Modelling of natural sounds by time–frequency and wavelet representations*

R. KRONLAND-MARTINET, PH. GUILLEMAIN and S. YSTAD

CNRS, Laboratoire de Mécanique et d'Acoustique, 31 Chemin Joseph Aiguier, 13402 Marseille cedex 20, France
E-mail: kronland@lma.cnrs-mrs.fr guillem@lma.cnrs-mrs.fr ystad@lma.cnrs-mrs.fr

Sound modelling is an important part of the analysis–synthesis process since it combines sound processing and algorithmic synthesis within the same formalism. Its aim is to make sound simulators by synthesis methods based on signal models or physical models, the parameters of which are directly extracted from the analysis of natural sounds. In this article the successive steps for making such systems are described. These are numerical synthesis and sound generation methods, analysis of natural sounds, particularly time–frequency and time–scale (wavelet) representations, extraction of pertinent parameters, and the determination of the correspondence between these parameters and those corresponding to the synthesis models. Additive synthesis, nonlinear synthesis, and waveguide synthesis are discussed.

1. THE SOUND MODELLING CONCEPT

Analysis–synthesis is a set of procedures to reconstruct a given natural sound and collect information about it. Different methods can be applied, and the success of each method depends on its adaptive possibilities and the sound effect to be produced. Figure 1 shows the most commonly used procedures. The three parts of the figure correspond to different processes. The central level corresponds to a direct analysis–synthesis process and consists of reconstructing a sound signal by inversion of the analysis procedure. This is a useful process which uses analysis to get information about a sound, and synthesis (inversion) to verify that no information is lost. The analysis will make it possible to classify and characterise audio signals (Kronland-Martinet, Morlet and Grossman 1987), but the result of the process will simply be a reproduction of the natural sound. From a musical point of view, a representation of sounds by analysis is useful when one intimates sound modifications. This sound transformation process corresponds to the upper path in figure 1 and consists of altering the coefficients of the representation between the analysis and the synthesis procedures. According to the analysis method used, different aspects of the sound can be altered. The energy distribution and/or the frequency of the partials can, for example, be manipulated through spectral analysis. Time–frequency analysis allows the separation of the time and frequency characteristics associated with the sound and is of great interest (Kronland-Martinet 1988). However, this approach conflicts with a very important mathematical principle which states that one cannot arbitrarily modify a time–frequency representation of a signal. This constraint is due to the existence of the so-called 'reproducing kernel' which takes into account the redundancy of such representations (Kronland-Martinet *et al.* 1987). It corresponds to the uncertainty principle which states that one cannot be as precise as one wishes in the localisation of both the time and the frequency domains. This principle imposes the concept that the time–frequency domain corresponding to the uncertainties (time–frequency atoms) be considered instead of isolated values of the representation. It is then natural to make the transformations act on these domains in order to conserve the necessary correlations between close representation values. This is done by using a mathematical operation known as a projection, which can transform any image into a time–frequency representation. The constraints limit the time–frequency transformation processes and make it difficult to determine the correspondence between the altered values and the obtained sounds. Nevertheless, very interesting sounds can be obtained by carefully using such altering procedures (Arfib and Delprat 1993). In this article we will pay special attention to the lower part of figure 1 which corresponds to sound modelling. In this part, the representations obtained from the analysis provide parameters corresponding to the synthesis models. The concept of the algorithmic sampler (Arfib, Guillemain and Kronland-Martinet 1992) consists of simulating natural sounds through a synthesis process that is well adapted to algorithmic and realtime manipulations. The resynthesis and the transformation of natural sounds are then part of the same concept.

The paper is organised as follows. We describe the most commonly used synthesis methods, analysis methods such as time–frequency and wavelet transforms, and the algorithms used for separating and

* This work has been partly supported by the Norwegian Research Council.

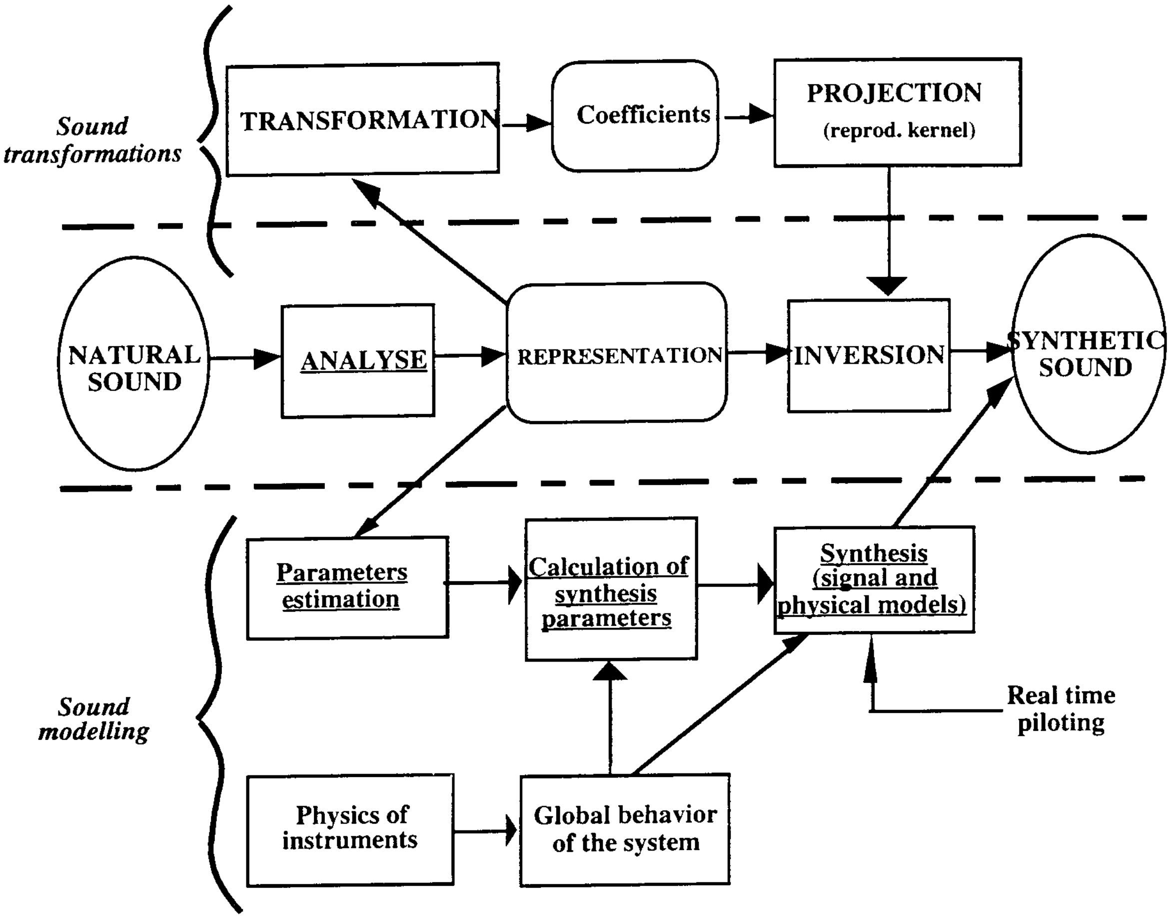

Figure 1. General organisation of the analysis–synthesis and modelling concept. Each underlined item corresponds to a section in the text.

characterising spectral components. We conclude by showing how the analysis of real sounds can be used to estimate the synthesis parameters of the signal models and of the physical models. Most of these techniques have been developed in our laboratory in Marseille, France.

2. SOUND SYNTHESIS MODELS

Digital synthesis uses methods of signal generation that can be divided into two classes:

- signal models aimed at reconstructing a perceptive effect without being concerned with the specific source that made the sound,
- physical models aimed at simulating the behaviour of existing or virtual sound sources.

2.1. Signal model synthesis

Signal models use a purely mathematical description of sounds. They are numerically easy to implement, and they guarantee a close relation between the synthesis parameters and the resulting sound. These methods are similar to shaping and building structures from materials, and the three principal groups can be classified as follows:

- additive synthesis,
- subtractive synthesis,
- global (or nonlinear) synthesis.

2.1.1. Additive synthesis

A complex sound can be constructed as a superposition of elementary sounds, generally sinusoidal signals modulated in amplitude and frequency (Risset 1965). For periodic or quasi-periodic sounds, these components have average frequencies that are multiples of one fundamental frequency and are called harmonics. The periodic structure leads to electronic organ sounds if one does not consider the microvariations that can be found through the amplitude and frequency modulation laws of the components of any real sound. These dynamic laws must therefore be very precise when one reproduces a real sound. The advantage of these synthesis methods is the potential for intimate and dynamic modifications of the sound.

Granular synthesis can be considered as a special kind of additive synthesis, since it also consists in summing elementary signals (grains) localised in both the time and the frequency domains (Roads 1978).

2.1.2. Subtractive synthesis

A sound can be constructed by removing undesired components from an initial, complex sound such as noise. This synthesis technique is closely linked to the theory of digital filtering (Rabiner and Gold 1975) and can be related to some physical sound generation systems such as speech (Flanagan, Coker, Rabiner, Schafer and Umeda 1970, Atal and Hanauer 1971). The advantage of this approach (excluding the physical aspects of physical modelling synthesis, discussed later) is the possibility of uncoupling the excitation source and the resonance system. The sound transformations related to these methods often use this property to make hybrid sounds or crossed synthesis of two different sounds by combining the excitation source of a sound and the resonant system of another (Makhoul 1975, Kronland-Martinet 1989). A well-known example of cross-synthesis is the sound of a talking 'cello obtained by associating an excitation of a 'cello string and a resonance system corresponding to the time-varying formants of the vocal tract.

2.1.3. Global synthesis

Simple and 'inert' signals can be dynamically modelled using global synthesis models. This method is nonlinear since the operations on the signals are not simple additions and amplifications. The most well-known example of global synthesis is audio frequency modulation (FM) updated by John Chowning (Chowning 1973) which revolutionised commercial synthesizers. The advantages of this method are that it calls for very few parameters, and that a small number of operations can generate complex spectra. These simplify numerical implementation and control. However, it is difficult to control the shaping of a sound by this method, since the timbre is related to the synthesis parameters in a nonlinear way and continuous modification of these parameters may give discontinuities in the sound. Other related methods have proved to be efficient for signal synthesis, such as waveshaping techniques (Arfib 1979, Le Brun 1979).

2.2. Synthesis by physical modelling

This is a more recent technique, which we will describe more precisely than signal model synthesis. Unlike signal models which use a purely mathematical description of sounds, physical models describe the sound generation system with respect to its physical behaviour. Such models can be constructed either from the equations describing the behaviour of the waves propagating in the structure and their radiation in air, or from the behaviour of the solution of these equations. The first approach is costly in terms of calculations and is generally used only in connection with research work (Chaigne 1995), unless one uses a simplified version consisting of modelling the structure by an association of simple elements (Cadoz, Luciani and Florens 1984) (masses, springs, dampers,. . .). Synthesis by simulation of the solution of the propagation equation has led to waveguide synthesis models (Smith 1992), which have the advantages of being easy to construct and of having a behaviour close to that of a real instrument. Such synthesis methods are, consequently, well adapted to the modelling of acoustical instruments.

We describe below the principles of these methods in order to reveal their parameters, together with their correlation to physical mechanisms. These parameters are related to the structure of the instrument as well as to the instrumental performance. If we consider a vibrating string, the Shannon theorem states that one can, without loss of information, split the movement into a succession of instantaneous cliches separated by an interval of time T called the sampling period. If c is the propagation speed of the waves in the string, this is equivalent to cutting the string into intervals of length $x = cT$ and considering the propagation as a passage from one elementary cell to another. This operation corresponds to a spatial 'discretisation' of the structure: one can then consider the wave propagation as the result of a succession of transformations or filterings of the initial excitation.

In the ideal case in which we neglect losses and nonlinearities, there is only a displacement of the waves (in two directions), and the result can thus be simulated by a succession of delay lines corresponding to the sampling period T, symbolised in digital signal processing by the variable z^{-1}. In the more realistic case in which the waves undergo an attenuation depending on the frequency, a filter P should be added between each delay. If in addition the medium is dispersive, a 'dephasor' or an all-pass filter D should be added (figure 2).

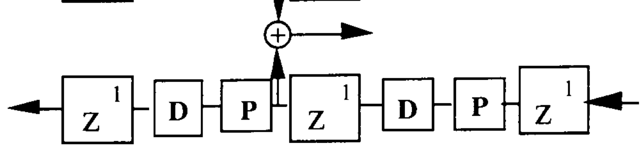

Figure 2. Discrete simulation of the wave propagation in a dissipative and dispersive medium.

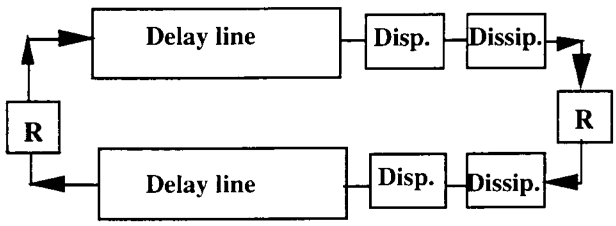

Figure 3. Propagation model in a bounded dissipative and dispersive medium.

The theory of digital filters allows elements of the same type to be gathered. Thus, the propagation medium can be represented by a succession of 3 elements, i.e. a delay line, an attenuating filter for simulating the dissipation, and an all-pass filter for simulating the dispersion. Real instruments have strings of finite length and the waves propagated through it are reflected at the ends. The reflections correspond to a return of the initial waves, with modifications depending on the boundary conditions. Thus one can simulate the wave behaviour corresponding to the solution of the equations. For that purpose, one uses a looped system which, in addition to the delay line, attenuating filter and the all-pass filter, also makes use of a filter corresponding to the reflections R (figure 3).

Synthesis models related to a particular digital filter are known as waveguide models. They can be used to simulate many different systems, such as a tube representing the resonant system in wind instruments (Cook 1992).

3. ANALYSIS OF REAL SOUNDS

The analysis of natural sounds calls for several methods giving a description or a representation of pertinent physical and perceptive characteristics of the sound (Risset and Wessel 1982). Even though the spectral content of a sound is often of great importance, the time course of its energy is at least as important. This can be shown by artificially modifying the attack of a percussive sound in order to make it 'woolly', or by playing the sound backwards. The time and frequency evolution of each partial component is also significant. The vibrato is a perceptively robust effect that is essential, for example for the synthesis of the singing voice. Another essential aspect that should be taken into account when creating a sound corresponding to a plucked vibrating string is the different decay times of the partials. These examples illustrate the need for analysis methods giving access to time and frequency variations of sounds. To solve this general analysis problem of signals, a collection of methods called joint representations has been designed.

The analysis methods of signals can be divided into two principal classes: parametric methods and nonparametric methods. The parametric methods require *a priori* knowledge of the signal, and consist of adjusting the parameters of a model. The nonparametric models do not need any knowledge of the signal to be analysed, but they often require a large number of coefficients.

3.1. Parametric methods

These techniques are generally optimal for the representation of signals adapted to the chosen parametric model. The most common method used for processing sounds is linear prediction (LPC). This technique is adapted to signals from sound production systems of the source–resonance type. The resonant filter should be modelled by a digital all-pass filter the coeffficients of which are related to the frequency and to the width of the formants. The applications of analysis–synthesis for speech signals are numerous, because of a good correspondence between the physics of the vocal tract and the linear filtering. The input signal of LPC systems is generally a broadband noise or a periodic signal adapted to a subtractive synthesis technique.

3.2. Nonparametric methods

Nonparametric techniques for analysis of sound signals generally correspond to representations with physically and/or perceptively meaningful parameters. The best known is the spectral representation obtained through the Fourier transform. In this case the signal is associated with a representation giving the energy distribution as a function of frequency. As mentioned earlier, this representation is not sufficient for characterising the timbre and the dynamic aspects of a sound. In what follows we describe the joint time–frequency representations considering both dynamic and frequency aspects. The time–frequency transformations distribute the total energy of the signal in a plane similar to a musical score in which one of the axes corresponds to the time and the other to the frequency. Such representations are to sound what musical scores are to melodies. There are two ways of obtaining this kind of representation depending on whether the analysis acts on the energy of the signal or on the signal itself. In the first case the methods are said to be nonlinear, giving, for instance, representations from the so-called 'Cohen's class'. The best known example of transformations within this class is the Wigner–Ville distribution (Flandrin 1993). In the other situation the representations are said to be linear, leading to the Fourier transform with a sliding window, the Gabor transform, or the wavelet transform. The linear methods have, at least as far as sound signals are concerned, a great advantage over the nonlinear methods. Linear methods make the resynthesis of signals possible and they ensure that no spurious data cause confusion during the interpretation of the analysis. These spurious data can occur in nonlinear analysis as a result of cross-terms appearing in the development of the square of a sum. This is why we shall focus on the linear time–frequency methods.

The linear representations are obtained by decomposing the signal into a continuous sum of elementary functions having the same properties of localisation both in time and in frequency. These elementary functions correspond to the impulse response of bandpass filters. The central frequency of the analysis band is related to a frequency parameter for time–frequency transformations and is related to a scaling parameter for wavelet transforms. The choice of the elementary functions gives the shape of the filter.

3.2.1. Gabor transform

In the case of the Gabor transform, the elementary functions, also called time–frequency atoms, are all generated from a 'mother' function (window) translated in time and in frequency. The 'mother' function is chosen to be well localised in time and frequency and to have finite energy (for instance a Gaussian function) (figure 4).

Each value of the transform in the time–frequency plane is obtained by comparing the signal to a time–frequency atom. This comparison is mathematically expressed by a scalar product. Each horizontal line of the Gabor transform then corresponds to a filtering of the signal by a bandpass filter centred at a given frequency with a shape that is constant as a function of frequency. The vertical lines correspond to the Fourier transform of a part of the signal, isolated by a window centred on a given time. The transform obtained this way is generally complex, since the atoms themselves are complex, giving two complementary images (Kronland-Martinet *et al.* 1987). The first one is the modulus of the transform and corresponds to a classical spectrogram, the square of the modulus being interpreted as the energy distribution in the time–frequency plane. The second image corresponding to the phase of the transform is generally less well known and less used, but it nevertheless contains a lot of information. This information concerns mainly the 'oscillating part' of the signal (figure 5). Actually, the time derivative of the phase has the dimension of a frequency and leads to the frequency modulation law of the signal components (Guillemain and Kronland-Martinet 1996).

3.2.2. Wavelet transform

The wavelet transform follows a principle close to that of the Gabor transform. Again the horizontal lines of the wavelet transform correspond to a filtering of the signal by a filter, the shape of which is independent of the scale, but whose bandwidth is

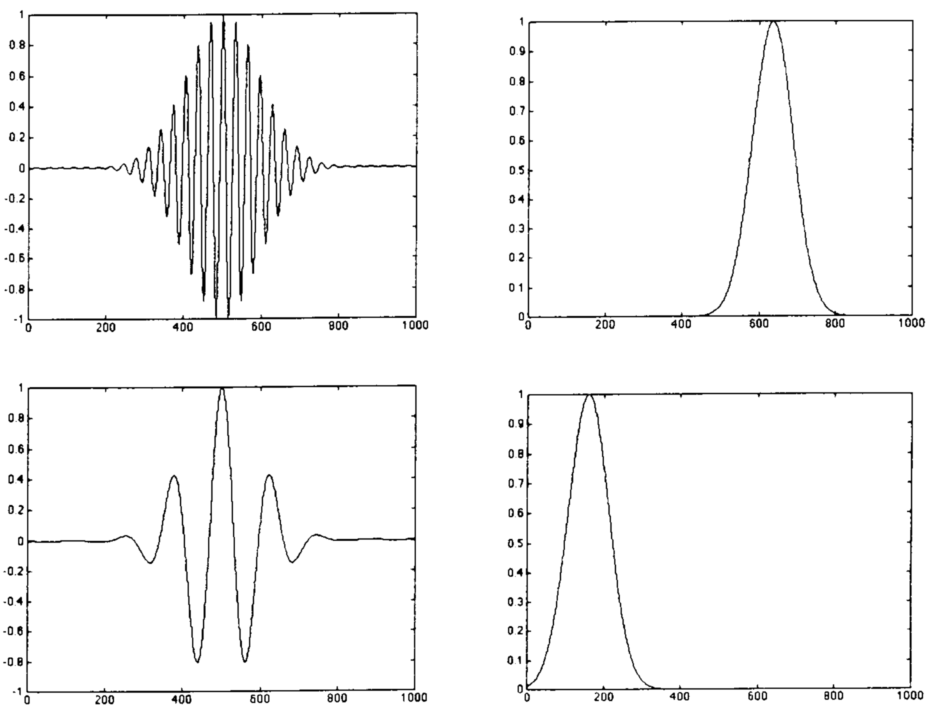

Figure 4. Two Gabor functions in the time domain (left), and their Fourier transform (right). In the Gabor representation, all the filters are obtained by shifting a 'mother' function in frequency, yielding a constant absolute bandwidth analysis.

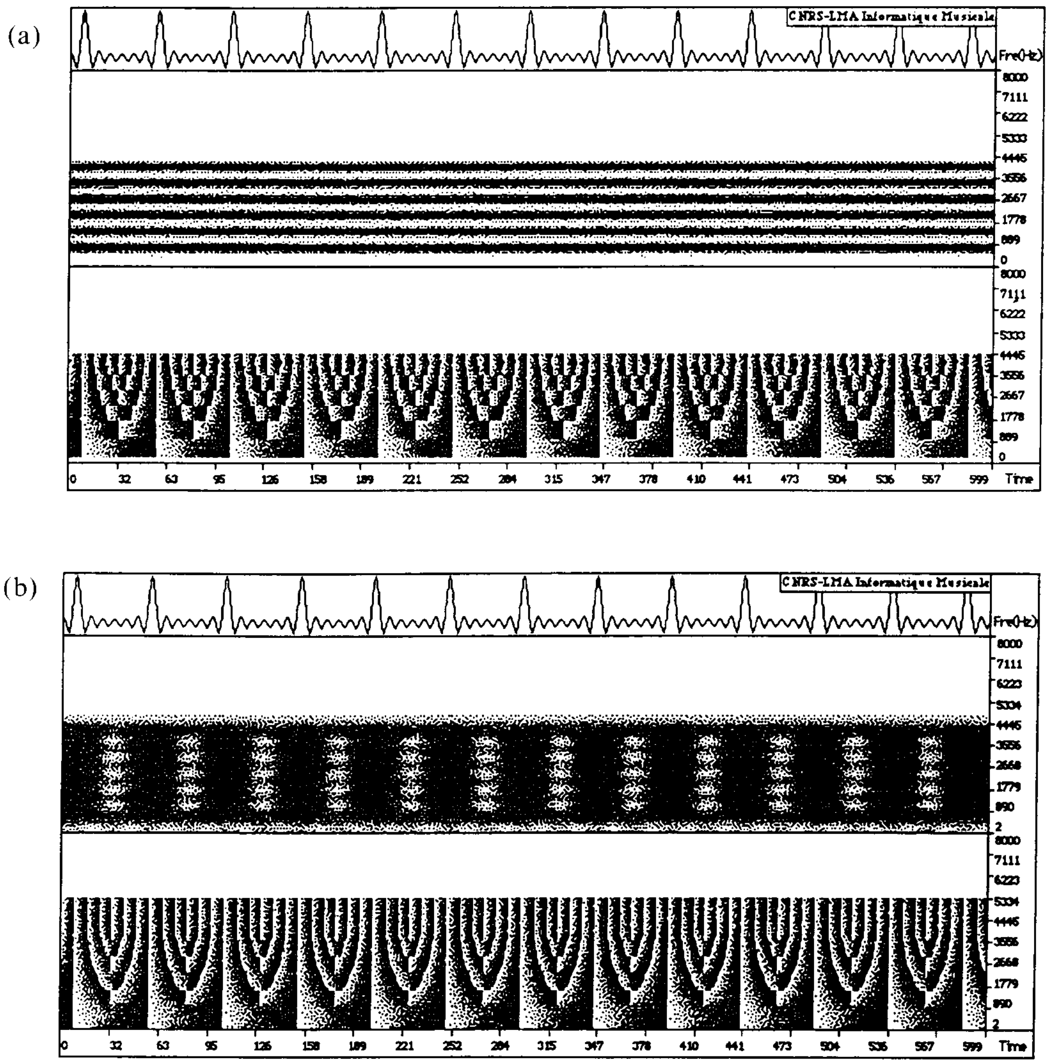

Figure 5. Gabor transform of the sum of six harmonic components analysed with two windows; the horizontal axis is time, the vertical axis is frequency. The upper picture is the modulus, the lower is the phase, represented by modulo-2π; their values are coded with a greyscale. In (a) the window is well localised in frequency, allowing the resolution of each component. In (b) the window is well localised with respect to time, leading to a bad separation of the components in the frequency domain, but showing impulses in time because the signal can also be considered as a filtered Dirac comb. In both figures, the phase behaves similarly, showing the periodicity of each component. This property has been used to estimate the frequencies of the components accurately.

inversely proportional to the scale. The analysis functions are all obtained from a 'mother' wavelet by translation and change of scale (dilation) (figure 6).

The 'mother' wavelet is a function with finite energy and zero mean value. These 'weak' conditions offer great freedom in the choice of this wavelet. One can, for example, imagine the decomposition of a speech signal in order to detect the word 'bonjour' pronounced at different pitches and with different durations. By using a 'mother' wavelet made of two wavelets separated, for example, by an octave, one can detect octave chords in a musical sequence (Kronland-Martinet 1988). This corresponds to a matched filtering at different scales. One important aspect of the wavelet transform is the localisation. By acting on the dilation parameter, the analysing function is automatically adapted to the size of the observed phenomena (figure 7). A high-frequency phenomenon should be analysed with a function that is well localised in time, whereas a low-frequency phenomenon requires a function well localised in frequency. This leads to an appropriate tool for the characterisation of transient signals (Guillemain *et al.* 1996). The particular geometry of the time–scale representation, where the dilation is represented according to a logarithmic scale (in fractions of octaves) permits the transform to be interpreted like a musical score associated with the analysed sound.

3.3. Parameter extraction

The parameter extraction method makes use of the qualitative information given by the time–frequency

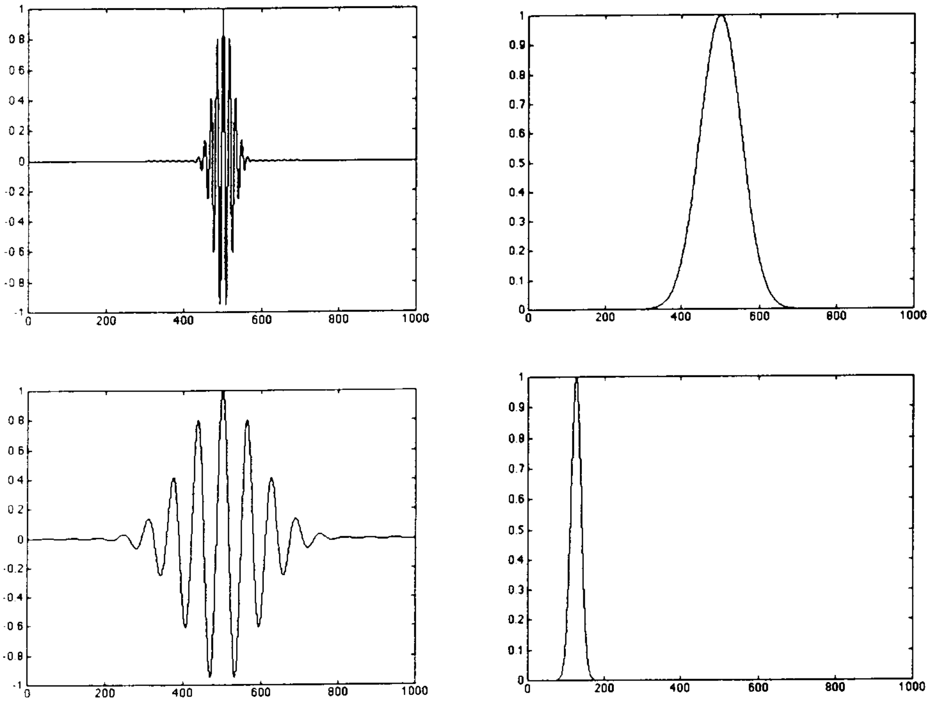

Figure 6. Two wavelets in the time domain (left), and their Fourier transform (right). In the wavelet representation, all the filters are obtained through dilation of a 'mother' function in time, yielding a constant relative ($\Delta\omega/\omega$) bandwidth analysis.

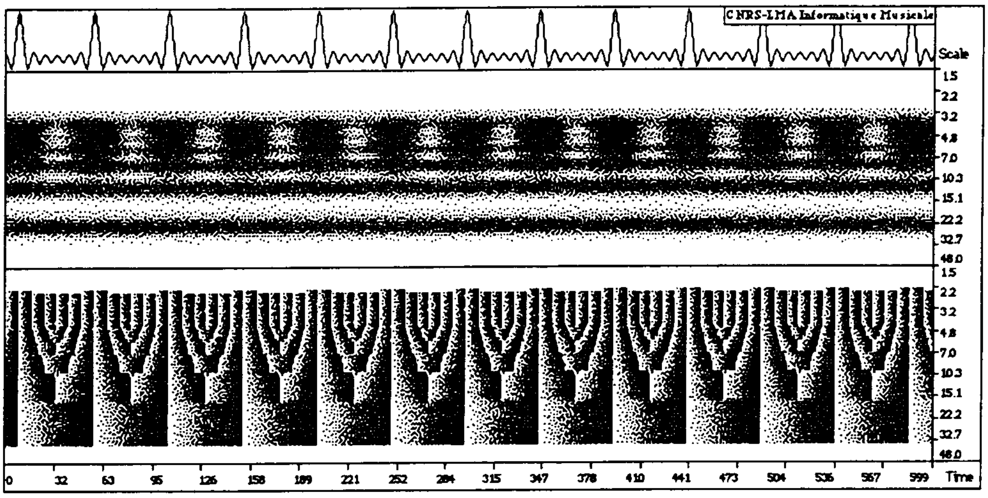

Figure 7. Wavelet transform of the same sum of six harmonic components. In contrast with figure 5 obtained through the Gabor transform, the wavelet transform privileges the frequency accuracy at low frequency (large scales) and the time accuracy at high frequency (small scales).

quantitative information from the signal. Even though the representations are not parametric, the character of the extracted information is generally determined by the supposed characteristics of the signal and by future applications. A useful representation for isolated musical instrument sounds is the additive model. It describes the sound as a sum of elementary components modulated in amplitude and in frequency, which is relevant from a physical and a perceptive point of view (figure 8).

Thus, to estimate parameters for an additive resynthesis of the sound, amplitude and frequency modulation laws associated with each partial should be extracted from the transform. Of course, this process must be efficient even for extracting components that are very close to each other and have rapidly changing amplitude modulation laws. Unfortunately, all the constraints for constructing the representation make this final operation complicated. The justification is of the same nature as the one given in the introduction in connection with sound transformation through modifying the representations. Absolute accuracy both in time and in frequency is impossible because of a mathematical relation between the transform at a point of the time–fre-

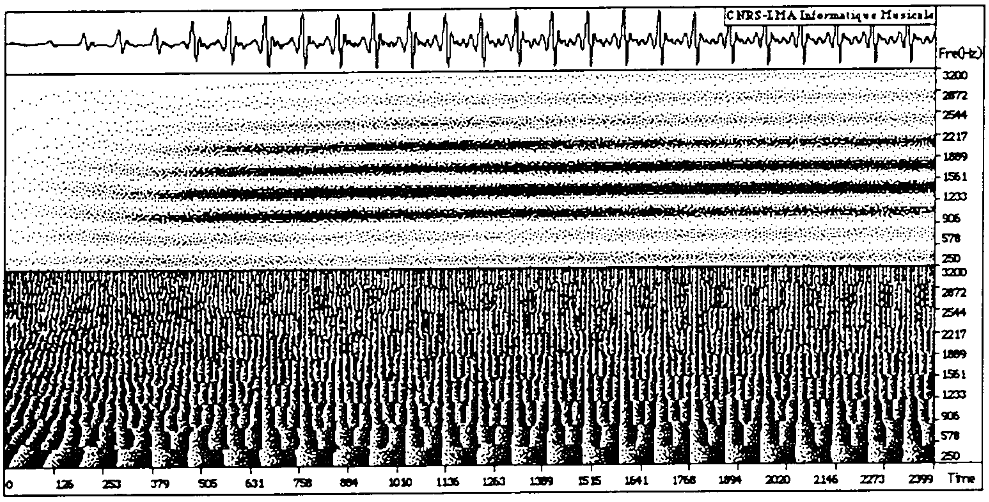

Figure 8. Gabor representation of the first 75 ms of a trumpet sound. Many harmonics with different time dependencies are visible in the modulus picture. The phase picture shows different regions, around each harmonic, where the phase wraps regularly at the time period of each harmonic, as in the previous figure.

Human hearing follows a rather similar 'uncertainty' principle: to identify the pitch of a pure sound, it must last for a certain time. The consequences of these limitations on the additive model parameter estimation are easy to understand. A high-frequency resolution necessitates analysis functions that are well localised in the frequency domain and therefore badly localised in the time domain. The extraction of the amplitude modulation law of a component from the modulus of the transform on a trajectory in the time–frequency plane smoothes the actual modulation law. This smoothing effect acts in a time interval with the same length as the analysis function. Conversely, the choice of well-localised analysis functions in the time domain generally yields oscillations in the estimated amplitude modulation laws, because of the presence of several components in the same analysis band. It is possible, however, to avoid this problem by astutely using the phase of the transform to precisely estimate the frequency of each component and by taking advantage of the linearity in order to separate them, without a hypothesis on the frequency selectivity of the analysis (figure 9).

The procedure uses linear combinations of analysis functions for different frequencies to construct a bank of filters with a quasi-perfect reconstruction. Each filter specifically estimates a component while conserving a good localisation in the time domain. Different kinds of filters can be designed, which permit an exact estimation of amplitude modulation laws locally polynomial on the time support of the filters (Guillemain *et al.* 1996) (figure 10).

The strict limitations of the wavelet transform or of the Gabor transform can be avoided by optimising the selectivity of the filter as a function of the density of the frequency components. The use of continuous transforms on the frequency axis is of great importance, since the central frequencies of the filters can be precisely calibrated at the frequencies of the components. Another important aspect of the musical sound is the frequency modulation of the components, in particular during the attack of the sound. Here the judicious use of the time derivative of the transform phase offers the possibility of developing iterative algorithms tracking the modulation laws, thus precluding the computation of the whole transform. These algorithms use frequency-modulated analysis functions, the modulations of which are automatically matched to the ones of the signal (Guillemain *et al.* 1996).

4. FEEDING THE SYNTHESIS MODELS

The extraction techniques using the time–frequency transforms directly provide a group of parameters which permit the resynthesis of a sound with the additive model. In addition, they can be used for identification of other synthesis models. The direct parameter identification techniques for the nonlinear models are difficult. Generally they do not give an exact reproduction of a given sound. The estimation criteria can be statistical (minimisation of nonlinear functions) (Horner 1996) or psychoacoustic (centroid of spectrum) (Beauchamp 1975). The direct estimation of physical or subtractive model parameters requires techniques like linear prediction, used, for instance, in speech synthesis (Markel and Gray 1976). Another solution consists in using parameters from the additive synthesis model to estimate another set of parameters corresponding to another synthesis model. In what follows we shall see how this operation can be done for the most common models.

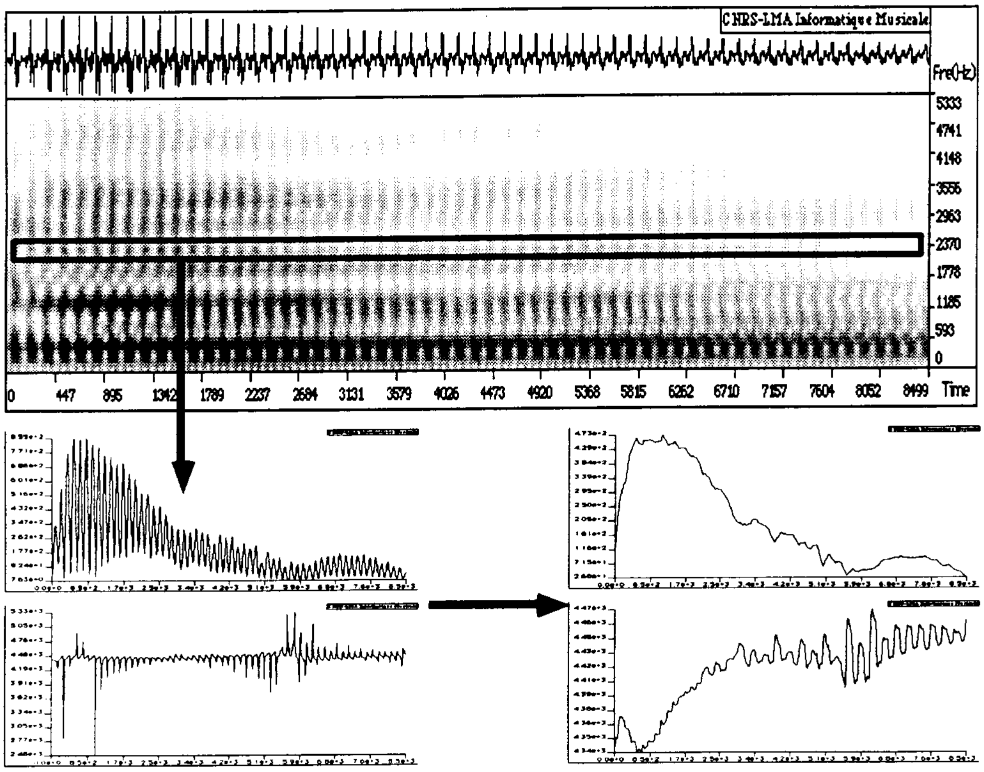

Figure 9. Estimation of the amplitude modulation law of a partial of a saxophone sound. The curves on the left show the estimated amplitude and frequency modulation laws using a straightforward Gabor transform. Several harmonics are present on the frequency support of the analysing function, yielding strong oscillations. The curves on the right show the estimated modulation laws using the filter bank displayed in figure 10. Although the time support remains the same, the oscillations are automatically cancelled by the algorithm.

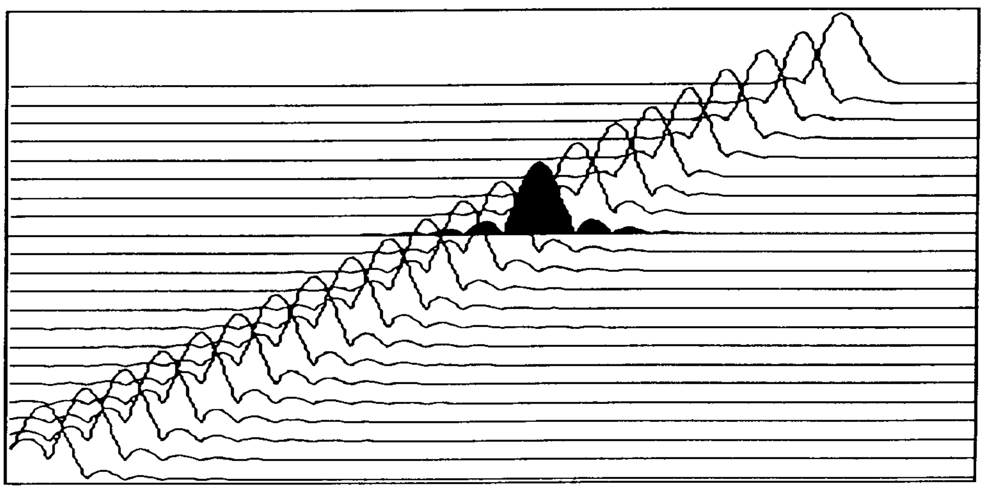

Figure 10. Filter bank in the frequency domain, allowing the estimation of spectral lines; one of the filters is darkened. The Fourier transform of each filter equals unity for the frequency it estimates, and zero for all the others. Its first derivative equals zero for all the frequencies. One can prove that this kind of filter allows exact estimation of locally linear amplitude modulation laws.

4.1. Additive synthesis

The parameter estimation for the additive model is the simplest one, since the parameters are determined in the analysis. The modelling of the envelopes can greatly reduce the data when one uses only perceptive criteria. The first reduction consists of associating each amplitude and frequency modulation law with a piecewise linear function (Horner and Beauchamp 1996) (figure 11). This makes it possible to automatically generate, for example, a Music V score associated with the sound.

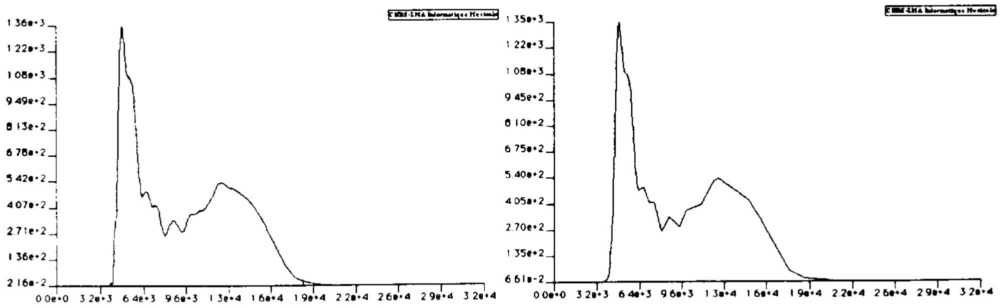

Figure 11. Original and modelled envelopes of a saxophone sound. The modelled curve is defined with thirty-five breakpoints and linear interpolation between them, while the original is defined on 32,000 samples.

Another possible reduction consists in grouping the components from the additive synthesis (group additive synthesis) (Kleczkowski 1989, Oates and Eagleston 1997). This can be done by statistical methods, such as principal component analysis, or by following an additive condition defined as the perceptual similarity between the amplitude modulations of the components (Kronland-Martinet and Guillemain 1993). This method offers a significant reduction in the number of synthesis parameters, since several components with a complex waveshape have the same amplitude modulation laws (figures 12 and 13).

4.2. Subtractive synthesis

An evolutive spectral envelope can be built by creating intermediate components obtained from the modulation laws of the additive modelling. Their amplitude modulation laws are obtained by interpolation of the envelopes of two adjacent components in the frequency domain (figure 14). These envelopes can then be used in order to 'sculpt' another sound (crossed synthesis). As we have already mentioned, physical modelling is sometimes close to subtractive synthesis. This aspect will be developed later.

4.3. Waveshaping and frequency modulation synthesis

From the parameters corresponding to the group additive synthesis (complex waves and their associated amplitude laws), one can deduce nonlinear synthesis parameters (Kronland-Martinet and Guillemain 1993). The technique consists of approaching each complex wave shape by an elementary nonlinear module. In the case of waveshaping, the knowledge of the complex wave allows the calculation of an exact distortion function. In the case of FM, the spectral components should be grouped, not only according to a perceptive criterion, but also according to a condition of spectral proximity. This condition is meaningful because real similarities between envelopes of neighbouring components are often observed. To generate the waveform corresponding to a group of components by an elementary FM oscillator, the perceptive approach is best suited. In that case, one can consider the energy and the spectral extent of the waveforms which are directly related to the modulation index. Other methods based on the minimisation of nonlinear functions by the simulated annealing or genetic algorithms have also been explored (Horner 1996). Attempts at direct estimation of the FM parameters by extraction of frequency modulation laws from the phase of the

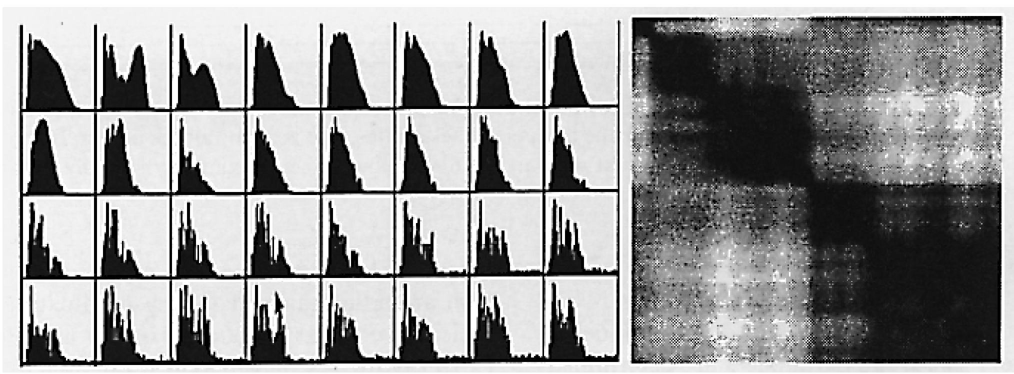

Figure 12. A whole set of envelopes of a violin sound, and the matrix showing the correlation between them. The dark regions around the diagonal correspond to curves that look similar and that correspond to components that are close in the frequency domain.

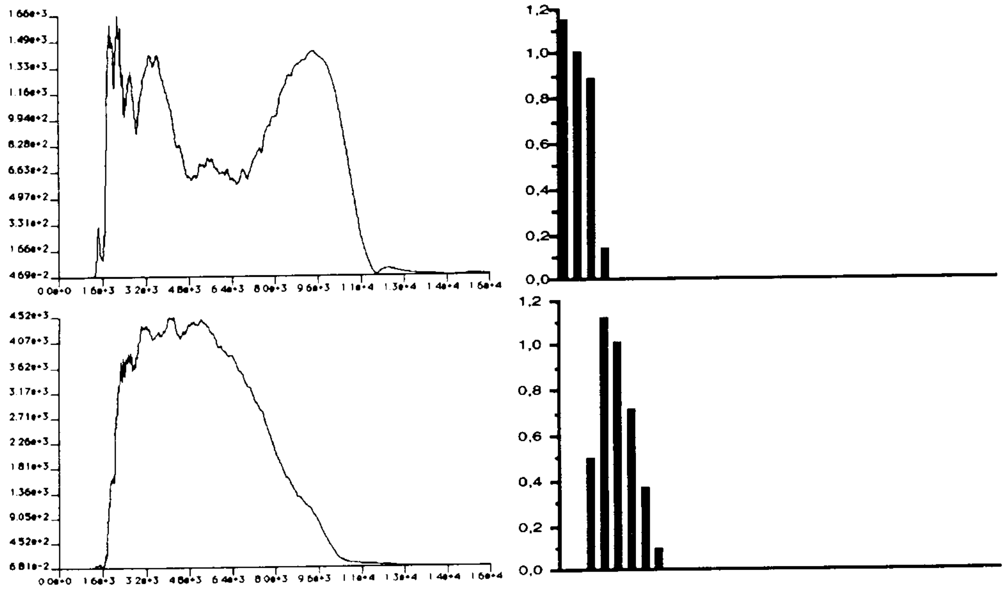

Figure 13. Two main envelopes of the group additive synthesis model, with the spectrum of their associated waveform. Psychoacoustic criteria can be used to generate a perceptively similar spectrum with nonlinear techniques.

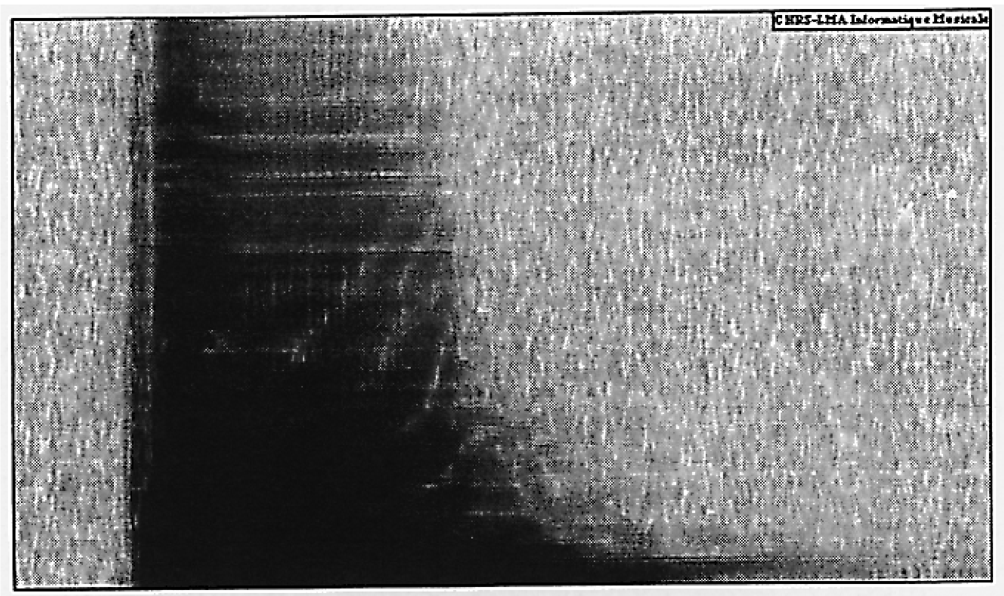

Figure 14. Spectral envelope of a saxophone sound built from the additive synthesis parameters. This envelope can be used to 'sculpt' the modulus of the Gabor transform of another sound in order to perform a crossed synthesis.

analytic signal related to the real sound have led to interesting results (Justice 1979, Delprat, Guillemain and Kronland-Martinet 1990).

4.4. Waveguide

The waveguide synthesis parameters are of a different kind. They characterise both the medium where the waves propagate and the way this medium is excited. From a physical point of view, it is difficult to separate these two aspects: the air jet of a wind instrument causes vortex sheddings interacting with the acoustic pressure in the tube (Verge 1995); the piano hammer modifies the characteristics of a string while it is in contact with it (Weinreich 1977). These source–resonator interactions are generally nonlinear and often difficult to model physically. However, a simple linear waveguide model often gives satisfactory sound results. In a general way, the study of linear wave propagation equations in a bounded medium shows that the response to a transient excitation can be written as a sum of exponentially damped sine functions. The inharmonicity is related to the dispersive characteristics of the propagation medium, the decay times are related to the dissipative characteristics of the medium, and the amplitudes are related to the spectrum of the excitation. In the same way, the impulse

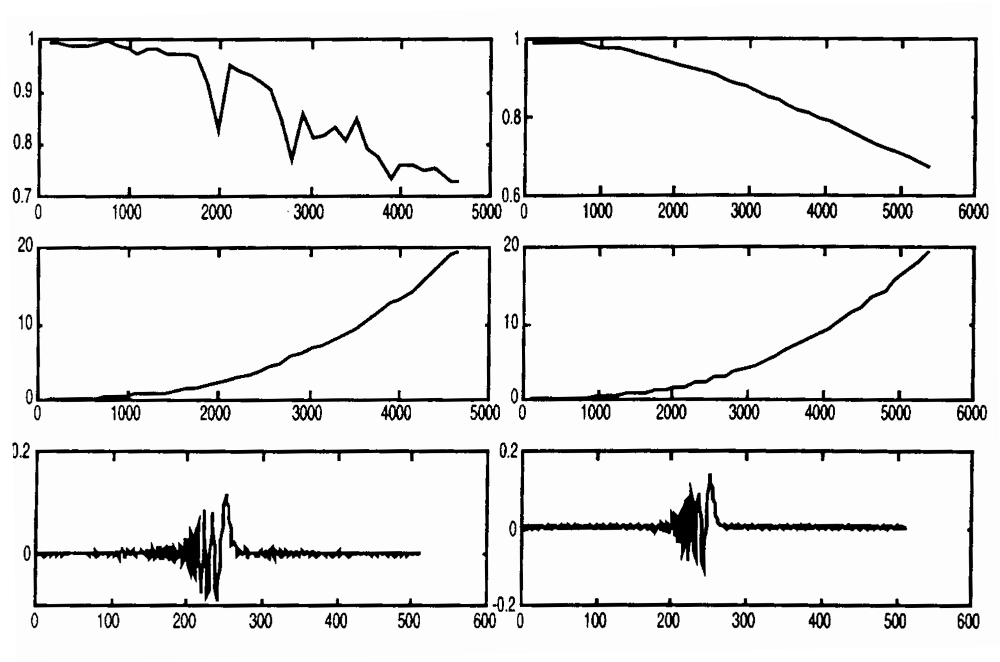

Figure 15. Parameter estimation for the waveguide model can be performed either from the solution of the partial differential equations of the string movement, or from the estimated damping factors and frequencies of the partials. Pictures on the left show the data from the estimation. Pictures on the right show the data from the movement equation of a stiff string. The figure shows, from top to bottom: modulus (related to losses during the propagation), phase derivative (related to the dispersion law of the propagation medium) of the Fourier transform of the filter inside the loop, and impulse response of the loop filter. The good agreement between theory and experimentation in this case can be used to fit mechanical parameters of the string from the experimentation.

approximated by a sum of exponentially damped sinusoids whose frequencies, amplitudes and damping rates are related in a simple way to the filter coefficients (Ystad, Guillemain and Kronland-Martinet 1996). Thanks to the additive synthesis parameters one can, for the percussive sound class, determine the parameters of the waveguide model, and also recover the physical parameters characterising the instrument (Guillemain, Kronland-Martinet and Ystad 1997) (figure 15). For sustained sounds, the estimation problem of the exciting source is crucial and necessitates the use of deconvolution techniques. This approach is entirely nonparametric, but it is also possible to use parametric techniques. Indeed, the discrete time formulation of the synthesis algorithm corresponds to a modelling of the so-called ARMA type (AutoRegressive Moving Average).

5. CONCLUSION

The modelling of sounds brings together the algorithmic synthesis process and the shaping of natural sounds. Such modelling may serve to develop an appropriate ‘algorithmic sampler’ to make all the intimate modifications offered by a mathematical description of the sounds. Time–frequency and time–scale representations of signals are helpful to extract relevant parameters describing the sounds. Both signal synthesis models and physical synthesis models can be fed in order to resynthesise and shape a given musical sound. Even though most musical sounds can be modelled from additive synthesis data, stochastic or very noisy sounds still remain difficult to model. Work is being conducted to fill this gap in order to offer in the near future a genuine sound simulator to musicians.

REFERENCES

Allen, J. B., and Rabiner, L. R. 1977. A unified approach to short-time Fourier analysis and synthesis. *Proc. of the IEEE* **65**: 1558–64.

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**: 757–68.

Arfib, D., and Delprat, N. 1993. Musical transformations using the modifications of time–frequency images. *Computer Music Journal* **17**(2): 66–72.

Arfib, D., Guillemain, P., and Kronland-Martinet, R. 1992. The algorithmic sampler: an analysis problem? *Journal of the Acoustical Society of America* **92**: 2451.

Atal, B. S., and Hanauer, S. L. 1971. Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America* **50**: 637–55.

Beauchamp, J. W. 1975. Analysis and synthesis of cornet tones using non-linear interharmonic relationships. *Journal of the Audio Engineering Society* **23**: 778–95.

Cadoz, C., Luciani, A., and Florens, J. L. 1984. Responsive input devices and sound synthesis by simulation of instrumental mechanisms. *Computer Music Journal* **8**(3): 60–73.

Chaigne, A. 1995. Trends and challenges in physical modeling of musical instruments. *Proc. of the ICMA*, Vol. III, pp. 397–400. Trondheim, Norway, 26–30 June.

Cheung, N. M., and Horner, A. 1996. Group synthesis with genetic algorithms. *Journal of the Audio Engineering Society* **44**: 130–47.

Chowning, J. 1973. The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society* **21**: 526–34.

Cook, P. R. 1992. A meta-wind-instrument physical model controller, and a meta-controller for real-time performance control. *Proc. of the 1992 Int. Computer Music Conf.*, pp. 273–6. San Francisco: International Computer Music Association.

Delprat, N., Guillemain, P., and Kronland-Martinet, R. 1990. Parameter estimation for non-linear resynthesis methods with the help of a time–frequency analysis of natural sounds. *Proc. of the 1990 Int. Computer Music Conf.,* pp. 88–90. Glasgow: International Computer Music Association.

Depalle, P., and Rodet, X. 1992. A new additive synthesis method using inverse Fourier transform and spectral envelopes. *Proc. of the 1992 Int. Computer Music Conf.*, pp. 161–4. San Francisco: International Computer Music Association.

De Poli, G., Picciali, A., and Roads, C. (eds.) 1991. *The Representation of Musical Signals*. Cambridge, MA: MIT Press.

Dolson, M. 1986. The phase vocoder: a tutorial. *Computer Music Journal* **10**(4): 14–27.

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

Flandrin, P. 1993. *Temps-frequence*. Hermes. Traite des nouvelles technologies, serie traitement du signal.

Gabor, D. 1947. Acoustical quanta and the nature of hearing. *Nature* **159**(4044): 591–4.

Grossman, A., and Morlet, J. 1984. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal of Mathematical Analysis* **15**: 723–36.

Guillemain, Ph., and Kronland-Martinet, R. 1996. Characterisation of acoustics signals through continuous linear time–frequency representations. *Proc. of the IEEE*, Special Issue on Wavelets, **84**(4): 561–85.

Guillemain, Ph., Kronland-Martinet, R., and Ystad, S. 1997. Physical modelling based on the analysis of real sounds. *Proc. of the Institute of Acoustics*, Vol. 19, pp. 445–50. Edinburgh: ICMA97.

Horner, A. 1996. Double-modulator FM matching of instruments tones. *Computer Music Journal* **20**(2): 57–71.

Horner, A., and Beauchamp, J. 1996. Piecewise-linear approximation of additive synthesis envelopes: a comparison of various methods. *Computer Music Journal* **20**(2): 72–95.

Justice, J. 1979. Analytic signal processing in music computation. *IEEE Trans. on Speech, Acoustics and Signal Processing* **ASSP-27**: 670–84.

Kleczkowski, P. 1989. Group additive synthesis. *Computer Music Journal* **13**(1): 12–20.

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

Kronland-Martinet, R. 1989. Digital subtractive synthesis of signals based on the analysis of natural sounds. In *Etat de la Recherche Musicale (au 1er janvier 1989).* Ed. A.R.C.A.M., Aix en Provence.

Kronland-Martinet, R., and Guillemain, P. 1993. Towards non-linear resynthesis of instrumental sounds. *Proc. of the 1993 Int. Computer Music Conf.*, pp. 86–93. San Francisco: International Computer Music Association.

Kronland-Martinet, R., Morlet, J., and Grossman, A. 1987. Analysis of sound patterns through wavelet transforms. *International Journal of Pattern Recognition and Artificial Intelligence* **11**(2): 97–126.

Laroche, J. 1993. The use of the matrix pencil method for the spectrum analysis of musical signals. *Journal of the Acoustical Society of America* **94**: 1958–65.

Le Brun, M. 1979. Digital waveshaping synthesis. *Journal of the Audio Engineering Society* **27**: 250–66.

Makhoul, J. 1975. Linear prediction, a tutorial review. *Proc. of the IEEE* **63**: 561–80.

Markel, J. D., and Gray, A. H. 1976. Linear prediction of speech. *Communication and Cybernetics* **12**. Berlin, Heidelberg, New York: Springer-Verlag.

McAulay, R., and Quatieri, T. 1986. Speech analysis–synthesis based on a sinusoidal representation. *IEEE Trans. on Speech, Acoustics and Signal Processing* **ASSP-34**: 744–54.

Moorer, J. A. 1978. The use of the phase vocoder in computer music applications. *Journal of the Audio Engineering Society* **26**: 42–5.

Oates, S., and Eagleston, B. 1997. Analytic methods for group additive synthesis. *Computer Music Journal* **21**(2): 21–39.

Rabiner, L. R., and Gold, B. 1975. *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice Hall.

Risset, J. C. 1965. Computer study of trumpet tones. *Journal of the Acoustical Society of America* **33**: 912.

Risset, J. C., and Wessel, D. L. 1982. Exploration of timbre by analysis and synthesis. In D. Deutsch (ed.) *The Psychology of Music*, pp. 26–58. New York: Academic Press.

Roads, C. 1978. Automated granular synthesis of sound. *Computer Music Journal* **2**(2): 61–2.

Ruskai, M. B., Beylkin, G., Coifman, R., Daubechies, I., Mallat, S., Meyer, Y., and Raphael, L. (eds.) 1992. *Wavelets and their Applications*. Boston: Jones and Bartlett.

Smith, J. 1992. Physical modeling using digital waveguides. *Computer Music Journal* **16**(4): 74–91.

Verge, M. P. 1995. *Aeroacoustics of Confined Jets with Applications to the Physical Modeling of Recorder-like Instruments*. PhD Thesis, Eindhoven University.

Weinreich, G. 1977. Coupled piano strings. *Journal of the Acoustical Society of America* **62**: 1474–84.

Ystad, S., Guillemain, Ph., and Kronland-Martinet, R. 1996. Estimation of parameters corresponding to a propagative synthesis model through the analysis of real sounds. *Proc. of the 1996 Int. Computer Music Conf.,* pp. 19–24. Hong Kong: International Computer Music Association.