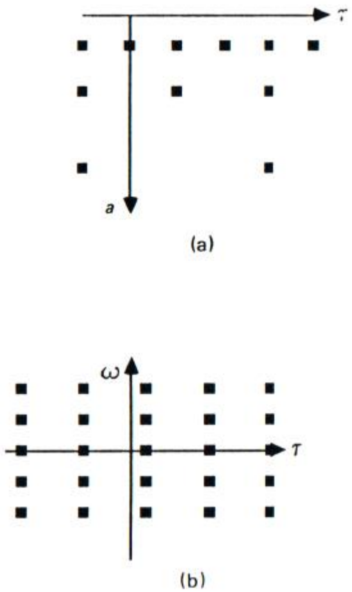


Figure 2.10
Comparison of analysis grids for the wavelet transform (A) and the short-time Fourier transform (B)

Behavior of the Transform under Translation and Dilation of the Signal
The wavelet transform of a signal changes in a simple way if the signal is translated or rescaled. Let $s(t) \to S(\tau,a)$, that is, let $S(\tau,a)$ be the wavelet transform of $s(t)$. Then $s(t-t_0) \to S(\tau - t_0, a)$ that is, the wavelet transform of the shifted signal can be obtained by a natural shift of the original transform. Also one has

$$\frac{1}{\sqrt{\lambda}}s\left(\frac{t}{\lambda}\right) \to S\left(\frac{\tau}{\lambda},\frac{a}{\lambda}\right).$$

The short-time Fourier transform is covariant under translation in frequency. Let $s(t) \to T(\tau,\omega)$, then $s(t)e^{i\omega_0 t} \to T(\tau,\omega-\omega_0)$. However, it is not covariant under translation in time $s(t-\tau_0) \to T(\tau-\tau_0,\omega)e^{-i\omega_0 t}$.

It is often preferable to be covariant under time translations rather than under frequency translations. This can be achieved by using the modified grain:

$$\tilde{h}_{\tau,\omega}(t) = h(t-\tau)e^{i(t-\tau)\omega}.$$

This is the grain to be used in the section "Analysis of Signals: Graphical Representations."

Localization One of the aims of time-frequency representations is to associate to the coefficients some information about the signal that is simultaneously localized in time and in frequency. In the case of wavelets let us assume that $g(t)$ and $\hat{g}(\omega)$ are negligible, respectively, outside an interval $[t_{\min}, t_{\max}]$ and $I\omega_{\min}, \omega_{\max}]$. Then $g_{\tau,a}(t)$ and $\hat{g}_{\tau,a}(\omega)$ have, respectively, the regions $[at_{\min} + \tau, at_{\max} + \tau]$ and $[\omega_{\min}/a, \omega_{\max}/a]$. Consequently the domain of influence in time of a value $s(t_0)$ of the signal is a cone pointing toward $\tau = t_0$ and defined by the lines illustrated in figure 2.11A, that is,

$$at_{\min} + t_0 = \tau, \qquad at_{\max} + t_0 = \tau.$$

The domain of influence in frequency is bounded by the two horizontal lines (figure 2.11B):

$$\omega_{\min}/a = \text{constant}, \qquad \omega_{\max}/a = \text{constant}.$$

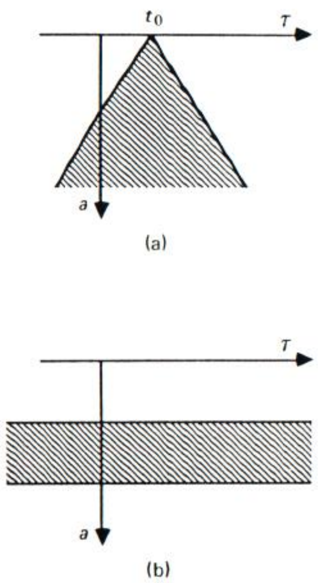


Figure 2.11
Wavelet domains of influence. (A) Domain of influence in time. (B) Domain of influence in frequency.

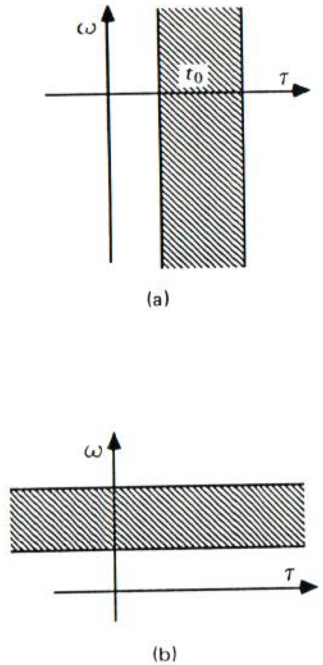


Figure 2.12
Short-time Fourier transform domains of influence. (A) Domain of influence in time. (B) Domain of influence in frequency.

In the case of short-time Fourier transform, let us suppose that $h(t)$ and $\hat{h}(\omega)$ are negligible, respectively, outside an interval $[t_{\min}, t_{\max}]$ and $[\omega_{\min}, \omega_{\max}]$, then $h_{\tau,\omega_0}(t)$ and $\hat{h}_{\tau,\omega_0}(\omega)$ have as a range, respectively, $[t_{\min} + \tau, t_{\max} + \tau]$ and $[\omega_{\min} + \omega_0, \omega_{\max} + \omega_0]$. Consequently the domain of influence of a point $s(t_0)$ is a strip given by (figure 2.12A)

$$\tau = t_{\min} + t_0, \qquad \tau = t_{\max} + t_0.$$

The domain of influence in frequency is bounded by the two horizontal lines (figure 2.12B)

$$\omega_{\min} + \omega_0 = \text{constant}, \qquad \omega_{\max} + \omega_0 = \text{constant}.$$

Algorithms for the Calculation of the Wavelet Transform

The short-time Fourier transform can be easily implemented with the help of fast Fourier transform (FFT) algorithms. Because the wavelet transform is a recent method, we describe the digital techniques used in its implementation and present methods that implement the wavelet transform in real time.

Our first implementation of wavelets was done in 1987 on the SYTER real-time signal processor (Allouis and Mailliard 1981). This system consists of a host processor and a real-time processor that can implement transverse filters. This classical filtering method has been used with impulse responses given by dilated and contracted versions of wavelets. For a fixed value of the dilation parameter, equation 7 consists of a convolution between the signal and the time reversed wavelet $g(-t)$. The calculation is discretized:

$$Y(n) = \sum_i g_i X(n - 1 - i)$$

(we suppose here that the sampling rate is 1).

Here $Y(n)$ is the nth output sample, g_i the ith value of the wavelet, and $X(n)$ the nth input sample. For each value of the dilation parameter a (for each "voice"), the host processor calculates the corresponding discrete values of the wavelet and transmits them to the processor, which then calculates the wavelet transform.

This method of calculation has the disadvantage of not being compatible with real-time operation. The convolutions have to be performed on an arbitrarily large number of points because—for a fixed sampling rate—the dilated wavelet is defined on a large number of samples. This is an important consideration in the case of sound signals. If the analysis takes into account n octaves (here n can be of the order of 10), then the ratio of the number of samples between the most dilated and the most contracted wavelets is 2^n. Consequently the number of operations (multiplications, additions) grows as 2^n.

A different algorithm based on the notion of decimation of signal has been developed and implemented on SYTER, called the *algorithme à trous* ("algorithm with holes"). It allows a parallel and hierarchical structure of calculations, while keeping fixed the number of operations as the scale is changed. We give here only a sketchy indication of the method of calculation and refer the reader to Holschneider et al. 1989. The scheme is shown in figure 2.13A. The information necessary for this algorithm is, on the one hand, discrete values of the analyzing wavelet G and, on the other hand, the coefficients of a filter f that performs a polynomial interpolation of the

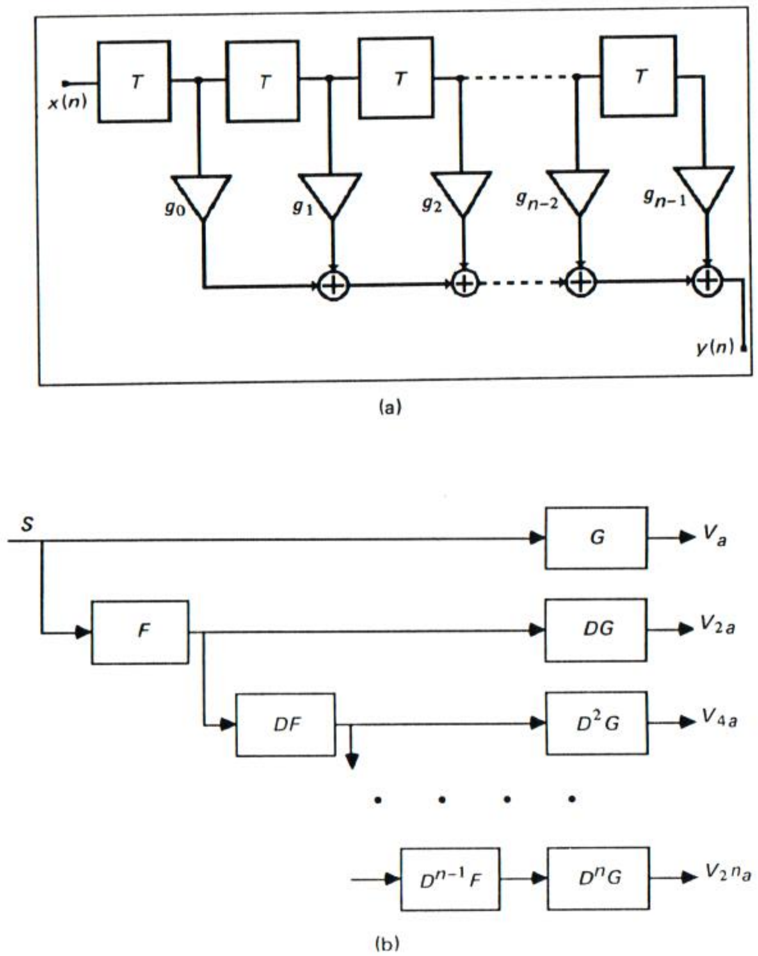


Figure 2.13
Implementation of the wavelet transform. (A) Transverse filter used for the calculation of the wavelet transform in the SYTER digital signal processor. T is the unit time shift. (B) schematic view of the *algorithme à trous*.

wavelet. One can so construct the filter $F = \mathbf{1} + TDf$, where T is the unit shift, D is the operation of dilation by 2:

$$Df(i) = f\left(\frac{i}{2}\right) \text{ for even } i,$$

$$= 0 \text{ for odd } i,$$

and $\mathbf{1}$ is the identity (the filter corresponding to an impulse response that is zero everywhere except for $i = 0$, where it is 1).

A hierarchical structure enables us to use partial results calculated for higher voices (figure 2.13B).

Analysis of Signals: Graphical Representations

Time-frequency analysis of audio signals has been the subject of numerous publications; we refer the reader to the bibliography. The aim is to extract detailed information about the behavior in time and in frequency (or scale) to help with the characterization of sound phenomena in view of reproduction or transformation of natural sounds. Although the wavelet transform and the short-time Fourier transform are conceptually closely related, they are nevertheless different in the way they cover frequency range. We have seen that the wavelet transform can be described as the output of a bank of filters with $\Delta f/f$ = constant, whereas the short-time Fourier transform performs a filtering with Δf = constant and so can separate the harmonics of the signal (provided that Δf is smaller than the fundamental frequency of the signal). By its very nature the short-time Fourier transform tends to decompose an arbitrary signal into harmonic components, whereas the wavelet transform allows free choice of the elementary function or grain of the decomposition.

Before visualizing the behavior of the two transforms, let us discuss the quantities to be represented. Energy conservation (see "Properties of the Wavelet Transform") is of capital importance, and we represent the modulus (or squared modulus) of the transform to obtain a physical interpretation of sound phenomena in terms of energy distribution. To obtain a realistic representation of the modulus of the transform, it is useful to separate positive and negative spectral components so as to avoid beats. This can be obtained by requiring the analyzing wavelet to the "progressive," that is, to contain only positive-frequency components.

$$\hat{g}(\omega) = 0, \qquad \text{if } \omega < 0.$$

Such a condition is easily satisfied for wavelet transform for all τ and a. It cannot be satisfied for the short-time Fourier transform because it can be destroyed by translations in frequency. Figure 2.14 shows the short-time Fourier transform and the wavelet transform of a triangle function. The starting analyzing function is the same for the two transforms. Only the construction of the families is different. The starting function is a modulated gaussian:

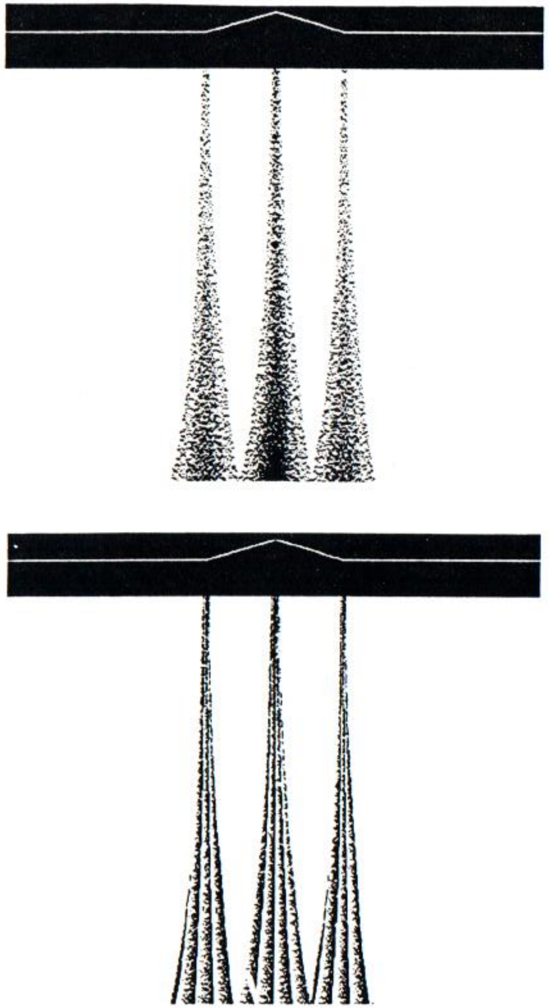
$$f(t) = \exp(-t^2/2)\exp(i\omega_0 t) + \text{small corrections.} \tag{19}$$

In the figure the modulus of the transform is proportional to a local density of black dots. The phase is represented in the same way; this allows an easy visualization of lines of constant phases corresponding to the transition between 2π and 0. The horizontal axis is time, and the vertical axis is frequency (respectively, the negative logarithm of the scale).

One can see in the figure that the two transforms suggest different interpretations of the phenomenon being studied. The wavelet transform allows a localization in time that improves as one progresses toward small scales. This is helpful for detection of discontinuities (Grossmann et al. 1987). On the other hand the short-time Fourier transform looks for harmonic structures in the signal showing horizontal equispaced striations in the figure.

From the point of view of psychoacoustics, wavelet transforms seem better adapted to the analysis of sound signals because they favor the temporal aspect at small scales and the frequency aspect at large scale. It is sometimes convenient to have a detailed description in terms of frequency and consequently to use a short-time Fourier transform. This aspect is developed by Arfib in chapter 3; we consequently focus more on the possibilities for analysis and synthesis provided by the wavelet transform.

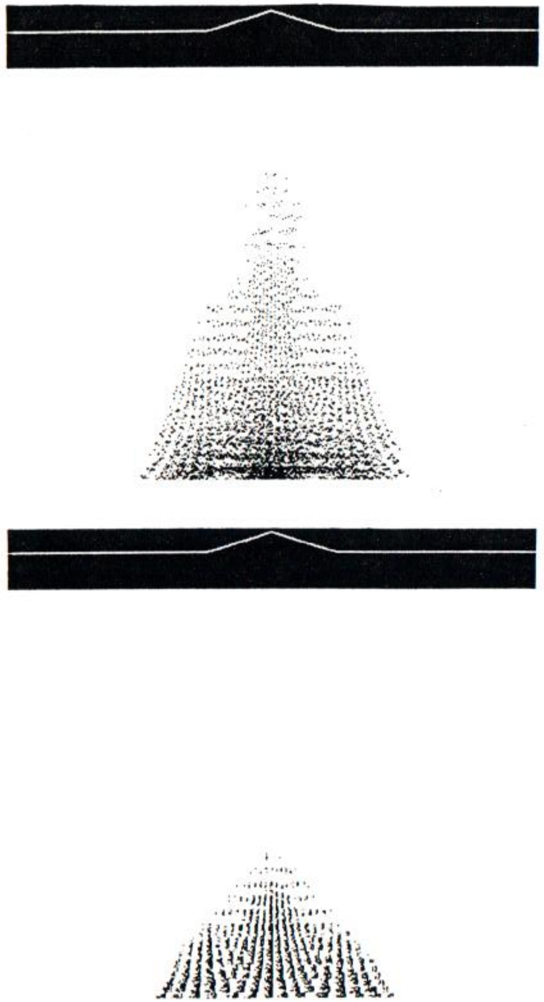
Five Examples Interested readers will find listed in the references many articles that interpret the behavior of coefficients of wavelet transforms. Here we restrict ourselves to five examples of the wavelet transform applied to real signals. A single figure can represent both the modulus and the phase of the transform. In the original images, the modulus is coded with the help of a palette of colors represented at the top-right corner of the image. The phase is coded by a density of black dots in a way that has already been discussed.



(a)

(b)

Figure 2.14
Transforms of a triangle function. (A) Modulus of the wavelet transform. (B) Phase of the wavelet transform. (C) Modulus of the windowed Fourier transform. (D) Phase of the windowed Fourier transform.



(c)

(d)

Figure 2.14 (cont.)

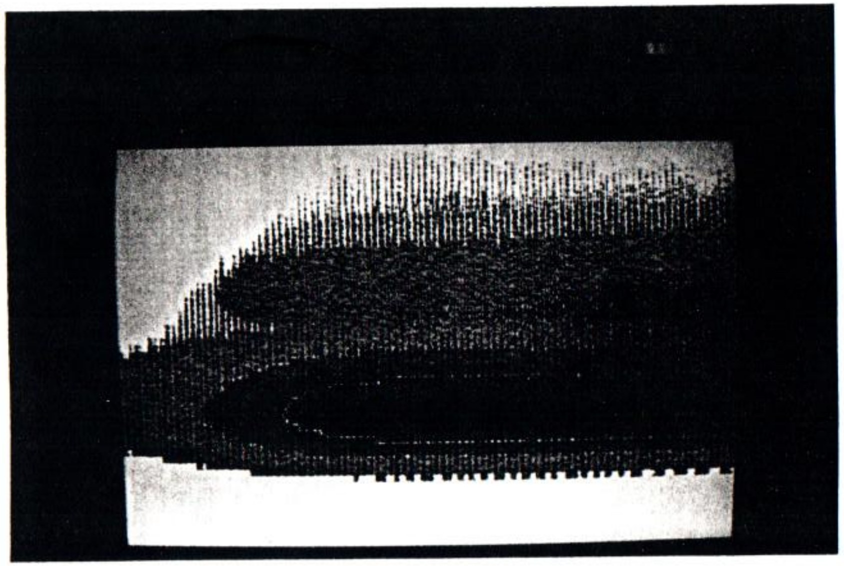


Figure 2.15
Wavelet transform image of the first 32 ms of a clarinet tone

Figure 2.15 represents the first 32 ms of a clarinet sound. Taking into account the domain of influence discussed in the section "Properties of the Wavelet Transform," one can conclude that successive harmonics appear at different times. There exist algorithms for the estimation of frequencies and of modulation laws associated with the components. These algorithms are based on the study of the phase of the transform (Guillemain et al. 1989) and are discussed later.

Figure 2.16 represents the same signal that has undergone a transformation realized by P. Dutilleux (Laboratoire de Mécanique et d'Acoustique, Marseille) in the framework of a study for a "sound sculpting machine" for the Museum of Music, Paris. The sound is cut into slices of equal length. To each slice one associates a linear amplitude modulation. The wavelet transform has turned out to be a very powerful tool for the visualization of the modifications so produced. In particular it provides a measure of the discontinuities introduced by the procedure and an estimation of the energy distribution generated by the modulation over the spectral components.

Figure 2.17 represents three successive notes on a trombone. Here again it is easy to interpret the distribution of the energy along the scale and time

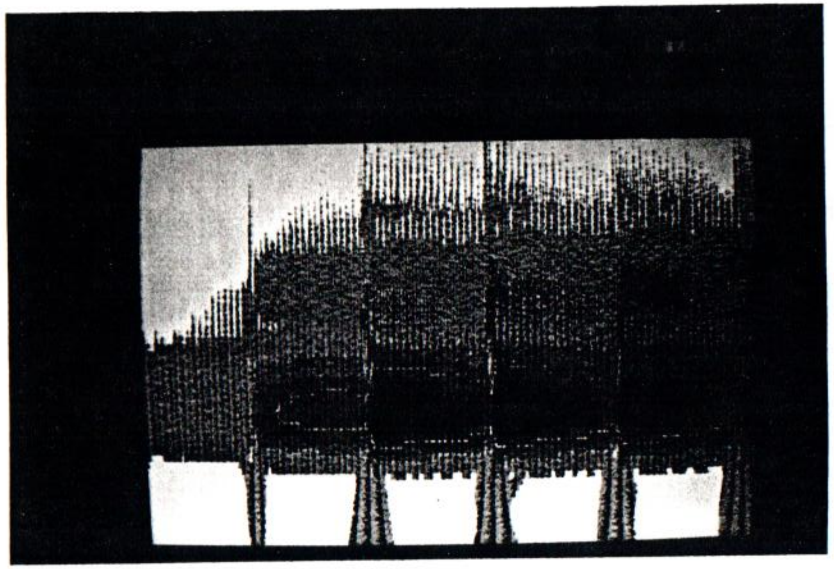


Figure 2.16
The same sound shown in figure 2.15, cut into equal-length slices and amplitude modulated

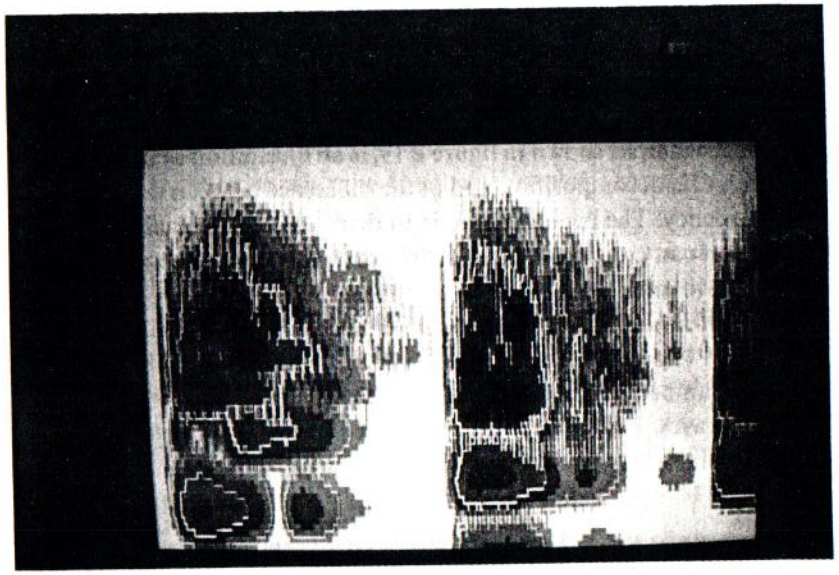


Figure 2.17
Wavelet transform image of three notes of a trombone

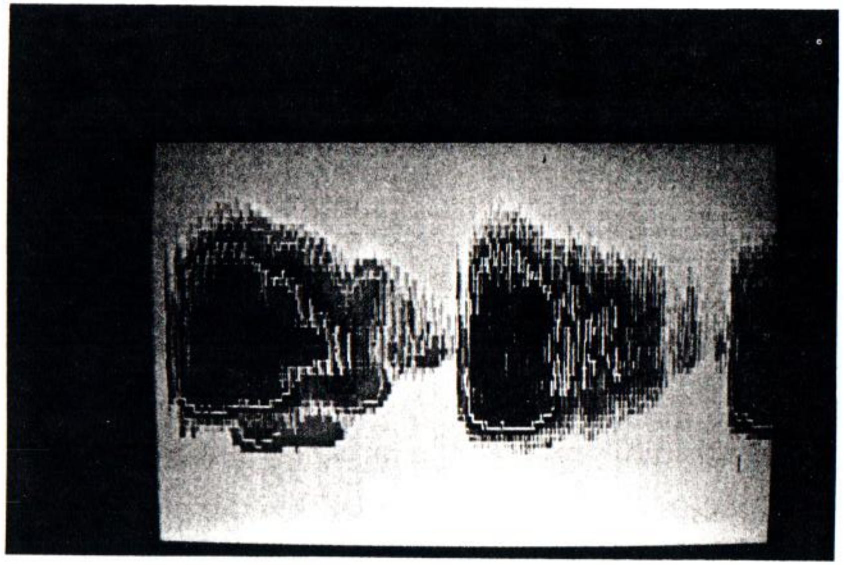


Figure 2.18
Same signal as in figure 2.17, processed by a bandpass filter

axes. This is shown clearly by considering the same signal filtered by a bandpass filter, shown in figure 2.18. One can so characterize the effect of the filtering on the signal and, most important, visualize possible transient effects in the case of time-varying filters.

The last example, shown in figure 2.19, is an illustration of the usefulness of a time-scale decomposition and of the difference between time-scale and time-frequency. The problem here is to detect within a signal a particular feature that can appear at an arbitrary scale. If, for instance, one wants to detect octave intervals appearing in a sound sequence, it is advantageous to decompose the signal in terms of elementary contributions that are adapted to this purpose. A wavelet with two bumps in the frequency domain at an octave distance to each other allows us to detect the occurrence of octaves. The local energy in the time-scale half-plane will be higher at times when the signal contains octaves than at other times (other parameters being equal). In such a way one can also construct wavelets adapted to detection of other chords or other predefined structures.

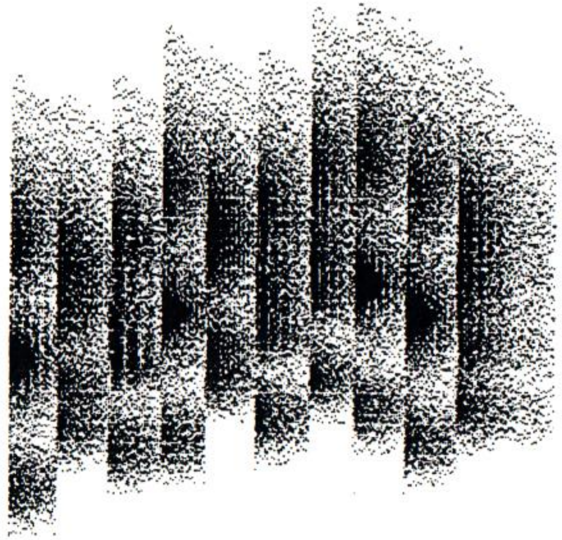


Figure 2.19
Wavelet transform adapted for octave detection. The four instances of octaves show up as dark spots in the wavelet transform.

Parameter Estimation with the Help of Wavelet Transforms

The information contained in the representations just described is sufficient for an exact reconstruction of the analyzed signal. It is often useful to extract from these representations some parameters that describe physical phenomena occurring in the signal. Without entering into mathematical details, we briefly describe some of the possibilities opened by wavelet transforms in the field of detection of important features of sound signals. A dominant role among those features is occupied by spectral components and by frequency and amplitude modulations. They allow synthesis of the signal by additive techniques or by frequency modulation. So far these methods have produced convincing simulations and characterizations of signals. The parameters so measured have not yet been used for resynthesis of sound.

Estimation of Spectral Lines

The aim here is to find on the wavelet transform side the information necessary for the extraction of monochromatic components modulated in amplitude. The extraction is performed automatically, that is, it does not require a visual inspection of the two-dimensional representation described previously.

A *spectral line* is defined as a function of the form $s(t) = A(t)e^{i\omega t}$, where $A(t)$ is the amplitude modulation law, and ω is the frequency.

The search for spectral lines consists of the identification of the function $A(t)$ and of the number ω. We discuss here the detection of such lines contained in signals of the form

$$s(t) = \sum_{j\,\text{finite}} A_j(t)e^{i\omega_j t} + b(t),$$

where $b(t)$ is an arbitrary signal that does not contain discrete spectral components (noise).

Before we describe an iterative method that detects these parameters, let us examine the case of a monochromatic signal.

Consider the signal $Ae^{i\omega_1 t+\phi}$, whose wavelet transform is given by

$$S(\tau, a) = \sqrt{a}A\bar{\hat{g}}(a\omega_1)e^{i\omega_1\tau+\phi}, \tag{20}$$

where $\bar{\hat{g}}$ is the Fourier transform of the wavelet.

To facilitate the interpretation of equation 20, we suppose that $\hat{g}(\omega)$ is real; this condition is satisfied, for example, by the wavelet equation 19. The phase of the wavelet transform is then equal to the phase of the signal itself, for every value of the scale parameter a. The modulus of the transform is constant on all lines of the fixed-scale parameter. In the case of real-valued monochromatic signals, the previous statement remains true, provided that $\hat{g}(\omega) = 0$ on the negative axis, so that interference phenomena are avoided. This is the motivation for the introduction of "progressive" wavelet in the section on graphical representations.

It is useful to define the instantaneous frequency of a voice $S_a(\tau)$ (the restriction of the transform to a fixed value of the scale parameter) by

$$\Omega_a(\tau) = \frac{d\varphi_a(\tau)}{d\tau},$$

where $\varphi_a(\tau)$ is the phase of $S_a(\tau)$, $S_a(\tau) = M_a(\tau)\exp(i\varphi_a(\tau))$, $M_a(\tau) = |(S_a(\tau))|$.

If the signal is monochromatic, then $\Omega_a(\tau) = \omega_1$. Consequently the instantaneous frequency obtained from the wavelet transform of such a signal is equal to the frequency of the signal and allows its determination. Furthermore the modulus of the transform of the monochromatic signal in equation 19 is given by

$$M_a = \sqrt{a}A\,|\bar{\hat{g}}(a\omega_1)|$$

Consequently, knowing the analyzing wavelet, we can also determine the value of A from the wavelet transform.

More generally one can show (Guillemain et al. 1989) that it is possible to estimate the frequency of spectral components contained in a signal by taking an average of the instantaneous frequency of its wavelet transform over a suitably chosen interval. The actual method of estimation cannot be given here, and we refer the reader to Guillemain et al. 1989. Figure 2.20 represents the wavelet transform of a sum of three spectral lines. The reduced frequencies are 1, 1.75, and 2.5. One can see the instantaneous frequencies and amplitude modulation laws for values of the scale parameter that correspond to the fixed points defined previously. The estimation of frequencies and amplitude modulation laws is excellent and has already been used by P. Guillemain to extract lines in nuclear magnetic resonance (NMR) spectroscopy.

Estimation of Modulation Laws

The determination of the law of frequency modulation in a signal can be of major importance, particularly if one wants to use methods of resynthesis such as FM (Chowning 1973). The main idea of the method that we describe is that the information contained in the signal is not distributed uniformly over the time-scale half-plane. Consequently, for purpose of reconstruction, certain parts of the half-plane carry the essential information, and it is crucial to be able to determine these parts.

Under some mathematical assumptions that are not discussed here, one can extract from the wavelet transform its essential points (corresponding mathematically to points of stationary phase) (Escudié et al. 1989). They describe in the half-plane a trajectory called the *ridge*.

The usefulness of the ridge stems from the fact that the restriction of the wavelet transform to it is sufficient to obtain a good approximate reconstruction of the signal up to a known amplitude factor and up to a specific phase. This restriction will be called the *skeleton* of the transform. As an

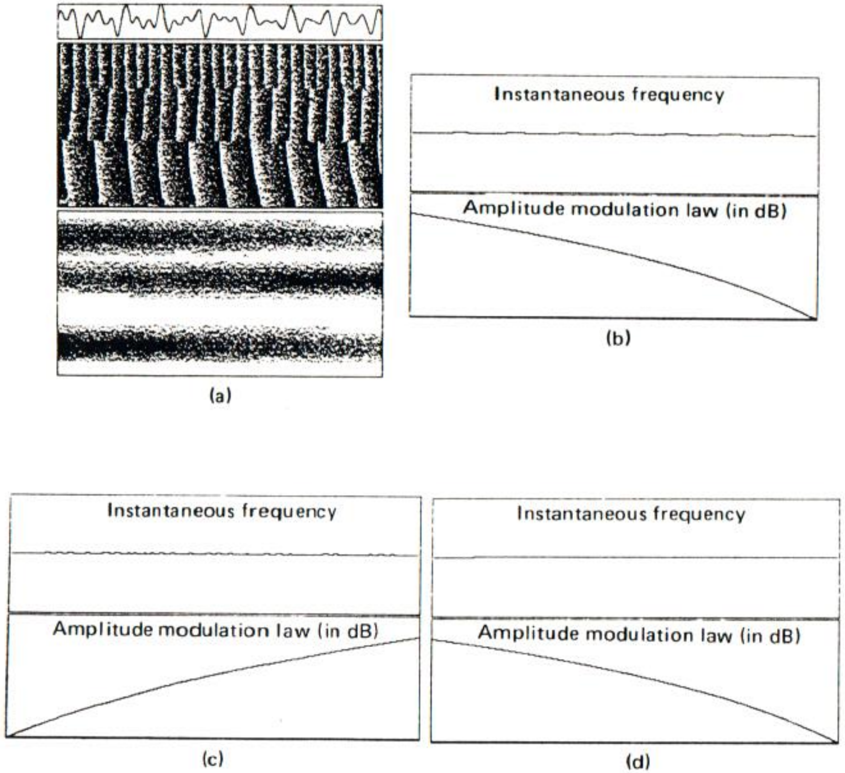


Figure 2.20
Extraction of spectral lines. The signal is a sum of three components with frequencies 1, 1.75, and 2.5. The amplitude modulation laws are linear. The first and the third lines decrease; while the second increases. (A) Wavelet transform of the signal (modulus and phase). (B, C, D) Instantaneous frequency and amplitude modulation law for each component.

example, figure 2.21A is the ridge of a signal obtained by FM synthesis:

$s(t) = A\sin\{\alpha t + I(t)\sin(\beta t)\}$.

Here the modulation index $I(t)$ increases linearly.

Figure 2.21B gives the phase of the skeleton of the transform, calculated by N. Delprat (Laboratoire de Mécanique et d'Acoustique, Marseille). The linear part can easily be estimated by the linear regression method, giving the parameter α (the carrier). Figure 2.21C represents the phase of the skeleton after subtraction of the carrier. It is thus possible to identify the modulating frequency β and the law of variation of the modulation index $I(t)$.

Resynthesis by Means of Wavelets

The reconstruction formulas equations 11 and 11′, show that there are two natural methods for resynthesizing a natural sound starting from its analysis by a wavelet transform: granular synthesis and generalized additive synthesis. This section contrasts these two approaches.

Granular Resynthesis

Equation 11 suggests an approach to resynthesis by summation over a set of dilated and translated wavelets, with complex-valued coefficients derived from the transform (figure 2.22). The digital implementation consists of associating a wavelet with an appropriate weight to each point on the analysis grid. These wavelets are *elementary grains* with overlap in time. Synthesis is then carried out as a sum of those contributions. Notice that because the coefficients associated with wavelets are in general complex, the *resynthesis grains* will in general have a different form due to phase shifts. The basic module of the instrument that perform this kind of synthesis is given by figure 2.23. The analysis is performed here with a displacement step equal to a quarter of the length of the wavelet. This requires four wavelet generators that are successively activated. The multiplexer distributes the values of the transform to the four generators of the wavelets.

F. Boyer has developed at LMA special unit generators and integrated them into the program Music V (Boyer and Kronland-Martinet 1989) to be able to use this method of synthesis. The results are very good and allow a reconstruction of almost perfect audio and digital quality using coeffi-

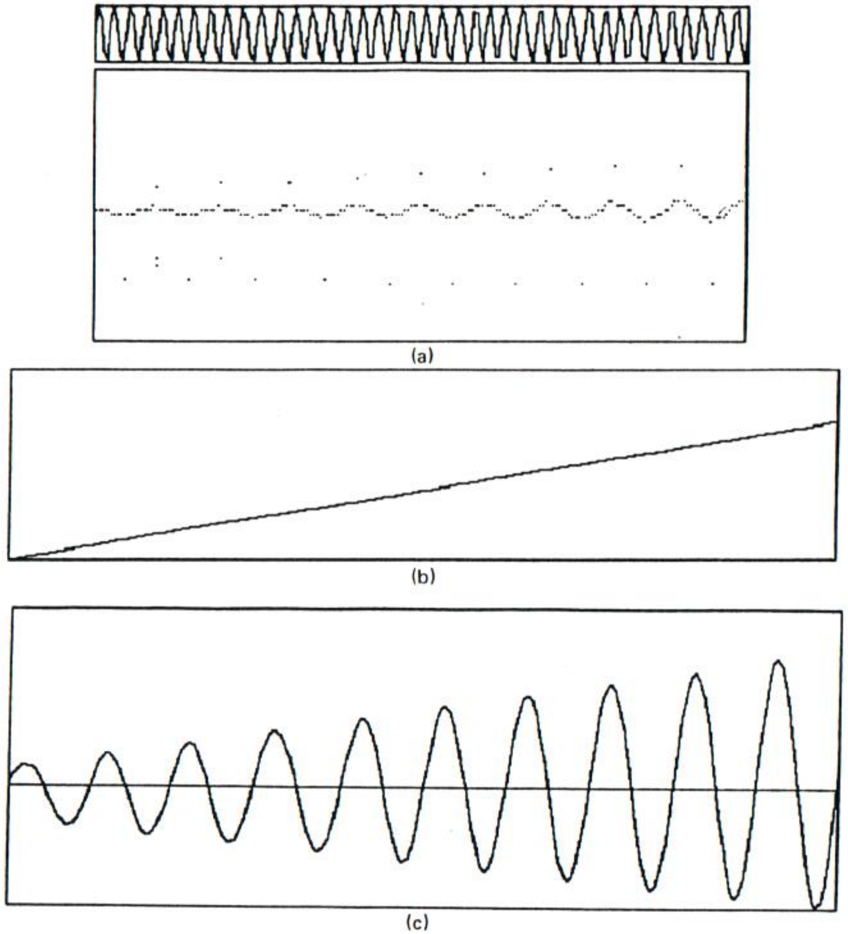


Figure 2.21
Estimation of modulation laws. (A) Ridge corresponding to a signal obtained by frequency modulation synthesis. The modulation index increases linearly. (B) Unwrapped phase of the skeleton of the transform. (C) Unwrapped phase of the skeleton after subtraction of the carrier. It is possible to identify the modulating frequency β and the law of variation of the modulating index $I(t)$.

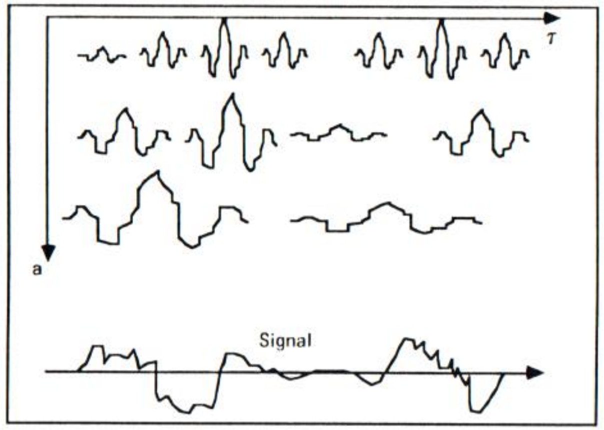


Figure 2.22
Granular resynthesis by summation over a set of dilated and translated wavelets with complex-valued coefficients derived from the transform

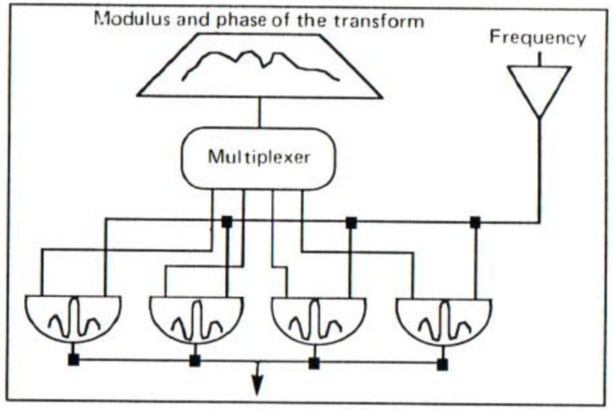


Figure 2.23
Module for granular resynthesis. The frequency depends only on the dilation parameters associated with the voice under consideration. The complete instrument requires as many modules as there are voices in the analysis.

cients on an appropriate grid, for example, a dyadic grid and the wavelet of equation 19.

Generalized Additive Synthesis

Equation 11′ allows the reconstruction of a signal at time t by summation of the values of the transform for fixed $\tau = t$. If the wavelet is complex valued, the values of the transform are also complex, and every voice (for fixed a) can be defined as

$$S_a(t) = x_a(t) + iy_a(t) = A_a(t)e^{i\varphi_a(t)},$$

where $M_a(t)$ is the modulus $M_a(t) = (x_a^2 + y_a^2)^{1/2}$, and $\varphi_a(t)$ is the instantaneous phase $\varphi_a(t) = \arctan(y_a/x_a)$.

We are here in a situation similar to the one encountered with the phase vocoder. The same reconstruction method can be used and the problem is reduced to the identification of instantaneous frequency or, more precisely, to the determination of the reading increment for a sine table. Those procedures are described by Moorer (1978).

The discretization of equation 11′ on the grid $a = 2^j$ compatible with the wavelet of equation 19 gives

$$s(t) = k_g \sum_{a=2^j} S(t, a)a^{-1/2}$$

$$= k_g \sum_{a=2^j} A_a(t)\cos(\varphi_a(t))a^{-1/2} \qquad \text{(taking the real part).}$$

The resynthesis is then of generalized additive type because the oscillators are modulated both in amplitude and in phase. The basic module of the instrument that realizes this kind of synthesis is given in figure 2.24.

Transformation of Sound Signals

The resynthesis of sound signals has enabled us to verify the power of the method by numerical tests as well as by listening. However, the main interest of the decompositions that we studied lies in the possibilities of characterization (by images), of modeling signals, of data reduction, and of intimate transformation of sounds. This last aspect is very attractive for musicians who are searching for new auditory sensations. Now we show how wavelet transforms can be used for the modification of audio signals

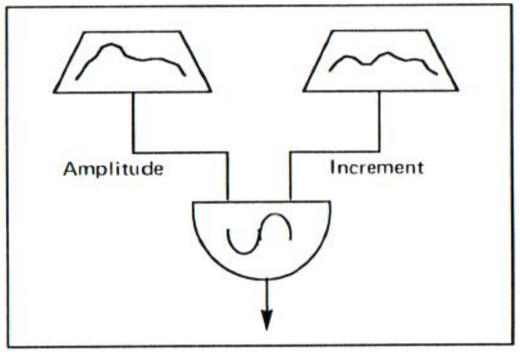


Figure 2.24
Module for generalized additive resynthesis. The complete instrument requires as many modules as there are voices in the analysis.

by altering the values (coefficients) between analysis or synthesis or by a deformation of the time-scale half-plane.

Among possible modifications one can distinguish between those that are obtained by linear action on the values of the transform and those that require nonlinear operators (Kronland-Martinet 1989). The latter are likely to be more interesting, but they require a more complicated mathematical formalism. We present here only the main ideas and an intuitive interpretation of the values of transform (the coefficients). These allow us to get a feel for the audio results obtained through modification.

The most obvious linear modification consists of performing a partial resynthesis by restricting the original analysis grid in time or scale or both. For instance, resynthesis that takes into account only the voices correspondig to small scales has the effect of a highpass filter, provided that the analyzing wavelet corresponds to a filter with a single resonance, or to a "high-scale" filter. More generally the interpretation of wavelet coefficients at a fixed scale as the output of a filter allows us to understand the effect of a constant gain on each of the voices. We can so obtain a "scale equalizer" with gains that can vary in time without stability problems.

Another interesting linear transform consists of time-shifting the analyzing voices with respect to each other. The physical interpretation here is that of an *acoustical wave*. The components of the wave propagate with different speeds, depending on their scale (dispersion). The resulting signal propagates in clusters, yielding an "aquatic" effect.

The nonlinear transformations we consider consist of separate modifications of the modulus and the phase of the wavelet coefficients. The results so obtained can be understood intuitively if one realizes that the modulus (or more precisely its square) plays the role of a density variable for energy. Thus it gives the distribution of the total energy in time and in scale (resonant aspects). On the other hand the phase gives the "oscillation," that is, the excitation. Such an interpretation of the coefficients allows us to make guesses concerning modifications that perform changes of the following types:

- Transpositions without change of duration (or conversely, say, stretching in time without changes in pitch)
- Harmonization or brightness effects
- Hybridization or cross-synthesis of signals.

Let us consider in a little more detail the effect of frequency transposition without changes of duration. To do this, consider first the monochromatic signal

$$s(t) = A\cos(\omega t)$$

As we saw previously, the values of its wavelet transform are given by

$$S(\tau, a) = \sqrt{a}A\bar{\hat{g}}(a\omega)e^{i\omega\tau}.$$

We have seen that the phase of the coefficients is the phase of the signal itself, provided that $\hat{g}(\omega)$ is real. A frequency transposition of ratio n can thus be obtained simply by multiplying the phases of all coefficients by n.

Although the mathematics does not allow us to do this, we can generalize this procedure to more general signals and multiply the phases of coefficients by n to obtain a transposition effect.

$$s_n(t) = k_g \sum_{a=2^j} A_a(t)\cos(n\varphi_a(t))a^{-1/2}.$$

Notice that n may depend on time; one obtains time-varying transpositions.

Similarly one can obtain brightness effects by adding harmonics at each scale in a synchronous way. One has to perform a multiple transposition with an integer ratio:

$$s_h(t) = k_g \sum_{a=2^j} A_a(t) \sum_n C_n \cos(n\varphi_a(t))a^{-1/2}.$$

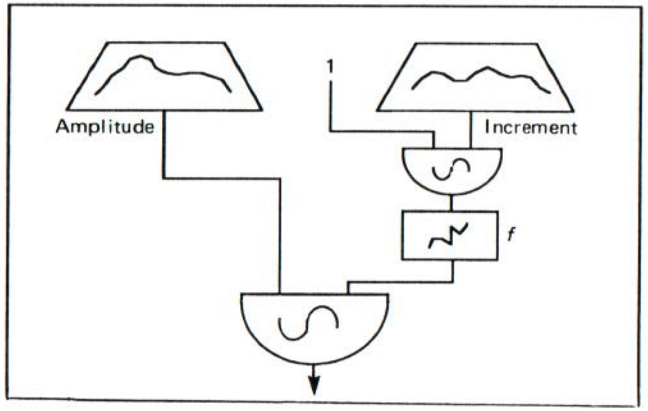


Figure 2.25
Resynthesis module for "brightness" effect using nonlinear distortion

The algorithm can be simplified if one notices that one is performing here a nonlinear distortion f, which can be written, with the help of Chebychev polynomials, as (Arfib 1979)

$$f(x) = \sum_k C_k T_k(x),$$

with $T_k(\cos\varphi(t)) = \cos(k\varphi(t))$.

The corresponding instrument Music V is represented in figure 2.25.

For a final example, we present a modification that is directly related to the cross-synthesis of two signals. The interpretation of the phase as excitation and of the module as resonance brings us naturally to "sound hybridization," by synthesis using the modulus and the phase of two distinct sounds. It is of course possible to perform other transformations as intermediate steps. Some of the examples can be heard on the soundsheet in Kronland-Martinet 1989. Almost all of the examples were realized with just one voice per octave (a total of eight voices).

Conclusion

Wavelet analysis provides an extremely powerful tool for the analysis of audio signals. The degree of refinement of such an analysis makes it a privileged tool in many domains where it is necessary to characterize transient phenomena. The results obtained so far have confirmed the power of the method in the analysis of discontinuities and in the detection of patterns in a given signal. It has also been shown to be extremely efficient, allowing resynthesis and transformations with a small number of voices.

The use of wavelets in the synthesis and modification of signals still needs to be explored in a much more systematic way. It is reasonable to expect that the remarkable mathematical properties of the wavelet transform will turn out to be keys to many classes of applications, particularly because a fast algorithm allows us to envisage real-time approaches.

Although the nonparametric aspect of wavelet transform is seductive by its generality, it is nevertheless clear that certain parametric approaches can give rise to results that are interesting and more specific. This is proved by cross-synthesis of signals, where linear prediction gives rise to simplified implementations.

One can conclude with the truism that in matters of analysis-synthesis there is no universal panacea and that each case has its own optimal approach. However, the results presented here on wavelet transforms give a general framework for the handling of time-varying signals that corresponds to a need of musicians: a "direct" chain in which the processing of real sounds could be standardized, just as musical performance is captured by MIDI.

References

Allouis, J. F., and B. Mailliard. 1981. "Simulation par ordinateur du studio électroacoustique et applications à la composition musicale." *Conférence des journées d'études, Festival international du son haute fidélité.* Paris: Editions SDSA, pp. 43–55.

Arfib, D. 1979. "Digital synthesis of complex spectra by means of multiplication of non-linear distorted sine waves." *Journal of the Audio Engineering Society* 27(10): 757–768.

Boyer, F., and R. Kronland-Martinet. 1989. "Granular resynthesis and transformation of sounds through wavelet transform analysis." In T. Wells and D. Butler, eds. *Proceedings of the 1989 International Computer Music Conference.* San Francisco: Computer Music Association, pp. 51–54.

Chowning, J. 1973. "The synthesis of complex audio spectra by means of frequency modulation." *Journal of the Audio Engineering Society* 21: 526–534.

Daubechies, I. 1988. "Orthonormal bases of compactly supported wavelets." *Communications in Pure and Applied Mathematics* 41: 909–996.

Daubechies, I. 1990. "The wavelet transform, time-frequency localization and signal analysis." To be published in *IEEE Transactions on Information Theory.*

Escudié, B., A. Grossmann, R. Kronland-Martinet, and B. Torrésani. 1989. "Représentation en ondelettes de signaux asymptotiques: emploi de la phase stationnaire." In *Proceedings Colloque GRETSI.* Juan les Pins.

Flanagan, J. L., C. Coker. L. Rabiner, R. Schafter, and N. Umeda. 1970. "Synthetic voices for computer." *IEEE Spectrum* 7: 22–45.

Gabor, D. 1946. "Theory of communication." *Journal of the IEE* 93: 429–441.

Grossmann, A., M. Holschneider, R. Kronland-Martinet, and J. Morlet. 1987. "Detection of abrupt changes in sound signals with the help of wavelet transforms." *Advances in Electronics and Electron Physics. Supplement* 19: *Inverse Problems.* Orlando: Academic Press.

Grossmann, A., and J. Morlet. 1985. "Decomposition of functions into wavelets of constant shape, and related transforms." In L. Streit, ed. *Mathematics + Physics, Lectures on Recent Results.* Singapore: World Scientific.

Guillemain, P., R. Kronland-Martinet, and B. Martens. 1989. "Application de la transformée en ondelettes en spectroscopie RMN." Note Interne 112. Marseille: Laboratoire de Mécanique et d'Acoustique.

Holschneider, M., R. Kronland-Martinet, J. Morlet, and P. Tchamitchian. 1989. "A real-time algorithm for signal analysis with the help of the wavelet transform." In J. Combes, A. Grossmann, and P. Tchamitchian, eds. *Wavelets, Time-Frequency Methods and Phase Space.* New York: Springer-Verlag, pp. 286–297.

Kronland-Martinet, R. 1988a. "Digital subtractive synthesis of signals based on the analysis of natural sounds." In *Etat de la Recherche Musicale (au ler janvier* 1989). Aix en Provence: Editions A.R.C.A.M.

Kronland-Martinet, R. 1988b. "The use of the wavelet transform for the analysis, synthesis and processing of speech and music sounds." *Computer Music Journal* 12(4): 11–20. (Sound examples on soundsheet with 13(1) 1989.)

Kronland-Martinet, R. 1989. "Analyse, synthèse, et transformation de signaux sonores: application de la transformée en ondelettes." Marseille: Thèse d'Etat-Sciences, Faculté des Sciences de Luminy.

Makhoul, J. 1975. "Linear prediction, a tutorial review." *Proceedings of IEEE* 63: 561–580.

Meyer, Y. 1989. "Orthonormal wavelets." In J. Combes, A. Grossmann, and P. Tchamitchian, eds. *Wavelets, Time-Frequency Methods and Phase Space.* New York: Springer-Verlag.

Moorer, J. A. 1978. "The use of the phase vocoder in computer music applications." *Journal of the Audio Engineering Society* 26: 42–45.

Moorer, J. A. 1979. "The use of linear prediction of speech in computer music applications." *Journal of the Audio Engineering Society* 27: 134–140.

Rabiner, R. R., R. Schafer, and J. Flanagan. 1971. "Computer synthesis of speech by concatenation of formant-coded words." *Bell System Technical Journal* 50: 1541–1558.

Risset, J. C., and D. Wessel. 1982. "Exploration of timbre by analysis and synthesis." In D. Deutsch, ed. *The Psychology of Music.* Orlando: Academic Press, pp. 26–54.