UNIVERSITE DE LA MEDITERRANEE - AIX-MARSEILLE II
FACULTE DES SCIENCES DE LUMINY

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE AIX-MARSEILLE II

Discipline: Acoustique, Traitement du signal et Informatique Appliqués à la Musique

présentée et soutenue publiquement par

YSTAD Sølvi

le 11 mars 1998

SOUND MODELING USING A COMBINATION OF PHYSICAL AND SIGNAL MODELS

JURY

Dr. RISSET Jean-Claude	Président
Pr. DE POLI Giovanni	Rapporteur
Pr. KROKSTAD Asbjørn	Rapporteur
Dr. ARFIB Daniel	Co-directeur de thèse
Pr. TRO Jan	Co-directeur de thèse
Dr. GUILLEMAIN Philippe	
Pr. KRISTIANSEN Ulf	

Acknowledgments

I would like to thank the Norwegian Research Council for having supported this work. I also thank NoTAM -Norsk nettverk for Teknologi, Akustikk og Musikk- and Ny Musikk -the Norwegian section of ISCM (International Society for Contemporary Music)- for having encouraged this project.

A mutual agreement in connection with this work was signed between the Université de la Mediterranée, Aix-Marseille II in Marseille, France and the NTNU -Norges Teknisk-Naturvitenskapelige Universitet- in Trondheim, Norway. The aim of this agreement was to favor a collaboration between the two universities, and more precisely between the research group « Informatique Musicale » at the CNRS, LMA (Centre National de la Recherche Scientifique, Laboratoire de Mécanique et d'Acoustique) and the Acoustics Group at NTNU. Consequently, this thesis has been validated by both universities, and the work has been co-supervised by Daniel Arfib, CNRS Marseille and Professor Jan Tro, NTNU Trondheim. I want to thank my supervisors for having trusted my project and my work. I also thank the French Ministry of Education and Scientific Research for having supported this collaboration.

The main part of the project has been effectuated at the CNRS, LMA in the research group Informatique Musicale, and I would therefore like to thank Jean-Claude Risset the chairman of this research group, as well as the members of his group for having accepted to let me work with them. Special thanks to Thierry Voinier (Real-time GURU), Philippe Guillemain (Spectral Line Estimation GURU) and Richard Kronland-Martinet (GURU of brilliant ideas) for having so generously shared their knowledge and their precious time working with me on this project.

I further would like to thank Professor Asbjørn Krokstad and Professor Giovanni De Poli for having accepted to review this work. Thanks also to Jean-Claude Risset, Philippe Guillemain and Professor Ulf Kristiansen for having accepted to be members of the jury of my thesis (together with the supervisors and the reviewers).

SOUND MODELING USING A COMBINATION OF PHYSICAL AND SIGNAL MODELS

1. Introduction: the sound modeling concept

Sound modeling 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 realize sound simulators using synthesis methods based on signal models and/or physical models.

Analysis-synthesis consists of a set of procedures the purpose of which is to collect information about a natural sound and to reconstruct it. Different methods can be applied, and the success of each method depends on their adequacy with the sound effect to be produced. *Figure 1.1* shows the most commonly used procedures in a schematic way. 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 makes it possible to classify and characterize audio signals [Kronland-Martinet et al., 1987], but the result of the process is simply a reproduction of the natural sound. From a musical point of view, a representation of sounds by analysis is useful for intimate modifications of a sound. This sound transformation process corresponds to the upper path in *Figure 1.1* and is performed by altering the representation between the analysis and the synthesis procedures. The way in which the sound can be altered depends upon the choice of the analysis method. The energy distribution and/or the frequencies of the spectral components can, for example, be manipulated through spectral analysis. The time-frequency analysis allows a separation of the time and frequency characteristics associated with the sound, and can be very useful. However, this approach conflicts with a very important mathematical principle which states that one can not 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 stating that one cannot be as precise as one wishes in the localization of both the time and the frequency domains. This constraint strongly limits the number of applicable time-frequency transformation processes, and makes the search for adequacy between the altered values and the obtained sounds difficult. Anyway, very interesting sounds can be obtained by carefully using such altering procedures [Arfib et al., 1993]. I shall in this document specially pay attention to the lower part of *Figure 1.1* which corresponds to the sound modeling. In this part, the representations obtained from the analysis provide parameters corresponding to the synthesis models. We now get to the concept of the algorithmic sampler [Arfib et al., 1992] which consists of simulating natural sounds through a synthesis process that is well adapted to algorithmic and real-time manipulations. The resynthesis and the transformation of natural sounds are then part of the same concept.

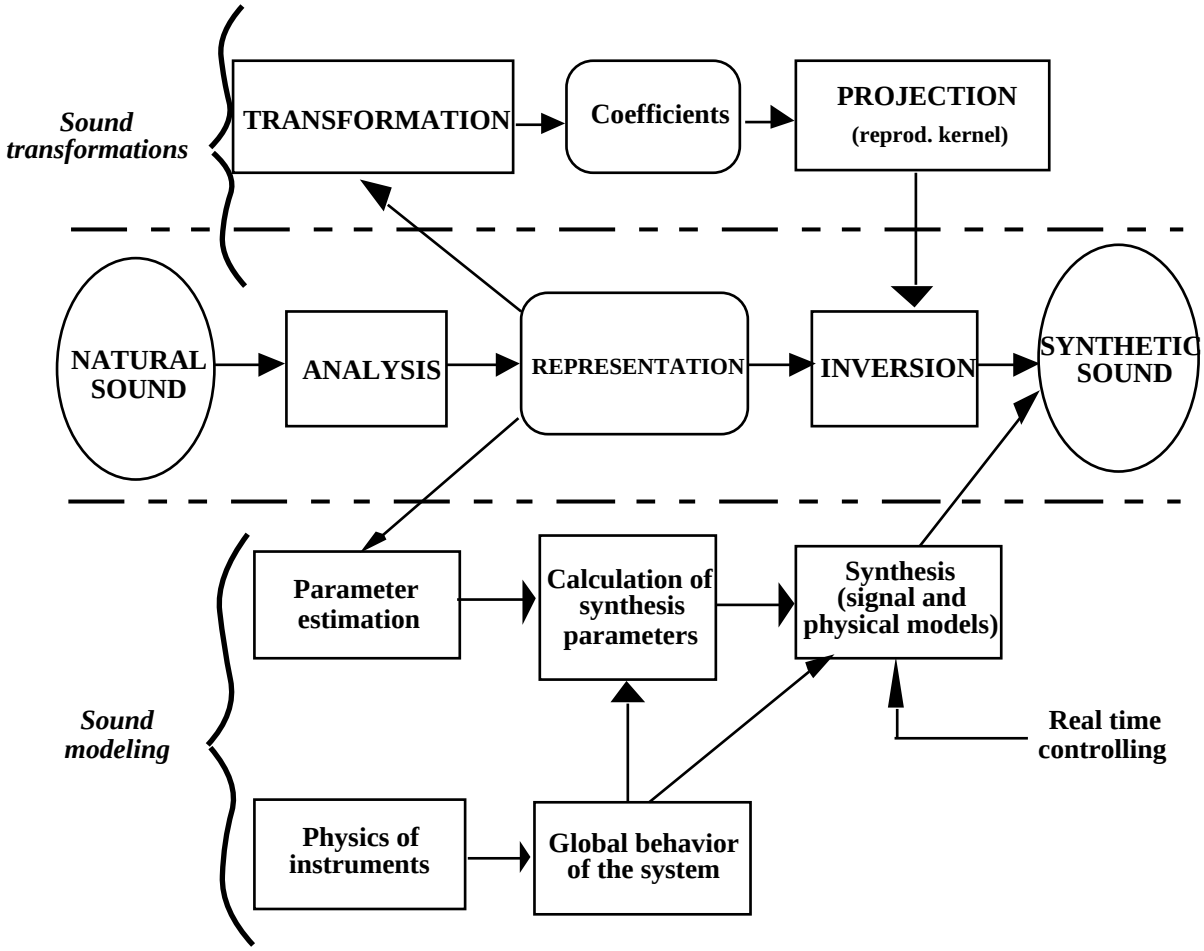


Figure 1.1: General organization of the analysis-synthesis and modeling concept.

This document is organized as follows:

In chapter 2, I describe the main synthesis techniques with particular emphasis on the so-called « physical synthesis ». This rather new approach is interesting when taking into account the control part of the model since the parameters used in the model are physically meaningful. In chapter 3, I briefly describe the analysis techniques that have proved to be useful when focusing on sound signals and how parameters can be extracted from the analysis for synthesis purposes. I particularly describe a « matched time-frequency » analysis technique I have designed to analyze resonant systems which takes into account the specificities of transient sounds. Chapter 4 gives an example of estimation of parameters corresponding to a group additive synthesis process modeling a flute sound. The difficulties in controlling such a model and the lack of important aspects of the sound like for instance the noise necessitates the use of a more physical description of the sound generator. Chapter 5 addresses this problem by connecting a physical description of one-dimensional systems and « propagative synthesis » models simulating phenomena such as dissipation and dispersion occurring during wave propagation in the medium. This approach is original and gives way to the characterization and the construction of the elements of the digital model. Further on, in order to simulate sustained sounds, I

present in chapter 6 a so-called hybrid model where the source and the resonator are modeled separately by respectively signal and physical models. In this approach, the source is obtained by a deconvolution of the sound with respect to the resonator. It has further been splitted into a deterministic and a stochastic part which have been modeled separately. The deterministic part is modeled by a non-linear synthesis model adjusted through psychoacoustic criteria such as the tristimulus. The stochastic part takes into account both the density of probability and the power spectral density provided by the analysis process and can be modeled by linear filtering of a white noise. Even though the methods described can be applied to numerous sound generators, most of the methods are illustrated using the flute signal. This signal represents a good way of testing the model, since it is generated by a very complicated system involving resonant systems coupled to a turbulent air jet. In chapter 7 it is shown how such a model can be controlled in real time using an adequate interface. A flute interface devoted to the real-time control of a flute model is described. This interface makes it possible to pilot the proposed model by a real flute offering possibilities for making sound transformations from a traditional instrument. This interface may be useful to musicians.

2. Digital synthesis of sounds

Digital synthesis uses methods of signal generation that can be divided into two classes:
- signal models aiming at reconstructing a perceptive effect without being concerned with the specific source that made the sound.
- physical models aiming at simulating the behavior of existing or virtual sound sources by taking into account the most important physical features.

2.1. Signal synthesis models

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 edification of structures from materials, and the three principal groups can be classified as follows:
- additive synthesis
- subtractive synthesis
- global (or non-linear) synthesis

2.1.1. Additive synthesis

Like a building being built by piling up materials (bricks, stones,...), 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 doesn't consider the micro variations that can be found from the amplitude and frequency modulation laws of the components of any real sound. These dynamic laws must therefore be very precise when reproducing a real sound. The advantage of these synthesis methods is essentially the possibilities of 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 up elementary signals (grains) localized in both the time and the frequency domains [Roads, 1978].

The parameters defining an additive synthesis model are given by the amplitude modulation laws A_k and the frequency modulation laws f_k of the components. The synthesis process is then obtained by:

$$s(t) = \sum_{k=1}^{K} A_k(t)\cos\left(2\pi \int_0^t f_k(u)du\right),$$

where K is the number of the spectral components.

A real sound can contain up to one hundred significant components. This means that additive syntesis models make use of a great number of parameters which are difficult to control without constructing an intermediate software layer which makes it possible to act on the parameters in a global way.

2.1.2. Subtractive synthesis

Like a shapeless stone being sculpted by removing unwanted parts, a sound can be constructed by removing undesired components from an initial complex sound such as a noise. This synthesis technique is closely linked to the theory of digital filtering [Rabiner et al., 1975] and can be related to some physical sound generation systems like for instance the speech signal [Atal et al., 1971] [Flanagan et al., 1970]. The advantage of this approach (if we omit the physical aspects which will be discussed when describing synthesis models by physical modeling) is the possibility of uncoupling the excitation source and the resonance system. The sound transformations related to these methods often use this property in order 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, 1988]. A well-known crossed synthesis result is for example the sound of a talking cello obtained by associating an excitation corresponding to the string and a resonance system corresponding to the time-varying formants (spectral bumps related to the modes) of the vocal tract.

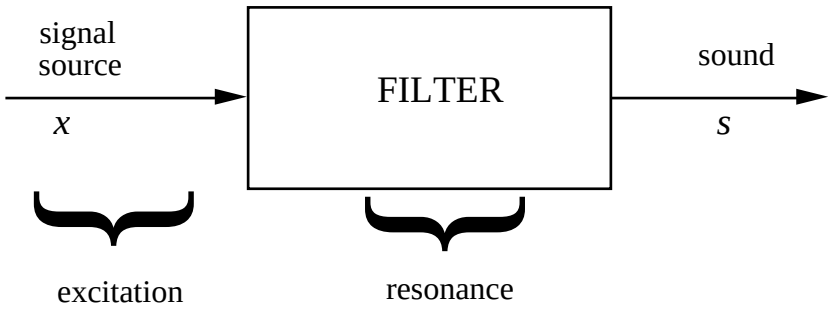


Figure 2.1 Illustration of a subtractive synthesis process.

The general form of the synthesis process is given by the discrete temporal equation obtained through an ARMA (autoregressive moving average) model [Max, 1987].

$$s(n) = \sum_{k=0}^{K} a_k x(n-k) + \sum_{l=1}^{L} b_l s(n-l),$$

where the coefficients a_k are related to the zeros of the filter, and the coefficients b_k to the poles.

Actually the frequency response of the system is given by:

$$H(\omega) = \frac{\sum_{k=0}^{K} a_k e^{-i\omega k}}{1 - \sum_{l=1}^{L} b_l e^{-i\omega l}}$$

2.1.3. Non-linear synthesis

Like modeling different objects from a block of clay, a simple and "inert" signal can be dynamically modeled using global synthesis models. These methods are non-linear since the operations on the signals are not simple additions and amplifications. The most well-known example of global synthesis is undoubtedly the audio Frequency Modulation (FM) introduced by John Chowning [Chowning, 1973] which has led to a « revolution » for the commercial synthesizers. The elementary synthesis process for this method is given by:

$$s(t) = A(t)\sin(\alpha t + I(t)\sin\beta t),$$

where α is the carrier frequency, β is the modulant frequency and *I(t)* is the index of modulation. The spectrum of *s(t)* contains components at frequencies $\alpha \pm n\beta$ with amplitudes given by the Bessel functions $J_n(I)$. Actually,

$$s(t) = A(t)\sum_{n=-\infty}^{n=\infty} J_n(I)\sin(\alpha + n\beta)$$

One can also generate asymetrical spectra by slightly modifying this process [Palamin et al., 1988]. The advantage of the FM is that it makes use of very few parameters, and that a small number of operations can generate complex spectra. This simplifies the numerical implementation on one hand and the control on the other hand. 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 non-linear way and the continuous modification of these parameters may give fast changes in the timre of the sound.

Other related methods have proved to be efficient for signal synthesis, such as the waveshaping technique [Arfib, 1979] [Lebrun, 1979]. I shall briefly recall the mathematics of this method since it will be used to model the deterministic part of a source signal in chapter 6.

The synthesis by waveshaping is given by:

$$s(t) = \gamma(I(t)\cos(\omega_0 t)),$$

where $\gamma(x)$ is a non-linear function and

$I(t)$ is the index of distortion.

The non-linear function can be decomposed in different bases. When the purpose is to reproduce a spectral evolution with spectral amplitudes that evolute as a function of the index, the Chebyshev's polynomials of the first kind are often used. These polynomials are defined by [Gradshteyn et al., 1980]:

$$T_n(x) = \frac{1}{2}\left[\left(x + i\sqrt{1-x^2}\right)^n + \left(x - i\sqrt{1-x^2}\right)^n\right] = \cos(n \arccos x)$$

and satisfy the recursive relation

$$T_{n+1}(x) - 2xT_n(x) + T_{n-1}(x) = 0$$

with $T_0(x) = 1$ and $T_1(x) = x$.

The Chebyshev's polynomials are orthogonal with respect to the measure $\frac{dx}{\sqrt{1-x^2}}$ on the subset [-1,1], that is

$$\int_{-1}^{1} \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}}dx = \begin{cases} \delta_{mn}\dfrac{\pi}{2} & m \neq 0 \\ \delta_{mn}\pi & m = 0 \end{cases}$$

One can then write:

$$\gamma(x) = \sum_{k=0}^{K} \alpha_k T_k(x)$$

When the index of distortion is *I(t)=1*, the output is given by:

$$\gamma(\cos\omega_0 t) = \sum_{k=0}^{\infty} \alpha_k T_k(\cos\omega_0 t) = \sum_{k=0}^{\infty} \alpha_k \cos k\omega_0 t$$

In this case the α_k's represent the amplitudes of the output spectral components. This means that the non-linear function has « created » harmonic components of the fundamental frequency ω_0. For other values of the index, the spectrum of the generated signal varies accordingly to:

$$\gamma(I(t)\cos\omega_0 t) = \sum_{k=0}^{\infty} \alpha_k T_k(I(t)\cos\omega_0 t)$$

leading to different sounds. This method shall be used when generating the deterministic part of the source signal in the proposed flute model (chapter 6).

2.2. Physical synthesis models

Physical synthesis is a more recent technique that I shall describe more precisely than signal model synthesis since it will widely be used in chapter 5.

Unlike signal models using a purely mathematical description of sounds, physical models describe the sound generation system using physical considerations. Such models can either be constructed from the equations describing the behavior of the waves propagating in the structure and their radiation in air, or from the behavior of the solution of the same equations. The first approach is costly in terms of calculations and is generally only used in connection with research work [Chaigne, 1995], unless one uses a simplified version consisting of modeling the structure by an association of simple elements (masses, springs, dampers...) [Cadoz et al., 1984]. Synthesis by simulation of the solution of the propagation equation has led to waveguide synthesis models [Smith, 1992], which have the advantage of being easy to construct with a behavior close to that of a real sound generator. Thus such synthesis methods are well adapted to the modeling of resonant sound generators such as musical instruments. The principles of these methods and the models they lead to are described, and it is shown how the parameters of these models relate to physical phenomenas. These parameters are related to the structure of the sound generator as well as to the control during performance. If we consider for example a vibrating string, the Shannon's sampling theorem states that one can, without loss of information, split the movement into a succession of instantaneous clichés 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 consider the propagation as a passage from one elementary cell to another. This operation corresponds to a spatial discretization of the structure, and it makes it possible to consider the wave propagation as the result of a succession of transformations or filtering of the initial solicitation.

In the ideal case where we neglect losses and non-linearities, the excitation of the medium will lead to 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, symbolized in digital signal processing by the variable z^{-1}. In a more realistic case where 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 » (an all-pass filter) D should be added (*Figure 2.2*).

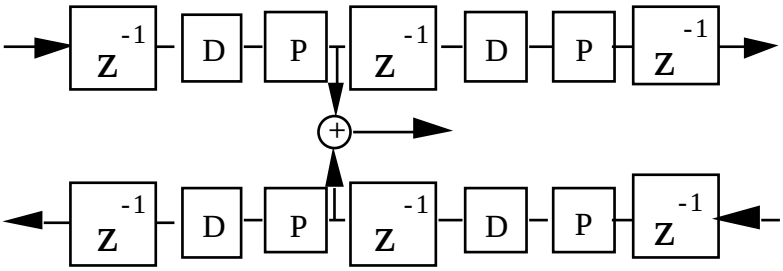


Figure 2.2 Discrete simulation of the wave propagation in a dissipative and dispersive medium.

Since the system is linear, the theory of digital filters allows the gathering of elements of the same type. Thus, the propagation medium can be represented by a succession of 3 elements, that is a delay line, an attenuating filter accounting for the dissipation and an all pass filter accounting for the dispersion. Real musical instruments represent media of finite length which means that the propagative waves are reflected at the ends. The reflections correspond to returning the initial waves with a modification depending on the boundary conditions. These waves are solutions of a boundary value problem. They can be simulated thanks to a looped system with a delay line, an attenuating filter, an all pass filter, and a filter corresponding to the reflections R (*Figure 2.3*).

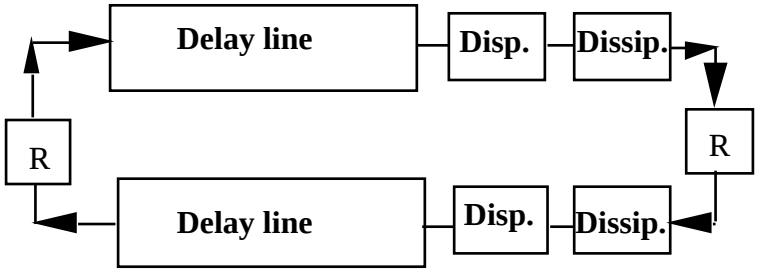


Figure 2.3: Propagation model in a bounded dissipative and dispersive medium.

Synthesis models related to a particular digital filter such as a second order IIR filter are known as waveguide models. They can be used to simulate many different systems, such as for instance finite length strings or tubes representing the resonant system in wind instruments [Cook, 1992].

3. Analysis of sounds

The analysis of natural sounds requires several methods to describe or represent pertinent physical and perceptive characteristics of the sound [Risset et al., 1982]. Even though the spectral content of a sound is often of great importance, the time evolution 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. When creating a sound mimicking a plucked vibrating string the different decay times of the partials are important. These examples illustrate the need for analysis methods giving access to time and frequency variations of sounds, and for this purpose 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 non-parametric methods. The parametric methods require an a-priori knowledge of the signal, and the desired sound is obtained by adjusting the parameters of a model. The non-parametric models work without any a-priori knowledge of the signal to be analyzed, but they often require a large number of coefficients to be determined.

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 probably the linear prediction (LPC) [Makhoul, 1975]. This technique is adapted to signals from sound production systems of the source-resonance type. The resonant filter should be modeled by a digital all pole filter whose coefficients are related to the frequency and to the width of the formants. The applications to the analysis-synthesis of 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 broad-band noise or a periodic signal adapted to a subtractive synthesis technique.

Non-parametric methods

Non-parametric techniques used in the analysis of sound signals generally correspond to representations with physically and/or perceptively meaningful parameters. The best known representation is undoubtedly the spectral representation obtained through the Fourier transform. The signal is in this case associated to a representation giving the energy distribution as a function of frequency. As mentioned earlier, this representation is not sufficient for characterizing the timbre and the dynamic aspects of a sound. I shall in what follows 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 where one of the axes corresponds to the time and the other to the

frequency. Such representations are to the sound what the musical scores are to the 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 non-linear, 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 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 non-linear methods since they make the resynthesis of signals possible and since they ensure that no spurious terms could cause confusion during the interpretation of the analysis. These spurious terms occur in non-linear analysis as a result of cross terms which appear in the development of the square of a sum. This is why I focus on the very important class of linear time-frequency methods.

3.1. Time-frequency and time-scale analysis

The linear representations are obtained by decomposing the signal into a sum of elementary functions having good properties of localization 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 to a scaling parameter for wavelet transforms. The choice of the elementary functions gives the shape of the filter and consequently changes the characteristics of the data obtained through the analysis.

3.1.1. Time-frequency analysis

In the case of the Gabor transform, the elementary functions, also called time-frequency atoms or grains, are all generated from a mother function (window) *W(t)* translated in time and in frequency. The mother function is chosen to be well localized in time and frequency and to have finite energy (for instance a gaussian function) (*Figure 3.1*). One can then construct a family of grains given by:

$$W_{\omega\tau}(t) = W(t-\tau)e^{i\omega(t-\tau)},$$

where ω is the parameter of frequency translation and τ the parameter of time translation.

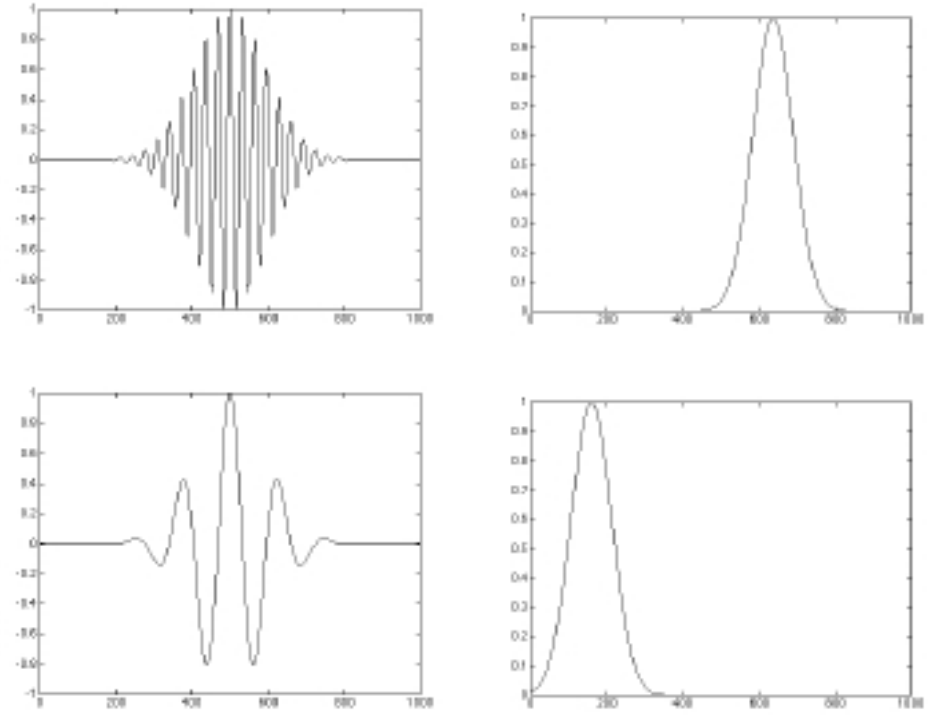


Figure 3.1 Two different « grains » of Gabor in the time domain (left), and their Fourier transform (right). In the Gabor representation, all the filters are obtained through translation in frequency of a mother function, yielding a constant absolute bandwidth analysis

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 and leads to the Gabor transform:

$$S(\omega,\tau) = \langle s, W_{\omega,\tau} \rangle = \int_{-\infty}^{\infty} s(t)\overline{W}(t-\tau)e^{-i\omega(t-\tau)}dt$$

The transform is a two dimensional function and can be represented as an image where the time t is along the horizontal axis and the angular frequency ω is along the vertical axis. For a fixed ω, this formula corresponds to a convolution between $s(t)$ and $W(-t)e^{i\omega t}$. Each horizontal line of the transform then corresponds to a filtering of the signal by a band-pass filter centered 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 centered at a given time. The transform obtained this way is generally complex-valued, since the atoms themselves are complex-valued, giving two complementary images of the representation [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 can be interpreted as the energy distribution in the time-frequency plane since the isometry relation gives

$$E_s = \int |s(t)|^2 dt = \frac{1}{2\pi} \int\int |S(\omega, \tau)|^2 d\omega d\tau$$

where E_s is the energy of the signal. The second image of the representation corresponding to the phase of the transform is generally less well-known and less used. Nevertheless it contains a lot of information. This information mainly concerns the "oscillating part" of the signal (*Figure 3.2* and *Figure 3.3*). The derivative of the phase with respect to time has the dimension of a frequency and leads to the frequency modulation law of the signal components [Guillemain et al., 1996].

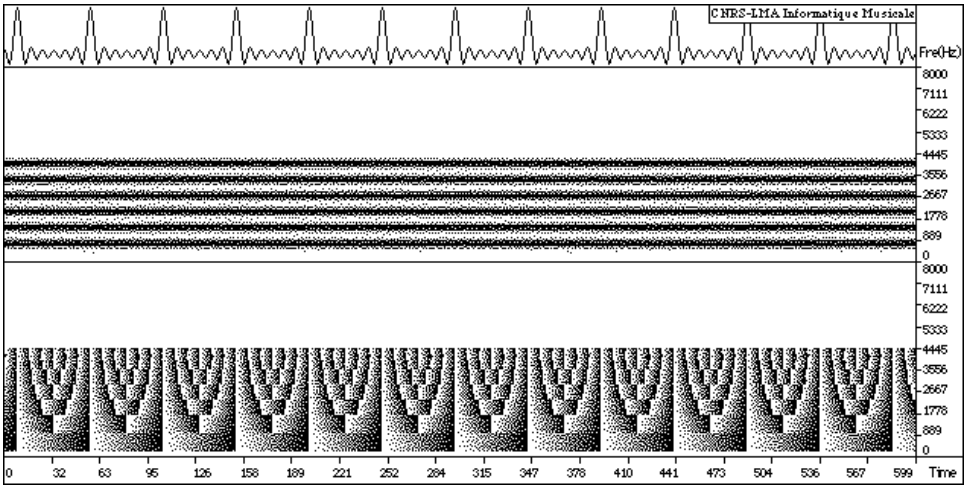


Figure 3.2: Gabor transform with a good frequency resolution (and thus a bad time resolution).

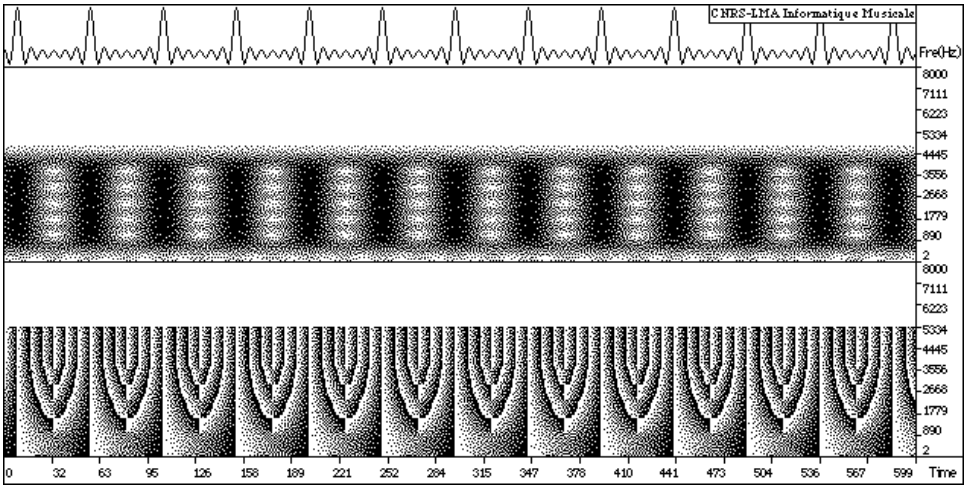


Figure 3.3: Gabor transform with a good time resolution (and thus a poor frequency resolution)

Figure 3.2 and *Figure 3.3* show the Gabor transform of the sum of six harmonic components analyzed with two different windows. The horizontal axis is time. The vertical axis is frequency. The upper part of each picture is the modulus, the lower is the phase, represented modulo-2π. Their values are coded with a gray scale. In *Figure 3.2*, the window is well localized in frequency, allowing the resolution of each component. In *Figure 3.3*, the window is well localized with respect to the time, leading to a bad separation of the components in the frequency domain, but showing impulses in time due to the fact that the signal can also be considered as a filtered Dirac comb. In both pictures, the phase behaves similarly, showing the periodicity of each component. This property can be used to accurately estimate the frequencies of the components.

The Gabor transform is invertible and one can resynthesize the signal from the analysis coefficients using the formula:

$$s(t) = \frac{1}{2\pi} \iint S(\omega, \tau) W(t-\tau) e^{i\omega(t-\tau)} d\omega d\tau$$

3.1.2. Time-scale analysis

The time-scale analysis, or wavelet transform, follows a principle close to that of the Gabor transform. Again, the restriction to an horizontal line of the wavelet transform corresponds to a filtering between the signal and a filter whose shape is independent of the scale, but whose bandwidth is inversely proportional to the scale. The analysis functions are all obtained from a mother wavelet g(t) by translation in time and change of scale (dilation) (*Figure 3.4*). The « grains » are given by

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

where τ is the time translation parameter and a is the scale or dilation parameter ($a > 0$). The wavelet transform is then given by

$$S(\tau, a) = \frac{1}{\sqrt{a}} \int s(t) \bar{g}\left(\frac{t-\tau}{a}\right) dt$$

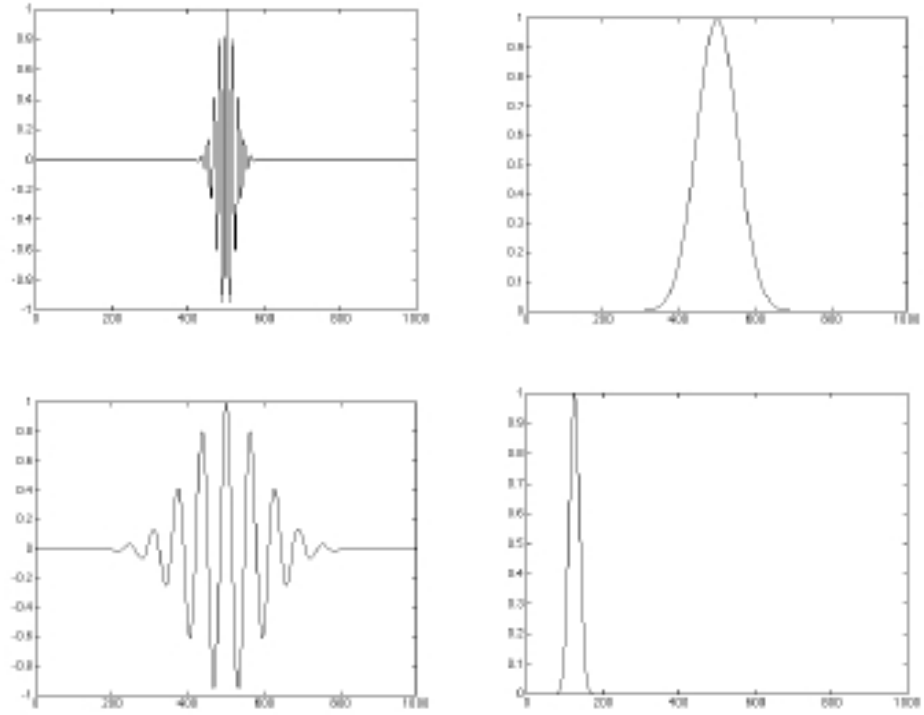


Figure 3.4: Two different wavelets in the time domain (left), and their Fourier transform (right). In the wavelet representation, all the filters are obtained through dilation in time of a mother function, yielding a constant relative $\Delta\omega / \omega$ *bandwidth analysis.*

The mother wavelet must satisfy an « admissibility » condition, namely

$$C_g = \frac{1}{2\pi}\int \frac{|g(\omega)|^2}{\omega} d\omega < \infty$$

This means in practice that *g(t)* is a function with finite energy and zero mean value. These "weak" conditions offer a great freedom in the choice of the mother wavelet. For example, by using a mother wavelet made with two wavelets separated by an octave, one can detect octave chords in a musical play [Kronland-Martinet, 1989].

Another important aspect of the wavelet transform is the localization. By acting on the dilation parameter, the analyzing function is automatically adapted to the size of the observed phenomena (*Figure 3.5*). A high frequency phenomenon should be analyzed with a function that is well-localized in time, while for a low-frequency phenomenon, this function should be well-localized in frequency. This leads to an appropriate tool for the characterization 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 fraction of octaves) permits the interpretation of the transform as a musical score associated to the analyzed sound.

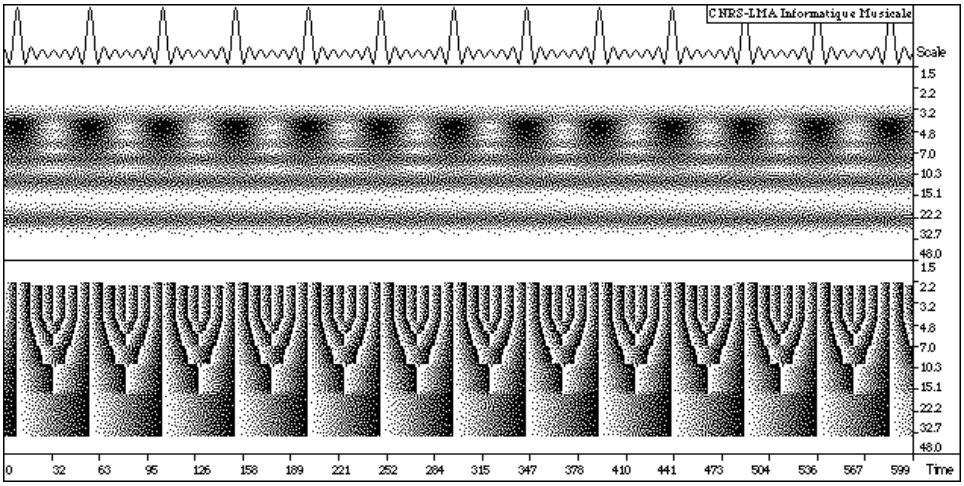


Figure 3.5 Wavelet transform of the same sum of six harmonic components. In contrast with Figure 3.3 and Figure 3.4 obtained through the Gabor transform, one can see that the wavelet transform privileges the frequency accuracy at low frequency (large scales) and the time accuracy at high frequency (small scales).

The wavelet transform is invertible and one can resynthesize the signal from the analysis using the formula:

$$s(t) = \frac{1}{c_g}\int\int \frac{1}{\sqrt{a}} S(\tau, a) g\left(\frac{t-\tau}{a}\right) \frac{da d\tau}{a^2}$$

This formula corresponds to a granular synthesis technique since it consists of reconstructing the signal by summing up elementary grains given by $g\left(\frac{t-\tau}{a}\right)$.

3.1.3. Matched time-frequency analysis

For certain kinds of signals, it can be interesting to use some a-priori knowledge in order to optimize the analysis process. This is the case for example of the wave packet corresponding to the propagation of a transient signal in a tube. This example will be precisely studied in chapter 5 where it is shown that the signal is composed of a sum of exponentially damped sinusoids. The damping factor is then given by $\alpha = K\omega^2$ where K is a constant depending on the medium. The Fourier transform of such a signal constitutes a sum of Lorentzian functions $S_n(\omega) = \frac{1}{\alpha_n + i(\omega - \omega_n)}$ and leads to « bumps » localized at $\omega = \omega_n$ whose width is proportional to α_n. In order to efficiently separate the components of

such a signal, one is tempted to decompose the signal in terms of contributions whose width follows the theoretical law $\alpha(\omega)$ (*Figure 3.6*).

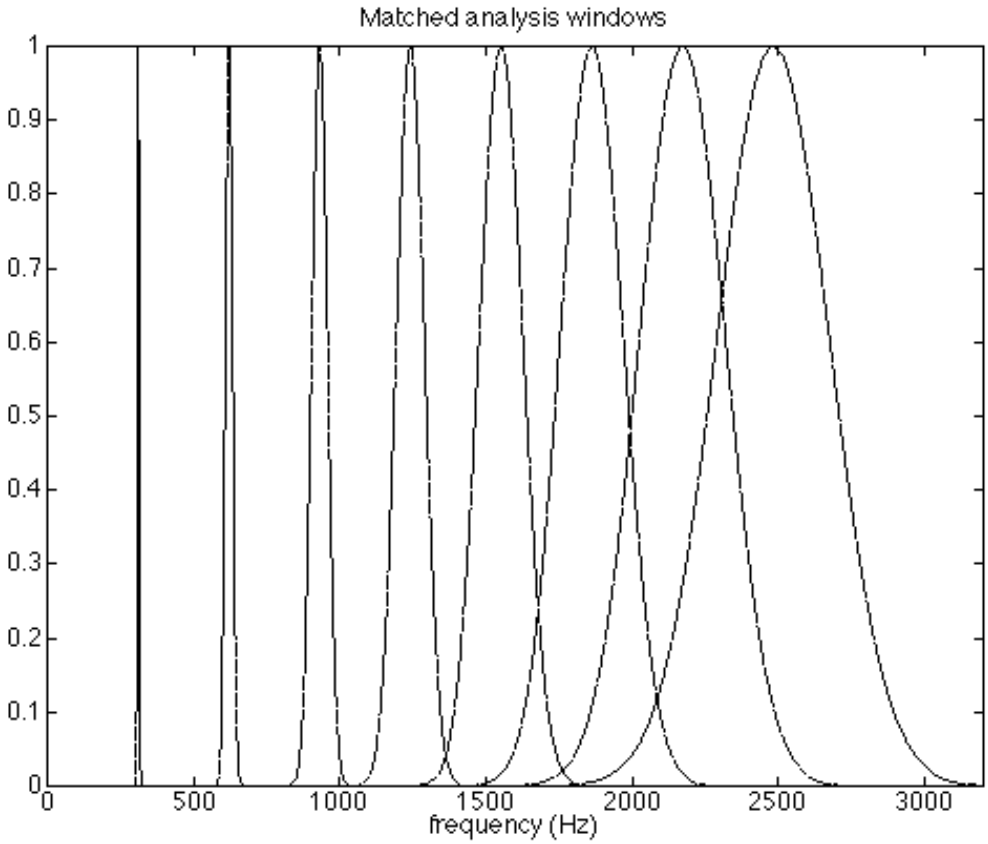


Figure 3.6: Filters corresponding to a matched time-frequency analysis. The relation between α *and* ω *is based on the mode behavior in a tube:* $\alpha = K\omega^2$ *where K has been chosen bigger than in reality for illustration purposes.*

This approach is closely related to the so-called continuous wavelet packet analysis [Torresani, 1995] and can be viewed as a compromise between time-frequency and time-scale analysis methods.

The method consists then in adjusting the analyzing grains to the components by linking their width to the frequency by a relation $\alpha = f(\omega)$.

The analysis process then reads:

$$M(\tau,\omega) = \int s(t)W\big((t-\tau)\alpha(\omega)\big)e^{i\omega(t-\tau)}dt$$

which, in the flute case leads to the expression:

$$M(\tau,\omega) = \int s(t)W\big((t-\tau)K\omega^2\big)e^{i\omega(t-\tau)}dt$$

Even though the pictures obtained this way are quite similar to the ones obtained by wavelet analysis, this matched analysis is of great importance for quantitative analysis, such as the extraction of modulation laws. This will be illustrated in the next section.

3.2. Extraction of modulation laws

The parameter extraction method makes use of the qualitative information given by the time-frequency, the time-scale or the matched analysis in order to extract 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 sounds from most resonant systems like musical instruments is the additive model. It describes the sound as a sum of elementary components modulated in amplitude and in frequency. This is pertinent from a physical and a perceptive point of view *(Figure 3.7)* .

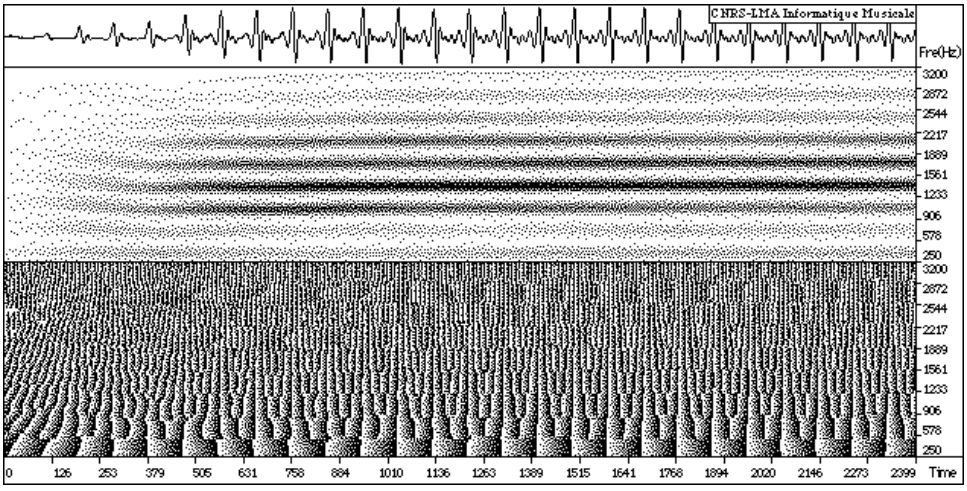


Figure 3.7: Gabor representation of the first 75ms of a trumpet sound. Many harmonics with different time dependency are visible on the modulus picture. The phase picture shows different regions, around each harmonics, where the phase wraps regularly at the time period of each harmonics.

Thus, in order to estimate parameters for an additive resynthesis of the sound, amplitude and frequency modulation laws associated to 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 which have rapidly changing amplitude modulation laws. Unfortunately all the constraints for constructing the representation make this final operation complicated. This is why absolute precision both in time and in frequency is impossible because of a mathematical relation between the transform in a point of a time-frequency plane and its close vicinity [Max, 1987] [Kronland-Martinet et al., 1987]. Human hearing follows a rather similar "uncertainty" principle: in order to identify the pitch of a pure sound, the sound must last for a

certain time. The consequences of these limitations on the additive model parameter estimation are easy to see. A high-frequency resolution requires the use of analysis functions that are well-localized in the frequency domain and therefore badly localized 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 on a time interval with the same length as the analysis function. Conversely, the choice of well-localized analysis functions in the time domain generally yields oscillations in the estimated amplitude modulation laws, due to the presence of several components in the same analysis band. It is possible to avoid this problem by astutely using the phase of the transform to precisely estimate the frequency of each component, and taking advantage of the linearity in order to separate them, without a hypothesis on the frequency selectivity of the analysis *(Figure 3.8).*

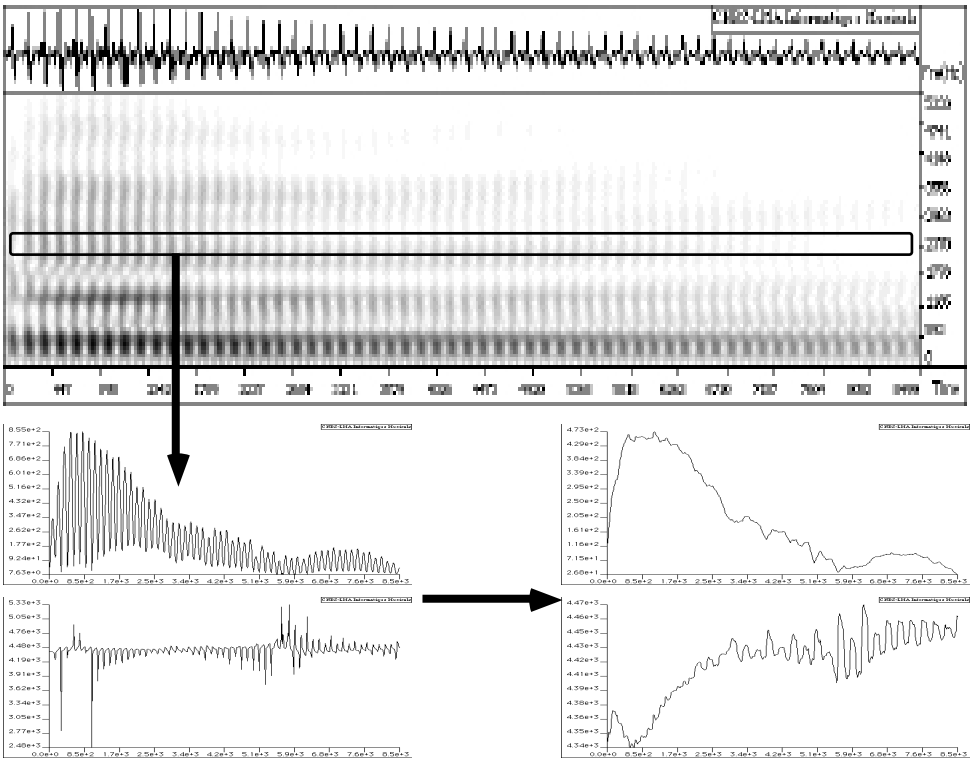


Figure 3.8 : Estimation of the amplitude modulation law of a partial of a saxophone sound. The curves on the left show the estimated amplitude and frequency modulation laws (as a function of time) using a straightforward Gabor transform. Several harmonics are present on the frequency support of the analyzing function, yielding strong oscillations. The curves on the right show the estimated modulation laws using the filter bank displayed on the next picture. Although the time support remains the same, the oscillations are automatically canceled by the algorithm.

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 localization in the time domain. Different kinds of filters can be designed, and it can be proved that they permit an exact estimation of amplitude modulation laws locally polynomial on the time support of the filters [Guillemain et al., 1996] *(Figure 3.9).*

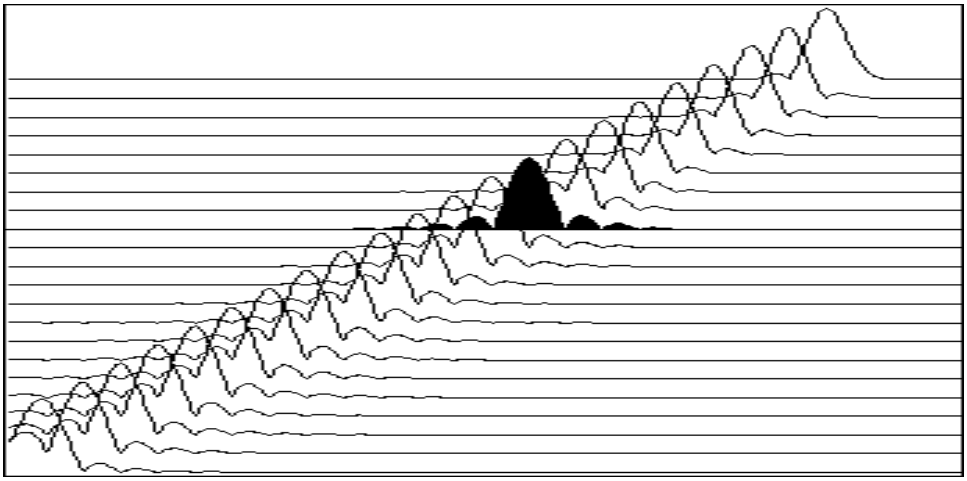


Figure 3.9: Filter bank in the frequency domain, allowing the estimation of spectral lines. One of the filters is darkened. The Fourier transform of each filter equals unity for the frequency it estimates, and zero for all the others. Its first derivative equals zero for all the frequencies.

The strict limitations of the wavelet transform or of the Gabor transform can be avoided by optimizing the selectivity of the filter as a function of the vicinity 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 phase of the transform offers the possibility of developing iterative algorithms tracking the modulation laws that does not require 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].

When the general behavior of the signal is known, matched time-frequency analysis, as described in chapter 3.1.3, can be directly used to estimate modulation laws of specific signals. In that case, one can use the same iterative algorithms as the one described above. To illustrate the advantage of this approach, we have compared the results obtained with estimations based on the matched time-frequency analysis to the ones based on the Gabor analysis. For that purpose, consider the signal:

$$s(t) = s_1(t) + s_2(t)$$

with $s_1(t) = \exp(-\alpha_1 t)\cos(\omega_1 t)$

$s_2(t) = \exp(-\alpha_2 t)\cos(\omega_2 t)$

and $\frac{\alpha_1}{\alpha_2} = \frac{1}{20}$, where $\alpha_1 > 0$, $\alpha_2 > 0$

Figure 3.10 represents the estimated amplitude modulation laws of each component $s_1(t)$ and $s_2(t)$ obtained using a Gabor analysis. The analysing window has been chosen to fit the first component. In that case, the window is not adapted to the second component, leading to a bias in the estimated amplitude modulation law. *Figure 3.11* represents the estimated amplitude modulation laws of the same components $s_1(t)$ and $s_2(t)$ obtained using a matched analysis. Here the analysing window is adapted to both components using the a priori knowledge of the ratio $\frac{\alpha_1}{\alpha_2}$. In that case, the width of the analyzing window used to estimate $s_2(t)$ is $\frac{\alpha_1}{\alpha_2}$ times the width of the one used for $s_1(t)$.

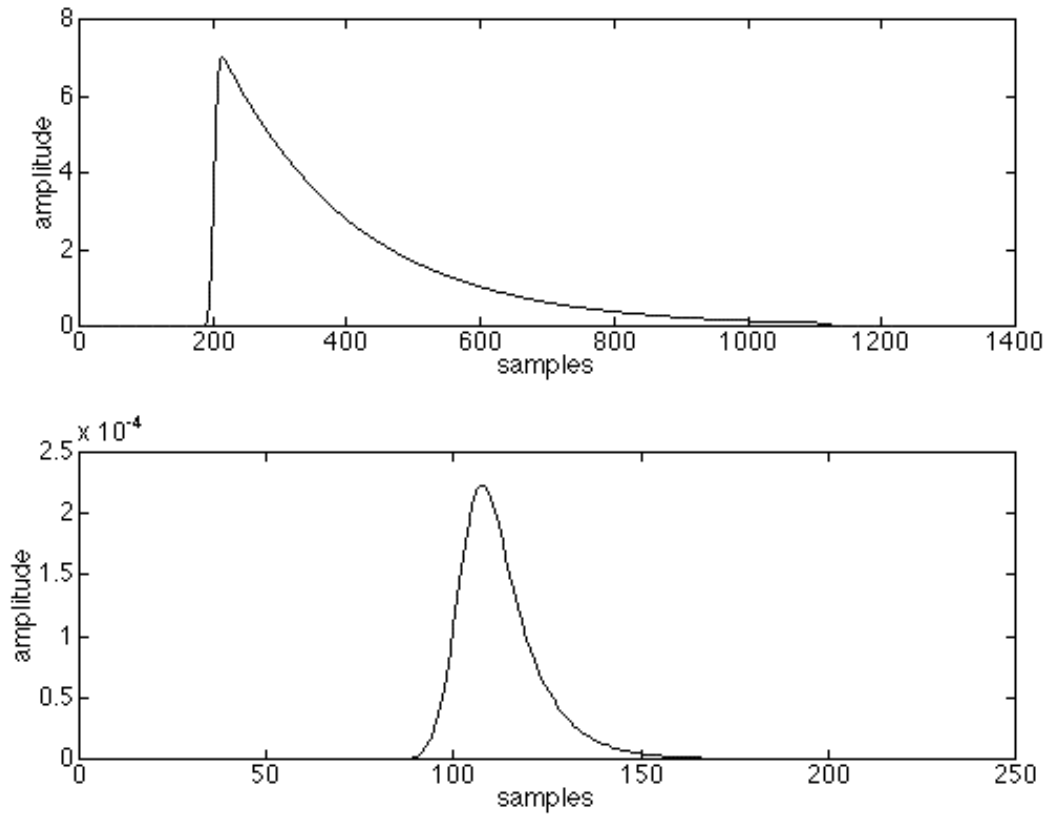


Figure 3.10: Amplitude modulation laws of two exponentially decaying components estimated using a Gabor analysis. Since the two damping factors are different, one can see the bias introduced by the constant width of the analyzing window. For representation purposes, the horizontal axis has been dilated in the second picture.

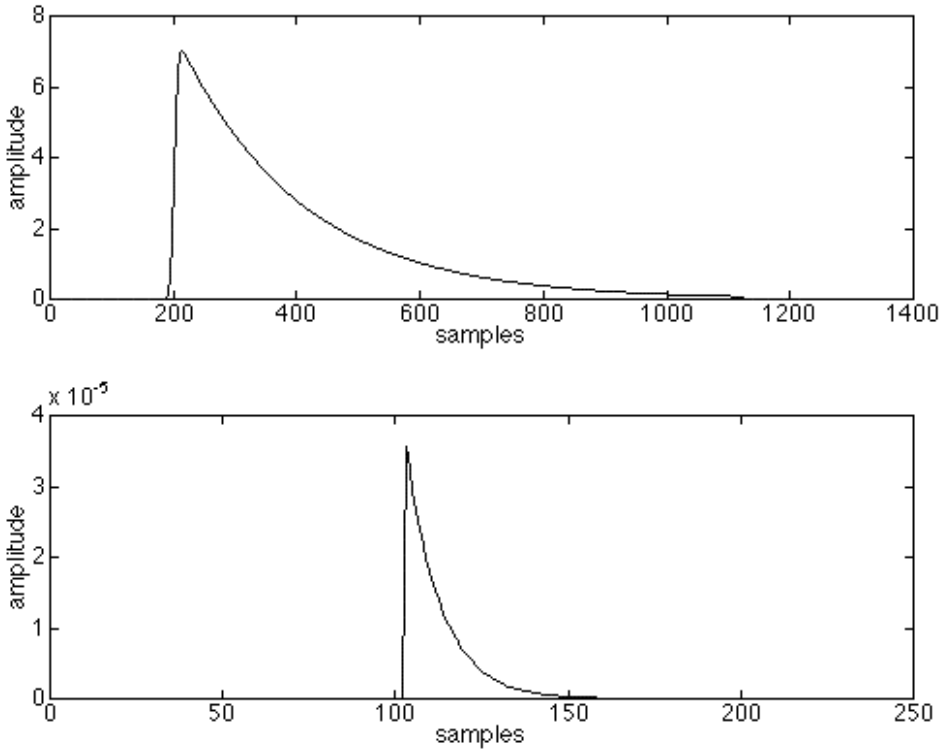


Figure 3.11: Amplitude modulation laws of two exponentially decaying components estimated using a matched analysis. The adaptation of the width of the analyzing window makes the estimation much more accurate than in the previous case.

From a practical point of view, the filters designed to extract a given component should be « analytic » in the sense that they only select positive frequency components. In that case, the transform is also analytic, allowing a direct computation of the amplitude and the instantaneous frequency. Actually, given a real signal $s(t)$, the complex valued analytic signal $z(t)$ is:

$$z(t) = s(t) + iH(s(t)) = M_z(t)e^{i\varphi_z(t)},$$

where H is the Hilbert transform. Then, one can show that [Picinbono, 1993] $M_z(t) = \sqrt{z_r(t)^2 + z_i(t)^2}$ is the amplitude envelope of the signal and $f(t) = \frac{1}{2\pi}\frac{\partial}{\partial t}\varphi_z(t) = \frac{1}{2\pi}\frac{\partial}{\partial t}\left[arctg\frac{z_i(t)}{z_r(t)}\right]$ is the instantaneous frequency of the signal.

The Fourier transform of the analytic signal satisfies the following relation:

$$z(\omega) = \begin{cases} s(\omega) & \text{for } \omega > 0 \\ \frac{1}{2}s(\omega) & \text{for } \omega = 0 \\ 0 & \text{for } \omega < 0 \end{cases}$$

confirming that the « analytic filters » only act on positive frequency components.

3.3. Examples of flute sound analysis

To illustrate the process described above, I give some examples of flute sound analysis. I present two types of sounds, namely transient and sustained flute sounds. The sounds have been recorded in an anechoïc room in order to avoid room acoustic effects (reverberation).

The transient sounds have been obtained by rapidly closing a fingerhole without exciting at the embouchure of the flute. The spectrogram of a transient sound shows that the low frequency components have a longer duration than the high frequency components.

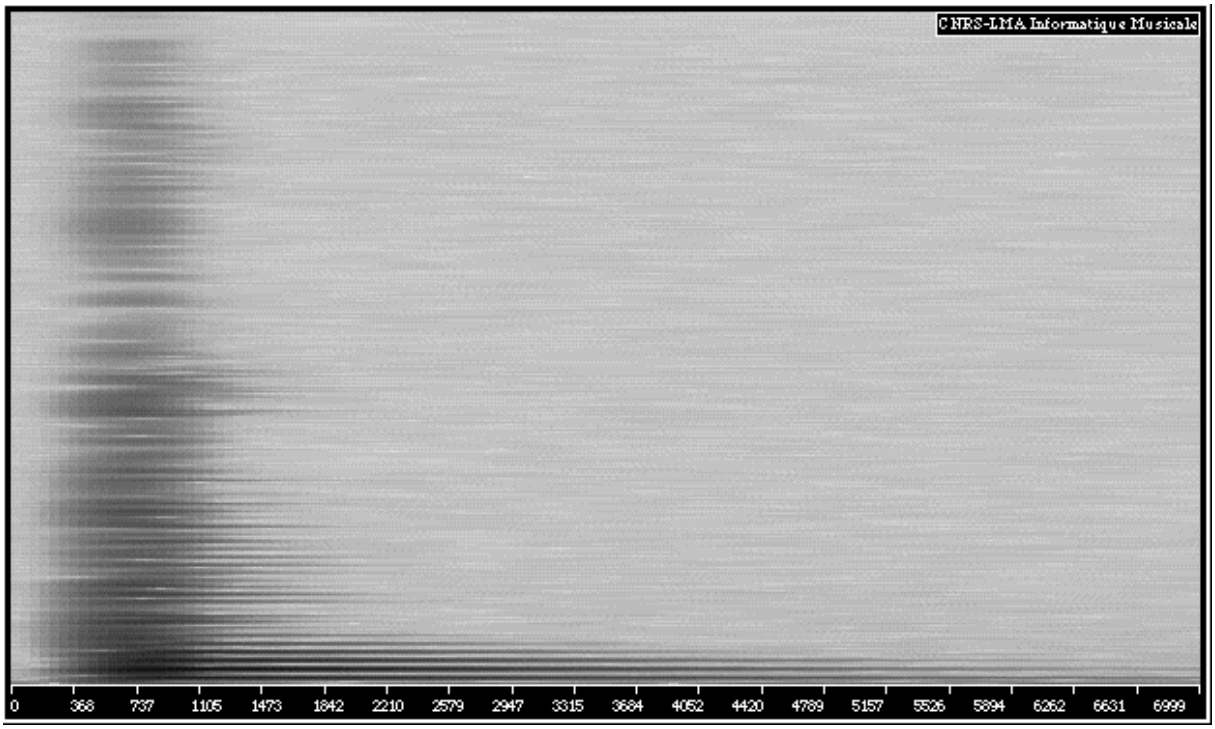


Figure 3.12:Spectrogram for a transient sound in a flute obtained by rapidly closing the fingerhole corresponding to the note D1. The horisontal axis corresponds to the time, here given in samples (f_s=48000Hz), and the vertical axis corresponds to the frequency.

In order to get a closer look at the behavior of each spectral component, their amplitude modulation law can be estimated by time-frequency techniques. In chapter 5 the behavior of the components of such a transient signal is discussed, and theoretical results are compared to real ones. From the representations of the spectrogram and extraction of the amplitude modulation laws it can be seen that the energy of the spectral components is exponentially decaying, and that high frequency components are more rapidly attenuated than low frequency components. *Figure 3.13* represents the amplitude modulation laws of the first and the sixth component of the transient flute sound obtained through the extraction technique based on the matched time-frequency analysis described in section 3.1.3.

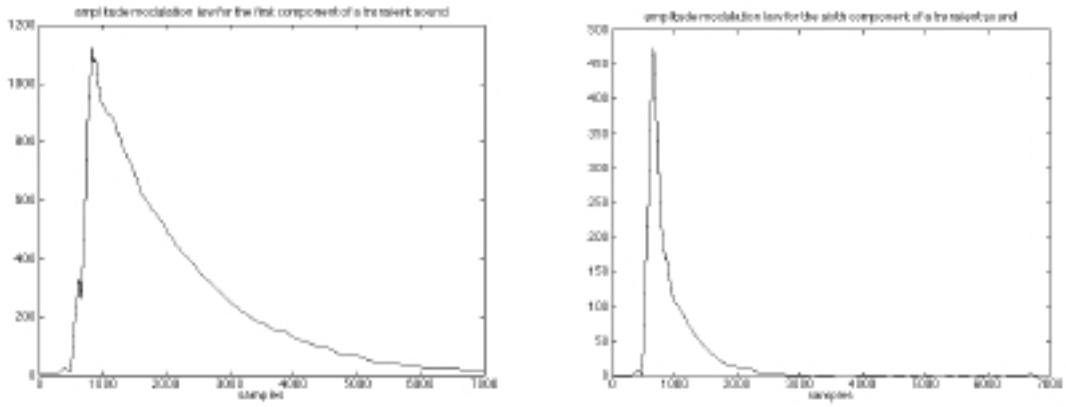


Figure 3.13:Amplitude modulation laws of the first and sixth components of the transient sound whose spectrogram is presented in Figure 3.12 (f_S=48000Hz).

From the analysis of continuous sounds several phenomena can be observed. As an example *Figure 3.14* shows the spectrogram of a transition of a flute sound from A1 to A2 (whose fundamental frequencies are respectively 440Hz and 880Hz). As expected the interval between the spectral components is doubled after octaviation. In addition, when studying the part of the spectrogram corresponding to the note A2, one can observe some remaining « under harmonics » corresponding to the odd harmonics of the note A1. This is due to the fact that the modes of the tube still are the same, since only the way of exciting them has changed. The source-resonant model of the flute in chapter 6 conveniently describes this behavior.

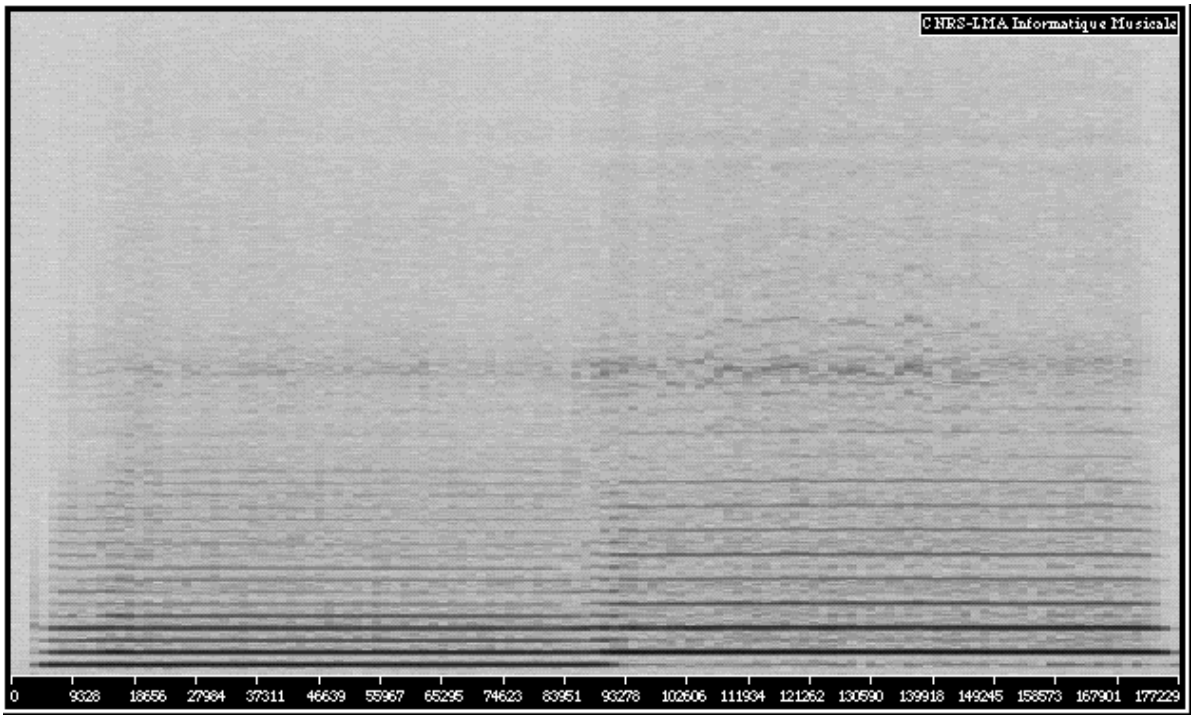


Figure 3.14:Spectrogram corresponding to a transition of a flute sound from A1 to A2 (f_S=48000Hz).

Using a spectral line estimation technique based on the matched time-frequency method, one can represent the amplitude and frequency modulation laws of each component of the sound. *Figure 3.15* and *Figure 3.16* represent such modulation laws for the third component of a 8 seconds long flute sound (whose fundamental frequency is 440 Hz). The oscillating part of the curve corresponds to natural fluctuations and leads to the so-called tremolo and vibrato which are strong « signatures » of the sound. Such data are of great importance for characterizing the sound and constructing synthesis models such as the additive model.

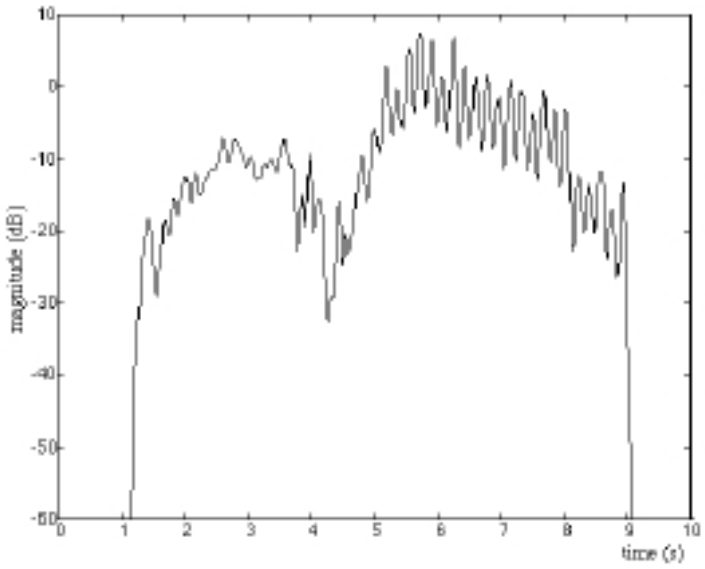


Figure 3.15: Amplitude modulation law of the 3rd component of A1 (f_0 =440Hz)

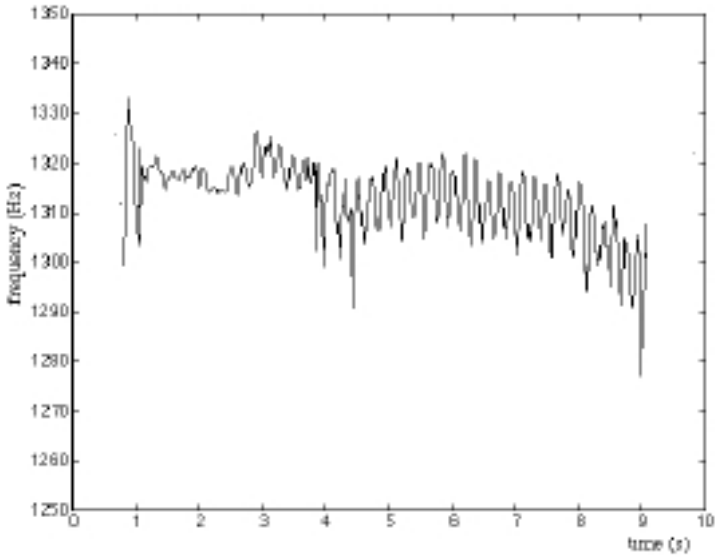


Figure 3.16: Frequency modulation law of the 3rd component of A1 (f_0 =440Hz)

4. Construction of a group additive synthesis model: Application to the flute

In this chapter we shall see how signal models can be constructed to simulate a given sound. The flute sound, or at least the deterministic part of it, will here be considered. As will be seen in chapter 6, the flute sound can be decomposed into a deterministic and a stochastic part. The deterministic part is mainly related to the modes of the tube while the stochastic part is related to the turbulence phenomenas at the excitation. To design a digital flute mirroring as closely as possible a real one and permitting in addition intimate transformations on the sound by altering the synthesis parameters, a signal model based on additive synthesis whose parameters are estimated by the analysis of real sounds is introduced. Some results concerning the attack transient and the vibrato obtained with the help of time-frequency techniques discussed in chapter 3 will be presented.

The aim of this chapter is to estimate the parameters $A_k(t)$ and $v_k(t)$ so that the sound can be simulated using the model:

$$s(t) = \sum_k A_k(t)\cos(\omega_k t + \int_0^t v_k(u)du)$$

This model can be dramatically simplified if one can group the $A_k(t)$ and the $v_k(t)$. In this case the model is called a group additive synthesis model

$$s(t) = A(t)\sum_k \alpha_k \cos(\omega_k t + k\int_0^t v(u)du)$$

4.1. Modeling of the attack transient of the flute sound

One of the key aspects characterizing flute-like instruments is the attack transient. Authors as Balibar [Balibar, 1981] and Rakowski [Rakowski, 1965] state that the attack transient is among the longest for wind instruments, and that the period of growth of a flute tone can be divided in two parts; one where the proper pitch cannot be detected, followed by a shorter part where the sensation of a pitch suddenly appears to the listener. It is accompanied by a sudden change in speed of growth of the harmonics. We will simplify this theory by modeling the attack transient with a unique law reflecting the growth of the harmonics.

To model the transient part of the sound (the attack), a correlation between the behavior of the amplitude modulation law of each component and the total energy of the sound (which is related to the jet pressure) must be determined. *Figure 4.1* shows the beginning of the amplitude modulation laws of the

first five harmonics corresponding to a flute sound whose fundamental frequency is 261Hz (note C1). These amplitude modulation laws have been calculated by time-frequency techniques discussed in chapter 3.

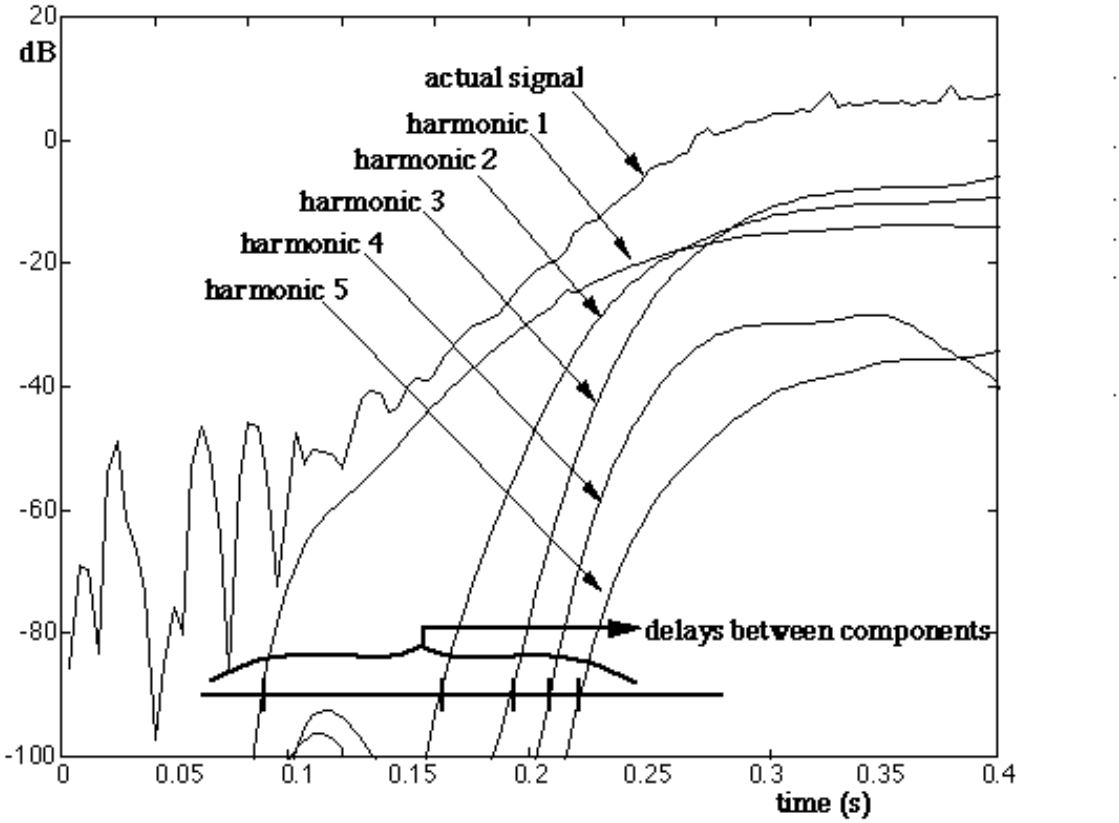


Figure 4.1: the first 0.4s of the amplitude modulation laws of the first 5 harmonics of a flute sound whose fundamental frequency is 261Hz (note C1).

The steepness of the amplitude modulation laws during the attack can be correlated to the amplitude of the total sound. This correlation can be visualized by representing the amplitude modulation law of each component versus the amplitude modulation law of the whole signal. In *Figure 4.2* the curves corresponding to the first and the third components of the flute sound are shown. The curves are getting steeper as the component rank increases. In addition a deviation due to the different attack and decay time of the energy of each component is observed. This deviation corresponds to an hysteresis effect which increases with the frequency of the components. This behavior has not been taken into account in the model.

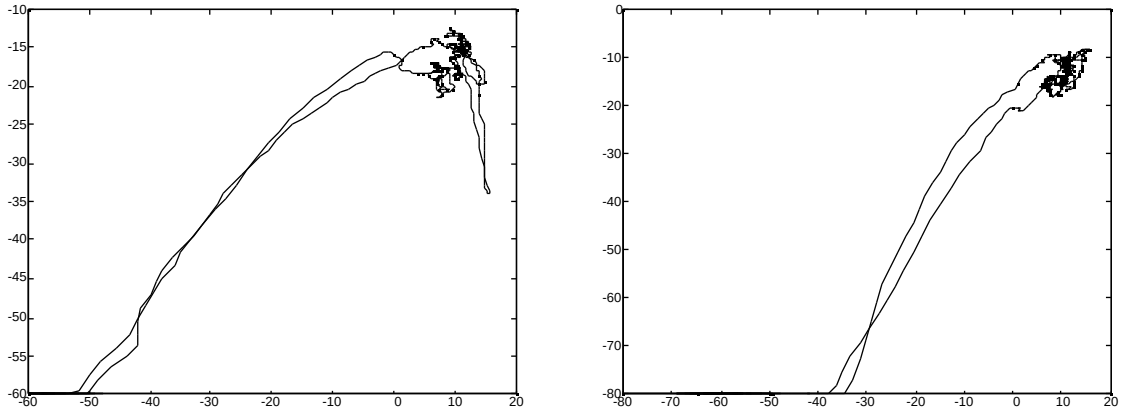


Figure 4.2: the energy density of components one and three versus the energy density of the whole sound

Thus the amplitude modulation law A_n of the n'th harmonic can be roughly linked to the total amplitude A_T of the signal by a linear law: $A_n = K_n + \alpha A_T$.

The constant K_n depending on the mode rank is related to the delays between the spectral components. Measurements show that K_n increases roughly proportional to *n*. Thus a delay between two successive components *(n-1)* and *n*, is proportional to $\frac{1}{n}$. This is in accordance with the measurements made directly on the modulation laws of the harmonics (*Figure 4.1*).

The steepness of the amplitude modulation laws of the n'th component during attack is found to increase like (0.75+0.3(n-1)) times the total amplitude of the sound. Thus the fundamental component is the only component whose steepness of the amplitude modulation law is less important than the steepness of the total amplitude. The law is then:

$$A_n = K_n + (0.75 + 0.3(n-1))A_T$$

As already mentioned, the observed hysteresis effect has not been taken into account in this formula. In order to model such an effect, the constant K_n should be adjusted in accordance with the "width" of the hysteresis, so that K_n-values applied during the attack of the sound are different from the ones applied during the decay.

The synthesized sound calculated from the amplitude modulation law is « poor » compared to a real flute sound. This is partly due to the absence of pressure fluctuations during the steady state part of the note. Grey [Grey, 1975] states that the steady state part of the sound is important in judging the timbre. The vibrato helps the identification of a sound. In the next section I therefore present a method for modeling the vibrato of the sound.

4.2. Modeling of the vibrato of the flute sound

By direct measurements, Fletcher [Fletcher et al., 1990] has shown that the pressure fluctuations producing the vibrato of a flute sound causes a frequency variation of 5 Hz and an amplitude variation of about 10% of the blowing pressure. By studying the admittance of the jet and the pipe when the pressure is fluctuating, Fletcher has found that an increase in blowing pressure gives a small increase in sounding frequency [Fletcher et al., 1990]. When the blowing pressure decreases, the frequency falls. This explains the variation in frequency when playing a vibrato. The player uses pressure fluctuations to obtain this effect.

To investigate the relation between the frequency variations and the pressure fluctuations, amplitude and frequency modulation laws have been calculated by time-frequency techniques that were presented in chapter 3. The flute sound that has been analyzed this way has a fundamental frequency of 440Hz, corresponding to an A1. The vibrato depth of each frequency modulation law can thus be measured together with the frequency of the vibrato and the correspondence between the pressure fluctuations and the frequency variations.

The measurement of the vibrato depth $\Delta\omega$ for each component have shown that it increases with the harmonic rank and that the ratio $\frac{\Delta\omega}{\omega}$ is constant as shown in *Figure 4.3*. This means that a harmonic sound remains harmonic when a vibrato is applied.

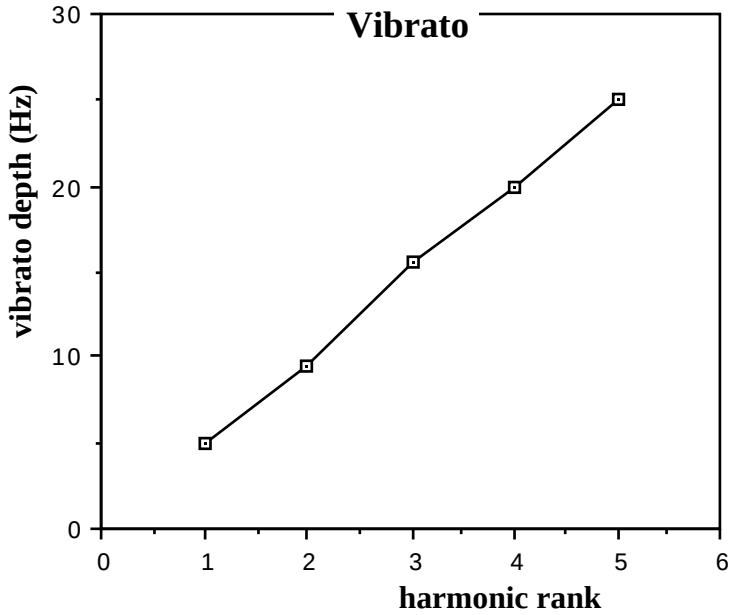


Figure 4.3: Vibrato depth versus harmonic rank for a flute sound whose fundamental frequency is 440 Hz.

As shown in *Figure 4.4* the vibrato is in phase with the amplitude variation (tremolo). This is due to the fact that flute players use pressure fluctuations to produce vibrato.

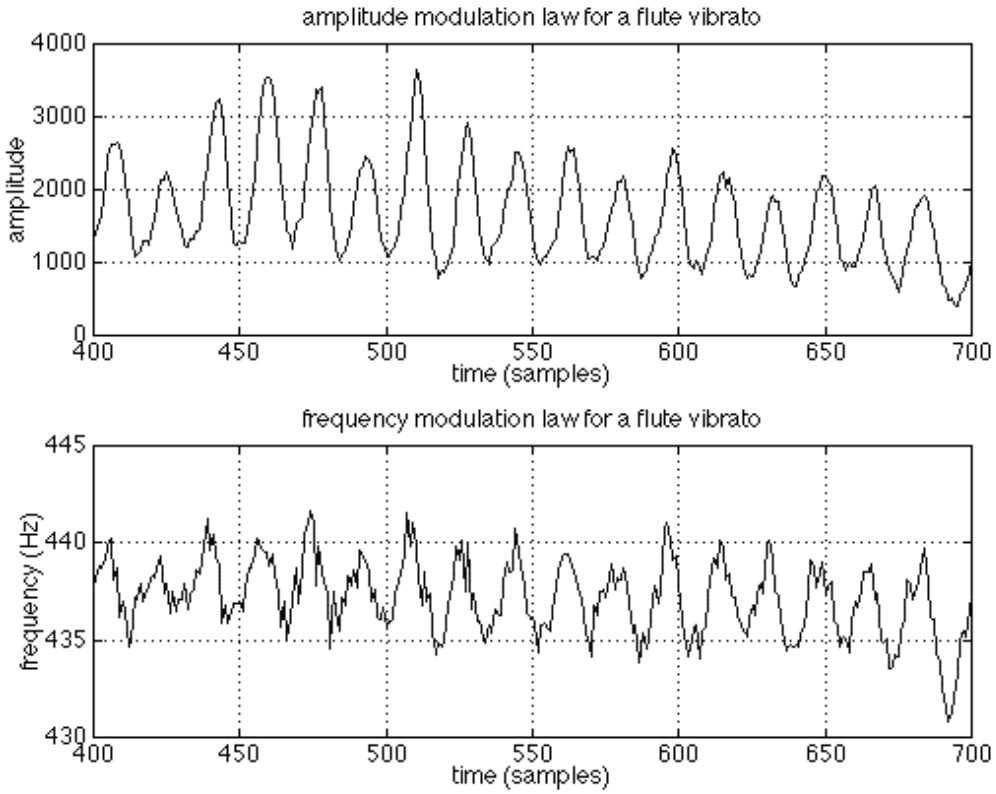


Figure 4.4: Amplitude and frequency modulation law of a flute sound (A1) with vibrato. The sampling frequency corresponding to this figure is f_s*=100Hz.*

One can therefore pilot the frequency modulation from measurements of the air jet pressure. The frequency variations of the vibrato is about 5Hz for all notes (and for any flute player). Fletcher [Fletcher, 1975] has explained this by calculating the resonant frequency of the system abdomen-diaphragm-lung given by

$$f = \frac{1}{2\pi}\sqrt{\frac{pA^2}{mV}} \ ,$$

where

V is the volume of the lung cavity (around $10^{-2}\,m^3$),

p is the air pressure (around $10^5\,Pa$),

A is the area of the diaphragm (around $3\ 10^{-2}\,m^2$) and

m is the mass of the content of the abdomen (around 10kg).

This resonant frequency is found to be approximately 5Hz which corresponds to the vibrato frequency. A second explanation also proposed by Fletcher is that the diaphragm is maintained in the correct state of tension by opposing sets of muscles controlled by neurological feedback loops whose oscillation loop is close to 5Hz.

From a practical point of view, the vibrato *v(t)* can be estimated by a bandpass filtering of the signal amplitude A_T. A filter selecting a narrow band centered around 5 Hz is appropriate. I used a Butterworth band pass filter made with two order 2 low and high pass filters. The Butterworth filter has been chosen for its good frequency behavior, especially its stability and high slope. A lot of filter design techniques have been developed [Hamming, 1989] and numerous filter design software exist. To design the filter, I used functions provided by the MATLAB signal processing toolbox which is described by T. W. Parks and C. S. Burrus. [Parks et. al., 1987].

The filtering satisfies the difference equation

$$v(n) = C\sum_{k=0}^{3} a_k A_T(n-k) + \sum_{k=1}^{4} b_k v(n-k),$$

where

$$a = \begin{bmatrix} 1.0000 \\ -3.9995 \\ 5.9986 \\ -3.9986 \\ 0.9995 \end{bmatrix} \qquad b = \begin{bmatrix} 0.2677 \\ 0.0000 \\ -0.5353 \\ 0.0000 \\ 0.2677 \end{bmatrix} 10^{-7}$$

and C is a constant adjusting the depth of the vibrato.

The frequency response of the filter is given in *Figure 4.5.*

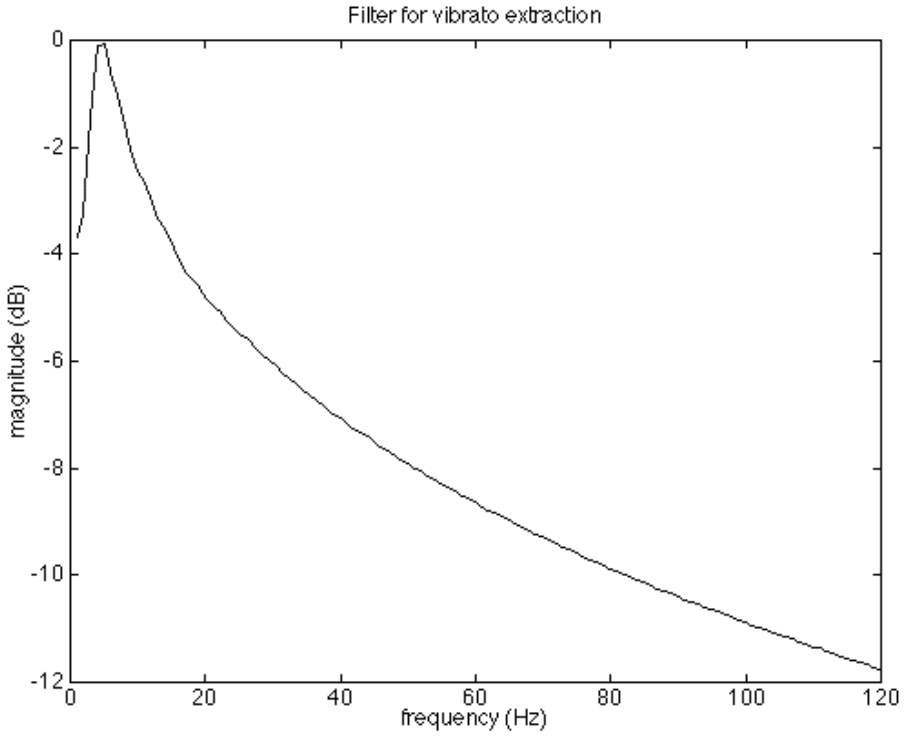


Figure 4.5: Frequency response of the filter used for extracting the vibrato from the amplitude A_T of the signal.

Figure 4.6 represents the envelope pressure measured at the cork level (close to the embouchure) of the flute and used at the entrance of the filter. *Figure 4.7* represents the extracted vibrato. To verify the auditive performance of this approach, the extracted vibrato was applied to a synthetic sound whose spectrum was close to that of a flute. The synthetic sound was presented to several listeners who found the frequency variations to be natural.

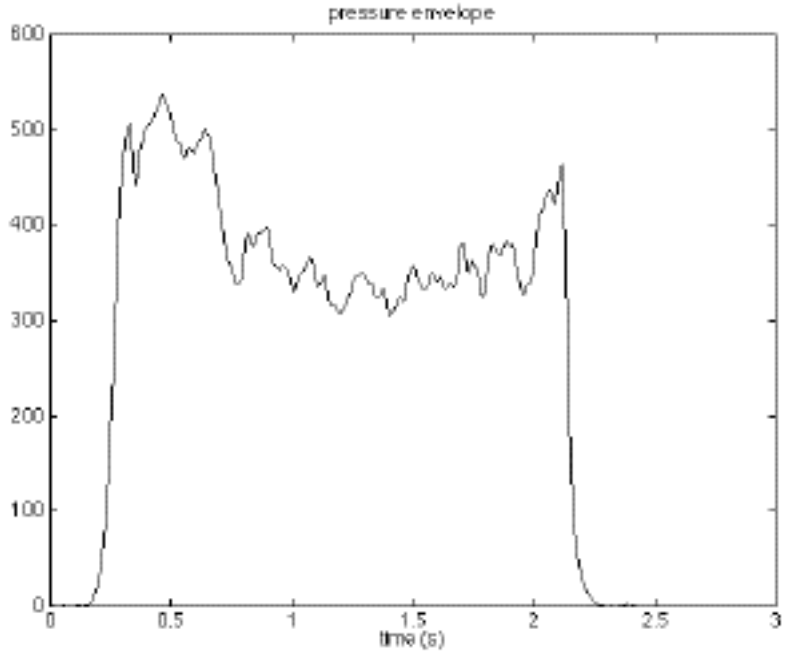


Figure 4.6: envelope of the pressure used at the entrance of the Butterworth filter.

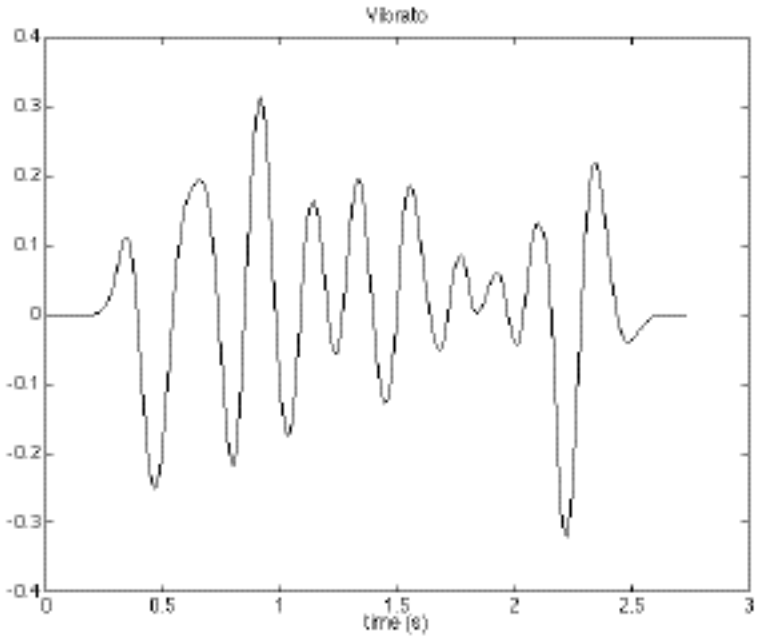


Figure 4.7: extracted vibrato obtained from the pressure envelope.

These investigations have led to a group additive synthesis model for the flute sounds given by

$$s(n) = \sum_{k=1}^{K} \left\{ K_k + (0.75 + 0.3(k-1))A_T(n) \right\} \cos(k\omega_0 \frac{n}{\omega_s} + k\phi(n)),$$

$$\text{with } \phi(n) = \phi(n-1) + \nu(n).$$

Even though the sounds generated this way are of good quality, this simplified model doesn't allow huge dynamic variations since the growth of the components is unbounded. Moreover, an important feature of the flute sound, the turbulent noise, is missing in this model. The state of this typical noise depends upon the physical characteristics of the tube, and is changing during the play (since the « effective length » of the tube changes). Thus physical considerations of the instrument's behavior should be taken into account to accurately model the sound.

5. Sound modeling using a physical approach

In chapter 2.2 different physical approaches of the synthesis processes were discussed. A detailed description of the waveguide synthesis model was given, leading to a propagation model in a bounded dissipative and dispersive medium. The parameters of such models can be found either from the solutions of the basic physical equations or from the analysis of real sounds. Theoretical parameters which are often biased since they are resulting from approximations, can thus be matched with the real parameters. In the case of the flute and the guitar, relevant parameters for a wave propagation model such as dispersion and dissipation can be estimated both by studying the solutions of the basic physical equations and by the analysis of real sounds. The dispersion introduces an « inharmonicity » between the partials, and the dissipation yields a different decay time of the components. These two effects can be measured on real sounds. To use the waveguide synthesis model in practice, I shall describe a way of building the filters of the model that will reproduce the dispersion and dissipation phenomena.

5.1. Physical description of 1-dimensional systems

In this section a physical description of the propagation of longitudinal waves in fluids (wind instruments) and transversal waves in solid media (string instruments) is given. These two systems can be treated similarly, and interesting differences between the two systems can be observed.

5.1.1. Propagation of longitudinal waves in a resonant tube

The one-dimensional wave equation of the acoustic pressure y inside a tube when visco-thermal losses are taken into account can be written [Kergomard, 1981]:

$$\frac{1}{c^2}\frac{\partial^2 y(x,r,t)}{\partial t^2} - \Delta y(x,r,t) - \frac{l_{hv}}{c}\Delta\frac{\partial y(x,r,t)}{\partial t} = \Psi(x,r,t) \qquad (5.1)$$

with boundary conditions $y(0,r,t)=y(L,r,t)=0$. l_{hv} represents characteristic lengths (l_{hv} is of order 10^{-8} m in free airfilled space). When the acoustic pressure is to be determined in a bounded airfilled domain (like a tube), l_{hv} depends upon the friction near the domain boundary. In this case l_{hv} is not easily determined. c is the velocity of sound, and Ψ is a time-dependent source.

When the pressure is to be determined inside the tube, the Laplacian of equation (5.1) should be represented in terms of cylinder coordinates. When the wavelength of the sound wave transmitted along the interior of the tube is large compared with the transversal dimension of the tube, the fluid motion is predominantly parallel to the tube axis, and the wave motion is very nearly one-dimensional. That is, for frequencies below the cut-off frequencies of the higher order modes, only the fundamental mode (0,0)

can propagate. The cut-off frequency of the (m,n)th mode can be found from the expression of the transverse eigenfunctions [Morse, 1968], and is given by

$$f_c = \frac{\alpha_{mn} c}{2b},$$

where

$\alpha_{m,n}$ depends on the mode ($\alpha_{10} = 0.5861, \alpha_{20} = 0.9722, \alpha_{01} = 1.2197, \alpha_{02} = 2.2331\ldots..$),

b is the radius of the cylinder, and

c is the velocity of sound in the medium.

For a tube with the dimensions of a flute resonator ($b=9.5mm$), the smallest cut-off frequency is $f_{c10} = 5.2kHz$. The next ones are $f_{c20} = 8.7kHz$ and $f_{c01} = 10.9kHz$. I shall therefore assume that the wave motion is one-dimensional, so that $\Delta = \frac{\partial^2}{\partial x^2}$ in (5.1) and $y=y(x,t)$.

To find a solution of equation (5.1), we search for a decomposition of the solution into an orthonormal basis whose elements are $\sin(n\pi\frac{x}{L})$, where L is the length of the tube. Thus, if the function $y(x,t)$ has only a finite number of maxima and minima in the interval $(0,L)$ and if it has only a finite number of finite discontinuities in $(0,L)$ and no infinite discontinuities (Dirichlet conditions), one can write

$$y(x,t) = \frac{2}{L}\sum_{n=1}^{\infty} y_n(t)\sin(n\pi\frac{x}{L})$$

where

$$y_n(t) = \int_0^L y(x,t)\sin(n\pi\frac{x}{L})dx$$

denotes the finite sine transform of $y(x,t)$.

By partial integration of the $y_n(t)$ term when the boundary conditions are assumed to be $y(0,t) = y(L,t) = 0$, we can calculate $\frac{\partial^2}{\partial x^2} y(x,t)$:

$$\int_0^L \frac{\partial^2}{\partial x^2} y(x,t)\sin(n\pi\frac{x}{L})dx = -\frac{n^2\pi^2}{L^2}\int_0^L y(x,t)\sin(n\pi\frac{x}{L})dx$$

and thus

$$\frac{\partial^2}{\partial x^2} y(x,t) = -\frac{n^2\pi^2}{L^2} y_n(t)$$

The homogeneous version of (5.1) can now be rewritten as

$$_n \qquad \frac{1}{c^2}\frac{\partial^2 y_n(t)}{\partial t^2} + \frac{n^2\pi^2}{L^2} y_n(t) + \frac{l_{hv}}{c}\frac{n^2\pi^2}{L^2}\frac{\partial y_n(t)}{\partial t} = 0$$

Let us now consider a source term $\Psi(x,t)$ applied to the cylinder at a point x_0. The finite sine transform of this source term is given by

$$\Psi_n(t) = \Psi(t)\int_0^t \delta(x - x_0)\sin(n\pi\frac{x}{L})dx = \Psi(t)\sin(n\pi\frac{x_0}{L})$$

Equation (5.1) with the source term can now be written:

$$_n \qquad \frac{1}{c^2}\frac{\partial^2 y_n(t)}{\partial t^2} + \frac{n^2\pi^2}{L^2} y_n(t) + \frac{l_{hv}}{c}\frac{n^2\pi^2}{L^2}\frac{\partial y_n(t)}{\partial t} = \Psi_n(t) \tag{5.2}$$

To solve equation (5.2), the Laplace transform can be applied. Actually, for certain functions *y(t)* the Fourier transform is not defined (when $\int_{-\infty}^{\infty}|y(t)|dt$ is not convergent), and one may consider the function $y_1(t) = e^{-\gamma t}y(t)$, where γ is a positive constant and $y_1(t) = 0$ if $-\infty < t < 0$. For a flute the pressure $y_n(t) = 0$ for $-\infty < t < 0$ and the initial conditions of the Laplace transform are zero, which simplifies the calculations.

It can be shown [Sneddon, 1951] that if $\int_0^{\infty}|y(t)|dt$ is not bounded but $\int_0^{\infty} e^{-kt}|y(t)|dt$ is bounded for some positive value of *k*, then the inversion formula $y(t) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\Phi(p)e^{pt}dp$ holds with $\gamma > k$.

When applying the Laplace transform to the wave equation one obtains

$$_n \qquad \frac{1}{c^2}p^2 y_n(p) + \frac{n^2\pi^2}{L^2} y_n(p) + \frac{l_{hv}}{c}\frac{n^2\pi^2}{L^2} p y_n(p) = \Psi_n(p)$$

and

$$y_n(p) = \frac{\Psi_n(p)}{\frac{1}{c^2}p^2 + \frac{l_{hv}}{c}\frac{n^2\pi^2}{L^2}p + \frac{n^2\pi^2}{L^2}} = \frac{\Psi_n(p)}{\frac{1}{c^2}(p+\alpha_n)^2 + \omega_n^2} \tag{5.3}$$

where

$\alpha_n = l_{hv}c\frac{n^2\pi^2}{2L^2}$ are damping factors corresponding to each mode, and

$\omega_n^2 = c^2 \frac{n^2\pi^2}{L^2} - l_{hv}^2 c^2 \frac{n^4\pi^4}{4L^4} = (n\omega_0)^2 - \alpha_n^2$ are the corresponding eigen frequencies. The term $l_{hv}^2 c^2 \frac{n^4\pi^4}{4L^4}$ in the frequency expression represents the "inharmonicity", and it depends on the value of l_{hv}. It is interesting to note the close relation between the inharmonicity and the damping factors. This is in accordance with the Kramers-Kronig relations stating that for a causal system there is no dispersion without dissipation [O'Donnell et al., 1981].

The theoretical damping factors are illustrated in *Figure 5.1* as a function of the mode number for a tube of length *L=58 cm*, corresponding to the length of the flute body. We can see that the damping factor alpha is increasing when the mode rank is increasing, which means that the attenuation of the components increases with the mode rank.

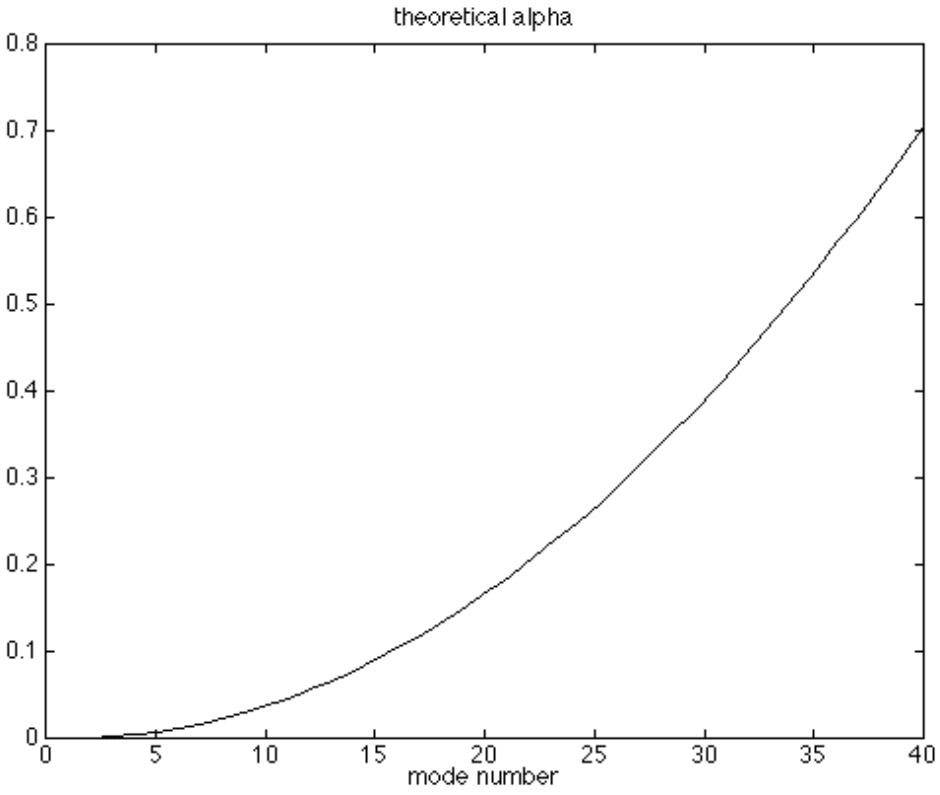


Figure 5.1: Theoretical damping factors as a function of the mode rank corresponding to a tube of length L=58cm.

In *Figure 5.2* the theoretical inharmonicity is plotted. In the harmonic case when $f_n = nf_1$, the value of the curve would be one for any mode. We can see that the "inharmonicity" in the flute case is very small and that $f_n < nf_1$, which means that the partials of the spectrum get closer as the frequency increases.

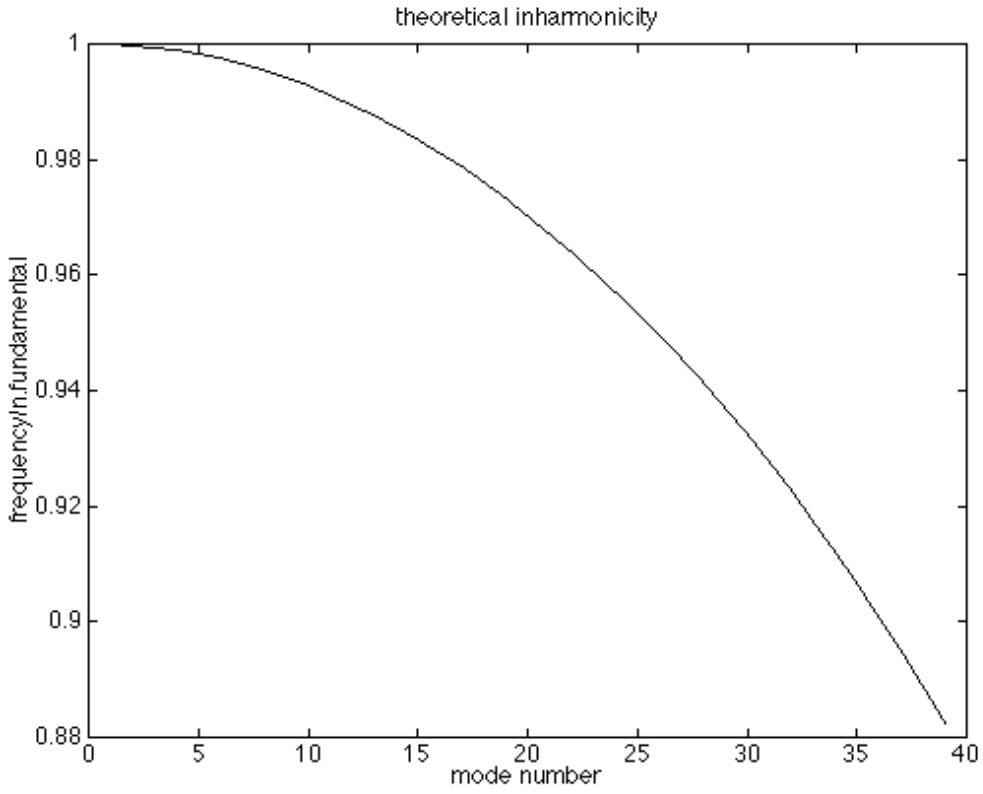


Figure 5.2: Theoretical inharmonicity of the modes in a tube.

A comparison between *Figure 5.1* and *Figure 5.2* shows that when the damping factors increase, the inharmonicity increases.

The inverse transform of the expression (5.3) can be calculated using the Faltung theorem [Sneddon, 1951]:

If $\Phi(p)$ and $\Psi(p)$ are the Laplace transforms of the functions *f(x)* and *g(x)*, then

$$\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\Phi(p)\Psi(p)e^{px}dp = \int_0^x g(y)f(x-y)dy$$

Applied to our case it gives

$$y_n(t) = \frac{\sin(n\pi\frac{x_0}{L})}{\frac{1}{c^2}\omega_n}\int_0^t \Psi(u)\sin(\omega_n(t-u))e^{-\alpha_n(t-u)}du \qquad (5.4)$$

To link the propagation to a synthesis model (signal processing), we seek an expression of the response when applying a punctual dirac source to the tube ($\Psi_n(u) = \delta(u)$). This response will be correlated to the impulse response of the system. This gives

$$y_n(t) = \frac{\sin(n\pi \frac{x_0}{L})}{\frac{1}{c^2}\omega_n} \sin(\omega_n t) e^{-\alpha_n t} \ ,$$

and the solution of the wave equation with a source term $\delta(x - x_0)\ \ \delta(t)$ is given by

$$y(x,t) = \frac{2}{L} \sum_{n=1}^{\infty} \frac{\sin(n\pi \frac{x_0}{L}) \sin(n\pi \frac{x}{L})}{\frac{1}{c^2}\omega_n} \sin(\omega_n t) e^{-\alpha_n t} \qquad (5.5)$$

This result shows that the response to a punctual dirac source in a tube is a sum of sinusoids at the frequencies ω_n exponentially damped with a damping factor α_n. This is an important result indicating the behavior of the modes in a tube. Nevertheless, this solution corresponds to a tube without holes and with only plane waves propagating and gives therefore only an idea of the behavior of the modes in a wind instrument. In spite of these simplifications, the spectrum of the theoretical impulse response as shown in *Figure 5.3* is rather close to the spectrum of an impulse response of a real wind instrument, as will be seen in section (5.3.1).

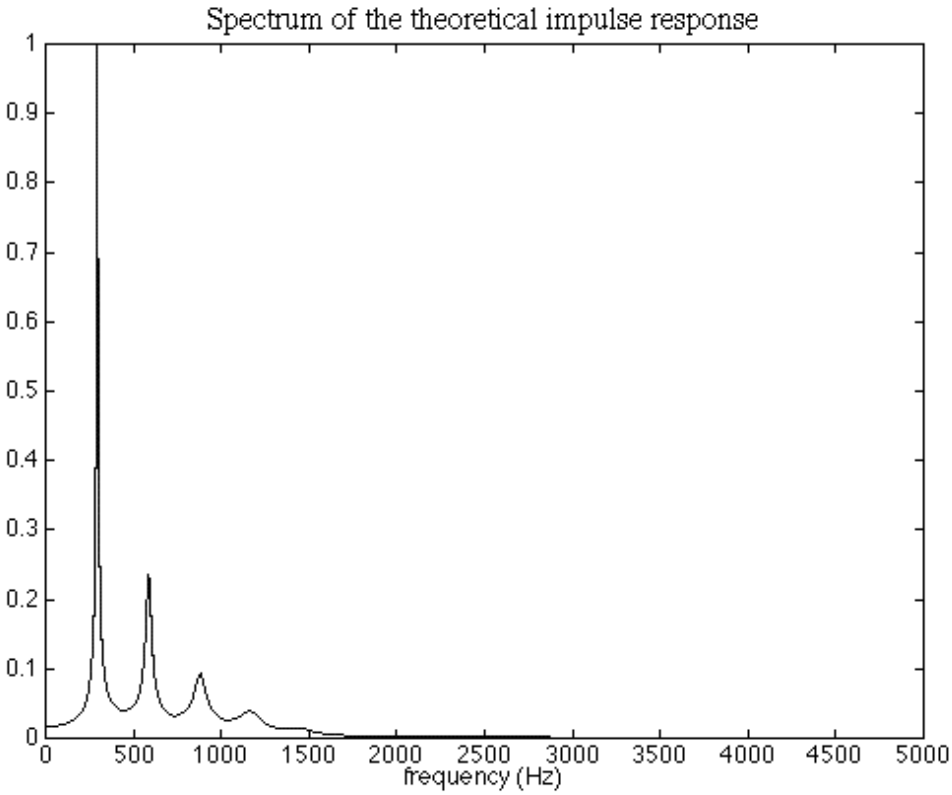


Figure 5.3: Spectrum of the solution of the wave equation for a tube of length L=0.58m, a characteristic length $l_{hv} = 4.5 \times 10^{-4}$ m and an excitation point x_0 =0.017m corresponding to the distance between the center of the embouchure hole of a flute and the cork.

5.1.2. Propagation of transverse waves in a stiff string

The transverse displacement *y(x,t)* of a stiff string is described by the equation [Valette, 1993]:

$$\rho_L \frac{\partial^2 y(x,t)}{\partial t^2} - T \frac{\partial^2 y(x,t)}{\partial x^2} + EI \frac{\partial^4 y(x,t)}{\partial x^4} + R \frac{\partial y(x,t)}{\partial t} = f(x,t),$$

where ρ_L is the mass per length unit (kg/m), *T* is the tension, *E* is the Young modulus, *I* is the quadratic moment (m^4) and *R* is the mechanical resistance per length unit(Nsm^2). It is convenient to assume that *R* takes into account the losses during the propagation (friction in the air, viscoelastic losses etc.):

$$R(\omega) = \frac{2\pi\eta + 2\pi d\sqrt{\eta\rho_{air}\frac{\omega}{2}}}{\rho\omega} + \rho_L EI \frac{\delta_{ve}}{T^2}\omega^2,$$

where η is the viscosity of the air (Pa s), *d* is the diameter of the string (m), ρ_{air} is the air density (kg/m^3) and δ_{ve} is the angle of viscoelastic losses. The admittance on the bound is supposed to be negligible as well as thermoelastic losses and internal losses due to friction in the string.

The boundary conditions are assumed to be: $y(0,t) = y(L,t) = \frac{\partial^2 y(0,t)}{\partial x^2} = \frac{\partial^2 y(L,x)}{\partial x^2} = 0$. By decomposing the solution into a sine basis (as in the tube case), and by applying a source term $\Psi(x,t)$ to the string at a point x_0 so that $\Psi_n(t) = \Psi(t)\sin(n\pi\frac{x_0}{L})$, the equation can be rewritten in a similar way as equation (5.2), namely:

$$_n \quad N, \qquad \rho_L \frac{\partial^2 y_n(t)}{\partial t^2} + R\frac{\partial y_n(t)}{\partial t} + (T\frac{n^2\pi^2}{L^2} + EI\frac{n^4\pi^4}{L^4})y_n(t) = \Psi_n(t)$$

By applying the Laplace transform on this equation, the solution can be written as:

$$y_n(p) = \frac{\Psi_n(p)}{\rho_L p^2 + Rp + (T\frac{n^2\pi^2}{L^2} + EI\frac{n^4\pi^4}{L^4})} = \frac{\Psi_n(p)}{\rho_L(p+\alpha_n)^2 + \omega_n^2},$$

where

$\alpha_n = \frac{R(\omega)}{2\rho_L}$ are damping factors corresponding to each mode, and

$\omega_n = \sqrt{\frac{T}{\rho_L}\frac{n^2\pi^2}{L^2} + \frac{EI}{\rho_L}\frac{n^4\pi^4}{L^4} - \frac{R(\omega)^2}{4\rho_L^2}}$ are the corresponding eigen frequencies.

In *Figure 5.4* the theoretical damping factors are plotted as a function of the mode rank for a string of length 0.9 m. The damping factor alpha increases when the mode number increases, but compared to the tube case the damping factors are very small. This states the fact that the modes in a string are much less attenuated than the modes in a tube.

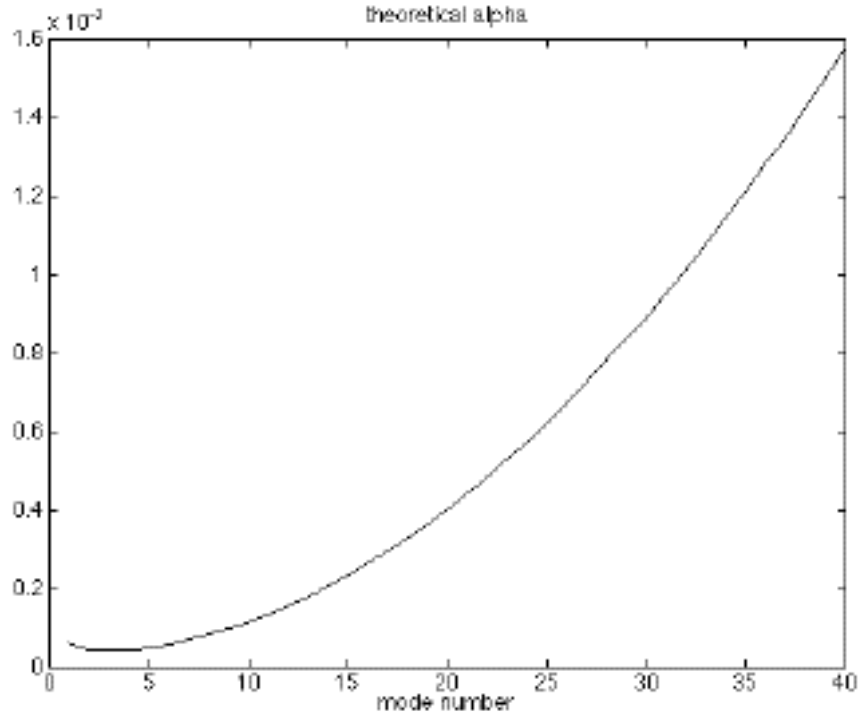


Figure 5.4: Theoretical damping factors as a function of the mode number.

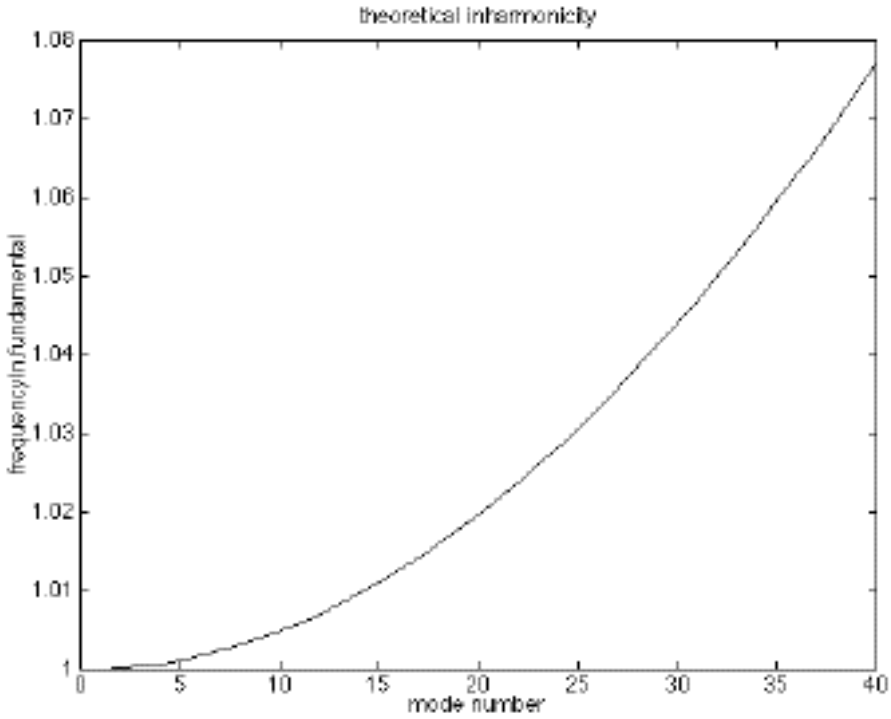


Figure 5.5: Theoretical inharmonicity of the modes in a string

Figure 5.5 shows the theoretical inharmonicity of the modes in a string. One can see that the inharmonicity of the modes in a string is much more important than in the tube case, and that $f_n > nf_1$, which means that the frequency interval between the partials of the spectrum gets bigger as the frequency increases.

The inverse transform of $y_n(p)$ is similar to the tube-case (5.4) namely

$$y_n(t) = \frac{\sin(n\pi \frac{x_0}{L})}{\rho_L \omega_n} \int_0^t \Psi(u) \sin(\omega_n (t-u)) e^{-\alpha_n (t-u)} du$$

And the general solution of the wave equation when applying a punctual dirac source at $x = x_0$:

$$y(x,t) = \frac{2}{L} \sum_{n=1}^{\infty} \frac{\sin(n\pi \frac{x_0}{L}) \sin(n\pi \frac{x}{L})}{\rho_L \omega_n} \sin(\omega_n t) e^{-\alpha_n t}$$

In *Figure 5.6* the spectrum of this expression is given. The length of the string is L=0.9m, the excitation point is x_0=0.2m and the density is ρ_L=0.0015kg/m.

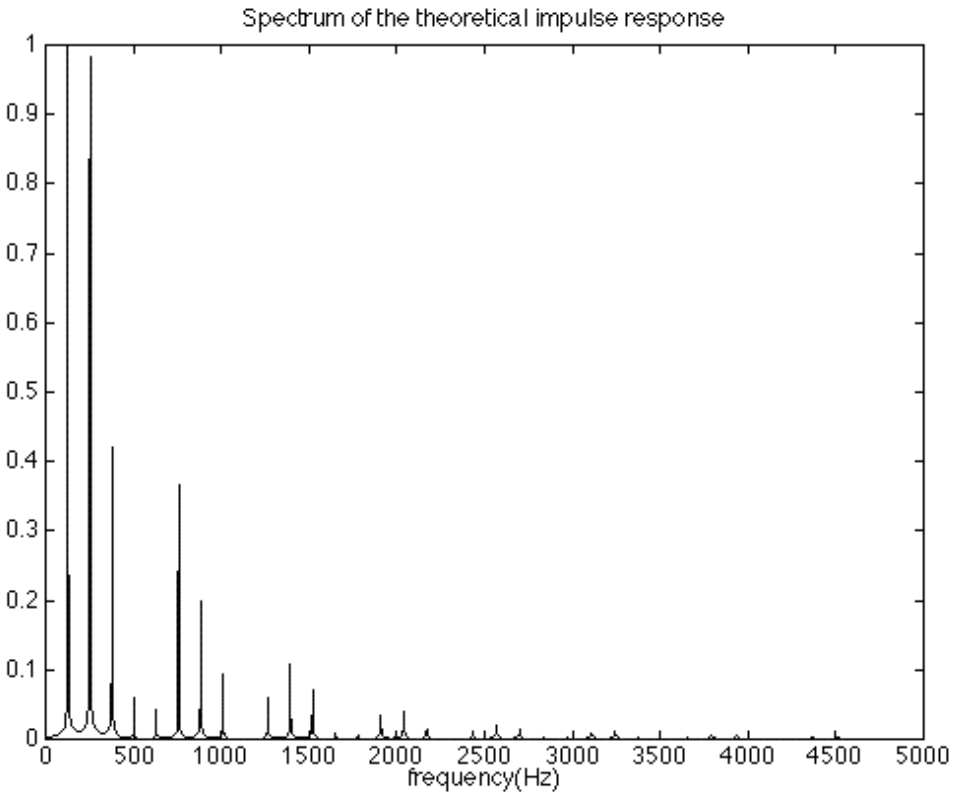


Figure 5.6: Spectrum of the theoretical impulse response of a string when the excitation point is x_0=0.2m.

Compared to the spectrum of the theoretical impulse response of a tube, the spectrum of the theoretical impulse response of a string is « richer ». The periodical spectral modulation is due to the choice of the excitation point x_0.

5.2. Construction of a propagative synthesis model

In chapter 2.2 a general propagation model of a bounded dissipative and dispersive medium is described. *Figure(2.3)* illustrates a propagation model with delay lines, dispersive and dissipative filters. In what follows this model is further simplified by grouping the delay lines into one delay line *d* corresponding to the time it takes for the wave to propagate back and forth in the medium (tube or string), and describe the construction of a filter F taking into account dispersion and dissipation phenomena as well as reflections at the extreme limits of the medium. The parameters being used for this purpose are the damping factors and the frequencies of the partials. These parameters can either be found from the theoretical equations or from the analysis of real sounds.

5.2.1. Characterization of the loop filter

Figure 5.7 illustrates our waveguide model, where E corresponds to the excitation signal and S corresponds to the output signal obtained at the point where the vibrations are measured.

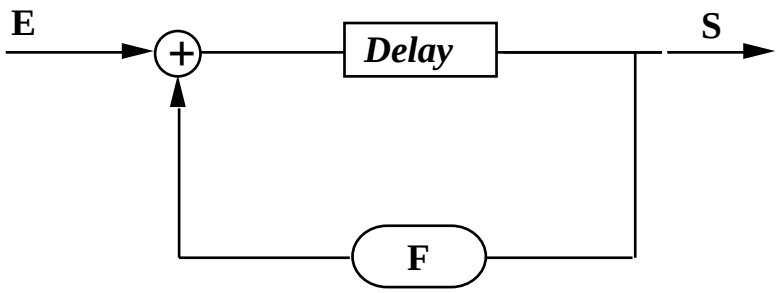


Figure 5.7: Waveguide model

The transfer function of the system is given by

$$T(\omega) = \frac{e^{-i\omega d}}{1 - F(\omega)e^{-i\omega d}} ,$$

where *d* is the delay given by $d = \frac{1}{f_1} = \frac{2L}{c}$, where f_1 is the fundamental frequency and *L* is the length of the string.

If this model is well adapted to the simulation of propagation given by the previous equations, we expect the power spectral density of the transfer function to be a sum of lorentzien functions $\zeta(x) = \frac{1}{\alpha + i(\omega - \beta)}$ like the power spectral density of the theoretical impulse response.

Let us write $F(\omega) = |F(\omega)|e^{i\varphi}$. The transfer function of the waveguide model can then be written as

$$T(\omega) = \frac{e^{-i\omega d}}{1 - |F(\omega)|e^{i\varphi}e^{-i\omega d}} = \frac{e^{-i\omega d}}{1 - |F(\omega)|e^{-i(\omega d - \varphi)}} = \frac{\cos(\omega d) - i\sin(\omega d)}{1 - |F(\omega)|(\cos(\omega d - \varphi) - i\sin(\omega d - \varphi))}$$

and the power spectral density of the system, namely $S(\omega)$ is given by:

$$S(\omega) = T(\omega)\overline{T}(\omega) = \frac{\cos(\omega d)^2 + \sin(\omega d)^2}{(1 - |F(\omega)|\cos(\omega d - \varphi))^2 + (|F(\omega)|\sin(\omega d - \varphi))^2}$$

$$= \frac{1}{1 + |F(\omega)|^2 - 2|F(\omega)|\cos(\omega d - \varphi)}$$

The maxima of $S(\omega)$ are obtained when $\cos(\omega d - \varphi) = 1$, that is for $\omega_n = \frac{2n\pi + \varphi}{d}$. These frequencies correspond to the "peaks" of the spectrum, and one can see that when $F(\omega)$ is real valued, $\varphi = 0$, and the spectrum is harmonic (equation without losses).

In order to approximate the behavior of these resonances, we suppose that the "peaks" are narrow-band enough to be considered as separated in the frequency domain. This assumption is generally reasonable for musical instruments.

Close to the resonance peaks, $\omega = \omega_n + \varepsilon$, where $\varepsilon << 1$. This gives

$$\cos(\omega d - \varphi) = \cos\left[(\omega_n + \varepsilon)d - \varphi\right] = \cos\left(\frac{2n\pi + \varphi}{d} + \varepsilon d - \varphi\right)$$

$$= \cos(2n\pi + \varphi + \varepsilon d - \varphi) = \cos(\varepsilon d)$$

Since $\varepsilon << 1$, we use a limited expansion of $\cos(\varepsilon d)$ in the neighborhood of 0, that is

$$\cos(\varepsilon d) \approx 1 - \frac{\varepsilon^2 d^2}{2}.$$

The power spectral density $S(\omega)$ for $\omega \approx \omega_n$ can now be written as

$$S(\omega_n) = \frac{1}{1 + |F(\omega_n)|^2 - 2|F(\omega_n)|(1 - \frac{\varepsilon^2 d^2}{2})} \qquad (5.6)$$

On the other hand, if one consider an exponentially damped sinusoid with amplitude C_n (corresponding to the theoretical impulse response)

$$s_n(t) = C_n e^{-\alpha_n t} e^{i\omega_n t}$$

with Fourier transform

$$s_n(\omega) = \frac{C_n}{\alpha_n + i(\omega - \omega_n)}$$

its power spectral density is given by

$$S_n(\omega) = s_n(\omega)\overline{s_n}(\omega) = \frac{C_n^2}{\alpha_n^2 + (\omega - \omega_n)^2}$$

Similarily, by considering the local behavior of $S_n(\omega)$ around the resonance frequency ($\omega = \omega_n + \varepsilon$), we get:

$$S_n(\omega) \approx \frac{C_n^2}{\alpha_n^2 + \varepsilon^2} \tag{5.7}$$

By comparing (5.6) and (5.7), and by assuming that the input of the waveguide model $E(\omega_n)$ takes into account the amplitudes of the modes, we can find an approximation of $|F(\omega_n)|$ depending on the damping factors α_n and the delay d.

$$\frac{|E(\omega_n)|^2}{1 + |F(\omega_n)|^2 - 2|F(\omega_n)|(1 - \frac{\varepsilon^2 d^2}{2})} = \frac{C_n^2}{\alpha_n^2 + \varepsilon^2}$$

by a term to term comparison

$$(1 - |F(\omega_n)|)^2 = |E(\omega_n)|^2 \frac{\alpha_n^2}{C_n^2} \tag{5.8}$$

and $$|F(\omega_n)|\varepsilon^2 d^2 = |E(\omega_n)|^2 \frac{\varepsilon^2}{C_n^2} \tag{5.9}$$

From equation (5.9) we see that $C_n^2 = \frac{|E(\omega_n)|^2}{|F(\omega_n)|d^2}$. Equation (5.8) can now be written as

$$(1 - |F(\omega_n)|)^2 = \alpha_n^2 |F(\omega_n)| d^2$$

which gives the expression of the modulus of the filter $F(\omega)$ at the "peaks":

$$|F(\omega_n)| = \frac{2 + d^2\alpha_n^2 \pm \sqrt{(2 + \alpha_n^2 d^2)^2 - 4}}{2}$$

Since higher harmonics decrease faster than lower harmonics, the minus is chosen in the solution of $|F(\omega_n)|$. The phase is:

$$\varphi(\omega_n) = \omega_n d - 2n\pi$$

The input signal $|E(\omega_n)|$ can be modeled as the impulse response of a filter whose response at ω_n is given by $C_n d\sqrt{|F(\omega_n)|}$.

Figure 5.8 and *Figure 5.9* illustrate the modulus of the loop filter F when using the theoretical values of α_n, ω_n and C_n.

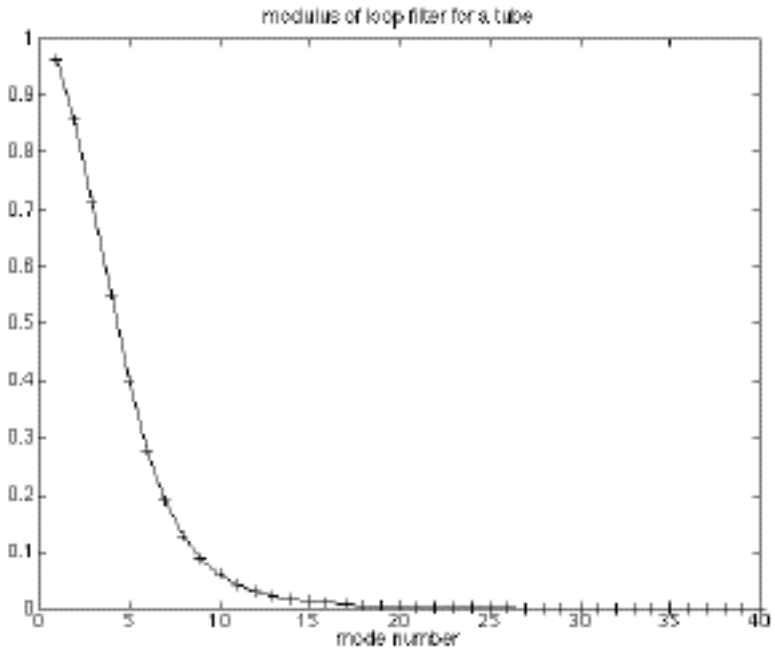


Figure 5.8: Modulus of the filter F in the tube case

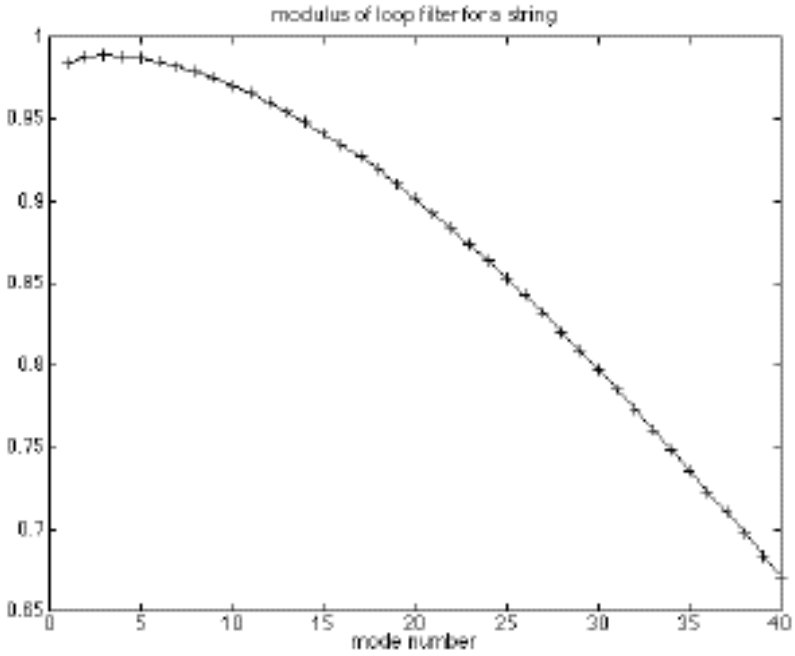


Figure 5.9: Modulus of the filter F in the string case

From equation (5.5) the amplitude in the tube case is

$$C_n = \frac{\sin(n\pi \frac{x_0}{L})\sin(n\pi \frac{x}{L})}{\frac{1}{c^2}\omega_n},$$

where x_0 is the excitation point.

As expected, the tube filter is much more low-pass than the string filter. This means that the modes in the tube are much more rapidly attenuated than in the string. In addition there are much more significant modes in the string case than in the tube case, which corresponds to what we expected, since the spectrum of a sound from a string (*Figure 5.6*) is much « richer » than the spectrum of a sound from a tube (*Figure 5.3*).

5.2.2. Construction of the loop filter

From the discrete values of the loop filter found in the previous section we now want to construct an impulse response which takes into account dispersion phenomena in addition to dissipation. Although several algorithms for constructing such filters already exist [Yegnanarayana, 1982], these algorithms are based on approximations and are therefore not precise enough for resynthesis where we want to

reproduce a sound which ressembles as closely as possible to the original one. This is the reason why a new way of constructing such filters based on the time-frequency representation of a transient sound must be designed. Let us consider the time-frequency representation of a transient propagating in a piano string (*Figure 5.10).* As we can see, the propagation of the transient in a dispersive and dissipative medium leads to a delocalized wave packet. As we can see the energy of the transform is localized along curves called the ridges of the transform [Guillemain et al., 1996]. The ridges are related to the group delay $\tau_g(\omega) = -\frac{d\varphi(\omega)}{d\omega}$ of the signal (where $\varphi(\omega)$ represents the phase of the Fourier Transform of the signal) [Max, 1987]. They are related to the way the propagation speed depends on the frequency.

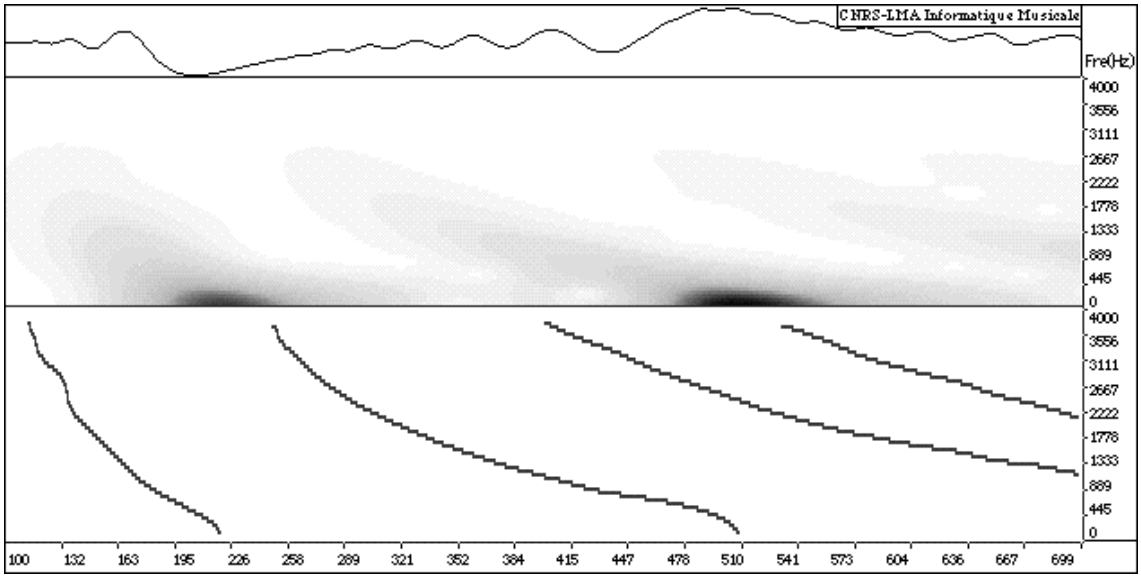


Figure 5.10: Modulus and ridge of the Gabor representation of the first 66ms of the ponctual speed of a piano string measured by laser vibrometry. The black curves correspond to the ridges related to the group delay of the wave. One can see that deformation increases with time.

The restriction of the time-frequency representation along the ridge contains most of the information concerning the signal and consequently allows an approximate reconstruction of it.

In the case of the Gabor representation the reconstruction consists in summing the grains $w(t-\tau)e^{-i\omega(t-\tau)}$ with weight given by the values of the transform along the ridge. In a similar way, the impulse response of the loop filter can be constructed by summing up elementary grains along the group delay curve (*Figure 5.11*).

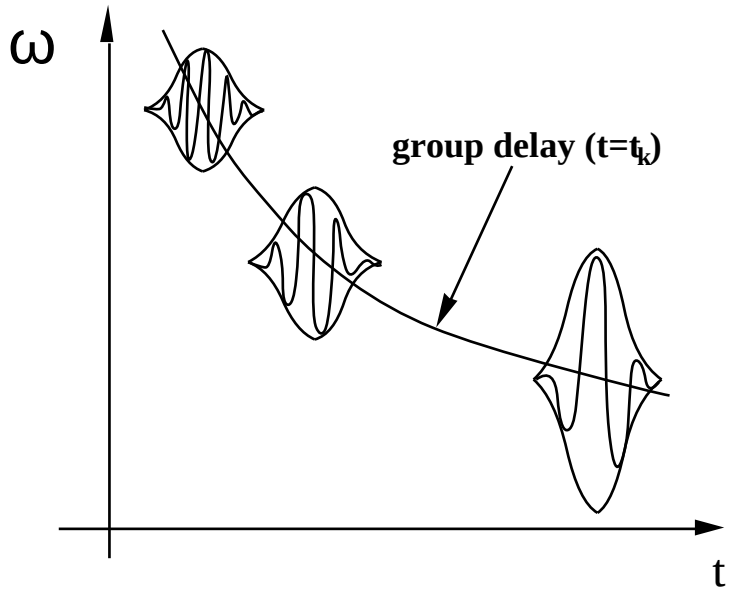


Figure 5.11: Construction of the impulse response of the filter F by adding wave packets along the curve corresponding to the group delay

The general expression of the impulse response of the filter can then be written:

$$f(t) = \sum_{k=1}^{K} \left\{ \alpha_k \cos(\omega_k (t - t_k)) + \beta_k \sin(\omega_k (t - t_k)) \right\} e^{-\frac{(t-t_k)^2}{\sigma^2}} .$$

In order to find the coefficients α_k and β_k corresponding to the amplitudes of the elementary grains, I chose modulated gaussian grains, which are invariant by Fourier transform. The Fourier transform of each grain is then given by a gaussian function centered at the frequency of the grain. The amplitudes of the elementary grains correspond to the coefficients α_k and β_k so that the sum of these grains makes a continuous curve passing exactly through the values of the modulus of the filter found in the previous chapter (*Figure 5.12*).

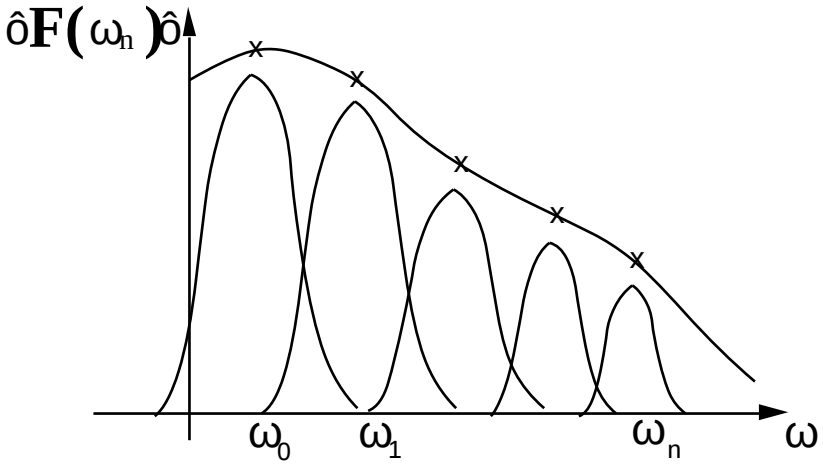


Figure 5.12: Construction of a continuous filter response by summing up gaussian functions

Modulated gaussian grains also have other advantages:

i) gaussian functions represent the best compromize of localization both in the time and the frequency domains,

ii) gaussian functions decrease faster than any polynomial functions,

iii) the Fourier transform of a real gaussian function is real valued, which means that it does not alter the phase of the functions,

iv) gaussian functions were used in the first applications of wavelets effectuated by Morlet and in the Gabor representation. They are easy to calculate, and have shown a good numerical stability.

The expression of the continuous filter response is now given by:

$$F(\omega) = \frac{\sigma\sqrt{2\pi}}{2}\sum_{k=1}^{K}\left[\alpha_k\left(e^{-\frac{1}{2}(\omega-\omega_k)^2\sigma^2} + e^{-\frac{1}{2}(\omega+\omega_k)^2\sigma^2}\right) - i\beta_k\left(e^{-\frac{1}{2}(\omega-\omega_k)^2\sigma^2} + e^{-\frac{1}{2}(\omega+\omega_k)^2\sigma^2}\right)\right]e^{-i\omega t_k}$$

where

K is the number of components,

α_k and β_k are the real and imaginary parts of the amplitude the k'th component,

ω_k is the frequency of the k'th component,

σ is the standard deviation of the gaussian functions (related to their width),

t_k is the group delay.

The standard deviation σ was chosen so that the intersection between the first gaussians arises when the value of the functions is $\frac{1}{\sqrt{2}}$. That is

$$\sigma = \frac{2}{\omega_0}\sqrt{\log(2)}$$

The group delay is due to the inharmonicity of the signal and to the different propagation velocity of each component. It is given by:

$$t_k = T_1 - kT_k = \frac{1}{f_1} - \frac{k}{f_k}$$

where T_1 is the period of the fundamental and T_k the period of the k'th component.

When imposing the continuous filter response to have the same values at the resonance peaks as the values of the modulus of the filter found previously, the coefficients α_k and β_k are given by:

$$\begin{pmatrix}\alpha_k\\ \beta_k\end{pmatrix} = \frac{2}{\sigma\sqrt{2\pi}}A(j,k)^{-1}F(\omega_k),$$

where A is a matrix defined by four matrixes:

$$A(j,k) = \begin{pmatrix}A1(j,k) & A2(j,k)\\ A3(j,k) & A4(j,k)\end{pmatrix}$$

with

$$A1(j,k) = (e^{-\frac{1}{2}(\omega-\omega_k)^2\sigma^2} + e^{-\frac{1}{2}(\omega+\omega_k)^2\sigma^2})\cos(\omega_j t_k)$$

$$A2(j,k) = (e^{-\frac{1}{2}(\omega-\omega_k)^2\sigma^2} - e^{-\frac{1}{2}(\omega+\omega_k)^2\sigma^2})\sin(\omega_j t_k)$$

$$A3(j,k) = (e^{-\frac{1}{2}(\omega-\omega_k)^2\sigma^2} + e^{-\frac{1}{2}(\omega+\omega_k)^2\sigma^2})\sin(\omega_j t_k)$$

$$A4(j,k) = (e^{-\frac{1}{2}(\omega-\omega_k)^2\sigma^2} - e^{-\frac{1}{2}(\omega+\omega_k)^2\sigma^2})\cos(\omega_j t_k)$$

and

$$F(\omega_k) = \begin{pmatrix} F(\omega_k)\cos(\omega_k t_k) \\ F(\omega_k)\sin(\omega_k t_k) \end{pmatrix}.$$

This filter construction method allows an exact reconstruction of dispersion and dissipation at the resonance peaks giving a high quality resynthesis of transient sounds.

Figure 5.13 and *Figure 5.14* show the theoretical impulse response of the filter F calculated by the proposed method in the tube case and in the string case.

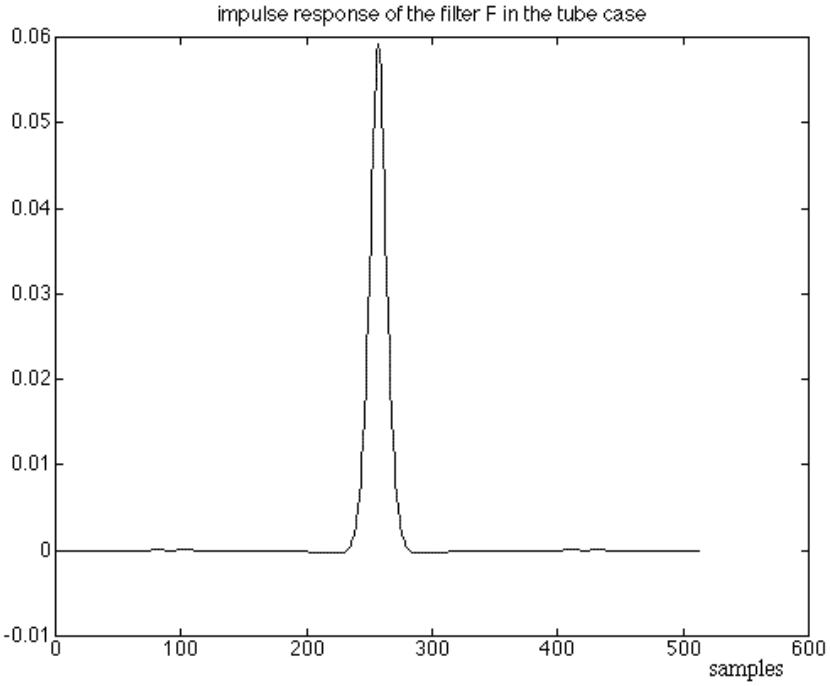


Figure 5.13: Impulse response of the filter F in the waveguide model using theoretical values for damping coefficients and eigen frequencies corresponding to the tube case (f_s=32000Hz)

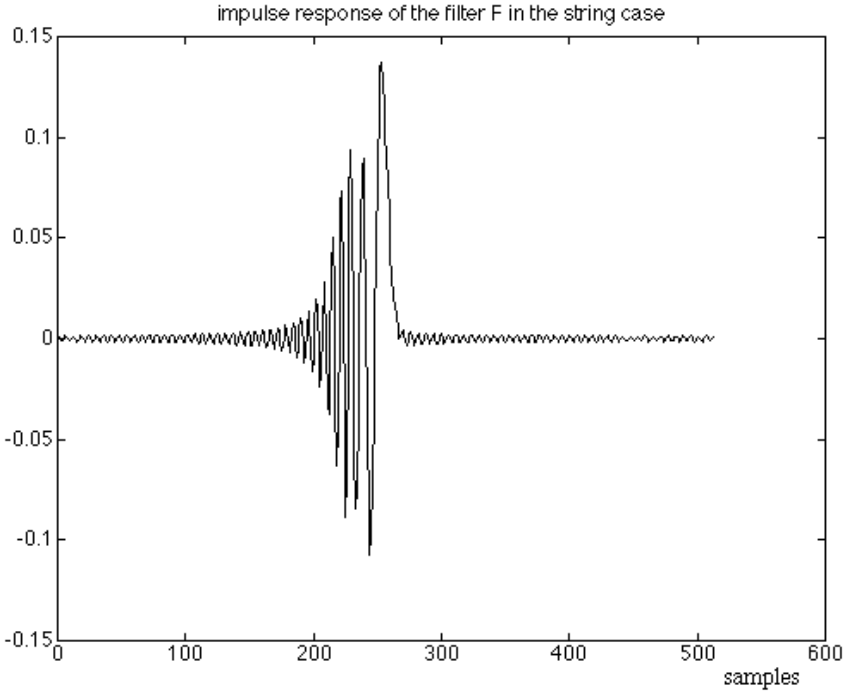


Figure 5.14: Impulse response of the filter F in the waveguide model using theoretical values for damping coefficients and eigen frequencies corresponding to the string case (f_s=32000Hz)

It is interesting to note that in the tube case the impulse response is almost symmetric due to the weak inharmonicity while it in the string case is strongly asymmetric due to the important inharmonicity of the modes in the string. In the tube case where $f_n < nf_1$, high frequency components are propagating « more slowly » than low frequency components. In this case the group delay t_k is negative. In the string case where $f_n > nf_1$, t_k is positive.

5.3. Examples of sound modeling

In order to compare theoretical and real values of the parameters characterizing the medium, real sounds from a flute and from a guitar string have been recorded. In the flute case transient sounds were measured in an anechoïc room by rapidly closing a finger hole without exciting at the embouchure. This sound is interesting since it can be heard during the instrumental play. In the string case the vibrations of a guitar string were measured by the use of a laser vibrometer.

5.3.1. Modeling of the flute

The theoretical equations show that the response to a punctual dirac source in a tube is a sum of sinusoids at the frequencies ω_n exponentially damped with a damping factor α_n. This is an important result indicating what the behavior of the modes in a tube may be like. The amplitude modulation laws of

the spectral components of a transient signal confirm the theoretical results. Thus the damping factors α_n and the frequencies ω_n can be found from the analysis of real sounds. *Figure 5.15* shows the amplitude modulation law of the first component of the spectrum of a transient flute sound obtained using the « matched » time-frequency analysis method. The damping factor can be found by measuring the slope of this amplitude modulation law in a logarithmic representation.

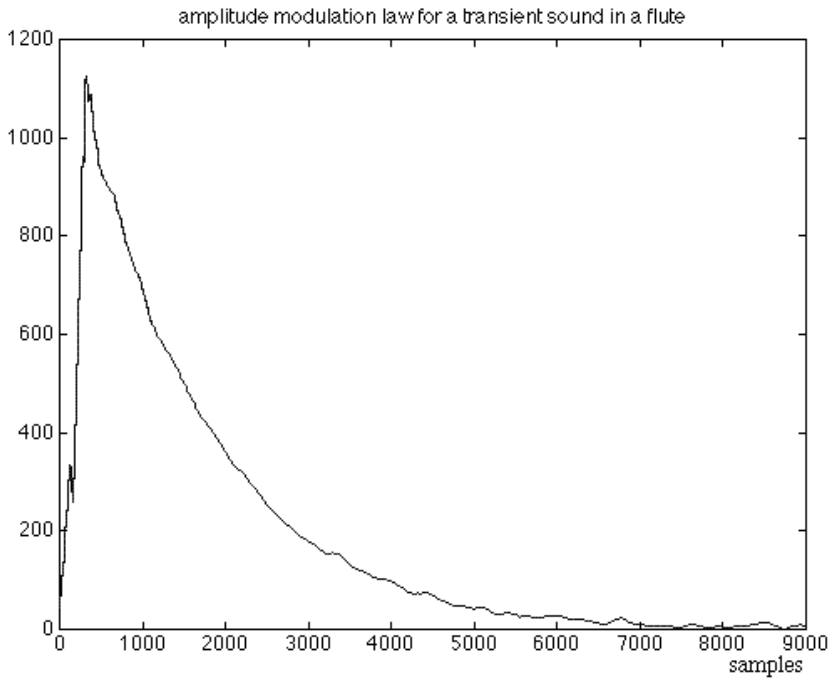


Figure 5.15 Amplitude modulation law of the first component of the spectrum of a transient flute sound (linear representation) f_s*=48000Hz*

The shape of the filter $F(\omega)$ is the same for different lengths of the tube (corresponding to different notes). As seen in section 5.2, the filter $F(\omega)$ depends on the delay d. *Figure 5.16* shows the theoretical filter for L, $L/2$, $L/4$ and $L/8$, where $L=0.58m$.

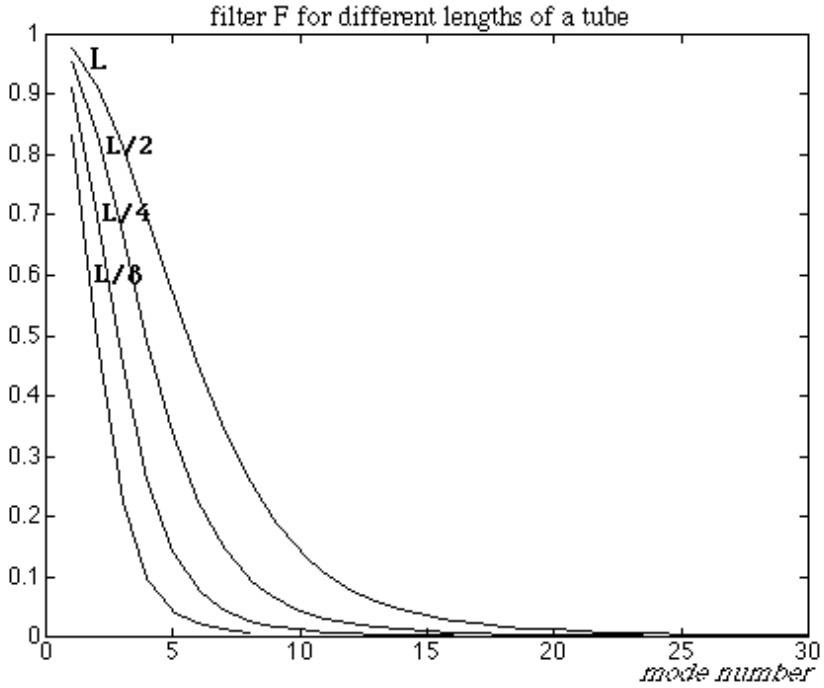


Figure 5.16: Evolution of the filter F due to changes of the tube length L.

The relation between the filter and the length is given by an homotecy (change of scale):

if $F(\omega)$ is the filter corresponding to the tube length L_0 and $F_L(\omega)$ the filter corresponding to the tube length L, then

$$F_L(\omega) = F\left(\frac{\omega}{\sqrt{n}}\right), \text{ where } n = \frac{L_0}{L}.$$

This is an important result for the real-time implementation of the model as is seen in chapter 7, since changing the delay in the physical model (changing the note in an instrument) only necessitates a dilation or a contraction of the filter.

As discussed in section 5.1, the value of l_{hv} in the wave equation of a tube, depends on the friction near the body of the instrument. It can be determined by comparing the theoretical damping factors given by $\alpha_n = l_{hv} c \frac{n^2\pi^2}{2L^2}$ and the real damping factors as seen in *Figure 5.17.*

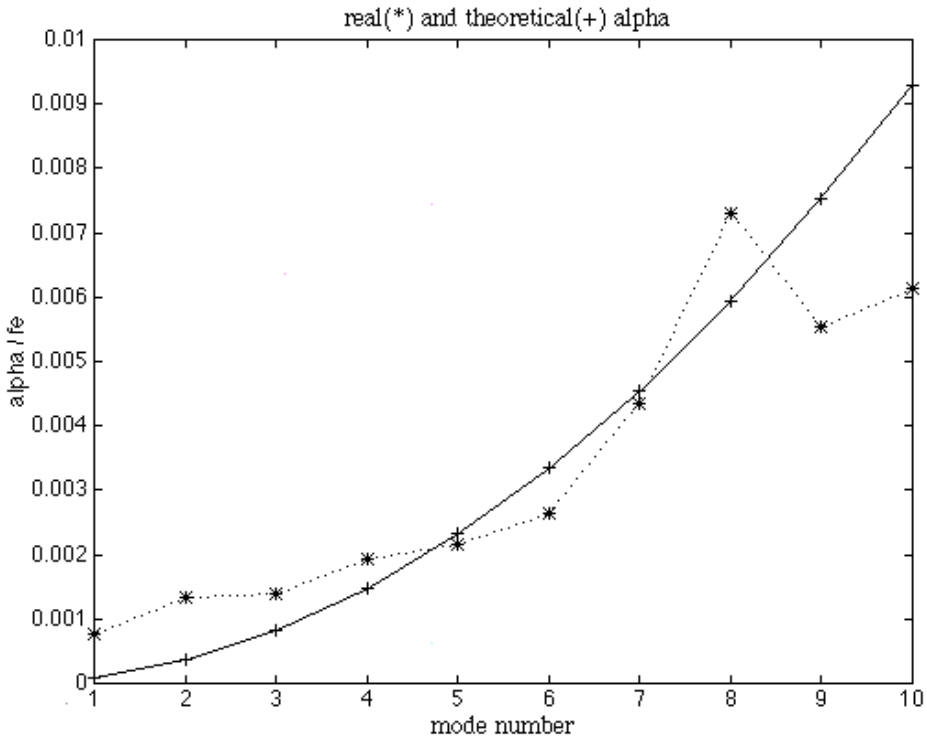


Figure 5.17: Damping factors in the real (*) and theoretical(+) cases

From the experience, the estimated l_{hv} leading to the best superposition of the theoretical and experimental curves is of order $10^{-3}\,m$. This value is much higher than the theoretical value in free field, which shows that the friction of the air near the body of the instrument can be seen as responsible of most of the losses.

Figure 5.18 displays respectively the theoretical and the estimated inharmonicity of the modes.

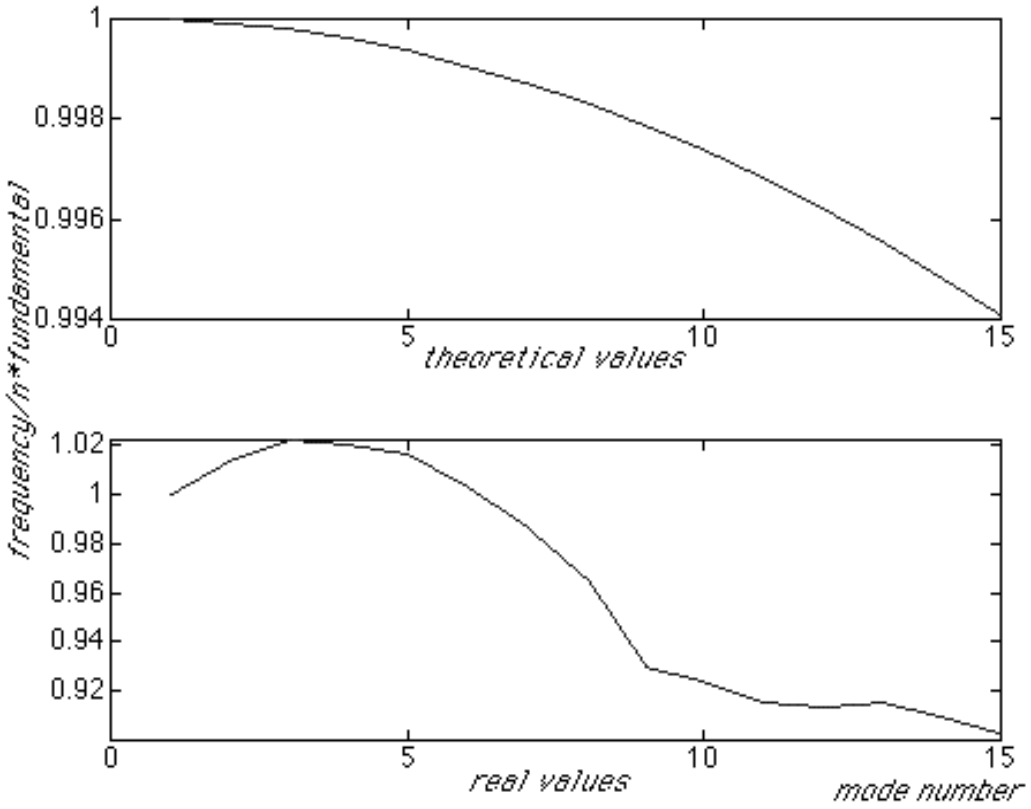


Figure 5.18: $\frac{f_n}{nf_0}$ *for a tube of length L=0.58m in the theoretical and real cases*

.

The inharmonicity corresponds to the ratio between the frequency and the perfectly harmonic case where $f_n = nf_0$ (f_0 being the fundamental). The estimated frequency shift is larger than the theoretical one. This is probably because of the simplifications of the theoretical model (plane wave, tube without holes...).

Figure 5.19 and *Figure 5.20* show the Fourier transforms of the input E and the loop filter F corresponding to the real response of the flute, measured as described previously. As mentioned in section 5.2.1., the input E of the waveguide model takes into account the amplitudes of the modes. One can see that the amplitudes of the first modes in a real signal are strong, and that the amplitudes of the higher harmonics are rapidly attenuated. The Fourier transform of the loop filter F shows that in the theoretical case the filter is much more selective than in the real case. This means that in the real case the high-frequency components of the spectrum are less attenuated than in the theoretical case. This may be due to the simplifications done in the theoretical study where only plane waves were assumed to propagate in a tube without fingerholes. In addition, the source was not the same in the real and in the theoretical cases, since a punctual dirac source was used to calculate the theoretical impulse response, while the « real » impulse response was obtained by rapidly closing a fingerhole.

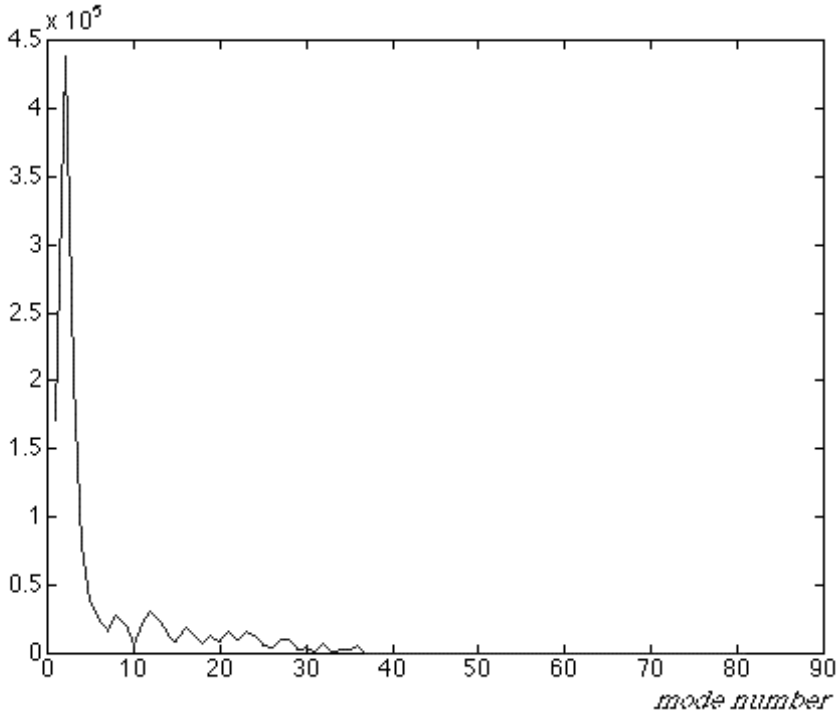


Figure 5.19: Fourier transform of the input signal E corresponding to a transient in a real flute

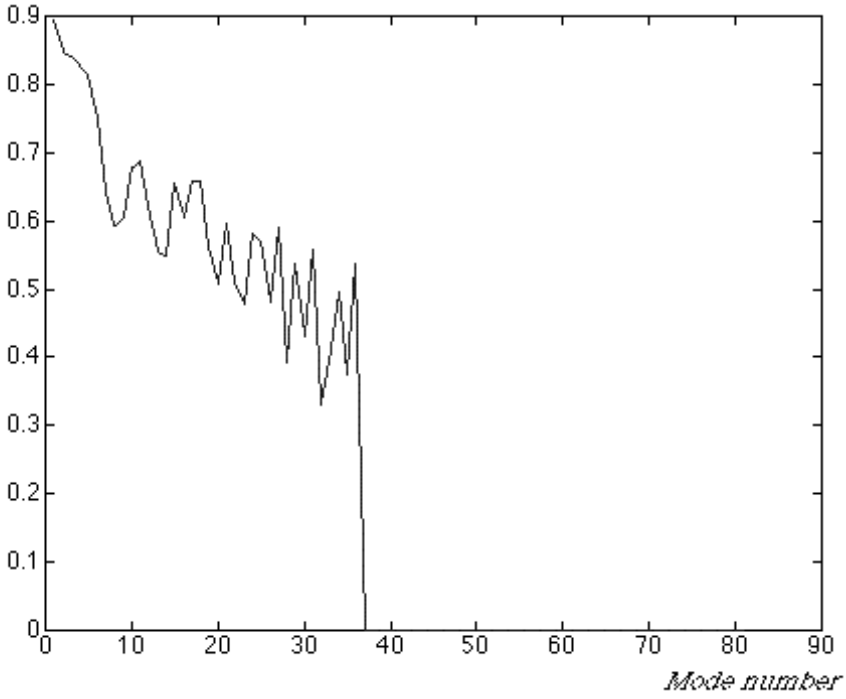


Figure 5.20: Fourier transform of the filter F corresponding to a transient in a flute

5.3.2. Modeling of the guitar string

The amplitude modulation laws corresponding to the partials of a guitar string show that the spectral components are exponentially damped, *Figure 5.21.* This confirms the result from the theoretical equations. The attenuation of the components is weak compared to the tube case . *Figure 5.22* displays respectively the theoretical and estimated inharmonicity of a steel guitar string. The estimated data are in good accordance with the model.

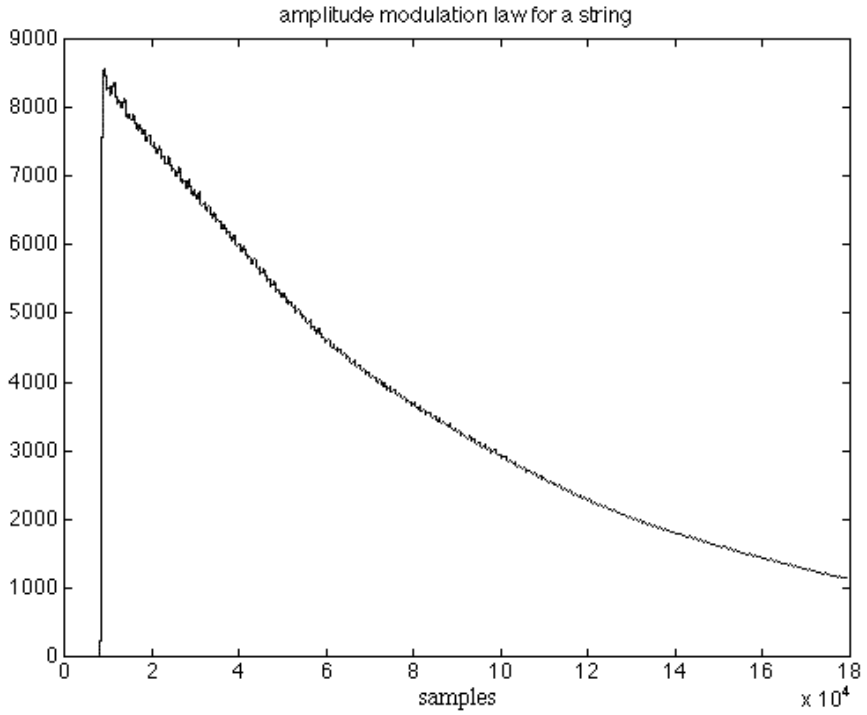


Figure 5.21: Amplitude modulation law of the first component of a guitar sound (f_s=32000Hz)

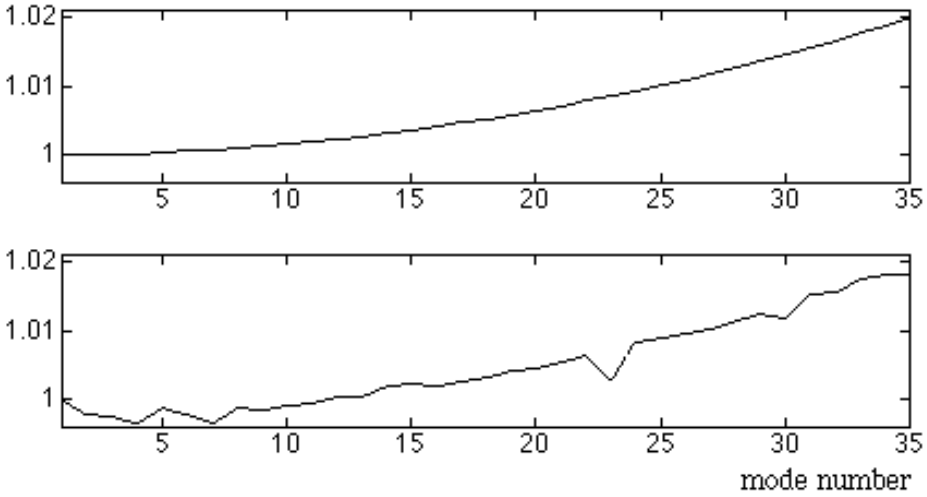


Figure 5.22: $f_n/(nf_0)$ of a steel guitar string in the theoretical and real cases

Figure 5.23 and *Figure 5.24* display the Fourier transforms of the input signal (which corresponds to the Fourier transform of the input filter E defined in section 5.2.1) and the loop filter F corresponding to a guitar sound. Since the filters have been estimated directly from the sound itself, E includes the frequency response of the soundboard mechanically coupled with the string, and its radiative characteristics. One can observe greater losses for the high frequencies than for the lower ones since high frequency components decrease faster than low spectral components. This phenomenon can be justified theoretically by a loss coefficient R being a function of the frequency (section 5.1.2).

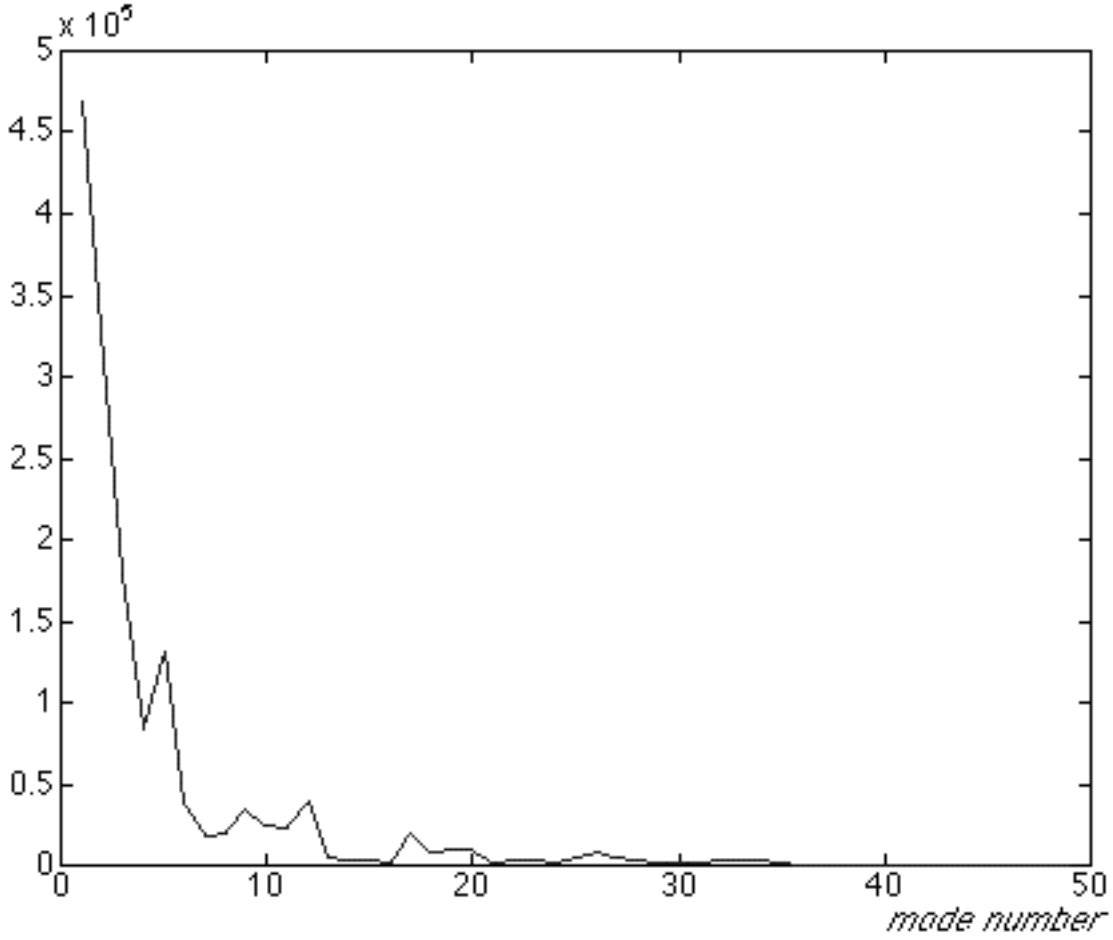


Figure 5.23: The Fourier transform of the input signal E corresponding to a guitar sound

The accordance between the loop filters in the theoretical and in the real case is better in the string case than in the tube case. This may be due to the fact that the excitation of the string in the real case is close to the punctual excitation used in the theoretical case, since the real string is excited by giving it an initial displacement. Moreover, it is easier to be more precise in the measurement of the string motion than in the measurement of the pressure in the tube.

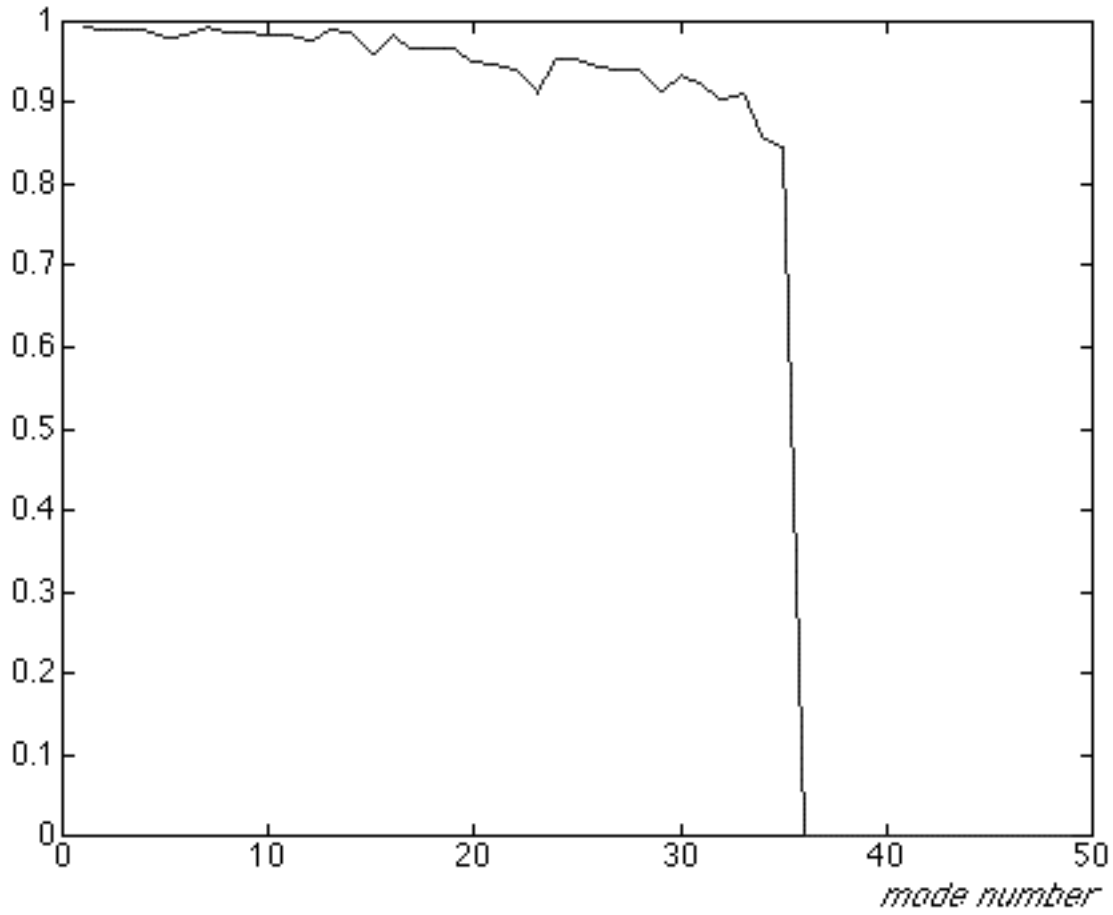


Figure 5.24: The Fourier transform of the filter F corresponding to a guitar sound

In both the tube -and the string cases, physical modeling gives a good description of the behavior of the sound produced by the propagation of a transient. In the guitar case where the source is not coupled to the resonator, the synthesized sounds are very close to real sounds. Nevertheless, for sound generator systems using sustained sources (like many musical instruments like wind instruments, string instruments excited by a bow etc.) the input signal has to be identified and modeled separately. In the next chapter, we shall see how one can extract the source from the signal, and how the source from a flute can be modeled. This will lead to a model which gives a good resynthesis of sustained sounds.

6. Sound modeling using a combination of physical and signal models

In this chapter I shall show how a hybrid model using both physical and signal models can be designed and how it can improve the realism of the generated, sustained sound.

For certain instruments the source and the resonator are closely coupled and can therefore not be considered separately. This can be illustrated by considering the flute case. For this instrument the source is the result of an air jet oscillating in response to a transverse acoustical flow due to the acoustical oscillation in the pipe. The interaction between the jet and the labium transfers energy to the acoustic standing waves in the resonator [Verge, 1995]. *Figure 6.1* shows, in a schematic way, how a recorder-like instrument works. The flute case is even more complicated since the distance from the flue exit (corresponding to the player's lips) to the labium as well as the width of the air jet can be varied by the player. By studying recorder-like instruments, Verge has found that these parameters and the jet velocity are of importance for the quality of the sound, since a turbulent or a laminar jet will be produced depending on these parameters. Flow visualizations of the interaction jet/labium have been effectuated, showing vortical structures appearing on each side of the jet. Fabre [Fabre et al., 1996] has found that vortex shedding at the labium is important for the generation of harmonics in the spectrum. The jet and the resonator are assumed to be linear elements, and one assumes that higher harmonics are generated by non-linearities in the interaction between the jet and the labium.

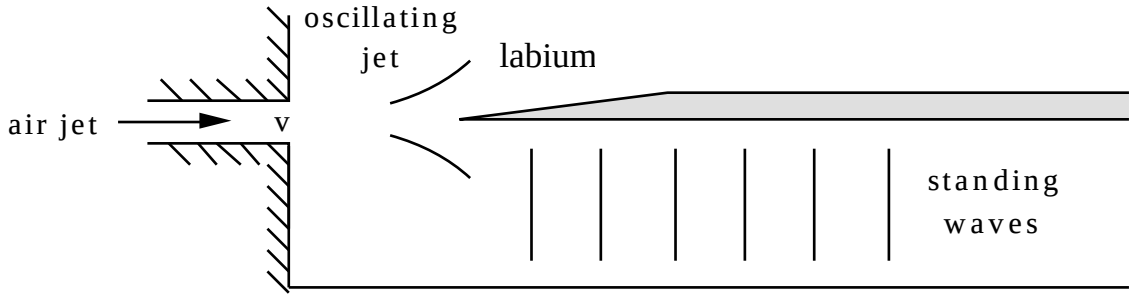


Figure 6.1: Schematic illustration of the excitation system of a recorder-like instrument

Synthesis techniques such as waveguide modeling involving non-linearities in the feedback loop have given interesting results [Cook, 1992] [Smith, 1987]. However, the parameters used in such non-linear systems are difficult to identify from the analysis.

I therefore propose another approach where the source and the resonator are separated, even though they are generally not separated in real systems. The resonator being considered as a linear system, the source can be identified by deconvolution of the sound with the impulse response corresponding to the resonator.

The source signal is decomposed in terms of a deterministic and a stochastic part. This approach is similar to the one proposed by Serra [Serra, 1989] but the separation methods used here are different. The deterministic part of the source signal is modeled by signal model synthesis techniques. These

techniques must behave non-linearly, taking into account the spectral variations as a function of the play (input parameters). The methods are chosen such that the models can be controlled by "measurable" data on the instrument. The final system is to be run in real-time.

6.1. Source identification by deconvolution

In chapter 5.2, it is shown that the resonator can be modeled using a recursive all-pole filter. The sound output y can be written:

$$y(t) = (x * h)(t),$$

where $h(t)$ represents the impulse response of the resonant system and $x(t)$ represents the entrance of the system called the source signal (*Figure 6.2*). The amplitudes of the modes related to the resonant system are here included in the source.

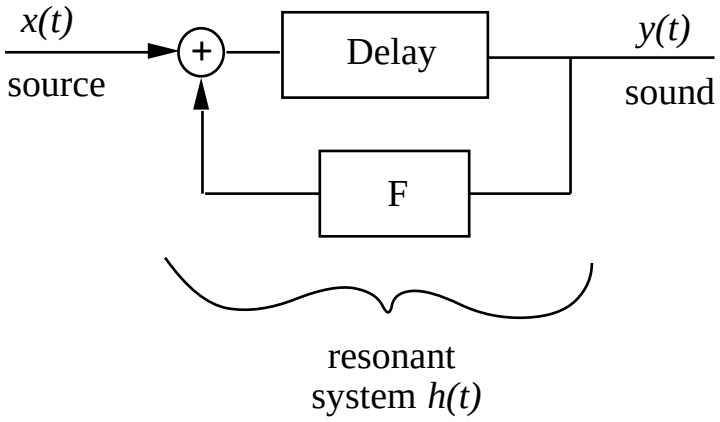


Figure 6.2: Filtering of a source by a resonant system

If $h^{-1}(t)$ such as $(h * h^{-1})(t) = \delta(t)$ exists, the source $x(t)$ can be obtained by deconvolution, that is:

$$x(t) = (y * h^{-1})(t)$$

This operation is legitimate since, as seen in chapter 5.2 the transfer function of the resonant system is given by:

$$H(\omega) = \frac{y(\omega)}{x(\omega)} = \frac{e^{-i\omega d}}{1 - F(\omega)e^{-i\omega l}}$$

and represents an all-pole filter. This means that $H^{-1}(\omega)$ is entirely defined and consequently that $h^{-1}(t)$ exists.

As an example *Figure 6.3* and *Figure 6.4* show respectively the resonant system's transfer function and its inverse found from the analysis of a transient flute sound.

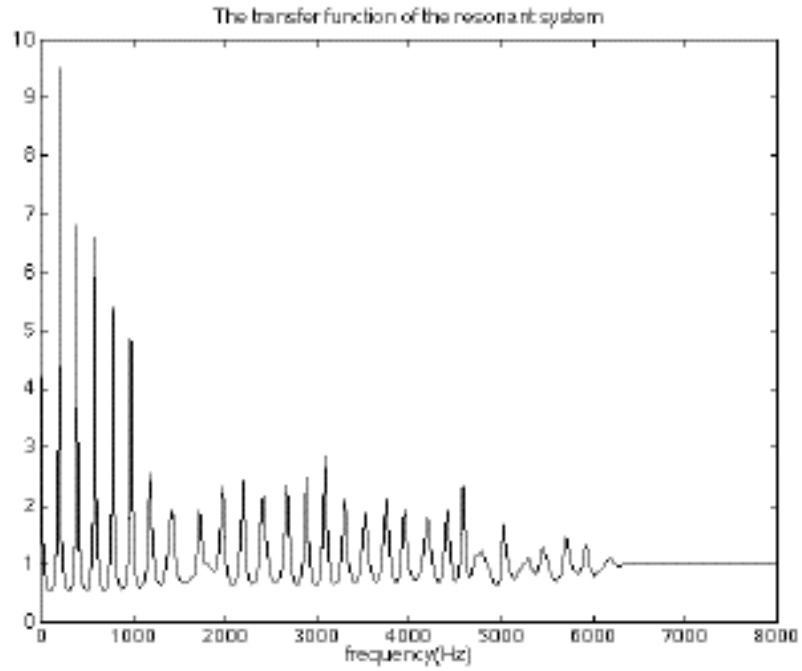


Figure 6.3: Transfer function $H(\omega)$ of resonance model corresponding to a flute

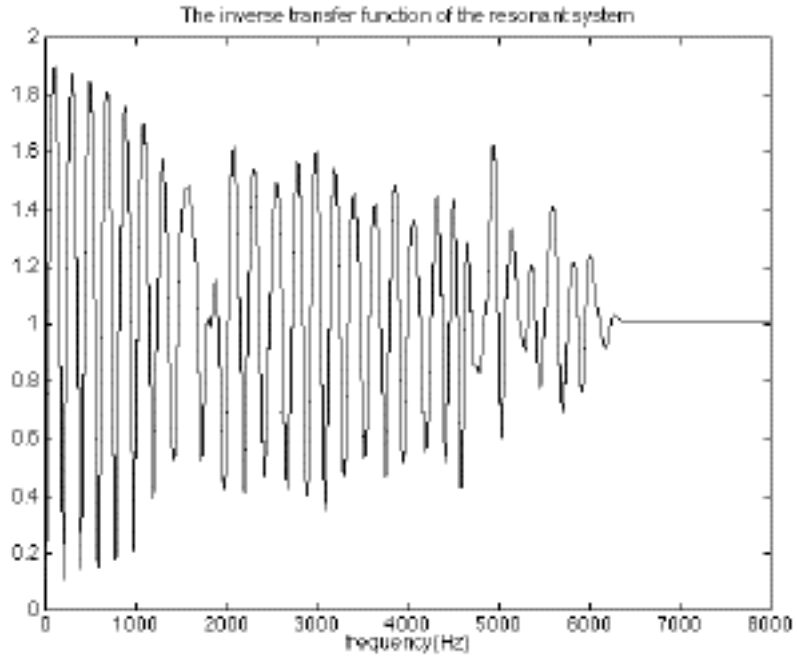


Figure 6.4: Inverse transfer function $H^{-1}(\omega)$ of the resonance model of a flute

Figure 6.5 shows the spectrum of the source obtained by deconvolution. In this case we can see that the source signal contains both spectral lines and a broadband noise (which in what follows respectively will be called the deterministic and the stochastic contributions).

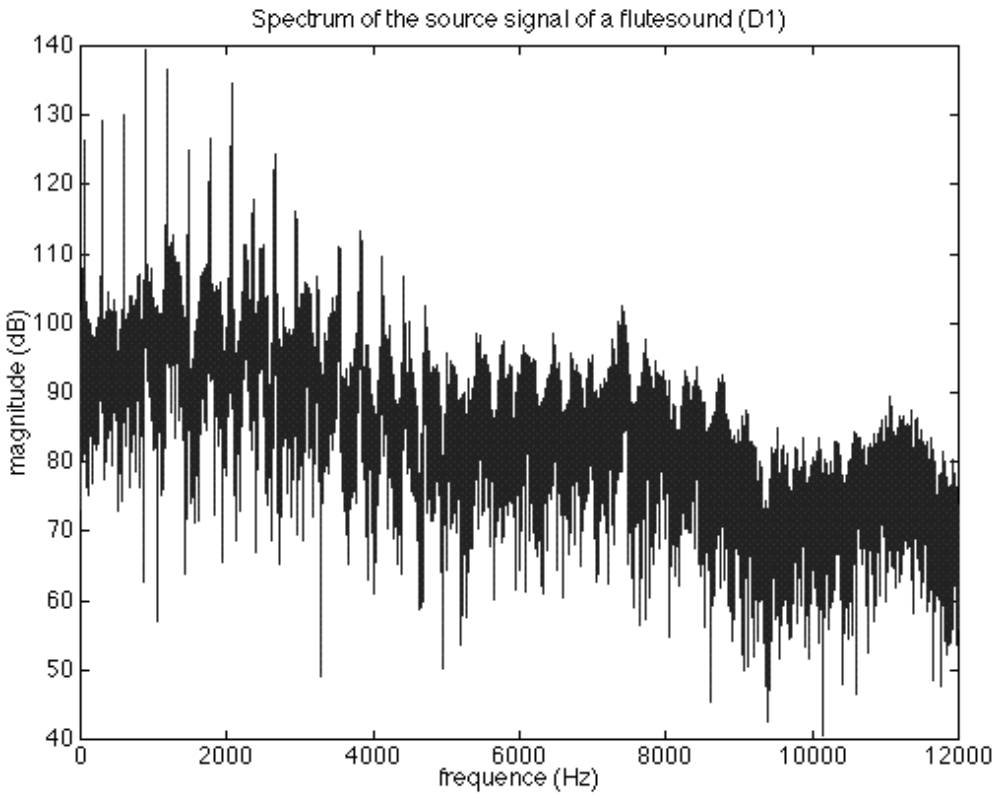


Figure 6.5: spectrum of the deconvoluted signal

6.2. Splitting the deterministic and the stochastic components

To separate the deterministic and the stochastic contributions of the source signal, I first construct an estimate of the deterministic part by selecting the corresponding spectral components. Further on, this deterministic contribution is used as a reference signal for an adaptive filter designed to extract the stochastic contribution.

6.2.1. Estimation of the deterministic contribution

The deterministic contribution of the source signal can be estimated by using a technique related to the one described in chapter 2.1.1. In this technique the deterministic components are selected by filtering in the frequency domain. The filtering is performed using well localized frequency windows $W(\omega - \omega_m)$ adapted to the spectral bandwidth of the components to isolate each of them. The spectral components $h_m(\omega)$ are given by:

$$h_m(\omega) = s(\omega)W(\omega - \omega_m)$$

By summing up the M spectral components one obtains the expression of the estimated deterministic signal, namely

$$\tilde{d}(\omega) = \sum_{m=1}^{M} h_m(\omega) = s(\omega) \sum_{m=1}^{M} W(\omega - \omega_m) = s(\omega) H_d(\omega)$$

This means that the estimated deterministic signal $\tilde{d}$ is found by filtering the source signal with a filter whose frequency response is

$$H_d(\omega) = \sum_{m=1}^{M} W(\omega - \omega_m)$$

If $W(\omega - \omega_m)$ is chosen in order to get an optimal selection of the component ω_m, the noisy part of the signal filtered by $W(\omega - \omega_m)$ can be neglected. In this case $\tilde{d}$ can be considered as an estimation of the deterministic signal d, differing by a linear filtering. This assumption will be of interest when extracting the stochastic part of the sound. *Figure 6.6* shows the spectral representation of the deterministic part of the source signal of a flute sound obtained this way.

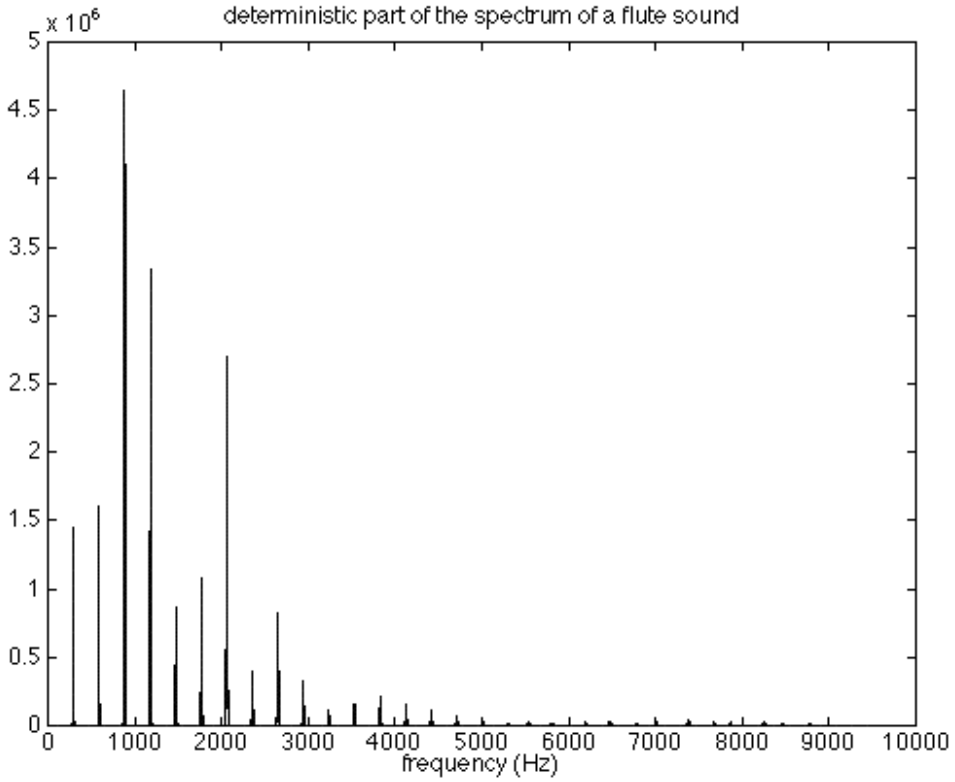


Figure 6.6: spectrum of the deterministic part of the source signal of a flute sound whose fundamental frequency is 293Hz (D1 for a flute)

6.2.2. Estimation of the stochastic contribution

The stochastic contribution can now be found by extracting the deterministic part from the source signal. This can be done by using an adaptive filter. Among numerous methods the LMS (Least Mean Square) algorithm has been chosen for its adequacy with the actual problem.

6.2.2.1. The LMS algorithm

The LMS algorithm is an adaptive method of minimization. It is more specific than Newton's method or the steepest descent method, since it uses a special estimate of the gradient [Widrow, 1985]. It is important because of its simplicity and ease of computation, and is generally the best choice if the adaptive system is a linear, finite impulse response filter. I shall briefly describe the algorithm in the case of a system identification problem. Let us suppose that $p(t) = (H * r)(t)$, and that the response $p(t)$ and the input $r(t)$ are available. The general discrete form of the LMS algorithm when searching for H is given in *Figure 6.5.* W_k represents the adaptive linear filter, while e_k represents the error signal.

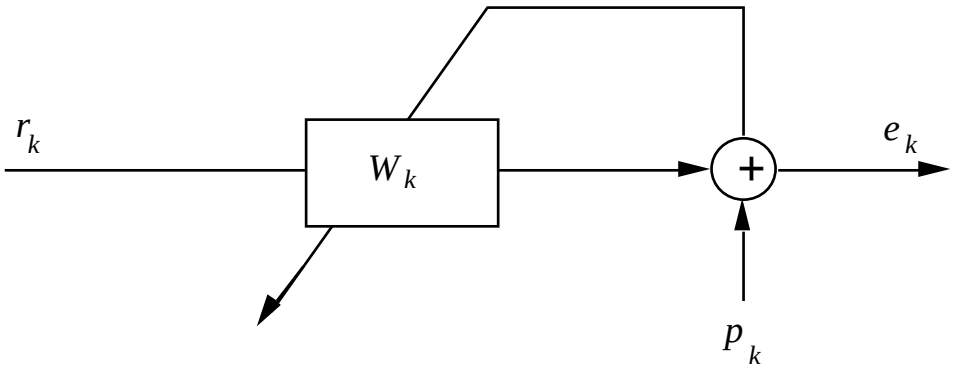


Figure 6.5: General discrete form of the LMS algorithm

The error signal that is to be minimized is given by

$$e_k = p_k - \underline{r}_k \underline{W}_k$$

The LMS algorithm uses e_k^2 as an estimate of the gradient [Widrow, 1985]. Thus at each iteration of the adaptive process the gradient estimate $\tilde{}_k$ is given by

$$\tilde{}_k = \frac{\partial e_k^2}{\partial w_k} = 2e_k \frac{\partial e_k}{\partial w_k} = -2e_k \underline{r}_k$$

The filter W_k should now be updated following a steepest descent type of adaptive algorithm giving

$$\underline{W}_{k+1} = \underline{W}_k - \mu \tilde{}_k = \underline{W}_k + 2\mu e_k \underline{r}_k$$

where μ is a constant gain that regulates speed and adaptation of the algorithm. The convergence of the weight vector $\underline{W}_k$ is assured by:

$$0 < \mu < \frac{1}{L.tr[R]} ,$$

where $R = E[\underline{r}_k \underline{r}_k^T]$ is the input correlation matrix and L is the length of the filter $\underline{W}$. In the case of a transverse filter, one can show that $tr[R] = E[\underline{r}_k^2]$. As a general rule the filter length is determined by the number of spectral components in the reference signal $\underline{r}_k$ and should be twice this number.

6.2.2.2. Application of the LMS algorithm

The stochastic contribution of a sound can be found by using the estimated deterministic part of the signal $\tilde{d}$ as a reference and remove the part of the original source signal that is correlated to this reference signal. In this case we suppose that the deterministic and the stochastic contributions of the source are not correlated. Let's call s_k the source signal.

$$s_k = d_k + b_k ,$$

where d_k is the deterministic contribution of the signal and b_k is the stochastic contribution. The estimated deterministic signal $\tilde{d}_k$ and d_k are related by a linear filter h_k;

$$d_k = \underline{\tilde{d}}_k \underline{h}_k$$

For this specific problem, the principle of the LMS algorithm can be illustrated as in *Figure 6.6*

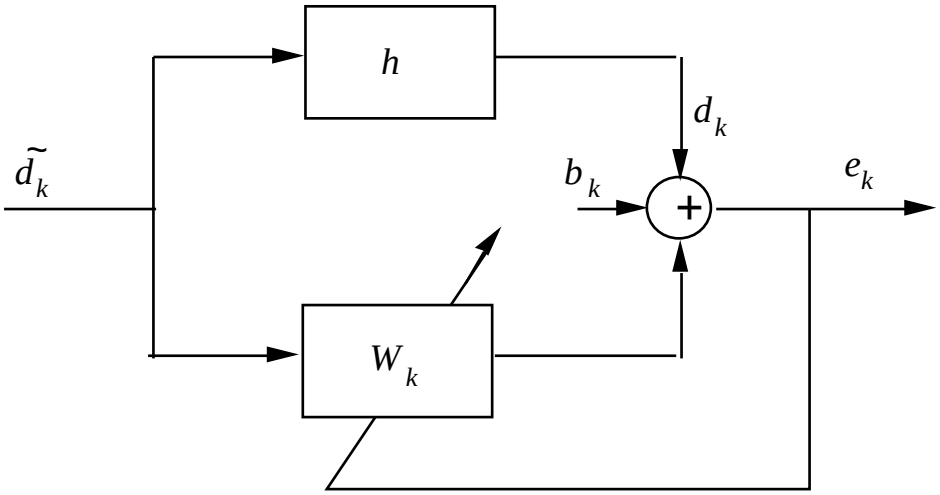


Figure 6.6 An illustration of the principle of the LMS algorithm used to separate the deterministic signal and the noise.

The signal that is to be minimized is now

$$\varepsilon_k^2 = (e_k - b_k)^2$$

and the LMS algorithm will make the adaptive filter w_k converge towards h and the error e_k towards b_k.

Another possibility yet not used here would consist in using a delayed version of the input signal as a reference [Widrow, 1985].

6.3. Modeling of the deterministic part

In many cases, sounds generated by an excited resonant system behave non linearly since the evolution of the spectrum is not a simple amplification. This is the case for most musical sounds for which the dynamic way of playing dramatically acts on the resulting sound. However, this non linear behavior is often related to the excitation, even though some non linearities sometimes appear during propagation [Gilbert et al., 1997]. I shall here only consider non linearities generated by the excitation system. In order to model the deterministic part of the non linear source signal, one can use group additive synthesis as described in chapter 4 or non-linear methods such as waveshaping synthesis. As discussed in chapter 2, global methods have the advantage of generating complex spectra from easy calculations using a small number of operations. A review of the waveshaping method is given in chapter 2.1.3. This method will be applied to model a source signal.

From a modeling point of view, one has to find a function γ and a corresponding index *I(t)* able to generate a given harmonic sound. As seen in chapter 2.1.3, the function γ can easily be linked to the spectrum of the generated sound for an index value *I=1*. In this case the coefficients of the decomposition of γ in the Chebyshev basis are given by the values of the modulus of the spectrum to be generated:

$$\gamma(\cos\omega_0 t) = \sum_{k=0}^{\infty} \alpha_k T_k(\cos\omega_0 t) = \sum_{k=0}^{\infty} \alpha_k \cos k\omega_0 t$$

In fact, the Chebyshev polynomials are completely defined if the argument is bigger than 1 or smaller than -1. But in this case we don't have the orthogonality properties described in section 2.1.3. This is why the index of distortion is said to be bounded ($-1 \le I(t) \le 1$), and why the waveshaping function will be chosen so that the synthetic signal obtained for *I=1* corresponds to the richest part of the real signal.

Resynthesizing the source signal with the help of a waveshaping method is possible if one can determine the index function *I(t)* so that the dynamic evolution of the spectrum coïncides with the one of the real sound. One can calculate the spectrum of the generated sound with respect to the value of the index *I*. The knowledge of the decomposition of the function γ in the Chebyshev basis allows the calculation of its decomposition in terms of powers of the argument [Gradshteyn et al., 1980]:

$$T_n(x) = \frac{1}{2}\left[(x + i\sqrt{1-x^2})^n + (x - i\sqrt{1-x^2})^n\right] = \sum_{j=0}^{\infty} (-1)^j \binom{n}{2j} x^{n-2j}(1-x^2)^j$$

The synthetic sound is then given by:

$$s(t) = \gamma(I(t)\cos(\omega_0 t)) = \sum_k \beta_k I(t)^k (\cos(\omega_0 t))^k = \sum_k \alpha_k T_k(I(t)\cos(\omega_0 t))$$

The decomposition of $\cos(\theta)^n$ in term of *cos(nθ)* reads:

$$\cos(\theta)^n = \cos^{2n}(\theta) + \cos^{2n-1}(\theta),$$

where

$$\cos^{2n}(\theta) = \frac{1}{2^{2n}} \left\{ \sum_{k=0}^{n-1} 2 \binom{2n}{k} \cos 2(n-k)\theta + \binom{2n}{n} \right\}$$

$$\cos^{2n-1}(\theta) = \frac{1}{2^{2n-2}} \sum_{k=0}^{n-1} \binom{2n-1}{k} \cos(2n-2k-1)\theta$$

and it allows the computation of the generated spectrum for a given index.

A great disadvantage of the global synthesis techniques is that the representation of signals is not complete, meaning that one can not reconstruct any spectral evolution by changing the index. Nevertheless, the index can be estimated so that the reconstructed signal satisfies perceptive criteria. In the following sections I give some examples of the non linearities measured on the flute signal and describe how the index of distortion can be estimated using perceptive criteria. We shall see that in the case of the flute, a simple perceptive criterion like the « brightness » (centroïd of the spectrum) cannot give a good description of the timbre evolution, and that more accurate perceptive attributes such as the tristimulus have to be used.

6.3.1. Consideration of the non-linear behavior

The spectral evolution of a sound depends on the sound generating source. Non-linearities in the sound producing system cause the spectral components to evolve differenly. This means that when the energy of the excitation increases, the energy of some spectral components may considerably increase, while the energy of others may hardly change.

The spectral evolution of a flute sound as a function of the driving pressure is not the same in the three registers (corresponding to different musical octaves). Fletcher has studied and analyzed the harmonic development of sounds produced by four flute players [Fletcher, 1975]. He found that in the lowest octave the fundamental frequency has a lower level than both the second and third harmonics and may even be weaker than the fourth and fifth harmonics. The level of the fundamental is the same for soft and for loud playing while the relative levels of all higher harmonics are increased for loud playing. In the second octave the fundamental becomes dominant for both loud and soft playing. It still changes little with dynamic level and most of the change is represented by changes in the upper partials. In the third octave the fundamental is clearly dominant and all upper partials are more than 10dB below it in relative

level. The fundamental changes considerably with dynamic level, though still not as much as do the upper partials.

The non-linear mechanism that generates harmonic components of the pressure in flutes is not fully understood. Fabre B., Hirschberg A. and Vijnands P.J. [Fabre et al., 1996] have studied the amplitude of the acoustic pressure in an organ pipe assuming that the non-linearity which generates harmonics and limits the amplitude of oscillation is concentrated at the jet/labium interaction. By flow visualization they have observed two vortices shed at the labium, one during the jet/edge interaction and the second after the jet has passed the labium. They found that both vortices are responsible for important energy losses of the fundamental frequency component of the acoustical pressure in the pipe, and that the second vortex appears to be an efficient source for higher harmonic components of the pressure. Verge has found that the ratio W/h, where h is the height of the flue and W is the distance from the flue exit to the labium is of importance for the generation of harmonics [Verge, 1995].

To illustrate the effect of the non-linearities, the note G1 (whose fundamental frequency is 392Hz) and the note C2 (whose fundamental frequency is 523 Hz) played by a flutist were recorded for five different dynamic levels in an anechoïc room. The flutist was asked to linearly increase the dynamic level for the five notes (from a perceptive point of view). It is interesting to note that the total energy of the sound increases in a logarithmic way when the dynamic level of the notes is increased in a perceptively linear way. This is probably due to the logarithmic sensibility of the hearing system and to the feedback occuring between the player and the resulting sound [Miskiewicz et al., 1990]. *Figure 6.7* shows the evolution of the total energy of the five sounds.

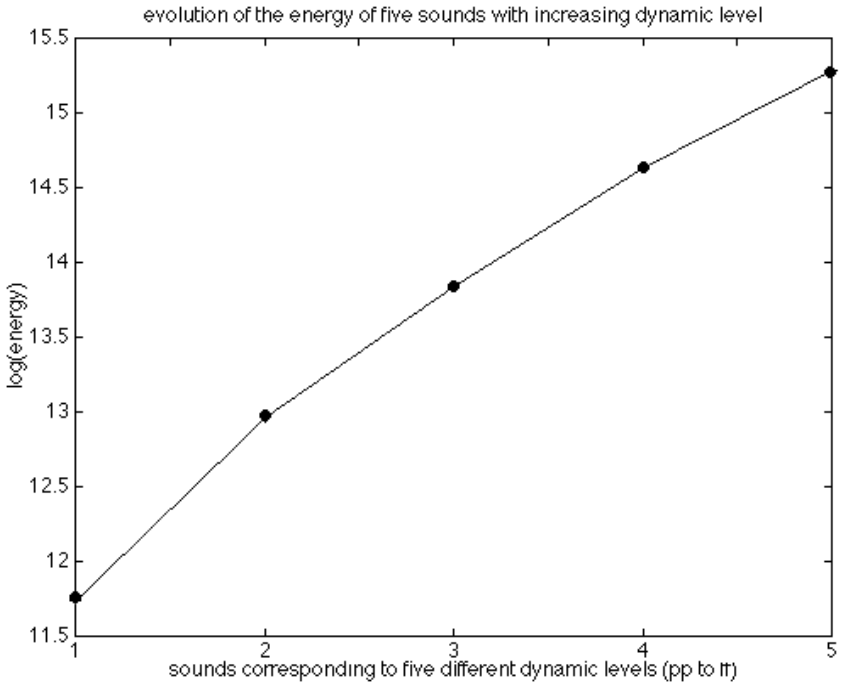


Figure 6.7: Energy of five different sounds with different dynamic levels. The sounds that are played have a fundamental frequency of 392Hz, corresponding to a G1.

The spectra of the deterministic part for three different dynamic levels of G1 are shown in *Figure 6.8* to *Figure 6.10.* For a sound played pianissimo, the fundamental component has the strongest level, and the third and fourth components have stronger levels than the second component. When the dynamic level is increased, the third component becomes the strongest one, its level being more than twice the level of the fundamental. The second harmonic is weaker than the third, the fundamental and the fifth harmonic for high dynamic levels, and the fourth harmonic that is important for weak dynamic levels is almost disappearing when the dynamic level increases. In general the relative level of the higher harmonics increases when the dynamic level increases (apart from the fourth harmonic which decreases between the mean and the strong dynamic level). Also the level of the fundamental frequency increases with the dynamic level, which is not in accordance with Fletcher's conclusions on the spectral evolution. This may be due to the different playing techniques used by the different flutists, since some flutists play with a high driving pressure even for weak dynamic levels, while others increase the driving pressure when increasing the dynamic level.

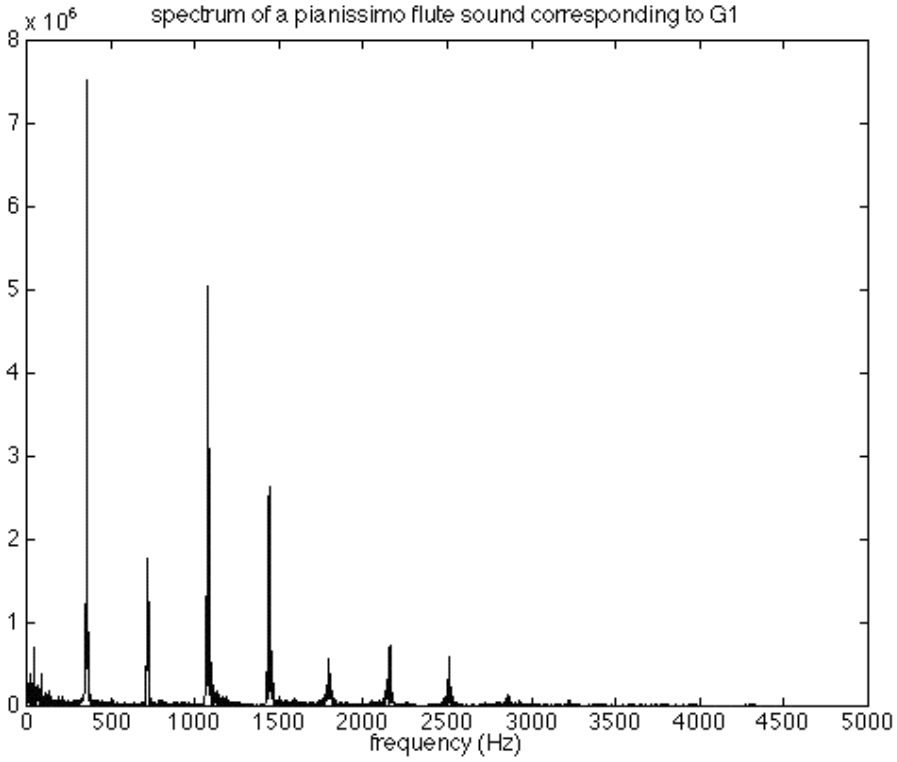


Figure 6.8: Spectrum of the flute sound G1 corresponding to the weakest dynamic level of the five notes.

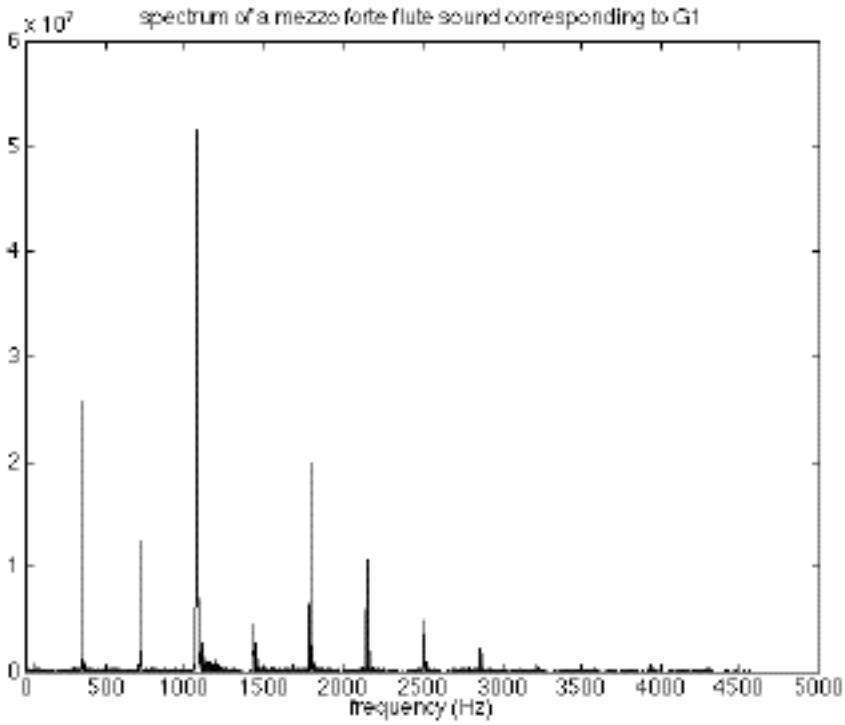


Figure 6.9: Spectrum of the flute sound G1 corresponding to a medium dynamic level of the five notes.

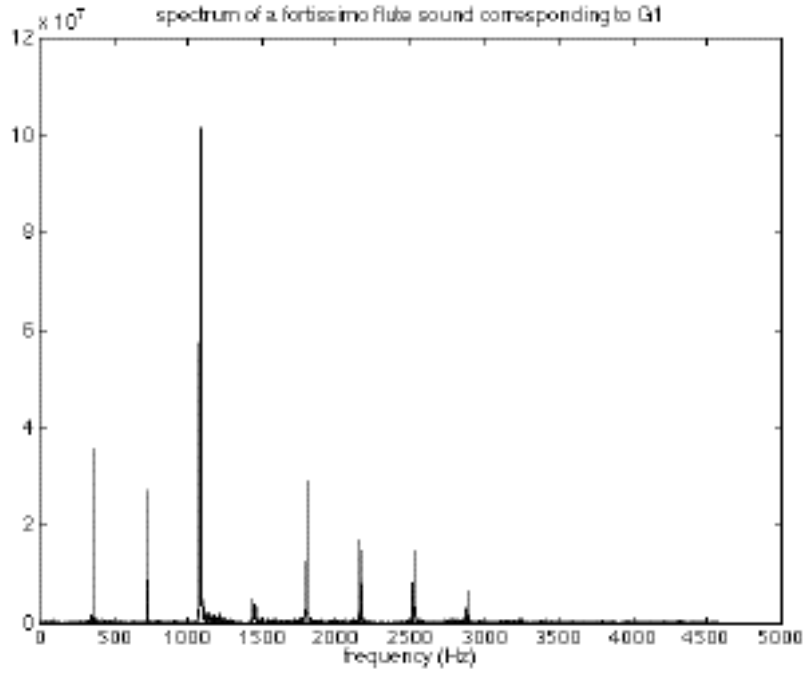


Figure 6.10: Spectrum of the flute sound G1 corresponding to the strongest dynamic level of the five notes.

The harmonic evolution as a function of energy for the first seven harmonics is shown in *Figure 6.11.*

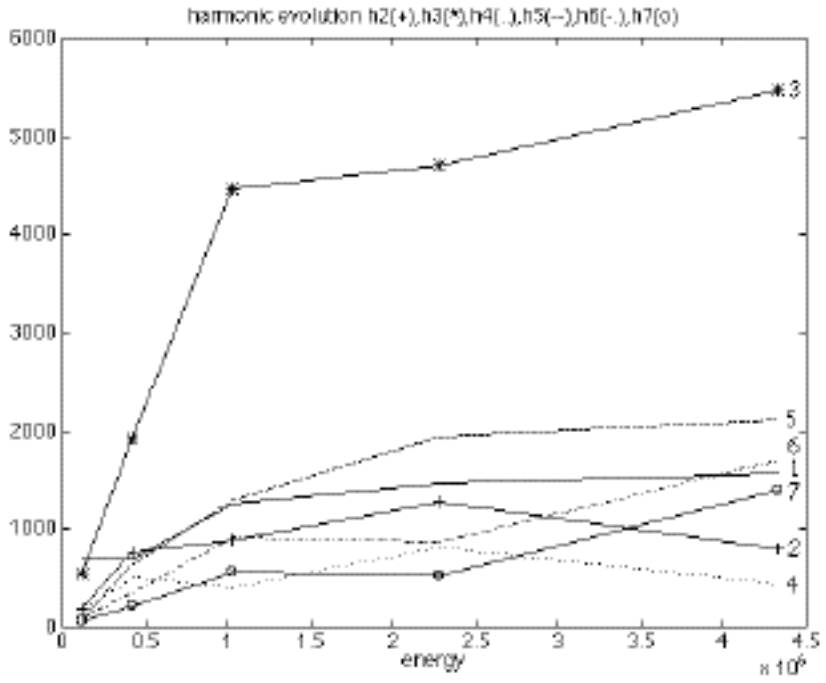


Figure 6.11: Harmonic development of a flute sound whose fundamental frequency is 392Hz (corresponding to G1) as a function of energy. The numbers at the end of the curves correspond to the harmonic rank of the component.

For the note played pianissimo (weak energy level), the fundamental is strongest, followed by the third(*) and the second(+) harmonics. As the energy increases the level of the third harmonic(*) rapidly increases, and for a mezzo forte note (medium driving pressure), the third harmonic is the strongest, followed by the fifth harmonic, the fundamental and the second harmonic. For a maximum driving pressure, the third harmonic is still the strongest, and both the fifth and the sixth harmonics are stronger than the fundamental in this case.

6.3.2. Estimation of the waveshaping index

By acting on the wave-shaping index *I(t),* a harmonic evolution of the synthesized signal can be simulated by waveshaping synthesis. As already mentioned, an arbitrary evolution of a spectrum cannot be synthesized with the help of a waveshaping method. Our purpose is to reconstruct a sound in such a way that it perceptively ressembles the original one. The problem then consists in approaching the spectral evolution of the real sound so that the resynthesis is perceptually similar to the original sound. In order to generate the spectral evolution by varying the waveshaping index, two criteria have been used: The spectral centroïd criterion and the tristimulus criterion. These criteria are described in the following sections, and a justification of the choice of the tristimulus criterion is given.

6.3.2.1. Spectral centroïd criterion

The spectral centroïd criterion is an important characteristic for the perception of the temporal behavior of a musical tone [Beauchamp, 1982]. It is related to the brightness of a sound and is given by the first order moment of the spectrum:

$$BR = \frac{\int_0^{\infty} \omega |h(\omega)| d\omega}{\int_0^{\infty} |h(\omega)| d\omega} \quad ,$$

where $h(\omega)$ represents the spectrum of the sound.

In *Figure 6.12* the spectral centroïd of the flute sound C2 (whose fundamental frequency is 523 Hz) is shown for five dynamic levels of playing. The spectral centroïd increases roughly linearly with the logarithm of the energy of the sound.

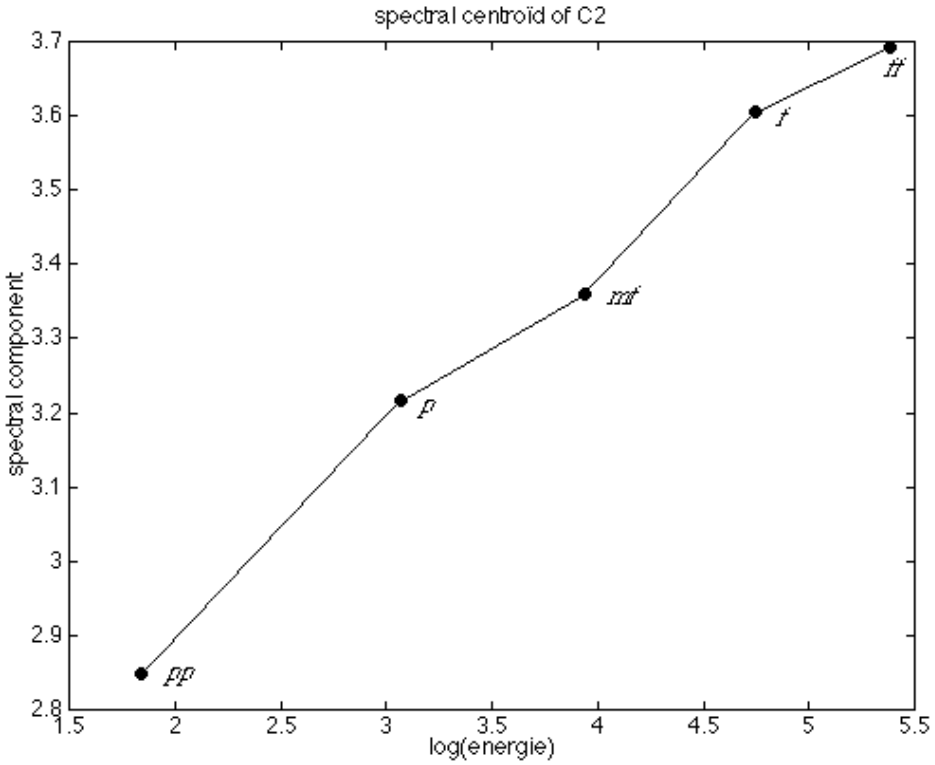


Figure 6.12: Spectral centroid of C2 (whose fundamental frequency is 523 Hz) as a function of the logarithm of the energy

6.3.2.2. Tristimulus criterion

A more accurate approach for representing musical timbre is the tristimulus method [Pollard et al., 1982]. This is a three-coordinate representation taking into account three dimensions that are supposed to give a satisfactory representation of the sound data [Grey, 1975], namely

1) the spectral energy distribution

2) the presence of low-amplitude, high frequency energy in the initial attack segment and

3) the synchronism in higher partials (whether the higher partials rise in level together at the beginning of the note or fall together at the end).

For that purpose, the loudness for three separate parts of the spectrum was chosen. The loudness is a perceptive measure which can be modeled by:

$$N = k\left[\sum_{m=1}^{M}\int_{t}^{t+T} q^2(t)dt\right]^{1/2},$$

where

N is the total loudness in sones,

k is a constant,

M is the number of critical bands,

T is an integration time (approximately 100ms for steady sounds, possibly as short as 10ms for transients),

$q(t)$ is the output of the auditory system taking into account auditive characteristics. It is given by

$$q(t) = \int_{-\infty}^{\infty} h(\tau)\, p(t-\tau)\, d\tau,$$

where

$h(\tau)$ is the impulse response of the auditory system and

$p(t)$ is the input sound pressure.

As one coordinate Pollard & Jansson decided to use the loudness of the fundamental. The two other coordinates were chosen by considering critical bands as well as the important effect that certain odd-numbered partials have on the timbre of the sound. These considerations led to a coordinate given by the loudness of the partials 2, 3 and 4 and a third coordinate corresponding to the loudness of the partials 5 to n. The loudness value of each group can be computed using Steven's formula:

$$N_i^n = 0.85N_{\max} + 0.15\sum_{i}^{n} N_i,$$

where

N_i^n is the required equivalent loudness, $N_{\max}$ is the loudest part of the group and $\sum_i^n N_i$ is the loudness of all the partials in the group. The total loudness N of the sound is now given by the sum of the three loudness groups, namely

$$N = N_1 + N_2^4 + N_5^n$$

By this method, the tristimulus can be represented in an acoustic tristimulus diagram, where

$$x = N_5^n / N$$

$$y = N_2^4 / N$$

$$z = N_1 / N$$

Since x+y+z=1, it is sufficient to use two of the coordinates to draw the tristimulus diagram as illustrated in *Figure 6.13*. When the values of x and y are small, the fundamental is strong, while a high y-value means strong mid-frequency partials and a high x-value means strong high-frequency partials. In this way a tristimulus diagram can represent the properties of both sustained sounds and time-varying transients.

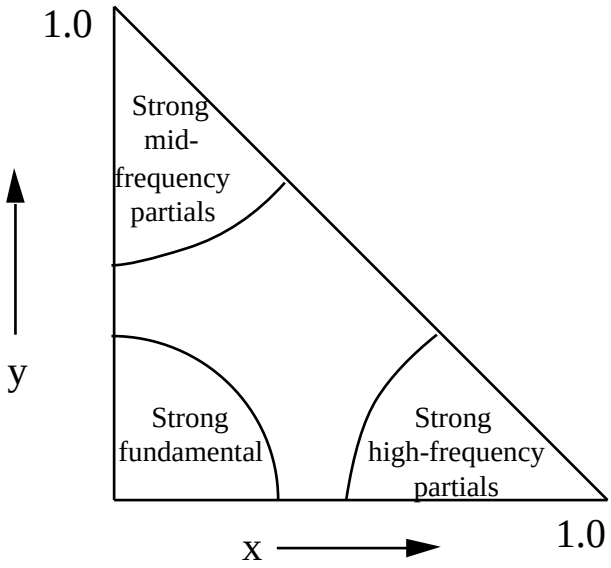


Figure 6.13:Acoustic tristimulus diagram

In *Figure 6.14* the tristimulus values corresponding to five different driving pressures of a real flute sound C2 (whose fundamental is 524Hz) are shown. The diagram shows that the flute sound contains strong mid-frequency partials. When the note is played pianissimo (pp) the fundamental is rather strong compared to the other partials, while it is weakened compared to the upper partials as the driving pressure increases towards fortissimo (ff).

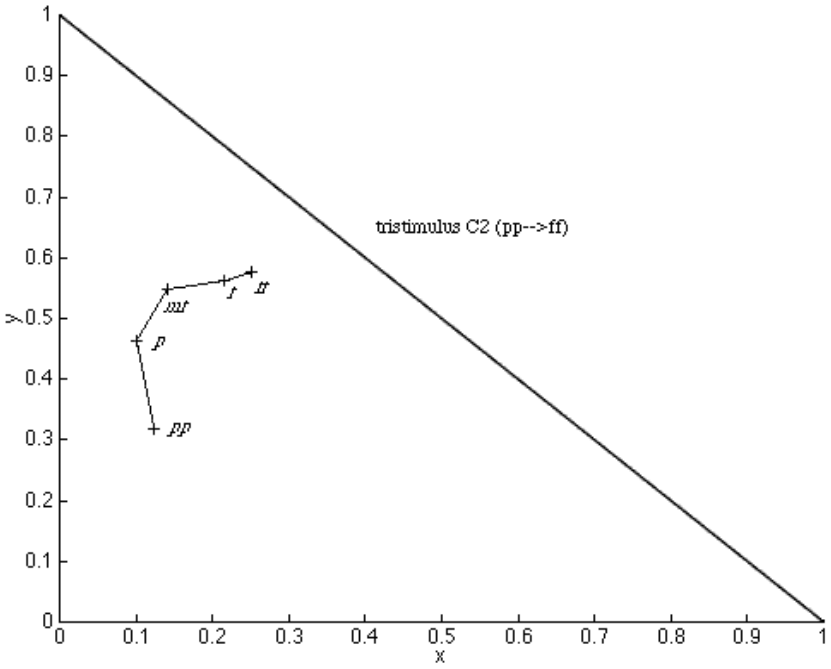


Figure 6.14: Tristimulus diagram for five flute sounds with different dynamic levels (pp to ff) whose fundamental frequency is 524Hz (C2).

6.3.3. Application to the flute

In this section, we apply the techniques described above to model the source signal corresponding to flute sounds.

The first step consists in constructing a non-linear function so that the « richest » spectrum obtained by playing the instrument corresponds to a wave-shaping index *I=1.* This spectrum is obtained for a fortissimo (ff) playing. In this example the values of a flute sound whose fundamental frequency is 524 Hz (C2) have been used. We here considered the 12 first components since the level of the higher components makes them inaudible, especially after filtering by the resonator. As seen in section 2.1.3, the amplitudes of these components coincide with the coefficients α_k of the non-linear function γ in the Chebyschev's polynomial basis. In this case, one obtains:

$$\gamma(x) = \sum_{k=0}^{k=12} \alpha_k T_k(x)$$

with

$\alpha_0 = 0$ $\alpha_1 = 17.9$ $\alpha_2 = 54.0$ $\alpha_3 = 22.9$
$\alpha_4 = 10.6$ $\alpha_5 = 22.2$ $\alpha_6 = 6.30$ $\alpha_7 = 2.18$
$\alpha_8 = 4.58$ $\alpha_9 = 1.80$ $\alpha_{10} = 1.43$ $\alpha_{11} = 1.89$
$\alpha_{12} = 1.45$

The function γ is represented in *Figure 6.15.*

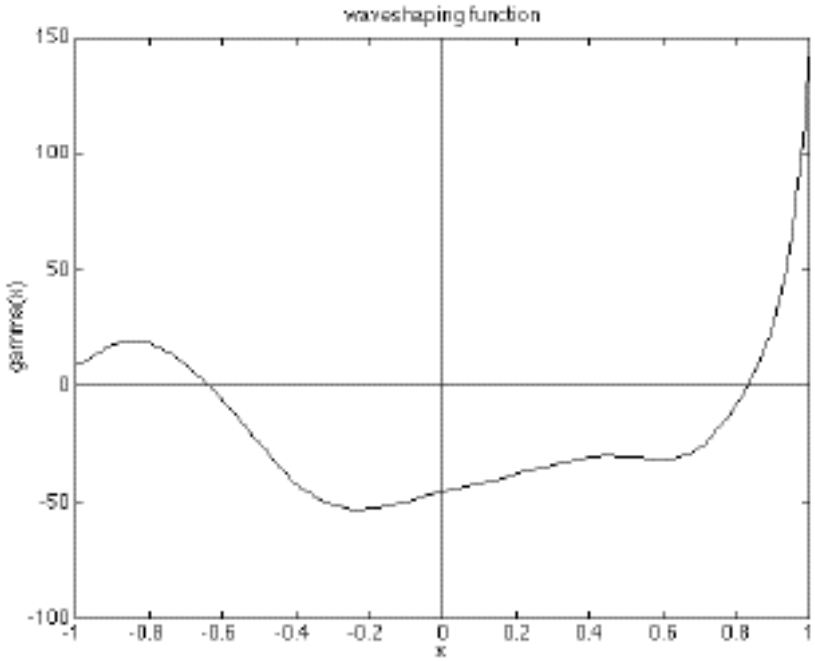


Figure 6.15: Waveshaping function corresponding to a fortissimo played flute sound whose fundamental frequency is 524Hz.

One note that the value of the function γ is different from zero when the argument is zero. This means that if the input of the non-linear system is zero, the output will be non-zero. This particularity is not a problem when the digital representation of the numbers allows a great precision (the DC bias will not be heard), which is the case with most floating point processors. Nevertheless, real-time implementation often necessitates a fast processor using a less precise representation of numbers (typically 16 bit integer). In such cases, the deviation of the signal due to a DC bias can cause saturations. One can avoid this difficulty by forcing the function γ to be zero when the argument is zero. This can be done by modifying the signs of the coefficients and the value of α_0. Actually,

$$\gamma(0) = \sum_{k=0}^{K} \alpha_k T_k(0) = \sum_{k=0}^{K} \alpha_k \cos\left(k.\arccos(0)\right) = \sum_{k=0}^{K} \alpha_k \cos\left(k(2k+1)\frac{\pi}{2}\right)$$

Only even coefficients have an influence on $\gamma(0)$, since the value of $T_k(0)$ is zero when k is odd, and $(-1)^{k/2}$ when k is even. One can then minimize the value of $\gamma(0)$ by modifying the signs of the even k's,

and then use the coefficient α_0 to cancel $\gamma(0)$. Altering the signs of the coefficients which corresponds to altering the phase of the sound does not modify the sound since it is periodic. Further on, the mean value of the signal should also be minimized to minimize the DC bias when the index of distortion is non- zero. Since this value is related to the integral of the non-linear function, this corresponds to minimizing this integral on the subset *(-I,I)* corresponding to the current index. To minimize this integral, one can make the function fluctuate as much as possible. This can be done by acting on the signs of the coefficients of the non-linear function. This time the signs of the odd coefficients should be modified, since the signs of the even coefficients were modified to cancel the DC bias. A criterion for finding the signs of the coefficients leading to a minimum DC bias is to minimize $\gamma(\pm 1)$. *Figure 6.16* represents the same non-linear function as above, obtained this way.

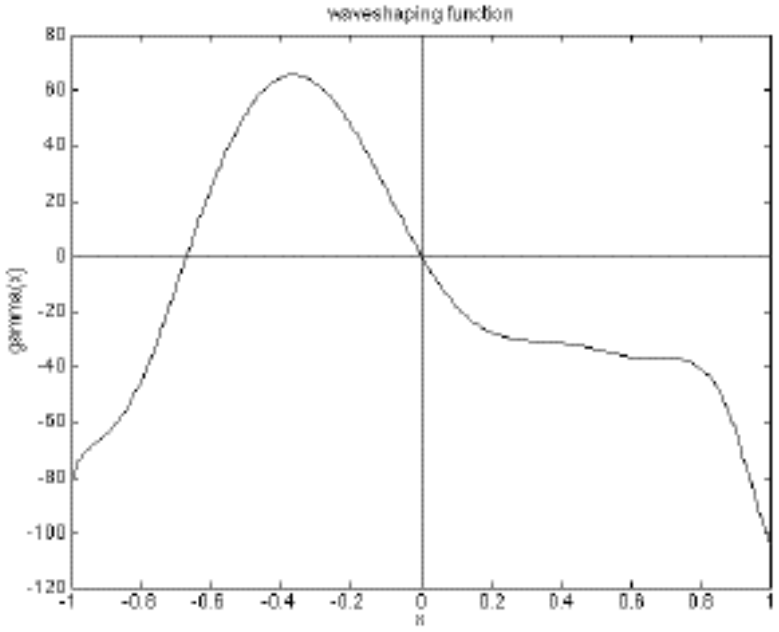


Figure 6.16:Waveshaping function with $\gamma(0)=0$ *and* $\gamma(\pm 1)$ *minimized.*

The second step in modeling the deterministic part of the source signal of a flute consists in estimating the wave-shaping index and to link it to a measurable value such as the driving pressure. In the first register there is an important exchange of energy between the first spectral components, while the other components increase rather monotonically with the energy of the driving pressure. This causes a problem when using the spectral centroïd criterion, since the brightness changes very little with increasing driving pressure, corresponding to small changes in the waveshaping index. This can be seen in *Figure 6.17* where the centroïd of the sound generated by waveshaping synthesis using the function γ defined above is plotted versus the value of the index.

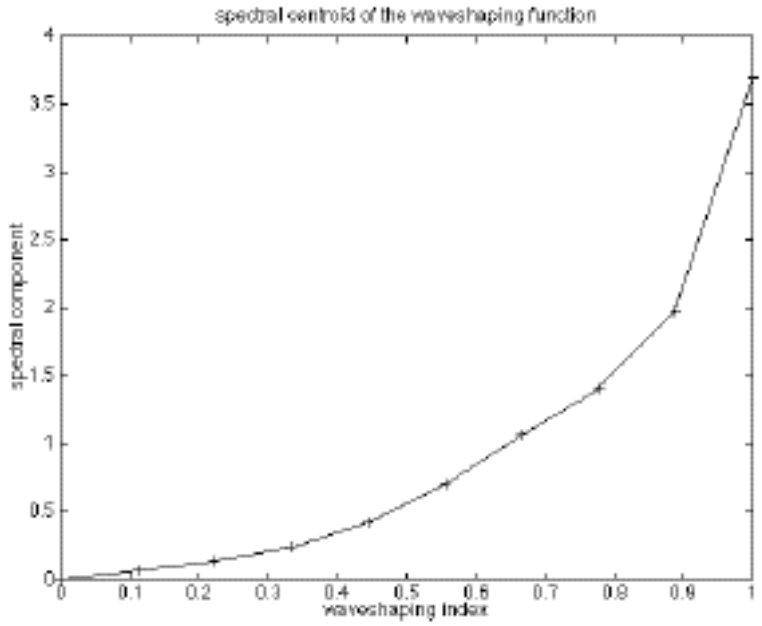


Figure 6.17: Spectral centroïd of the waveshaping function when the waveshaping for values of the waveshaping index from 0 to 1.

The spectral centroïd for five different dynamic levels (pp to ff) of C2 varies from 2.8 to 3.7 as shown in *Figure 6.12.* In the waveshaping case, a similar variation range can be obtained for an index varying from 0.97 to 1 (*Figure 6.18*).

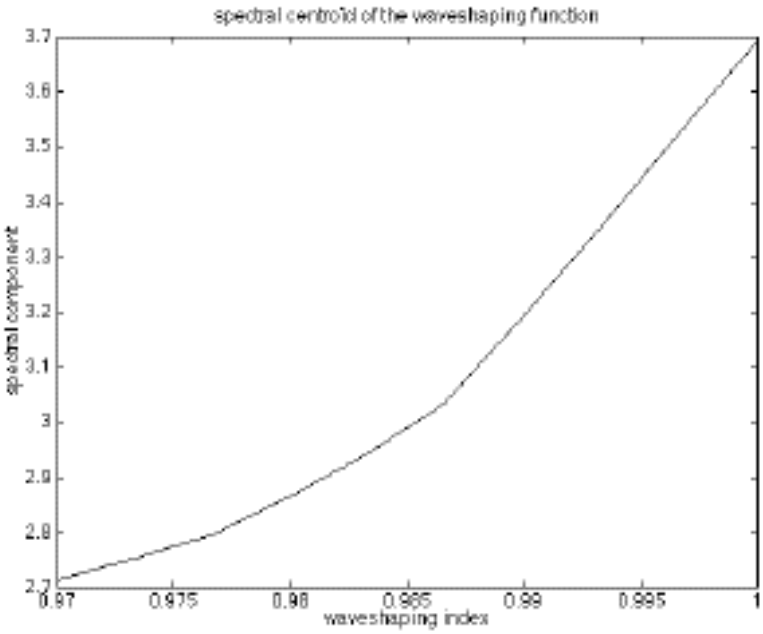


Figure 6.18: Spectral centroïd of the waveshaping function for I=0.97 to I=1.

Unfortunately, when the waveshaping index is *I=0.97,* the spectrum of the synthetic signal does not correspond to the spectrum of the real signal played piano (small energy in the driving pressure).This is shown in *Figure 6.19* and *Figure 6.20.*

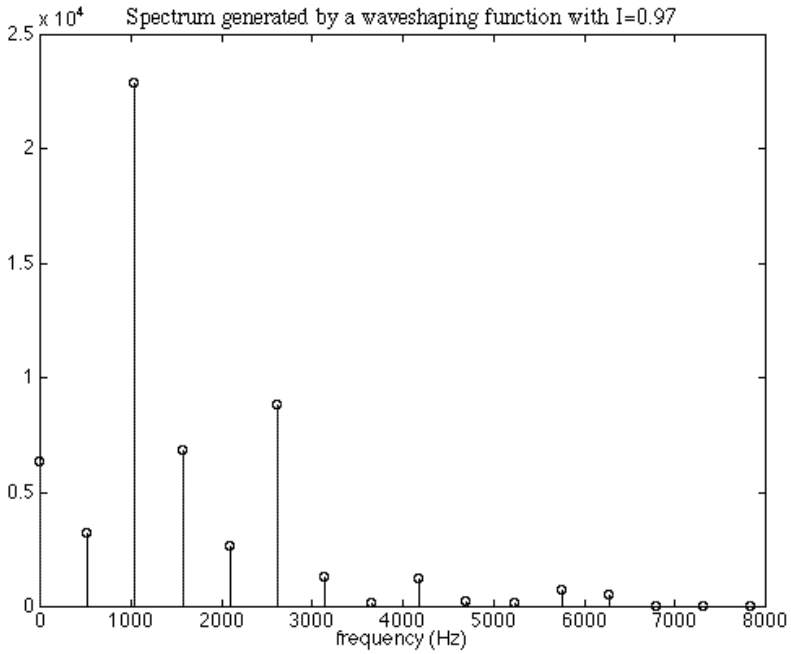


Figure 6.19: Spectrum generated by the waveshaping function for I=0.97

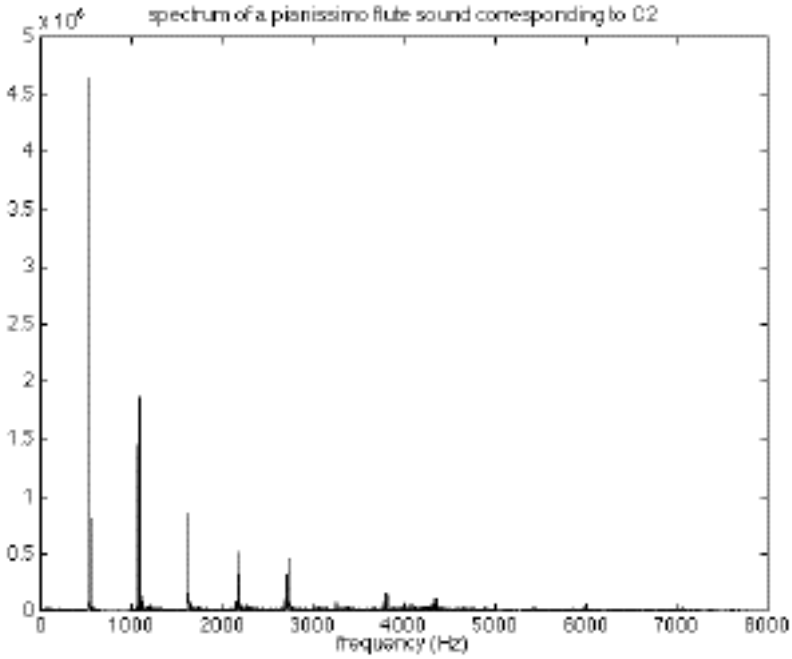


Figure 6.20: Spectrum of a pianissimo played flute sound whose fundamental frequency is 523Hz (corresponding to C2)

This means that the spectral centroïd criterion is not convenient for flute sounds. In fact, this criterion is suitable for sounds whose spectral components globally increase and not for sounds whose spectrum dramatically changes during the play.

Another criterion should therefore be used to find the waveshaping index of a flute as a function of the driving pressure. For a flute sound the most important exchange of energy appears between the first to the fifth or sixth components. More importance should therefore be given to these components than to the higher ones. The tristimulus criterion therefore turns out to be well adapted to a flute sound, since it divides the spectrum into three groups: one where the evolution of the fundamental is considered, one where the evolution of the second, third and fourth components is considered, and one where the rest of the components are considered. *Figure 6.21* represents the tristimulus diagram corresponding to a sound generated by waveshaping synthesis and using the function γ defined above. The index evolves from 0 to 1 by steps of 0.1.

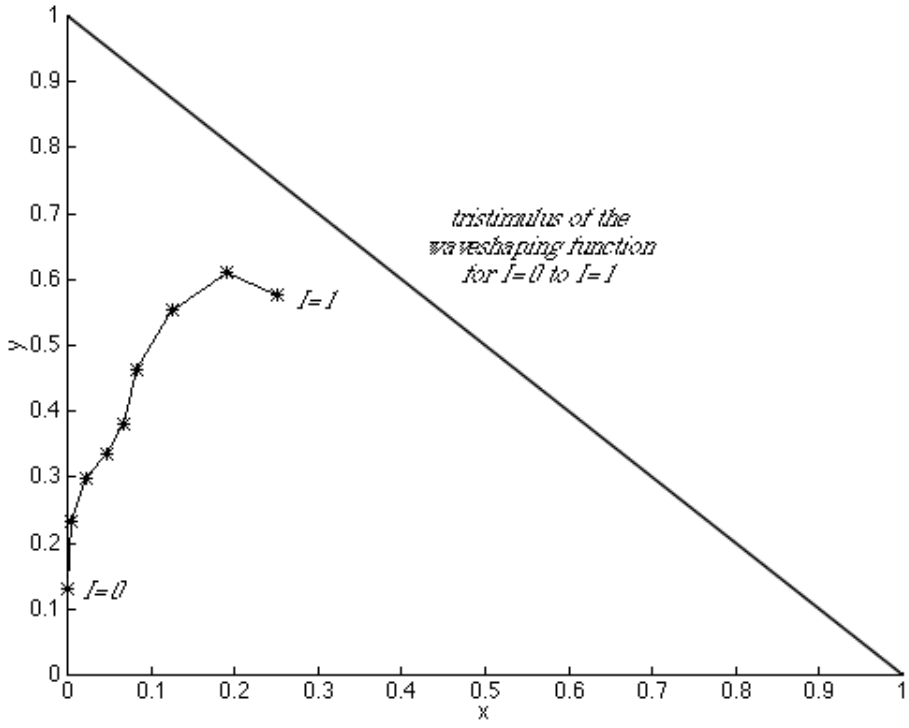


Figure 6.21: Tristimulus diagram of sounds generated by a waveshaping function with index values from 0 to 1.

To fit the tristimulus diagram of the real and of the synthetic sounds, the waveshaping index should vary from 0.5 to 1 (*Figure 6.22*). In contrast to the situation obtained with the centroïd criterion, this variation range gives a spectral evolution of the synthesized sounds close to the spectral evolution of real sounds with different dynamic levels.

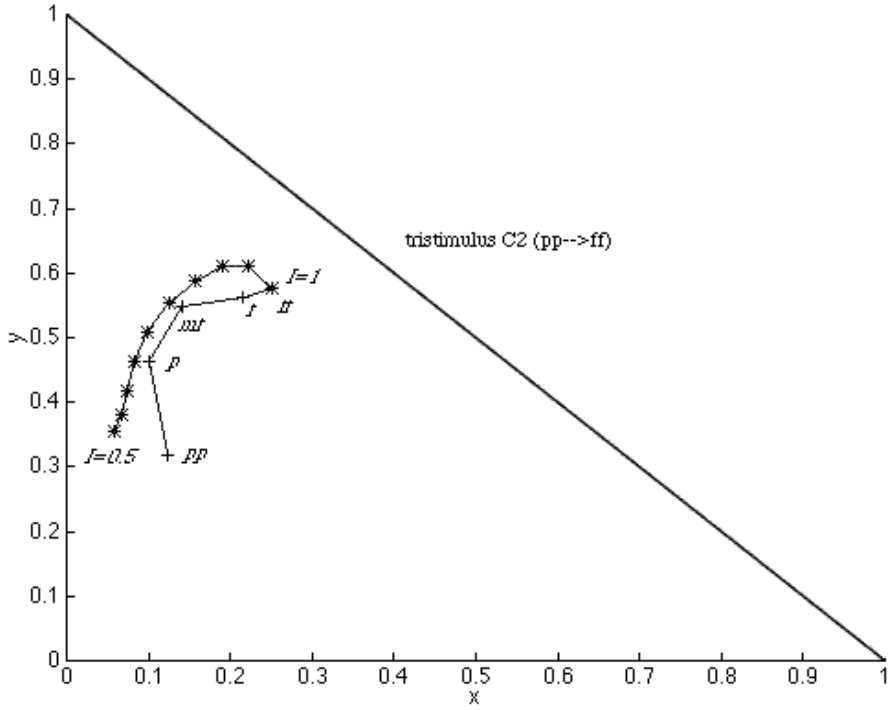


Figure 6.22: Tristimulus diagram of sounds generated by a waveshaping function with index values from 0.5 to 1 (*) and of five flute sounds (whose fundamental frequency is 523Hz) with different dynamic levels (pp to ff).

The modeling of the deterministic part of the source signal of a flute sound can thus be obtained using a waveshaping technique. The index of the waveshaping function should vary from *I=0.5* to *I=1* according to the logarithm of the driving pressure, *Figure 6.23.*

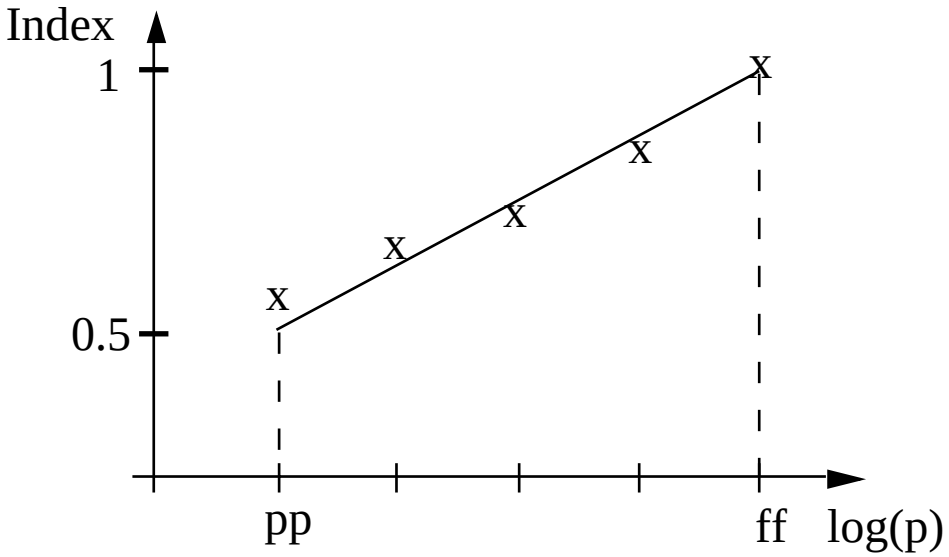


Figure 6.23: Relation between the waveshaping index and the driving pressure.

In order to equalize the changes in amplitude induced by the variations of the distorsion index, the output signal must be adjusted by a post correction (*Figure 6.23*).

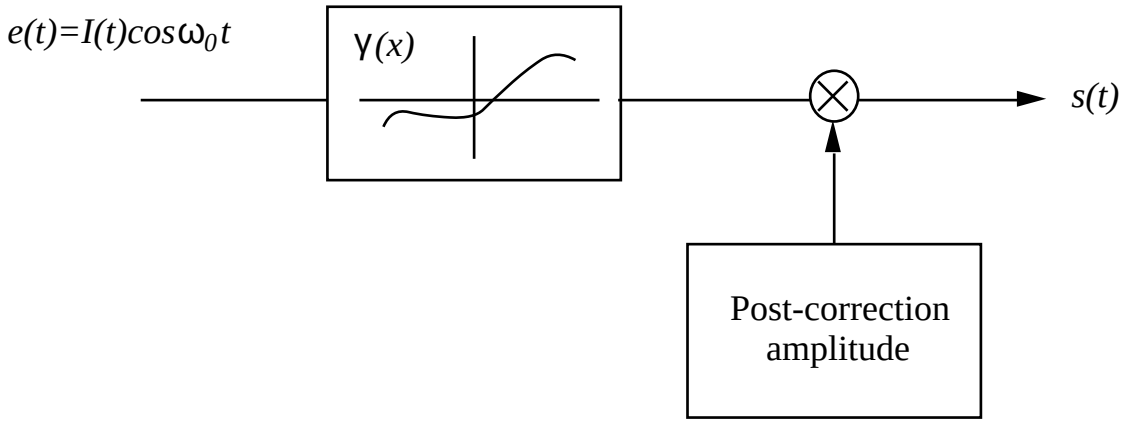


Figure 6.24: Equalization of amplitude changes induced by the variations of the distortion index.

The amplitude of the output signal generated using the function $\gamma(x)$ illustrated in *Figure 6.16* is shown below.

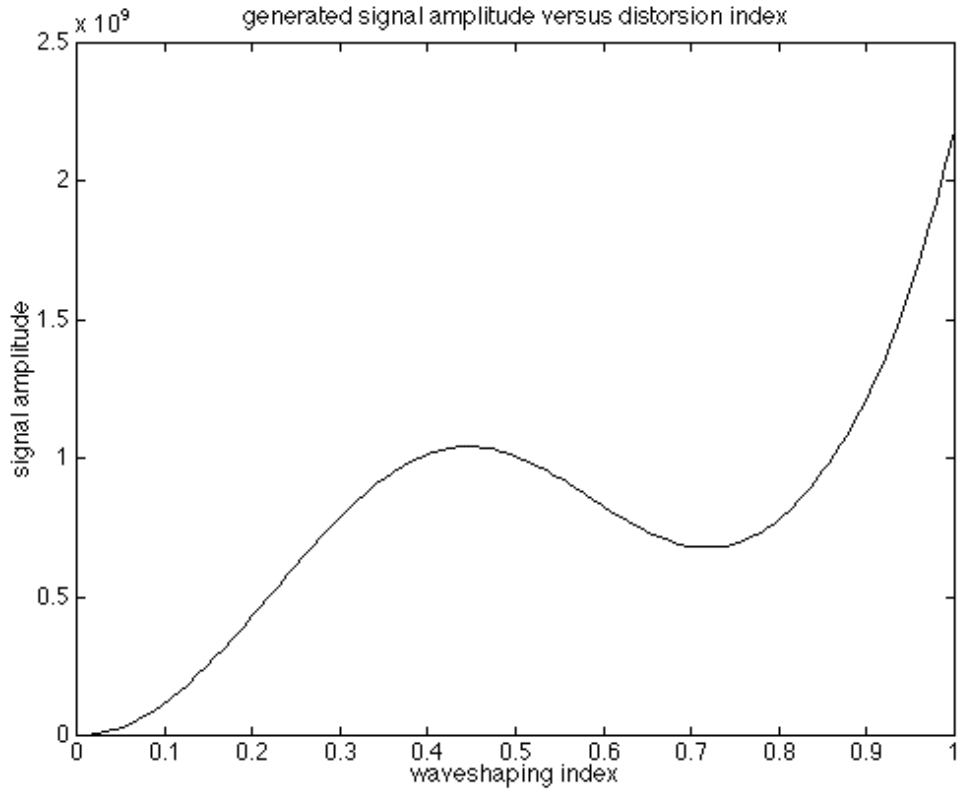


Figure 6.25: Amplitude of the generated signal.

The post correction amplitude corresponds to the inverse if this function.

6.4. Modeling of the stochastic part

In this chapter the stochastic part of the source signal will be characterized in order to be resynthesized so that is satisfies perceptive criteria. Stationary and ergodic processes are here considered, since such processes generally correspond to steady state sounds of musical instruments. As an illustration, *Figure 6.26* represents the spectrogram of a flute noise obtained using the LMS algorithm, as described in section 6.2.2. One can see that the spectrogram is roughly invariant with respect to time as soon as a steady state is established.

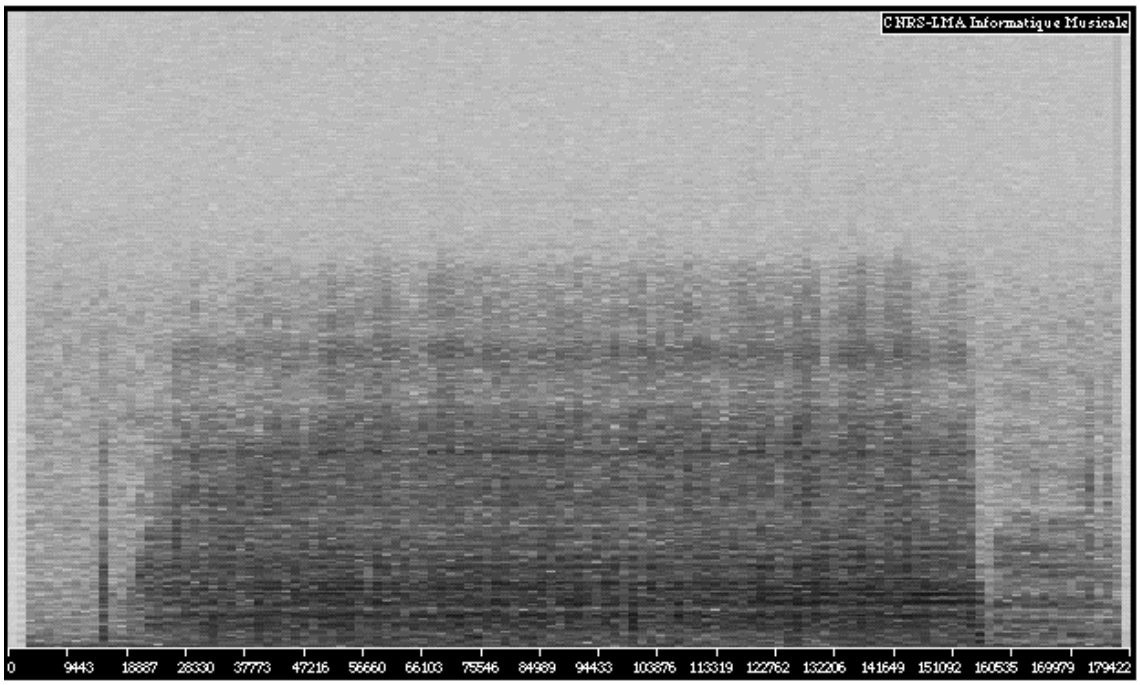


Figure 6.26: Spectrogram of a flute noise

The assumption of stationarity and ergodicity means that the noise can be described by its power spectral density and its probability density function. From a perceptive point of view, the « coloring » of the noise is mainly related to its Power Spectral Density (PSD). The probability density function can also be of importance. Differences in these functions can make the noise sound differently. The difference is easy to notice when comparing, for example, a noise with a uniform and a Gaussian probability density. However, the difference is very subtle and sometimes hard to notice when the probability density functions have more similar shapes. Since the probability density function may have a perceptive influence on the noise, I shall show how such a characteristic can be taken into account in the synthesis process.

6.4.1. The probability density function

The probability density function $f_B(x)$ is a positive valued function such that:

$$\int_{-\infty}^{\infty} f_B(x)dx = 1$$

The cumulative distribution function *F(x)* defined by:

$$F(\xi) = \int_{-\infty}^{\xi} f_B(x)dx$$

corresponds to the probability for the value of the random signal *B* to be less than or equal to some value ξ [Schwartz, 1970].

The probability density function $f_B(x)$ is related to the histogram of the values *x* from the noisy process *B*. Actually, by considering a random signal with n samples the values of which are separated by Δx, and by calling n_x the number of samples having the value x, the probability density function $f_B(x)$ is given by:

$$f_B(x) = \lim_{\substack{\Delta x \to 0 \\ n \to \infty}} \frac{n_x / \Delta x}{n}$$

As an example, *Figure 6.27* shows the histogram of a random Gaussian process containing *n=2000* values with Δx=16.

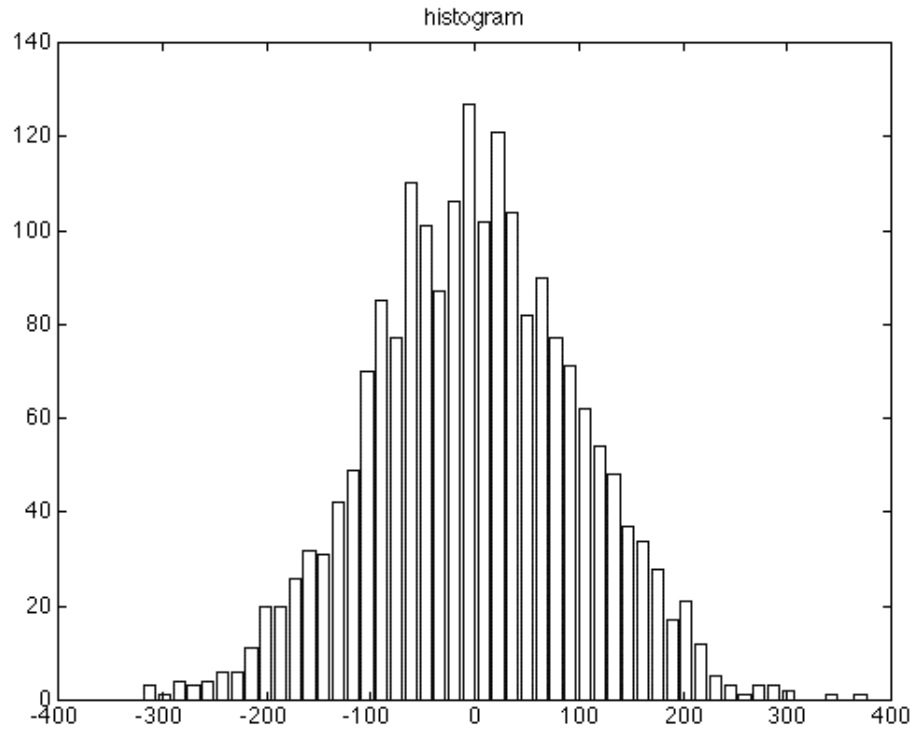


Figure 6.27: Histogram of a random gaussian process.

To generate a random process with a given probability density function $f_B(x)$, one can start with a process U having a uniform probability density function on the interval [0,1]. Actually, one can show [Haykin, 1983] that the process $B = F^{-1}(U)$ has the probability density function $f_B(x)$. This corresponds to a waveshaping algorithm, where the entrance is given by the uniform process *U*, and the non-linear function by the inverse of the cumulative distribution function *F(x)*.

6.4.2. The power spectral density

The Fourier transform of a random process is generally not defined since the process is either not integrable or square integrable. The Power Spectral Density (PSD) plays a role equivalent to the spectrum of a deterministic signal.

The PSD of a random process is given by the Fourier transform of its autocorrelation function [Schwartz, 1970]. The autocorrelation function of a stationary process *x(t)* is defined as:

$$R_x(\tau) = E[x(t+\tau)x(t)] = \lim_{T\to\infty} \frac{1}{T}\int_0^T x(t+\tau)x(t)dt$$

and the power spectral density is:

$$S_x(\omega) = \{R_x(\tau)\} = \int_{-\infty}^{\infty} R_x(\tau)e^{-i\omega\tau}d\tau$$

From a synthesis point of view, one can generate a random process having a given power spectral density by using a linear filter. Actually, let's respectively denote $S_x(\omega)$ and $S_y(\omega)$ the power spectral densities of the input *x(t)* and of the output *y(t)* of a linear filter, and $H(\omega)$ the transfert function of the linear filter. We then obtain:

$$S_y(\omega) = |H(\omega)|^2 S_x(\omega)$$

This means that one can generate the « colored » process by filtering a white noise with a filter whose transfert function satisfies $|H(\omega)|^2 = S_y(\omega)$.

6.4.3. Application to the flute noise

In this section, we describe the characteristics of the non deterministic part of the source signal corresponding to a flute sound. This non deterministic part has been extracted using the LMS algorithm described above.

As discussed in section 6.4.1, the probability density function $f_B(x)$ is related to the histogram of the values *x* taken by the noisy process *B*. It can be easily estimated as soon as the random process can be separated from the deterministic one, which is generally the case for source signals. The histogram of the stochastic part of a flute sound is shown in *Figure 6.28*. It gives an estimation of the « shape » of the probability density function.

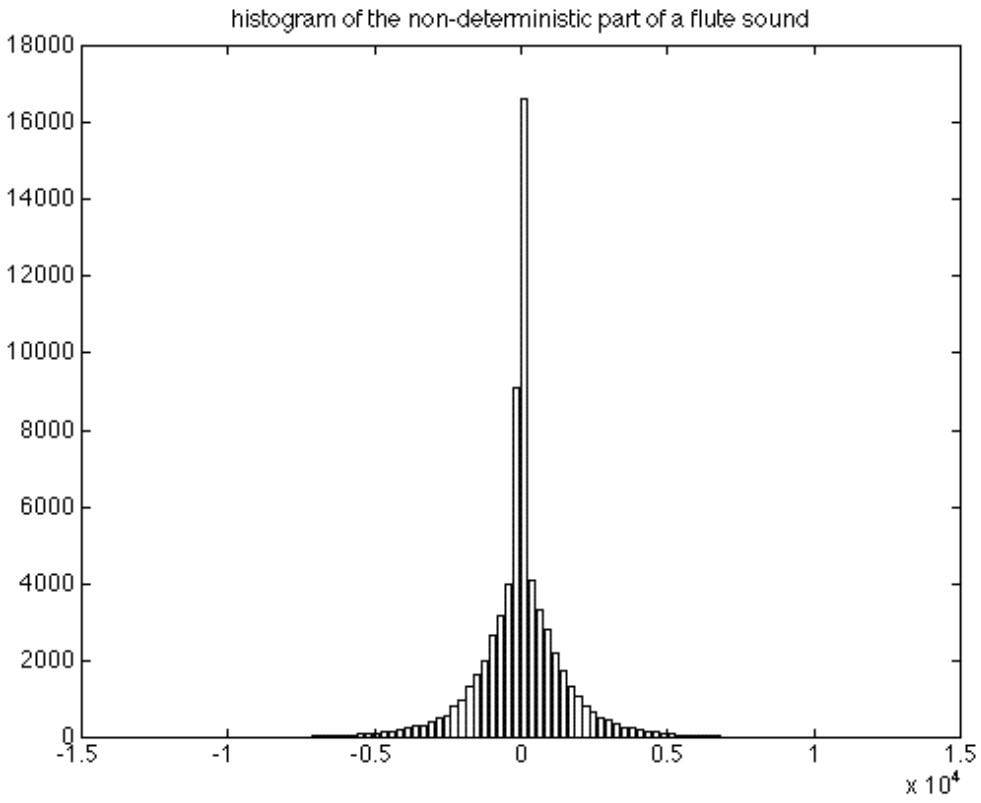


Figure 6.28: Histogram of the stochastic part of a source signal of a flute sound.

We see that the histogram is symmetric and that it follows an exponential law. This means that the noise to be generated when modeling a flute sound should have the following probability density function:

$$f_e(x) = \frac{\lambda}{2} e^{-\lambda|x|}$$

In the case of the flute, λ has been measured to be equal to 1.25×10^{-3}. This estimation has been done by fitting a straight line on the logarithmic representation of the estimated probability density function represented in *Figure 6.29.*

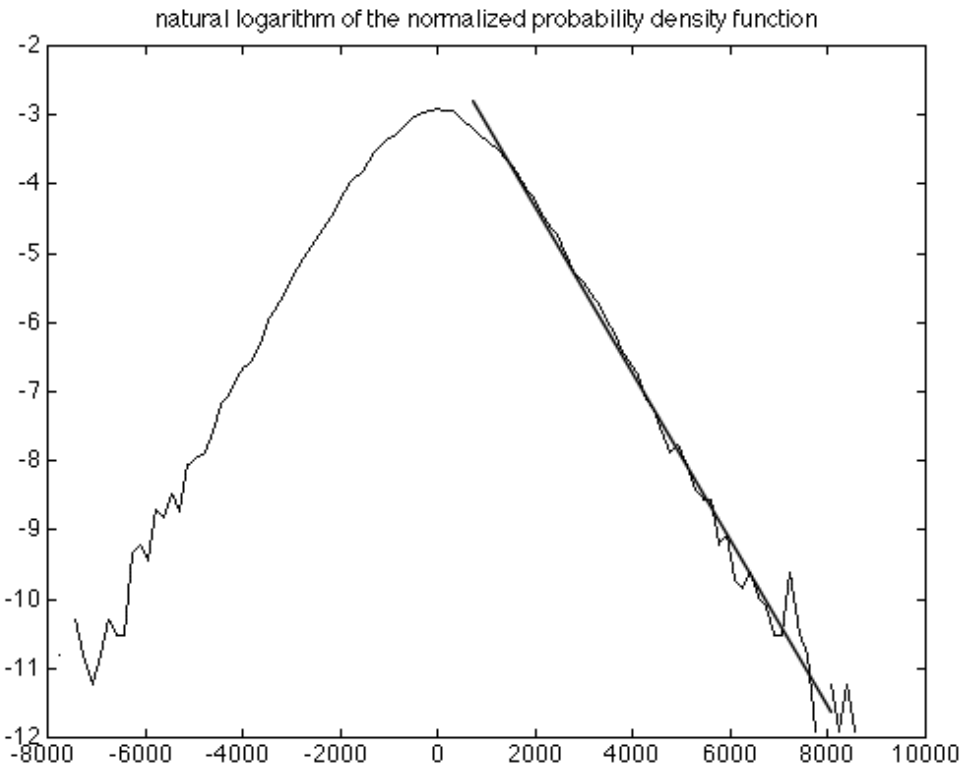


Figure 6.29: Natural logarithm of the estimated probability density function corresponding to the stochastic part of the source signal of a flute. The superposed straight line allows the estimation of λ.

To construct a process with a given probability density function from a process $u(t)$ with a uniform probability density function on the interval [0,1], we must calculate $F^{-1}(u)$, where $F(x)$ is the "repartition" function given by:

$$F(x) = \int_{-\infty}^{x} f_x(x)\,dx$$

it can be calculated as follows:

$$F(u) = \int_{-\infty}^{u} \frac{\lambda}{2} e^{-\lambda|x|} dx = \begin{cases} \displaystyle\int_{-\infty}^{u} \frac{\lambda}{2} e^{\lambda x} dx & \text{for} \quad u \le 0 \\ \displaystyle\int_{-\infty}^{0} \frac{\lambda}{2} e^{\lambda x} dx + \int_{0}^{u} \frac{\lambda}{2} e^{-\lambda x} dx & \text{for} \quad u \ge 0 \end{cases}$$

The result of the integration is

$$F(u) = \frac{1}{2} e^{\lambda u} \qquad \text{for} \quad u \leq 0$$

and

$$F(u) = 1 - \frac{1}{2} e^{-\lambda u} \qquad \text{for} \quad u \geq 0$$

This means that

$$\log(F(u)) = -\log(2) + \lambda u \qquad \text{for} \quad u \geq 0$$

and

$$\log(F(u) - 1) = -\log(2) + \lambda u \qquad \text{for} \quad u \leq 0$$

The inverse repartition function obtained writing $u = F^{-1}(y)$ is given by :

$$F^{-1}(y) = \begin{cases} \dfrac{\log(2y)}{\lambda} & \text{for} \quad \dfrac{1}{2} \leq y \leq 1 \\ -\dfrac{\log(2(y-1))}{\lambda} & \text{for} \quad 0 \leq y \leq \dfrac{1}{2} \end{cases}$$

Figure 6.30 shows the inverse repartition function corresponding to the non deterministic part of the source signal of a flute. It allows the generation of a random process with an exponential probability density function from a process having a uniform probability density function.

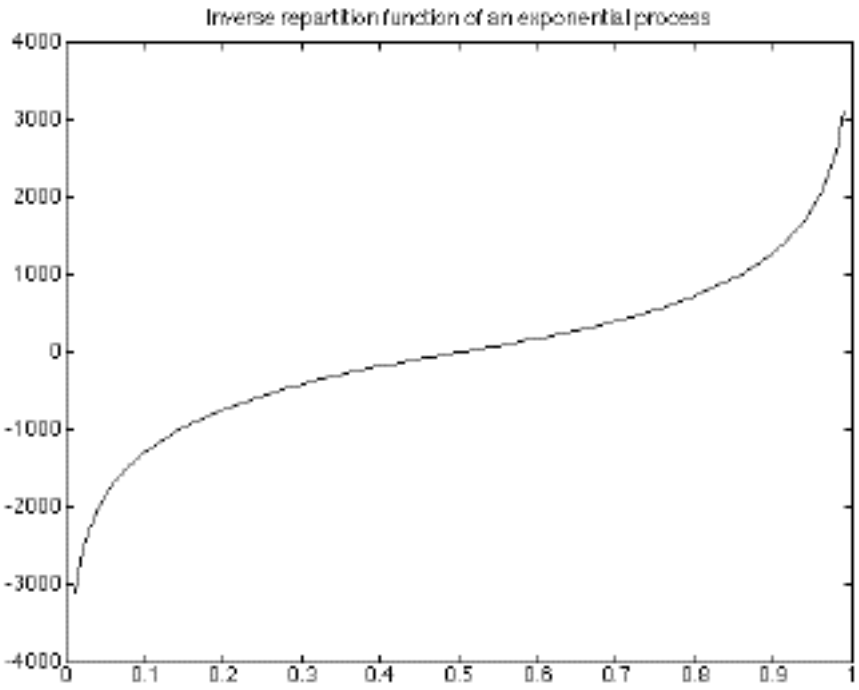


Figure 6.30: The inverse repartition function corresponding to the non deterministic part of the source signal of a flute.

At this stage, one may notice that the probability density function of a process is not invariant by linear filtering; except in the case of the normal law (Gaussian probability density function). Nevertheless, if the correlations induced by the filtering are weak, then the probability density function is almost unmodified. This is the case for the flute signal where the non deterministic part of the source corresponds to a slight low-pass filtering of a white noise.

As an example *Figure 6.31* shows the power spectral density of the stochastic part of the source of a flute sound.

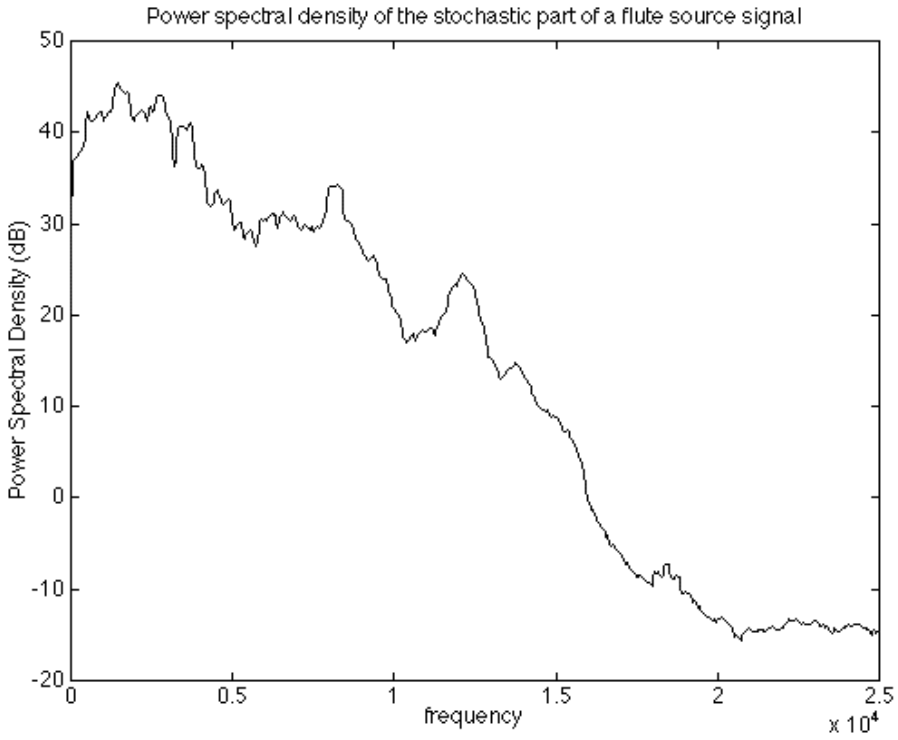


Figure 6.31: Power spectral density of the stochastic part of the source of a flute noise.

The stochastic part of the source signal can now be modeled by linear filtering of a white noise. This model together with the model of the deterministic part of the source signal gives a general model of the source signal based on signal modeling.

By combining the source model with the physical model simulating the behavior of the waves during propagation in the medium, very general sound models can be constructed. In the next section we shall see how this general model can be applied to the flute case.

6.5. A hybrid flute model

In *Figure 6.32* a hybrid flute model which makes use of the methods described in this document is shown. As mentioned in the introduction of this chapter, the physics behind the sound generating system of a flute is not fully understood. By flow visualisation, vortex phenomena have been observed during jet/edge interaction in recorder-like instruments [Verge, 1995]. Therefore a hybrid flute model is proposed, where the excitation system is modeled by signal synthesis models described in section 2.1, and the resonator is modeled by a physical synthesis model of the waveguide type described in section 2.2. The model can be piloted by the driving pressure and by the frequency. For this purpose a classical flute has been used. Electromagnetic sensors have been connected to each fingerhole in order to detect the note played. The cork of the flute situated close to the embouchure has been replaced by a microphone detecting pressure variations from the player. The finger position together with the pressure gives the « effective » length of the resonator and thus the delay D used in the waveguide model simulating the resonator. The frequency corresponding to this length feeds the oscillator generating the input signal of the non-linear waveshaping function. The pressure measured at the embouchure is related to the waveshaping index and gives the amplitude of the oscillator. It is obtained from the logarithm of the envelope (as described in section 6.3.3) and is in the flute case bounded between the $I=0.5$ and $I=1$. By band-pass filtering of the pressure, the vibrato is extracted and added to the frequency input of the oscillator.

By adding the vibrato to the source of the system, the resonance peaks of the source will fluctuate. This means that when the source is injected into the resonator, the resonance peaks of the source and of the resonator will not be tuned all the time. Thus the amplitude at the output of the system will fluctuate and be stronger when the two systems are tuned than when they are not tuned. In this way the amplitude fluctuations (tremolo) follow the frequency fluctuations (vibrato) like on a traditional flute.

The pressure is also related to the level of the noise generator. The output of the non-linear function in addition to the generated noise gives the whole source to be injected into the resonant system. The noise generated when closing the finger holes is also part of the input of the resonant system. The speed at which the key pad is closed can be calculated as explained in section 7.3.2. The resonant system, as described in chapter 5, takes into account dissipation and dispersion phenomena related to the wave propagation in the tube.

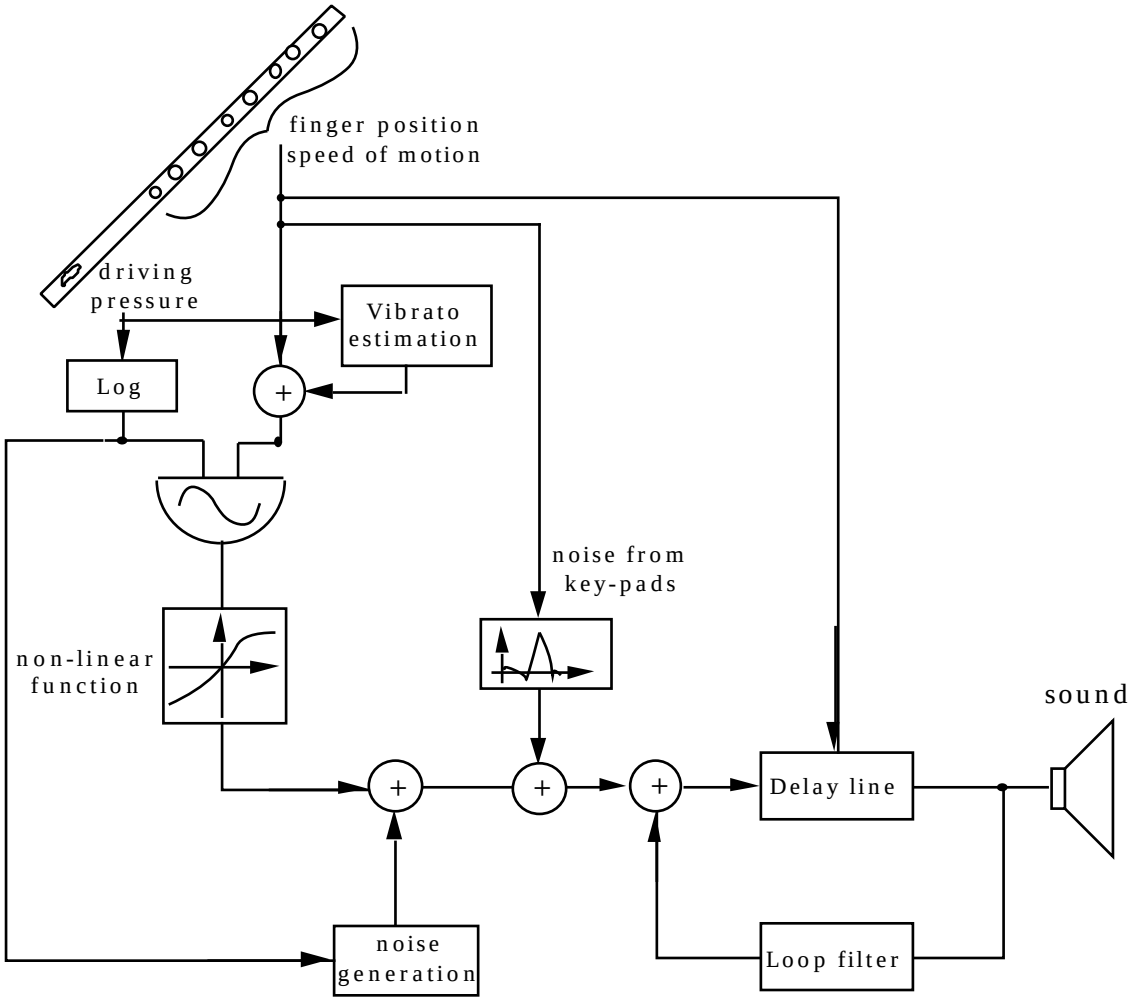


Figure 6.32: The hybrid flute model

A description of the real-time processor used for the realization of the hybrid flute model is given in chapter 7.

7. Real-time control of a sound model

This chapter addresses the control of sound models using appropriate interfaces. As seen in the previous chapters, a general sound model can be designed by splitting the sound generating system into a source and a resonator. Hybrid models can then be constructed using both physical and signal models. The digital implementation of such techniques necessitates the programming of units such as oscillators, waveshaping units, noise generators, recursive filters, delay lines etc.

The auditive bandwidth which determines the sampling frequency imposes the speed of the processor if it is to run in real-time. The Shannon (or Nyquist) theorem states that the sample rate should be at least twice the maximum frequency of the signal [Haykin, 1983]. Assuming that the auditive bandwidth is limited to 20KHz, the sampling rate for audio signals should be at least 40 KHz. Consequently, a real time implementation of a sound model necessitates a processor able to run the whole synthesis program within at most 1/40000 s. Even though computers unceasingly are getting more powerful, these constraints generally impose the use of specific architectures for digital signal processing. I shall describe how the GENESIS processor has been used for real time implementation of the hybrid flute model.

An important aspect of real-time implementation is the control of the model. In this case, digital interfaces must be designed in accordance with the aim of the sound simulator. The construction of these interfaces is mainly related to two aspects. First of all, one must take into account the problem of the correspondence between the synthesis parameters and the perception of the produced sounds. Actually these parameters should be controlled with the help of the interface. Then the physical interface should be designed as to allow a natural control of the sound. Two examples will be given, namely the radio baton of Max Mathews and a flute interface which was designed to control the hybrid flute model. I shall finally discuss the performance possibilities offered by such a realization and how a digital flute like instrument can be used for musical purposes.

7.1. The real-time processor GENESIS

The real-time processor GENESIS (GENErateur des SIgnaux Sonores), developed by the Steria Digilog company, is based on the same structure as the processor SYTER (SYstème TEmps-Réel). This processor was developed in the 80's by Jean-Francois Allouis [Allouis, 1982] at INA-GRM as a synthesis and sound processing tool. This machine was an extension of a host computer (DEC PDP 11/73), well adapted to the real-time execution of synthesis algorithms and to signal processing. The processor GENESIS has the same architecture as SYTER, but is, thanks to new technologies, faster with a bigger memory and a larger word length (32 bits). In addition, GENESIS is not related to a specific

host computer, but contains its own operating system and a hard disk. This makes the machine « independent », being able to receive program loading instructions from a serial port or an ethernet link.

A host station is however necessary when using development tools such as edition, compilation, updating of programs and wavetable calculations, and when there is to be an interaction with GENESIS via MIDI (Musical Instrument Digital Interface). A description of MIDI will be given in section 7.3.2. The host station could be arbitrarily chosen, but so far only implementations on UNIX workstations have been realized. A Silicon Graphics Indy computer has here been chosen because of its audio interface quality and of its computing power offering additional possibilities of synthesis and of real-time processing. The ease of using a MIDI interface is an additional advantage.

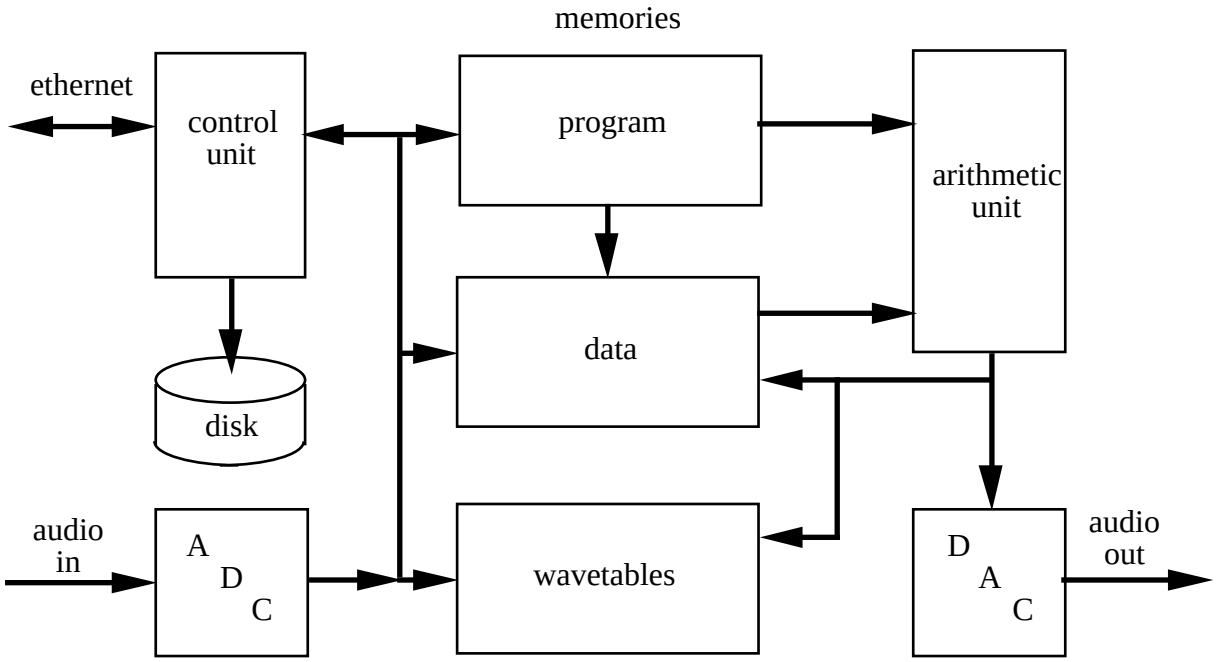


Figure 7.1: Synoptic scheme of GENESIS

The GENESIS processor is a complicated machine. The architecture is of the Harvard type in which the programs and the data are situated in different parts of the memory. The data are distributed in two physically different parts of the memory, one with a weak capacity (2kbytes), but with a rapid access containing the registers of the machine, the other with a larger capacity (1Mbytes), but with a slower access, intended for the wavetables.

The efficiency of the machine for classical synthesis algorithms is essentially due to the simplicity of the arithmetic unit which in most cases executes the same basic operation (*Figure 7.2*). Two operands A and B are read from the register memory and transmitted either to the multiplier or to the adder. The result of the operation can either be sent to the digital to the analog converter, stored in the data memory for further processing, or being used as an address to read or write to a wavetable.

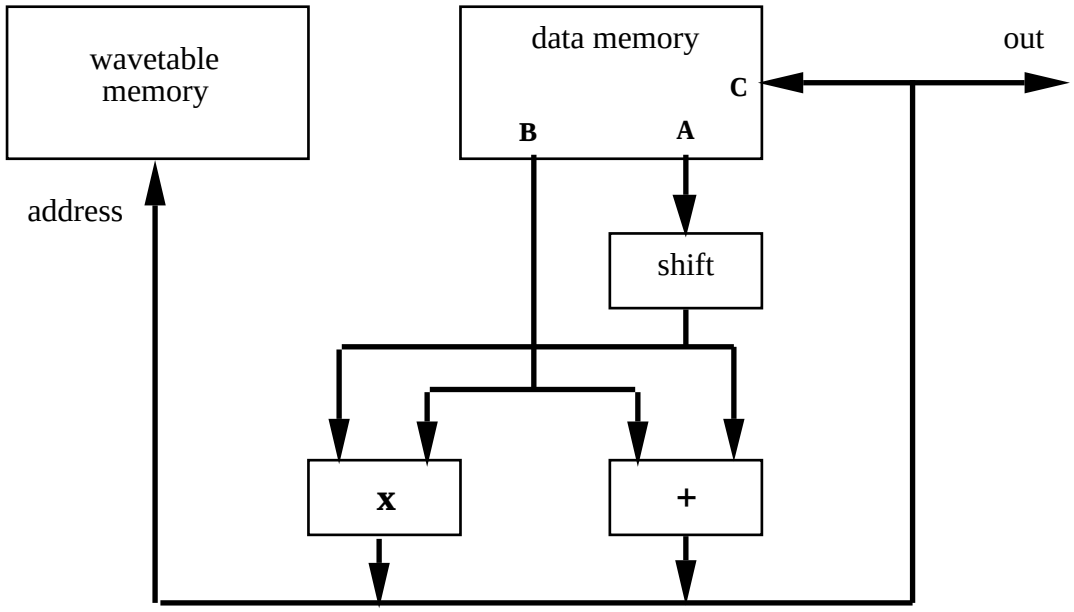


Figure 7.2: simplified organization of the GENESIS arithmetic unit

Each elementary stage of this operation necessitates an instruction cycle. The complete read-compute-store process therefore needs several cycles, but the pipeline structure makes it possible to execute all the elementary operations in parallel. While executing an operation, the processor stores the result of the previous operation and fetches the following operation.

Since the execution of an elementary operation is distributed over several cycles, the programmer has to be careful not to use the result of a previous operation before it has been stored in the memory. Thus, the programmer may have to insert wait cycles into the program. This rule represents the most severe constraint for the programming of the machine. The operator A can be shifted before addition or multiplication, and can for instance be used to generate a white noise (pseudo-random signal).

The adder can work in several modes; saturated, modulo or test mode. In the saturated mode, if the result goes beyond the capacity of the adder (32 bits), it is forced to be given the highest value on 32 bits. In modulo mode the calculus is done with one additional carry bit, but only the 32 least significant bits make the result. These two modes are especially interesting for computing wavetable addresses. In a saturated mode, the array is swept in a cyclic way: after having reached the end, one systematically gets to the beginning. The test mode makes it possible to compare the two operators. It is the only conditional instruction of the machine.

The multiplier is able to accumulate the results of successive multiplications, which means that algorithms such as convolutions are straightforward. Addition, multiplication and multiplication-accumulation are thus the basic operations. If other functions are necessary, they should be calculated off-line and stored in the wavetables. In each sampling period the complete program is executed. The higher the sampling frequency, the lower the capacity of calculus resources. These should therefore in priority be reserved for the synthesis.

7.2. Real-time implementation of the model

In this section a description of the implementation of the sound model proposed in chapter 6 is given. The source signal synthesized by a signal model and the resonator simulated by a physical model are respectively described.

7.2.1. Real-time implementation of the source model

The source signal of the flute model comprises four different contributions to be calculated and to be mixed before being processed by the resonator. These contributions correspond to the deterministic part, the stochastic part, the impulsive part corresponding to the noise generated by the key pads and a part provided by an external source. The implementation of the source model is illustrated in *Figure 7.3.*

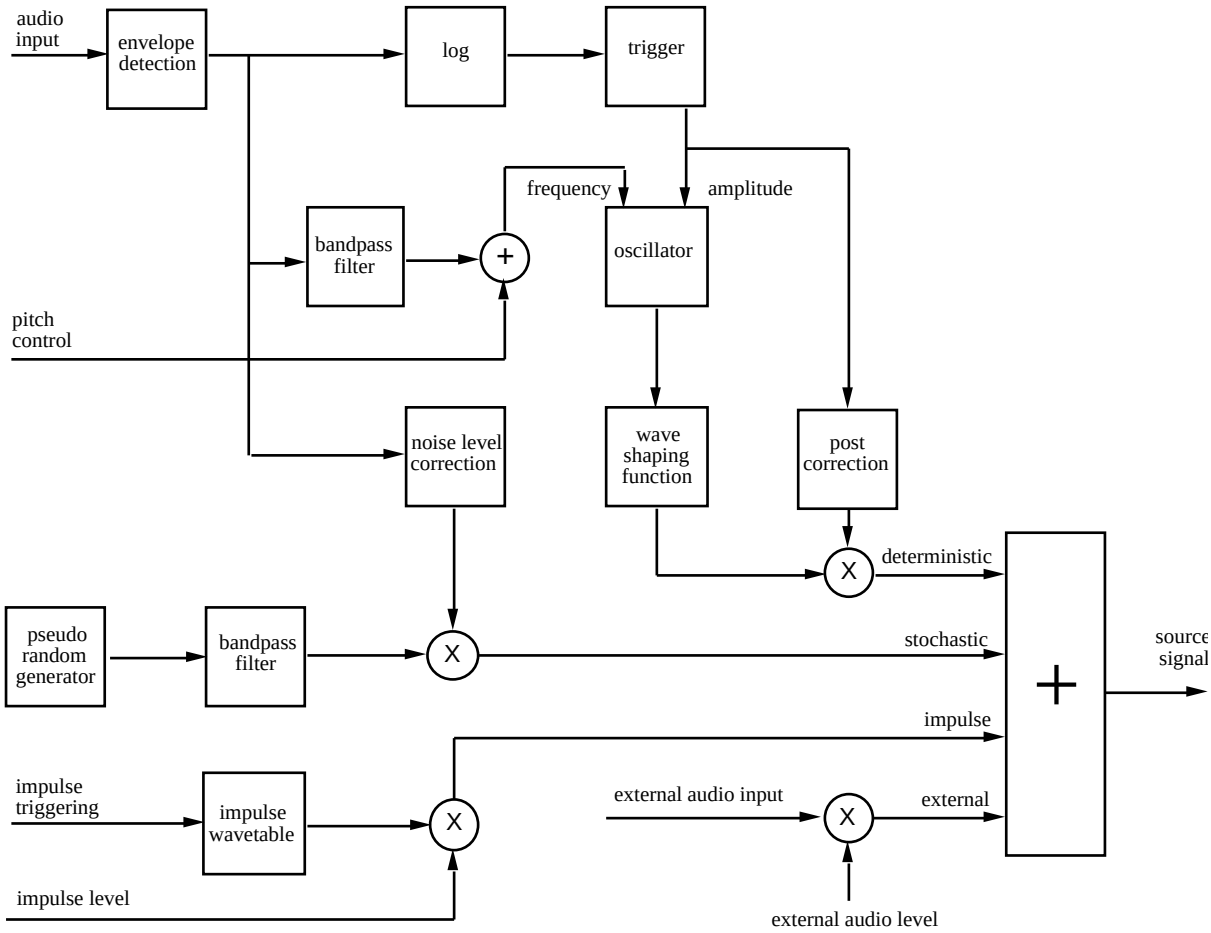


Figure 7.3: Scheme of the implementation of the source model.

In order to calculate the deterministic part of the signal, the signal measured by the microphone at the embouchure is processed by a peak detector the aim of which is to extract the « envelope » of the signal. This envelope is the main external actuator since the dynamic behavior of the sound will be driven

by it. It will be used to construct both the frequency fluctuations (vibrato) and the spectral evolution through the distortion index. The vibrato is correlated to the fluctuations of the envelope as seen in section 4.2. These fluctuations are extracted from the envelope using a bandpass filter (5-15Hz) implemented by two cascaded Butterworth IIR filters. These variations are then added to the pitch command which is given by the finger position in order to drive the frequency input of the oscillator. The amplitude input of the oscillator corresponding to the distortion index is obtained from the logarithm of the envelope according to section 6.3.3. This data is then processed by a trigger module to ensure that it is bounded between two values: $I_{\min}$ and $I_{\max}$. In the flute case these boundary values correspond to $I_{\min}=0.5$ and $I_{\max}=1.0$ as described in section 6.3.3. At the output of the waveshaping module the amplitude of the signal is adjusted using a post correction table equalizing the changes in amplitude induced by the non-linearity.

The generation of the stochastic part of the source signal is implemented using a gaussian pseudo random noise source band-pass filtered through a cascade of two Butterworth IIR filters (500-2500 Hz). To obtain the correct relationship between the envelope of the input signal and the level of the stochastic part of the source signal we make use of a correction table.

To make the model more realistic, an impulsive contribution has been added. This contribution corresponds to the noise produced by the key pad and is triggered when a key is pressed on the instrument. The impulse response of the « entrance filter » described in section 5.2.1 is stored in a wavetable that is read when a key is pressed. The level of the impulse noise is controlled by the closing speed of the key pad provided by the flute interface that will be described later.

Finally, an external input with its own level control has been added, offering a possibility of driving the resonator with an external signal. This will for example allow the various noises from the player's mouth to be taken into account.

7.2.2. Real-time implementation of the resonator model

The resonator constitutes a delay line and a loop filter. The fundamental frequency of the « instrument » can be adjusted varying the delay length. The implementation of the resonator model is illustrated in *Figure 7.4.*

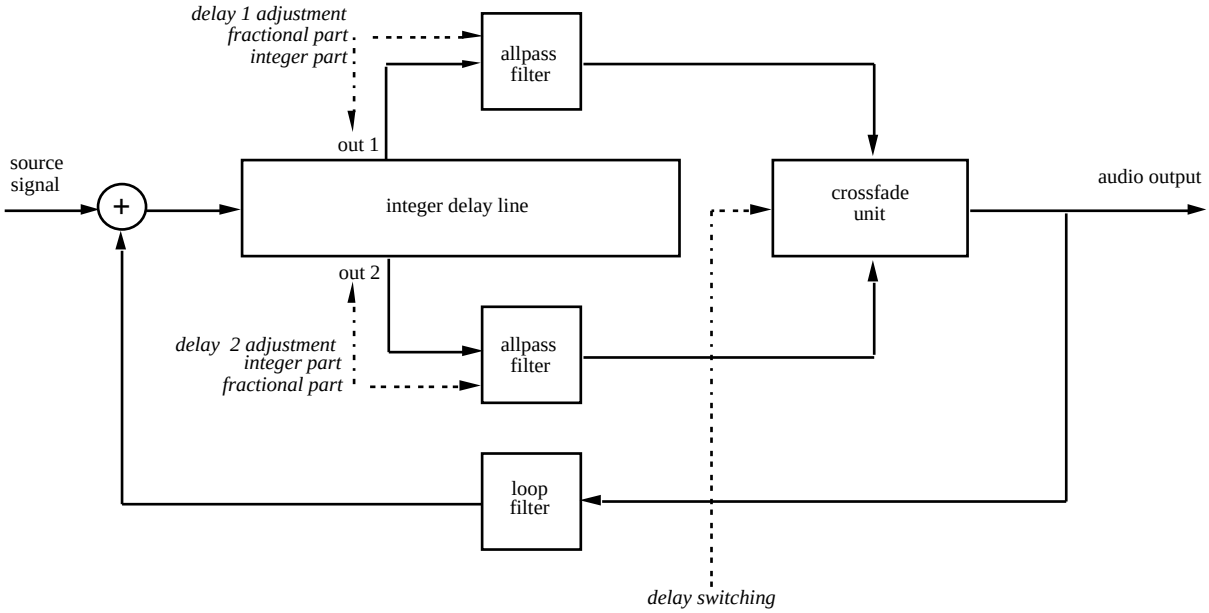


Figure 7.4: Scheme of the implementation of the resonator model

For musical purposes the resonator may be tuned so that the delay is not an integer number of the sampling period. This can be taken into account by splitting the delay line into an integer delay line realized with a table and a fractional delay line implemented as an all-pass filter according to the works of Jaffe and Smith [Jaffe et al., 1983]. Such an implementation may produce audible « clicks » when switching from one delay to another. A « clickless » implementation has been proposed by Van Duyne, Jaffe, Scandalis and Stilson [Van Duyne et al., 1997] and implemented here. It consists of two delay lines (which means two allpass filters) feeding a crossfade module to gradually switch from the output of one delay line to the output of the other.

7.3. The man-machine interfaces

In an interactive way, the computer can make sounds and sound sequences evolve in real-time. Musical interpretation can then make sense when disposing possibilities of expression adapted to such new digital instruments. Unlike for traditional instruments, an interface with a digital instrument is not restrained by the mechanics of the instrument. The relation computer-sound depends on synthesis parameters which have a direct influence on the generated sound. The synthesis techniques that are used determine the number of parameters to be controlled and whether or not they can be controlled independently. When using for instance an additive synthesis technique, a software layer between the controller and the parameters should be defined, so that the player could act on several parameters in a simple way. By acting directly on the synthesis parameters intervening in the construction of a sound, the musician would have the possibility of giving a personal touch to the sound.

Thanks to powerful computers, interesting realizations of new interfaces are possible nowadays. Even though completely new instruments can be made, the most common digital instruments use keyboard interfaces. Such interfaces detect the speed at which the keys are struck. To control a percussive instrument this information may be sufficient, but such an interface is not satisfactory for models corresponding to sustained instruments, since for such instruments the sound can be modified after the attack. The music industry has tried to solve this problem by adding so called breath controller or after touch systems. However, the instrumental playing is often closely related to the structure of the instrument. This means that for example the linear structure of a keyboard is not easily adapted to the trumpet play, and that the information given by a keyboard is poor compared to the possibilities that a flute player disposes when playing a sound. Some attempts have been done to make digital wind instruments like MIDI saxophones or flutes. Some instruments have completely failed because of the fingering system when it is completely different from the simulated instrument. Still when the fingering system corresponds to a known instrument, the blowing system remains too simple, giving no possibility for a musician to personalize the sound. This is the most difficult problem, since the excitation system of a wind instrument is generally very complicated to simulate and since the musician generally investigates a lot of time to generate a satisfying sound. In general physical models simulating the propagation of elastic waves (waveguide synthesis models) are well adapted to interfaces imitating traditional musical instruments. Additional commands can then be added extending the sound and/or playing possibilities.

Even though most digital instruments do not sufficiently take into account the personal touch a musician wants to give when playing, musicians are interested in sound effects and transformations. Many composers dream of adding new possibilities to instruments. Playing on non-realizable instruments like a 120 meter long flute or blowing into a string could be interesting. In addition, by modifying the timbre of traditional sounds, instruments that are generally associated to one special « kind » of music, can be used in any context.

To demonstrate the way a sound model can be controlled, two different interfaces are here described: The radio baton and a flute-like interface. The radio baton is a recent instrument designed by Max Mathews, originally meant to execute musical sequences by moving two batons over a square plate containing sensors. In spite of the initial aim of this instrument, we shall here see how it can be used to perform intimate transformations on the sound itself. The second interface is a flute equipped with electromagnetic sensors and a microphone at the embouchure level. The interface is devoted to the proposed hybrid flute model. In this case traditional playing techniques can be used together with the commands added to the instrument.

7.3.1. The radio baton

The conception of sensors detecting movements has been an important part of Max Mathews' research work. One of his realizations, the Sequential Drum, consists of a surface equipped with sensors. When an impact is detected on this surface, the apparatus indicates the intensity and the position of the

impact on the surface [Mathews et al., 1980]. A collaboration with Boie [Boie, 1988][Boie et al., 1989] led to the realization of the radio drum which detects the motion even when there is no contact between the surface and the emitters. The radio drum is in fact able to continuously detect the position of the extremities of the two drum sticks (emitters).

These first prototypes were connected to a microcomputer containing an acquisition card piloted by a conductor program allowing a real-time control of the execution tempo of a partition already memorized by the computer [Mathews 1991a], [Mathews 1991b], [Mathews 1997]. The radio drum is also called the radio baton, since the drum sticks launching musical sequences can be related to the baton used by a conductor of an orchestra.

Max Mathews realized the radio drum for two reasons;

- to make it possible to actively listen to a musical play by releasing musical sequences and thus give a personal interpretation of a musical work.

- to make it possible for a singer to control his or her own accompaniment.

The radio drum comprises two sticks (batons) and a receiver (a 55x45x7 cm parallelepiped) containing an electronic part. In a simplified way, each drum stick can be considered as an antenna emitting radio frequency waves. A network with five receiving antennas is placed inside the drum. It measures the coordinates of the placement of the extremities of the drum sticks where the emitters are placed.

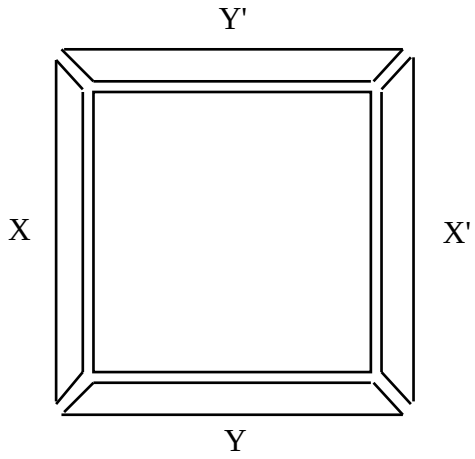


Figure 7.5: Placement of the receiving antennas of the Radio Baton

The intensity of the signal received by one of the antennas depends on its distance from the emitter: it gets stronger as the emitter gets closer to the antenna. To evaluate the position of the extremity of the baton as a function of the x axis, it is sufficient to calculate the difference in intensity between the antennas x and x' (*Figure 7.5*). In the same way the antennas y and y' give the position of the emitter as a function of the y-axis. In order to get information about the height of the baton (the z-coordinate), it is sufficient to add the intensities received by the five antennas.

Since each baton emits signals with different frequencies, it is relatively easy to discriminate between the signals sent by the two batons to each antenna.

The relation between the coordinates and the intensity of the received signals is not linear. A treatment of these data is necessary so that the system gives information proportional to the coordinates of the batons. This operation is realized by a microprocessor making it possible to add other functions to the instrument. The last versions contain an additional software either making it possible to transmit the coordinates of the batons (and information about the controllers) when requested from a micro computer, or to transmit the coordinates of one of the batons when it cuts a virtual plane parallel to the surface of the drum the same way as when one hits its surface. When the program detects the virtual plane, it calculates and transmits the velocity of the movement as a function of the z-axis. The height of the plane can of course be modified. This leads us back to the functioning of the Sequential Drum.

These working modes (transmission of the data when there is an external request or a strike detection) make possible a dialog between the instruments and the device to be controlled. The microprocessor of the radio baton makes the implementation of this communication possible by the use of the MIDI protocol and of a program like MAX [Puckette et al., 1985] to communicate with the radio baton. The control possibilities with instruments that can be programmed are then almost unlimited.

A presentation of the radio baton was given by our research team at an international colloquium on new expressions related to music organized by GMEM (Groupe de Musique Expérimentale de Marseille) in Marseille [Kronland et al., 1997][Ystad, 1997]. Intimate transformations on sound using the radio baton were here demonstrated. I shall briefly describe how the radio baton was used to perform sound transformations with additive and physical synthesis techniques.

The parameters defining an additive synthesis model are given by the amplitude modulation laws A_k of the components together with their frequency modulation laws f_k. The synthesis process is then obtained by:

$$s(t) = \sum_{k=1}^{K} A_k(t)\cos\left(2\pi \int_0^t f_k(u)du\right),$$

where K is the number of spectral components.

Additive synthesis methods give resynthesized sounds of high quality, but these are difficult to manipulate because of the high number of parameters that intervenes. Three parameters have been used for manipulating the sound using the radio baton, namely the duration of the note, its frequency and its amplitude. In each case the manipulations can be done independently for each modulation law or globally on all the modulation laws. In our case the duration of the sound and the frequency manipulations are done in a global way. This corresponds to a simple acceleration or slowing down of a note when the duration is altered, and to a simple transposition when the frequency is altered. The amplitude modulation laws have been modified differently, giving the possibility of effectuating a filtering or an equalization on the sound. In *Figure 7.6*, the control possibilities of the radio baton are illustrated. The sound is generated when one of the radio batons cuts a virtual plane the height of which is predefined by the user. The x-coordinate is related to the duration of the generated sound and the y-coordinate to the transposition factor. The second baton is used to control the note after excitation (aftertouch) and uses the y coordinate

to act on the frequency transposition (like for the first baton) and the z-coordinate to fix the slope α of a straight equalization line. This slope is positive when the baton is over a predefined plane (0 point in *Figure 7.6*) corresponding to an attenuation of low-frequency components and thus to a high-pass filtering. This is illustrated in *Figure 7.7*. When the baton is below the zero point, the slope is negative, corresponding to a low-pass filtering.

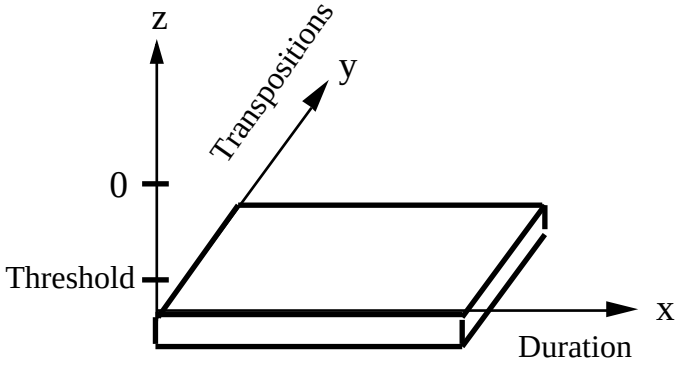


Figure 7.6: Control of the instrument by additive synthesis

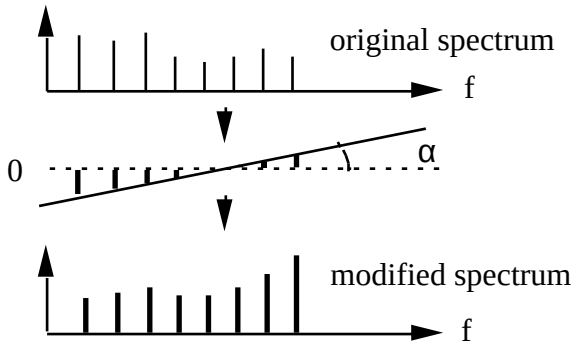


Figure 7.7: Equalization of the spectrum over the 0 point

The second baton could have controlled a third parameter corresponding to the x-coordinate, but since playing the radio baton was difficult with the possibilities already described, this parameter was not used.

When the radio baton acts on a waveguide model (shown in *Figure 7.8*), the parameters to be modified have a physical significance. The delay corresponds for instance to the effective length of the resonator and thus to the frequency of the note played.

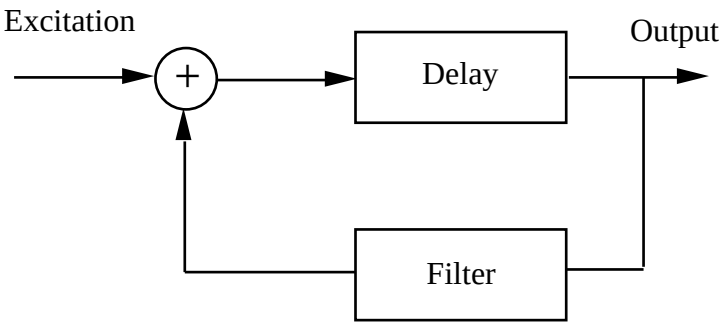


Figure 7.8: Physical synthesis model

The parameters that are to be modified by the batons in this example act on the delay and on the excitation of the model (the source). One of the batons acts on the choice of the excitation: each of the four corners of the receiver corresponds to a different excitation as shown in *Figure 7.9.* In the middle of the plate, the source corresponds to a mixture of the four sources with weights depending on the distance from each corner. The second baton acts on the delay of the resonator (thus on the frequency of the note played) given by the y-coordinate. In addition it acts on the frequency of the excitation signal when a saw tooth source is used.

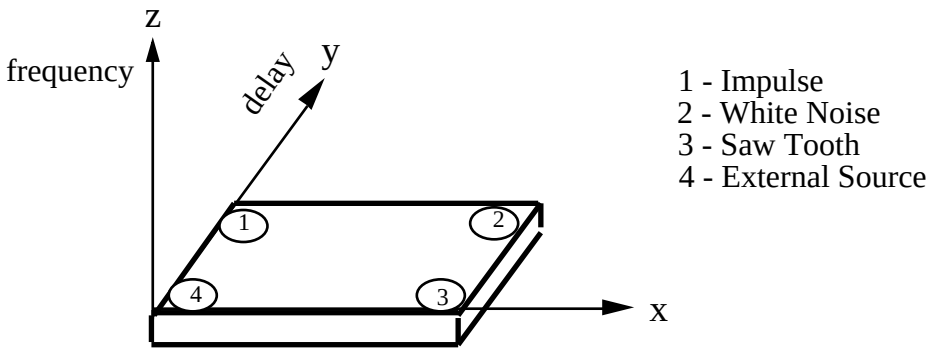


Figure 7.9: Control of the instrument with a waveguide synthesis model

The radio baton is an interesting instrument offering a great number of playing possibilities. However, certain playing effects like for instance the vibrato played on a wind instrument is hard to make with the radio baton. An interface where an other playing technique is used is therefore presented in the next section, namely a flute interface.

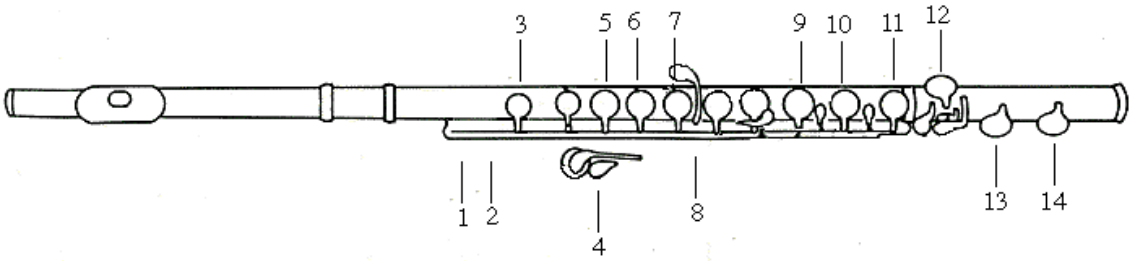
7.3.2. A flute interface

In this section, an example of the construction of an interface using a traditional instrument of the flute type is given. The flute has been connected to a computer by magnetic sensors detecting the finger position and a microphone at the embouchure level detecting the pressure variations. It is important to underline that the aim of such an interface is to add new possibilities to an already existing instrument. The « new » instrument will give musicians the possibility of making use of playing techniques already acquired. This means that the musician has to be familiar with the instrument and will in addition have to learn how to play with the additional commands. The general problem, when proposing new instruments

to musicians, is that they often are afraid of investigating a lot of time to learn to play on an instrument that may not be used in the future. That is why an instrument which conserves traditional playing techniques will hopefully be attractive to musicians.

As seen in section 6.5, the flute synthesis model is based on a hybrid model where the resonator of the instrument is modeled with a physical model simulating the propagation of the waves in a tube (waveguide model), and the source is modeled using a global synthesis model of the waveshaping type. As already mentioned, the physical phenomena observed at the embouchure of a flute are not fully understood, and even though some models describing the interaction between the air jet and the labium have been proposed [Verge, 1995], some of the parameters intervening are difficult to measure, and the equations are not often compatible with real-time implementation.

The flute has been equipped with linear Hall effect sensors detecting the distance to the magnets connected to each key pad. The state of each key pad then gives the frequency of the note played. The speed at which the key pad is closed can be calculated by measuring the distance between the sensor and the magnet at different instants. *Figure 7.10* shows the table of the state of the keys in the first octave.



hole	1	2	3	4	5	6	7	8	9	10	11	12	13	14
C1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C1#	1	1	1	1	1	1	1	1	1	1	1	1	1	0
D1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
D1#	1	1	1	1	1	1	1	1	1	1	1	0	0	0
E1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
F1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
F1#	1	1	1	1	1	1	1	1	0	0	1	0	0	0
G1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
G1#	1	1	1	1	1	1	1	0	0	0	0	0	0	0
A1	1	1	1	1	1	1	0	1	0	0	0	0	0	0
A1#	1	1	1	1	1	0	0	1	0	0	0	0	0	0
H1	1	1	1	1	0	0	0	1	0	0	0	0	0	0
C2	1	1	1	0	0	0	0	1	0	0	0	0	0	0

Figure 7.10: Table of the state of the keys in the first octave.

The linear Hall effect sensors measure a system's performance with negligible system loading while providing isolation from contaminated and electrically noisy environments. *Figure 7.11* shows that the sensor includes a linear amplifier in addition to the Hall sensing element.

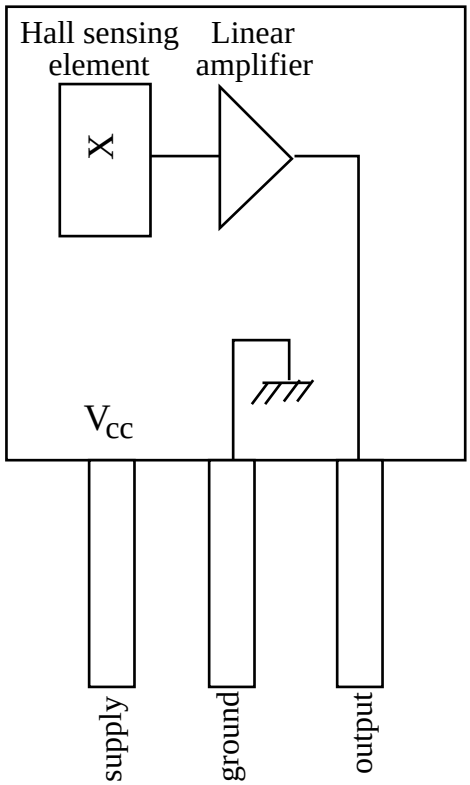


Figure 7.11: Linear Hall effect sensor

The magnetic sensors have been placed on an aluminum rail fastened to the flute. The magnets are fastened to the support rods where the key pads are fixed so that they approach the sensors when the key pads are closed as illustrated in *Figure 7.12* to *Figure 7.15.*

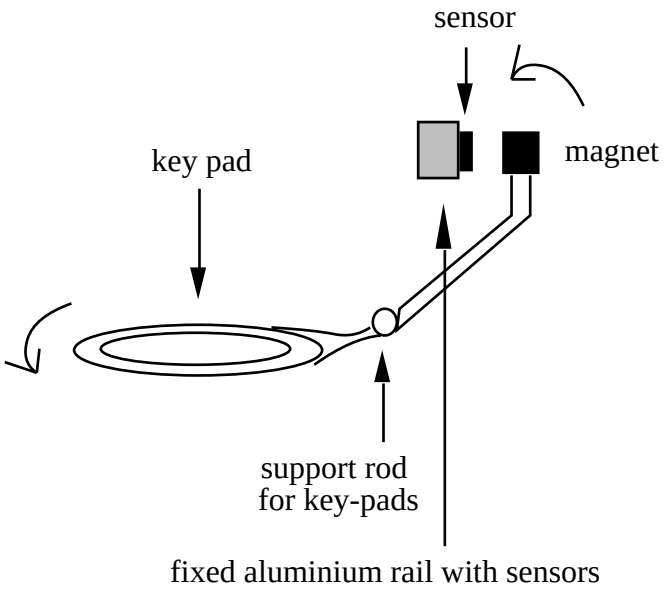


Figure 7.12: Schematic illustration of the connection between the key pad and the magnet.

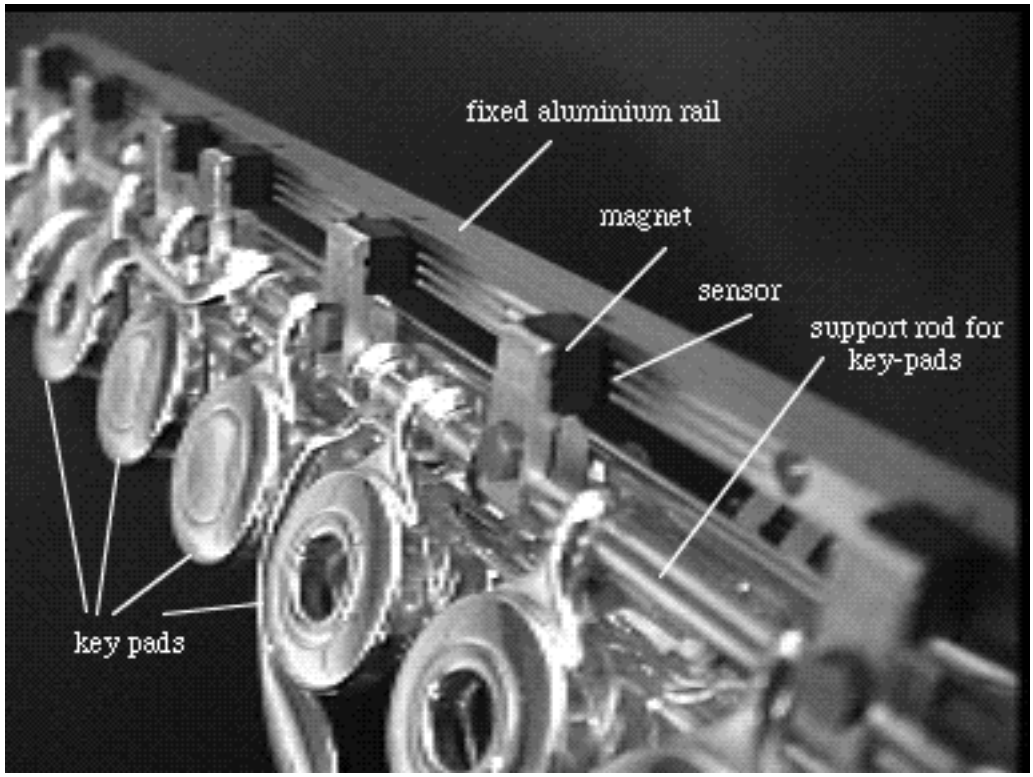


Figure 7.13: Close view of a flute with magnets and sensors

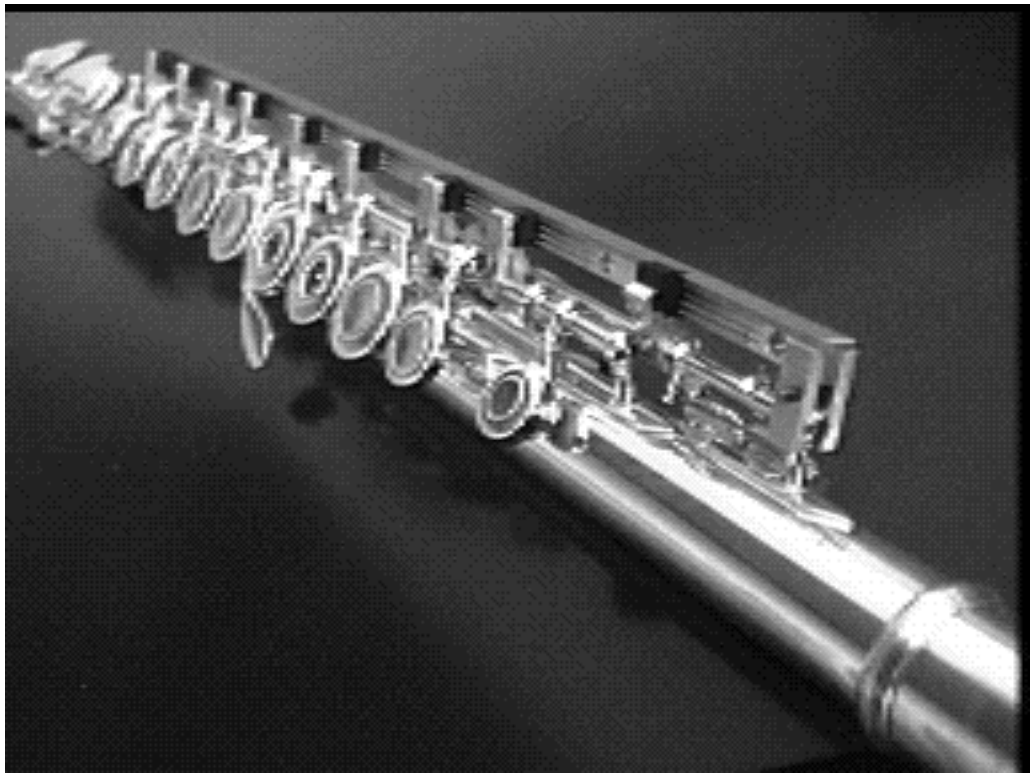


Figure 7.14: Flute with magnets and sensors.

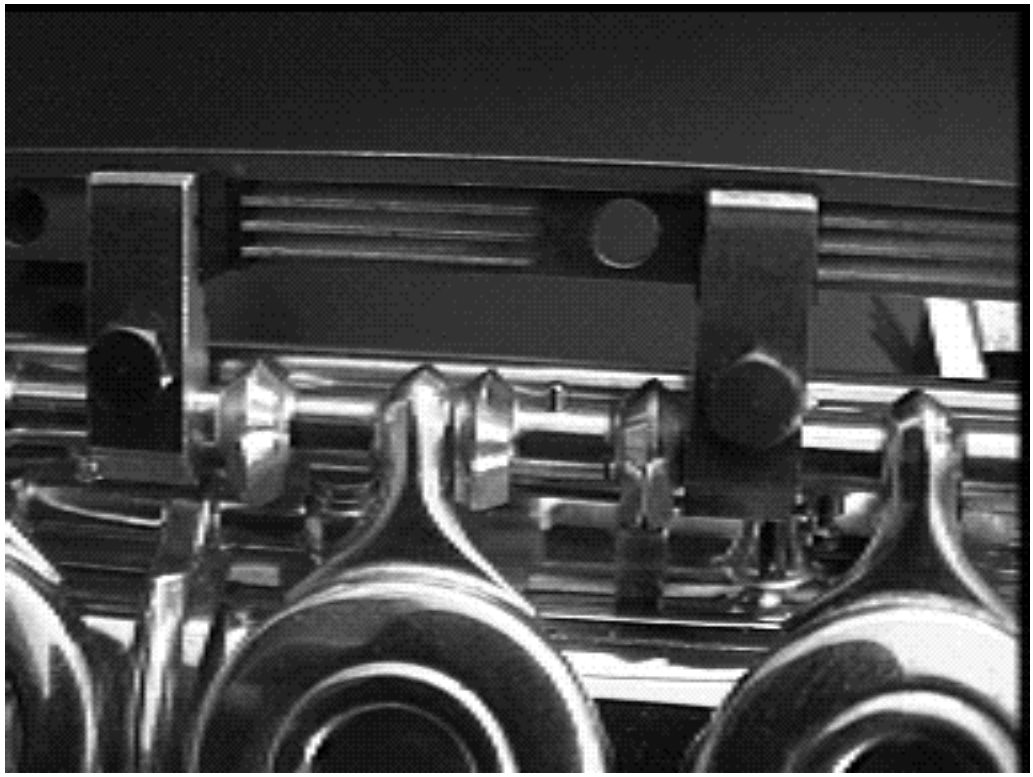


Figure 7.15: Close view of the fixation system of the magnets connected to the key-pad bar.

The state of the keys gives the equivalent length of the resonator, and thus the frequency of the note played. The corresponding delay line in the waveguide model can then be found. In some cases, the state of the keys does not change when the same note is played in different registers, and the measurements of the pressure level at the embouchure should in this case determine the frequency of the note played. This frequency is also used at the entrance of the sine generator feeding the non-linear function of the source model.

The speed at which the key-pad has been closed can also be detected, giving the level of the key-pad noise to be generated.

The pressure variations are detected by a microphone situated inside the flute at the cork position near the embouchure, as shown in *Figure 7.16.*

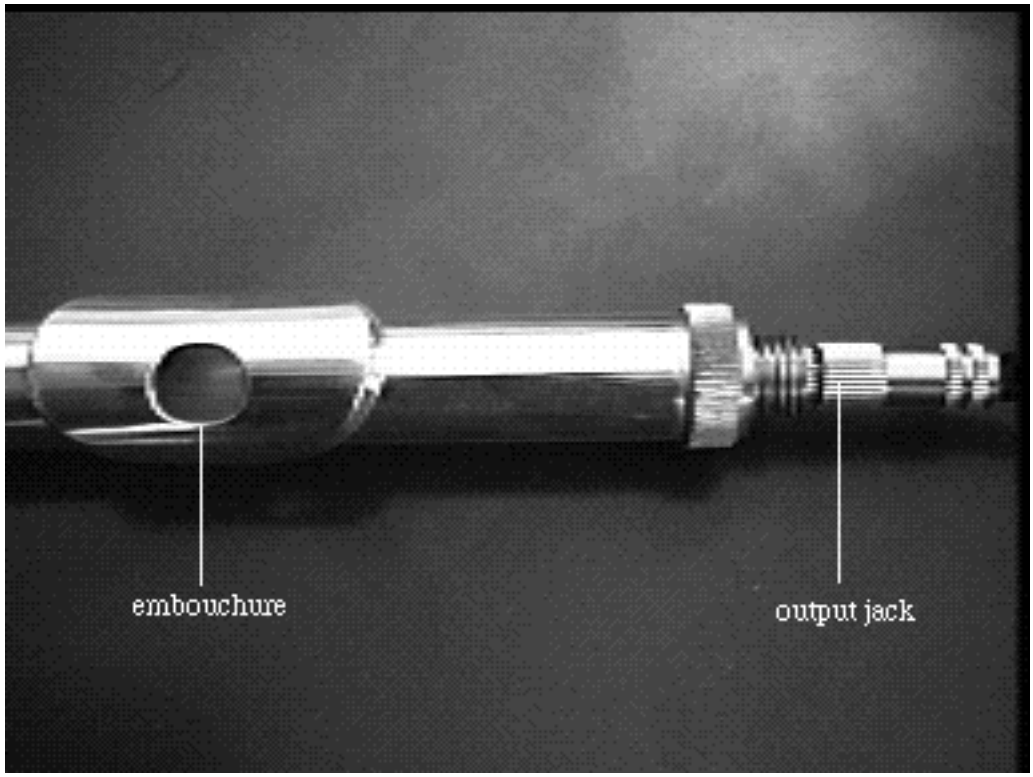


Figure 7.16: Placement of the microphone detecting the driving pressure at the embouchure level

This system does not represent a complete way of characterizing the playing, since for instance the angle at which the air jet hits the labium, or the position of the player's lips, are not taken into account. However, these important features influence the internal pressure which is measured and which acts on the generated sound. The detection of the pressure variations makes it possible to estimate the vibrato, which will be added to the frequency information and used at the entrance of the sine generator. The logarithm of the pressure envelope will further be used as an amplitude at the entrance of the sine generator. It will also be used to determine the level of the non-deterministic source signal that will be added to the deterministic source signal before the entrance of the resonator.

Figure 7.17 shows how the information from the sensors can be transformed and sent to the real-time processor. In this case the information is transformed into MIDI codes through a MIDI coding processor. MIDI (Musical Instrument Digital Interface) is the specification for a set of digital codes for transmitting music control and timing information in real time, and for the hardware interface through which the codes are transmitted [Moog, 1986]. MIDI can for instance be used to network synthesizers, sequencers and rhythm-machines on-stage. It is often used as a tool for composers, and represents an inexpensive link between large research computers and conveniently programmable music-oriented tone-producing instruments. Each MIDI-equipped instrument usually contains a receiver and a transmitter. The receiver accepts messages in MIDI format and executes MIDI commands. The transmitter originates messages in MIDI format and transmits them by a UART (Universal Asynchronous Receiver-Transmitter) and a line driver. The interface operates at 31.25 kBaud, asynchronous with a start bit, 8

data bits (D0 to D7) and a stop bit. This makes a total of 10 bits for a period of 320 microseconds per serial byte.

The information can be manipulated by the MAX program [Puckette et al., 1985] which can for example link the instrument to other MIDI systems. The program MAX can also be used to alter the sound by modifying the synthesis parameters.

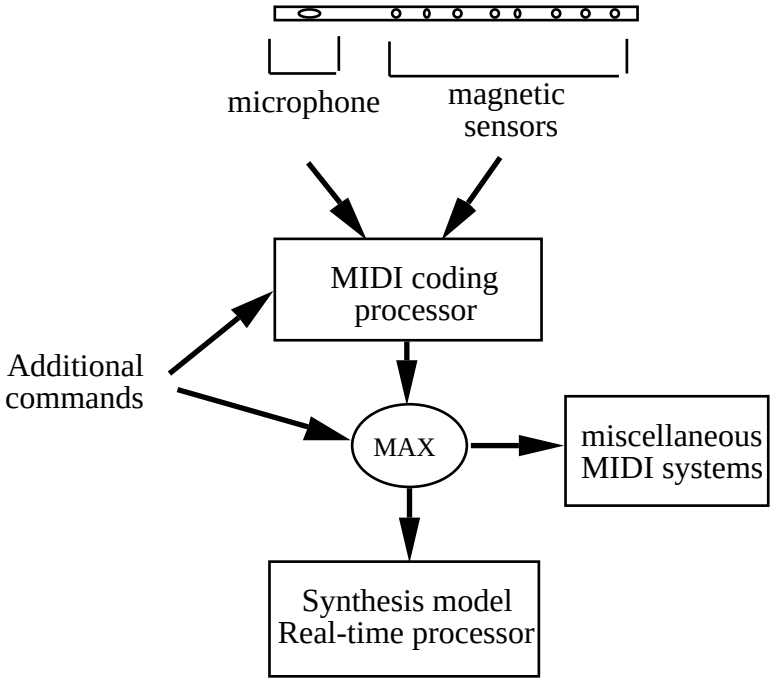


Figure 7.17: Synoptic of the connection of a digital interface of the flute type

In the next section, a discussion on the performance with this interface will be given. It will be seen how new commands can be added and how the characteristics of the instrument can be changed.

7.4. Performance using the flute-like interface

As mentioned in the introduction, sound modeling does not only consist of resynthesizing sounds, but also of doing intimate transformations on the sound. This makes it possible to use the digital flute both as a traditional instrument and as a digital instrument which makes use of additional commands to modify the sound. The modifications of the sounds can either be done at the MIDI interconnection level or by acting directly on the parameters of the model.

Since the flute-like interface generates MIDI codes to control the model, one can use these codes to pilot any kind of MIDI equipped instrument. This is why the MAX program is used to make the correspondence between the MIDI codes obtained from the interface and the parameters of the model.

Actually, the MAX program can be used to distribute information and commands to other MIDI systems. *Figure 7.18* shows the MIDI control in a schematic way.

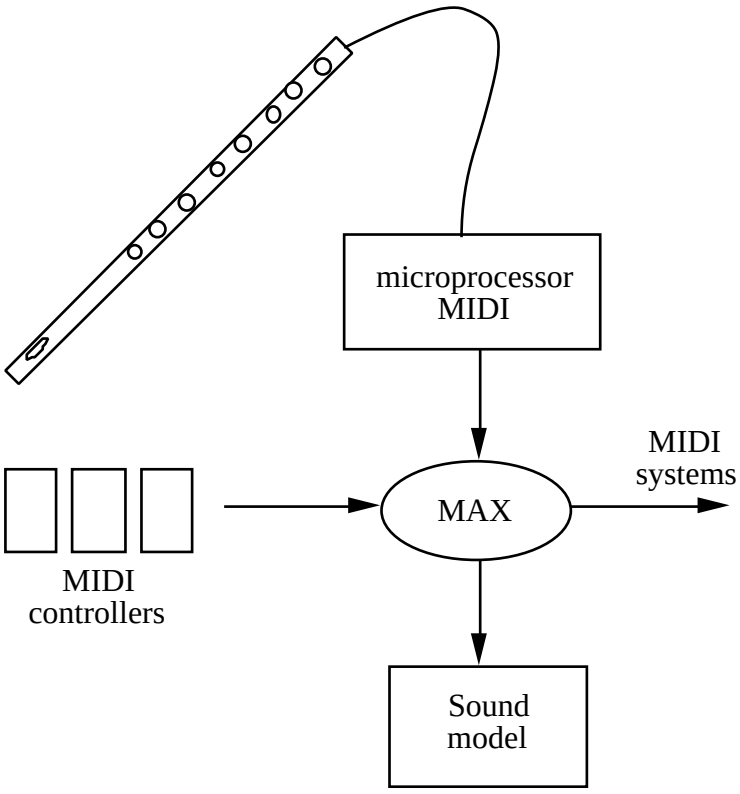


Figure 7.18: Schematic representation of the MIDI interconnection

From a musical point of view, one can for example imagine a musical piece where some MIDI sequences are launched when a specific note is played on the flute. In the same way, since the real-time processor is also considered as a MIDI instrument, the tune of the flute can be changed by assigning arbitrary values of the frequency for a given key state. At this point, the musical possibilities are only limited by the imagination.

The parameters of the model can dramatically act on the sound itself. Since the digital flute model has been designed to respond to MIDI codes, one can act on the parameters of the model using MIDI controllers such as pedals, sliders, etc. These controllers can act on different parameters of the model. In *Figure 7.19* the main locations on which it may be interesting to operate have been pointed out. Even though these possibilities are not exhaustive, I shall briefly describe their impact on the sound:

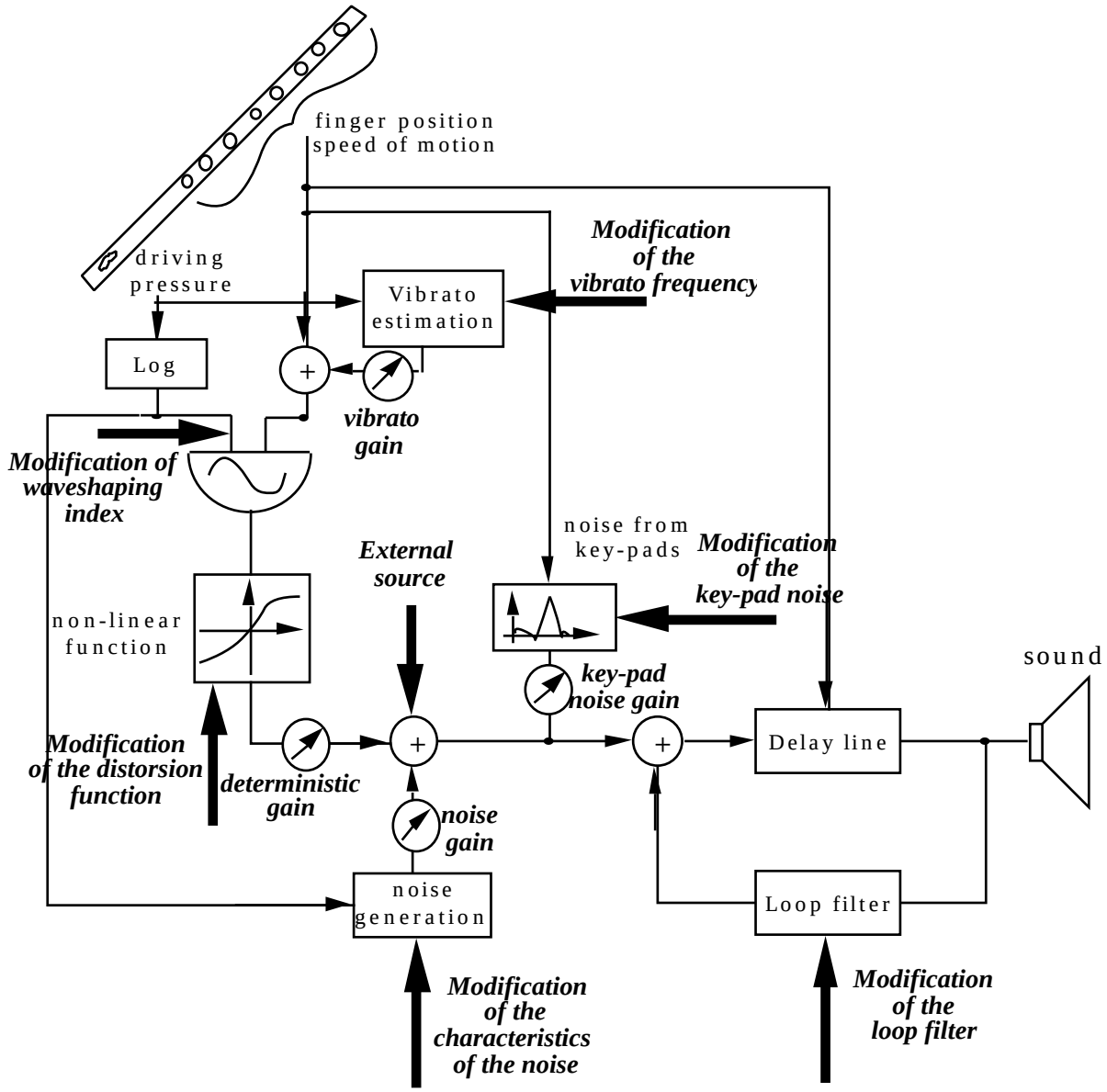


Figure 7.19: Control of the model's parameters. The bold and italic indications show the modification possibilities.

- Action on the vibrato. The frequency of the vibrato can be changed by acting on the filter selecting the fluctuations of the amplitude modulation law corresponding to the internal pressure. By changing the gain of the output of the filter, the depth of the vibrato can be changed. The vibrato can also be artificially generated.
- Action on the distortion index. The distortion index is a very sensitive parameter which has been estimated to fit the spectral evolution of the flute sound. Nevertheless, a change in the correspondence between the internal pressure and the distortion index can be imagined. A brass effect can then be given to a flute sound by increasing the variation domain of the distortion index.

- Action on the distortion function. Changing the characteristics of the distortion function dramatically changes the timbre of the deterministic part of the source signal. A distortion function with a decomposition which only contains odd Chebychev polynomials can for example be used to generate a clarinet- or pan flute- like source.
- Action on the noise. The characteristics of the noise can be modified by acting on the noise filter (power spectral density) and on the statistics (probability density function). The relation between the deterministic and the stochastic parts of the source signal can also be modified by acting on the noise gain. If the deterministic part is removed, then the resulting sound would be a noise filtered by the resonator.
- Action on key-pad noise. The level of the key-pad noise can be adjusted using a gain at the output of the key-pad noise table. If both the deterministic and stochastic parts of the source are removed, the resulting sound would correspond to the one obtained by closing the key-pads. The key-pad noise can also be altered by modifying the corresponding table and could for instance be replaced by any percussive sound.
- Action on the resonator. The loop filter characterizes the resonator and takes into account dissipation and dispersion phenomena. By altering this filter, the characteristics of the medium in which the waves are propagating will be altered. Cross-synthesis effects can be made by using parameters corresponding to a source of a certain instrument and the resonator corresponding to another instrument. Using a loop filter corresponding to a string with a flute excitation, a very particular sound would be generated, corresponding to blowing « into » a string. In a similar way external sources to be filtered by the resonator can be added. This would, for example, allow the generation of the noises made by the flutist while playing.
- Action on the delay line and the oscillator frequency. By changing the offset of these parameters, one can simulate instruments with unrealistic sizes, like for example a 100m long flute.

All these manipulations show the advantage of sound modeling, which enables a modification of a natural sound in the same way as synthetic sounds.

8. Conclusion

Sound modeling consists of designing synthesis methods the aim of which is to reproduce and manipulate a given natural sound. For that purpose, one first has to develop analysis methods adapted to the non-stationary nature of sounds. Further on, estimation procedures must be designed in order to extract parameters characterizing the sound from a physical and a perceptive point of view. The synthesis process can be designed to reproduce a perceptive effect and/or to simulate the physical behavior of the sound generating system. The correspondence between analysis and synthesis parameters is crucial and can be achieved only if the synthesis model is well adapted to the sound which is to be simulated. This work addresses the sound modeling using a combination of physical and signal models. This approach aims to take into account the most relevant aspects of the sound linked to the physical behavior of the sound generator and to the perceptive aspect which is sensitive to the extreme complexity of most natural sounds. Thus, it is possible to manipulate and to make intimate transformations on the sound while conserving the physical nature of the original sound. Real-time control of such models makes it possible to use specially designed interfaces mirroring already existing sound generators like traditional musical instruments.

The presentation of this work starts with a description of various synthesis techniques. These techniques can be divided into two main classes, namely signal models and physical models. Signal models use a purely mathematical description of sounds and are used to simulate a perceptive effect. Physical models describe the sound generating system using physical considerations. Their aim is to simulate the behavior of existing or virtual sound generators.

Analysis techniques able to characterize the timbre of a sound must take into account both its dynamic and spectral behavior. The analysis methods of the time-frequency type are shown to satisfy this requirement. After a description of the most famous methods such as the Gabor and the wavelet transform, a matched time-frequency analysis method using some a-priori knowledge on the signal has been designed. This method, which is adapted to the sound generated by resonant systems, optimize the analysis process. Examples of the extraction of quantitative values from these time-frequency representations are discussed to illustrate the specificity of each method.

In a first attempt to construct a sound model, an example of estimation of parameters corresponding to an additive synthesis method applied to the flute sound is given. It consists of grouping the amplitude and frequency modulation laws of the components to obtain a simple model based on a group additive synthesis. The sounds obtained using this method are satisfactory, but since the model makes use of a purely mathematical description of the sound, the relation between its parameters and the physics of the instrument is not obvious. This makes the control of such a model difficult. In addition such a model does not take into account the turbulent noise being an important part of a flute sound.

Another attempt of sound modeling has therefore been done using physical models. Such models are constructed to simulate the solution of the one-dimensional wave equation in a bounded dissipative

and dispersive medium. The parameters of the model can be found either from the explicit solution of the equation or estimated from the analysis of real sounds. Theoretical parameters which are often biased since they are resulting from simplified equations, can thus be adjusted thanks to experiments on natural sounds. The proposed physical model has been applied to the flute and to the guitar. The dispersion introduces inharmonicity between the partials and the dissipation leads to different decay times for the components. The comparison between the flute and the guitar shows that the modes in a guitar string are less attenuated than the modes in a tube. The dispersion is also very different in the two cases. In a tube the dispersion effect is very small and the frequency interval between the components gets smaller as the component rank increases. In the string case the dispersion effect is important and the frequency interval gets bigger as the component rank increases. A description of the construction of specific filters taking into account these important phenomena is given. It is based on the minimization of the energy difference in a neighborhood of the resonant peaks between the response of the model and the response of the real system. The corresponding impulse response is then constructed in a way similar to the inversion of a time-frequency representation. Actually it consists in summing up elementary functions adjusted and positioned in the time-frequency plane, along a curve representing the group delay. The model constructed this way leads to good quality sounds where physical considerations are taken into account. Hereby, for instance, a sound produced by a short string decays faster than a sound produced by a long string. In the case of a tube, a noise source produces an easily recognizable turbulent noise propagating into the tube.

Nevertheless, for sustained systems it is important to model the source of the system. This problem has been addressed by separating the source and the resonator using a deconvolution technique. Further on the source is modeled in terms of a sum of a stochastic and a deterministic part separated by adaptive filtering techniques such as the LMS algorithm. When the source behavior is assumed to be non-linear, the deterministic part can be modeled using a non-linear technique such as the waveshaping. To fit the spectral evolution of the synthetic and the real source, psychoacoustic criteria such as the tristimulus have been used. Such criteria makes it possible to separately treat components within the same critical band. The stochastic part has been modeled using linear filtering to reconstruct the power spectral density of the stochastic part of a real sound. The probability density function has also been taken into account. These methods have made the construction of realistic sounds using a combination of signal and physical models possible.

The real-time control problem of such a model has finally been addressed to propose a realization of an interface. The implementation on a real-time processor and the control of such a general model have therefore been described. Two different interfaces have been demonstrated; the radio baton and a flute designed especially to control the proposed flute model. Some applications to the sound modeling are finally given showing the advantage of sound modeling which enables a modification of natural sounds in the same way as synthetic sounds.

References

Allen, J. B. & Rabiner, L. R. A unified approach to short-time Fourier analysis and synthesis. Proceedings of the IEEE, 1977, 65, pp.1558-1564.

Allouis, J.F., & Bernier, J.Y. The SYTER project : Sound Processor Design and Software Overview. Proceedings of the International Computer Music Conference, Venice 1982.

Arfib, D. Digital synthesis of complex spectra by means of multiplication of non-linear distorted sine waves. Journal of the Audio Engineering Society, 1979, 27, pp. 757-768.

Arfib, D., & Delprat, N. Musical transformations using the modifications of time-frequency images. Computer Music Journal, 1993, 17 n° 2, pp. 66-72.

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

Atal B.S. & Hanauer S.L. Speech analysis and synthesis by linear prediction of the speech wave. Journal of the Acoustical Society of America, 1971, 50, pp. 637-655.

Balibar S. L'acoustique de la flûte. La Recherche, 1981, n°118, pp.36-44.

Beauchamp, J. W. Analysis and synthesis of cornet tones using non-linear interharmonic relationships. Journal of the Audio Engineering Society, 1975, 23, pp.778-795.

Beauchamp, J. W. Synthesis by Spectral Amplitude and « Brightness » Matching of Analyzed Musical Instrument Tones. Journal of the Audio Engineering Society, 1982, Vol. 30, No. 6.

Boie R.A. AT&T Bell Laboratories, 1988.

Boie, R.A., Ruedisueli, L.W. & Wagner, E.R. Gesture Sensing via Capacitive Moments. AT&T Bell Laboratories, 1989. Work Project N° 311401-(2099,2399).

Cadoz, C., Luciani, A., & Florens, J.L. Responsive input devices and sound synthesis by simulation of instrumental mechanisms. Computer Music Journal, 1984, 14 n° 2, pp. 47-51.

Chaigne, A. Trends and Challenges in Physical Modeling of Musical Instruments. Proceedings of the ICA 1995 in Trondheim, Volume III, pp.397-400.

Cheung, N.M., & Horner, A. Group synthesis with genetic algorithms. Journal of the Audio Engineering Society, 1996, 44, pp.130-147.

Chowning, J. The synthesis of complex audio spectra by means of frequency modulation. Journal of the Audio Engineering Society, 1973, 21, pp.526-534.

Cook, P.R. A meta-wind-instrument physical model controller, and a meta-controller for real-time performance control. Proceedings of the 1992 International Music Conference, San Francisco, 1992, pp.273-276.

Delprat, N., Guillemain, P., & Kronland-Martinet, R. Parameter estimation for non-linear resynthesis methods with the help of a time-frequency analysis of natural sounds. Proceedings of the 1990 International Music Conference, Glasgow, 1990, pp.88-90.

Depalle, P., & Rodet, X. A new additive synthesis method using inverse Fourier transform and spectral envelopes. Proceedings of the 1992 International Music Conference, San Francisco, 1992, pp.161-164.

De Poli, G., Picciali, A., & Roads, C. The representation of musical signals. Cambridge, Mass: M.I.T. Press, 1991.

Dolson, M. The phase vocoder: a tutorial. Computer Music Journal, 1986, 10 n° 4, pp. 14-27.

Fabre B., Hirschberg A., Wijnands P.J. Vortex Shedding in Steady Oscillation of a Flue Organ Pipe. Acta Acoustica, 1996, Vol. 82 n°6, pp. 863-877.

Flanagan J.L, Coker C.H, Rabiner P.R, Schafer R.W., & Umeda N. Synthetic voices for computer. I.E.E.E Spectrum, 1970 , 7, pp. 22-45.

Flandrin P. Temps-fréquence. Hermès, traité des nouvelles technologies, série traitement du signal, 1993.

Fletcher N.H. Acoustical correlates of flute performance technique. Journal of the Acoustical Society of America, 1975, Vol. 57 n° 1, pp. 233-237.

Fletcher N.H. & Rossing T.D. The Physics of Musical Instruments, Springer-Verlag, 1990.

Gabor, D. Acoustical quanta and the nature of hearing. Nature, 1947, 159, n° 4.

Gilbert J. & Petiot J.F. Brass Instruments, Some Theoretical and Experimental Results. Proceedings of the International Symposium on Musical Acoustics, Edinburgh, 1997, pp.

Gradshteyn, I.S. & Ryzhik, I.M. Table of Integrals, Series and Products. Academic Press, 1980.

Grey, J.M. An exploration of musical timbre. Dept for Music, Stanford University, Report No.STAN_M_2, 1975.

Grossman, A., & Morlet, J. Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM Journal of Mathematical Analysis, 1984, 15, pp. 723-736.

Guillemain, Ph., & Kronland-Martinet, R. Characterisation of Acoustics Signals Through Continuous Linear Time-frequency Representations. Proceedings of the IEEE, Special Issue on Wavelets, 1996, Vol. 84 n°4, pp. 561-585.

Guillemain, Ph., Kronland-Martinet, R., & Ystad, S. Physical Modelling Based on the Analysis of real Sounds. Proceeding of the ISMA, Edinburgh 1997, Vol 19, pp. 445-450.

Hamming, R.W. Digital Filters. Prentice-Hall International Editions, 1989, Third edition.

Haykin, S. Communication systems. McMaster University, John Wiley & Sons, 1983, Second Edition.

Horner, A. Double-Modulator FM Matching of Instruments Tones. Computer Music Journal, 1996, V. 20 (2), pp 57-71.

Horner, A., & Beauchamp J.W. Piecewise-linear approximation of additive synthesis envelopes: a comparison of various methods. Computer Music Journal, 1996, 20 n° 2, pp.72-95.

Jaffe, D. A. & Smith, J.O. Extensions of the Karplus-Strong Plucked-String Algorithm. Computer Music Journal, Vol. 7, No.2, summer 1983.

Karplus, K. & Strong, A. Digital Synthesis of Plucked String and Drum Timbres. Computer Music Journal, 1983, 2 (7): 43-55.

Kergomard, J. Champ interne et champ externe des instruments à vent. Thèse d'Etat, Université Paris IV, 1981.

Kleczkowski, P. Group additive synthesis. Computer Music Journal, 1989, 13 n° 1, pp. 12-20.

Kronland-Martinet R. Digital substractive synthesis of signals based on the analysis of natural sounds. In "Etat de la Recherche Musicale (au 1er janvier 1989)" 1988, Ed. A.R.C.A.M. , Aix en Provence.

Kronland-Martinet, R. The use of the wavelet transform for the analysis, synthesis and processing of speech and music sounds. Computer Music Journal, 1989, 12 n° 4, pp.11-20 (with sound examples on disk).

Kronland-Martinet, R., & Guillemain, P. Towards non-linear resynthesis of instrumental sounds. Proceedings of the 1993 International Music Conference, San Francisco, 1993, pp. 86-93.

Kronland-Martinet, R., Morlet, J., & Grossman, A. Analysis of sound patterns through wavelet transforms. International Journal of Pattern Recognition and Artificial Intelligence, 1987, 11 n° 2, pp. 97-126.

Kronland-Martinet, R., Voinier, T. & Guillemain, P. Agir sur le son musical avec la baguette radio. Colloque International sur les nouveaux gestes de la musique, organized by GMEM (Groupe de Musique Experimentale de Marseille), 4-5 April 1997.

Laroche, J. The use of the matrix pencil method for the spectrum analysis of musical signals. Journal of the Acoustical Society of America, 1993, 94, pp. 1958-1965.

Le Brun, M. Digital waveshaping synthesis. Journal of the Audio Engineering Society, 1979, 27, pp. 250-266.

Makhoul, Linear prediction, a tutorial review. Proceedings of Institute of Electrical and Electronics Engineers, 1975, 63, pp. 561-580.

Markel, J. D, & Gray, A. H. Linear Prediction of Speech. Springer-Verlag, Communication and Cybernetics (12), 1976.

Mathews, M.V. & Abbot, C. The Sequential Drum. Computer Music Journal 4 n°4 , 1980.

Mathews, M.V. The Conductor Program and Mechanical Baton. Mathews M.V. & Pierce J.R. eds, Current Directions in Computer Music Research, M.I.T. Press, Cambridge, Mass, 1991a.

Mathews, M.V. The Radio Baton and Conductor Program, or : Pitch, the Most Important and Least Expressive Part of Music. Computer Music Journal 15 n°4, 1991b.

Mathews, M.V. Exécution en direct à l'âge de l'ordinateur. H. Dufourt & J.M. Fauquet eds., "La musique depuis 1945 - matériau, esthétique, perception", Mardaga, Bruxelles, 1997.

Max, J. Méthodes et techniques de traitement du signal et applications aux mesures physiques. Masson, 1987, 4ième edition.

McAulay, R., & Quatieri, T. Speech analysis-synthesis based on a sinusoidal representation. IEEE Transactions on Speech, Acoustics and Signal Processing, 1986, ASSP-34, pp. 744-754.

Miskiewicz, A. & Rakowski A. The range of sound pressure level and the range of loudness in musical instruments. Proceedings of the international conference Subjective and Objective Evaluation of Sound, Czerniejewo (Poland) 1990.

Moog, B. MIDI: Musical Instrument Digital Interface. Journal of Audio Engineering Society, Vol. 34, No.5, 1986.

Moorer, J. A. The use of the phase vocoder in computer music applications. Journal of the Audio Engineering Society, 1978, 26, pp. 42-45.

Morse, P.M. & Ingard, K.U. Theoretical Acoustics. McGraw-Hill, 1968.

Oates S., & Eagleston, B. Analytic Methods for Group Additive Synthesis. Computer Music Journal, 1997, V. 21 (2), pp. 21-39.

O'Donnell, M., Jaynes, E.T. & Miller, J.G. Kramers-Kronig relationship between ultrasonic attenuation and phase velocity. Journal of the Acoustical Society of America, 1981, Vol. 69, No. 3.

Palamin J.-P., Palamin, P. & Ronveaux, A. A Method of Generating and Controlling Musical Asymmetrical Spectra. Journal of the Audio Engineering Society, 1988, Vol 36, No. 9.

Parks, T.W. & Burrus, C. S. Digital Filter Design. John Wiley & Sons, 1987.

Picinbono, B. Signaux aléatoires. Dunot, 1993, tome 1, 2 et 3.

Pollard H.F. & Jansson E.V. A Tristimulus Method for the Specification of Musical Timbre. Acoustica, 1982, Vol. 51.

Puckette, M. & Zicarelli, D. Max - an Interactive Graphic Programming Environnement, Opcode Systems, 1985.

Rabiner L.R. & Gold B. Theory and Application of Digital Signal Processing. Prentice Hall, 1975.

Rakowski A. Opening Trancients in Tones of the Flute, Bulletin de la Société des amis des sciences et des lettres de Poznan, 1965.

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

Risset J.C, & Wessel D.L. Exploration of timbre by analysis and synthesis, in D. Deutsch ed., The Psychology of Music, Academic Press 1982, pp. 26-58.

Roads, C. Automated granular synthesis of sound. Computer Music Journal, 1978, 2, No. 2, pp. 61-62.

Rothstein J. MIDI A Comprehensive Introduction. Oxford University Press, 1992.

Ruskai, M.B., Beylkin, G., Coifman, R., Daubechies, I., Mallat, S., Meyer, Y., & Raphael, L., ed. Wavelets and their applications. Boston: Jones and Bartlett, 1992.

Serra, X. A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition. PhD, Stanford University, October 1989.

Smith, J.O, Musical Applications of Digital Waveguides. Stanford University Center For Computer Research in Music and Acoustics, Report No. STAN-M-39, 1987.

Smith, J.O. Physical modeling using digital waveguides. Computer Music Journal, 1992, 16 n° 4, pp. 74-91.

Smith, J.O. & Rocchesso, D. Connections between Feedback Delay Networks and Waveguide Networks for Digital Reverberation. Proceedings from the ICMC Aarhus, Denmark, 1994, pp.376-377.

Sneddon, I.N. Fourier Transforms, McGraw-Hill Book Company, 1951.

Schwartz, M. Information, Transmission Modulation, and Noise. Second Edition, McGraw-Hill Kogakusha, LTD., 1970.

Torresani, B. Analyse continue par ondelettes. InterEditions/CNRS Editions, 1995.

Valette, C. & Cuesta, C. Mécanique de la corde vibrante, Hermès, 1993.

Van Duyne, S. & Smith, J.O. A simplified Approach to Modeling Dispersion Caused by Stiffness in Strings and Plates. Proceedings from the ICMC Aarhus, Denmark, 1994, pp.407-410.

Van Duyne, S., Jaffe, D. A., Scandalis, G. P. & Stilson, T. S., A Lossless, Click-free, Pitchbend-able Delay Line Loop Interpolation Scheme. Proceedings of the 1997 International Music Conference, Thessaloniki, Hellas, 1997.

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

Weinreich, G. Coupled piano strings. Journal of the Acoustical Society of America, 1977, 62, pp. 1474-1484.

Widrow B., & Stearns S.D. Adaptive Signal Processing. Englewood Cliffs, Prentice-Hall Inc., 1985.

Yegnanarayana B. Design of recursive group-delay filters by autoregressive modeling. IEEE Trans. of Acoust. Sp. and Sig. Proc. ASSP-30, 632-637.

Ystad S., Guillemain Ph., & Kronland-Martinet R. Estimation Of Parameters Corresponding To A Propagative Synthesis Model Through The Analysis Of Real Sounds. Proceedings of ICMC Hong Kong, 1996, pp. 19-24.

Ystad, S., De la facture informatique au jeu instrumental. Colloque International sur les nouveaux gestes de la musique, organized by GMEM (Groupe de Musique Experimentale de Marseille), 4-5 April 1997.

Summary

Sound modeling consists in designing synthesis models to reproduce and manipulate natural sounds. The aim of this work is to define sound models taking into account physical aspects linked to the sound source and their perceptive influence. For this purpose, a combination of physical and signal models has been used.

The document starts with a presentation of the most important synthesis methods. Further on, one searches for a correspondence between the synthesis parameters and the data obtained through the analysis. The non-stationary nature of sound signals necessitates the consideration and the adaptation of analysis methods like time-frequency representations. The parameters resulting from the analysis can directly feed an additive synthesis model. An application to flute sounds corresponding to this kind of modeling is presented.

Models simulating the wave propagation in the medium are further designed to give more importance to the physical characteristics of the sound generating system. Stretched strings and tubes were here considered. By comparing the solutions of the movement equations and the response of the so-called waveguide system, one constructs an estimation method for synthesis parameters. Dispersive and dissipative effects due to the medium in which the waves propagate are then taken into account.

For sustained sounds the source and the resonator have been separated by deconvolution. By using an adaptive filtering method, the source signal is decomposed in two contributions: a deterministic component and a stochastic component. The modeling of the deterministic part, whose behavior generally is non-linear, necessitates the use of global synthesis methods like waveshaping, and perceptive criteria such as the Tristimulus criterion. The stochastic component is modeled taking into account the probability density function and the power spectral density of the process.

An example of real-time control of a flute model is presented. A flute equipped with sensors is used as an interface to control the proposed model. Possibilities of intimate sound manipulations obtained by acting on the parameters of the model are discussed.

Sammendrag

Lydmodellering innebærer konstruksjon av syntesemodeller til bruk for gjenskaping og manipuliering av naturlige lyder. Målet med dette arbeidet er å definere lydmodeller som tar hensyn til fysiske aspekt knyttet til lydkilden og deres lyttemessige innflytelse. En kombinasjon av fysiske modeller og signal modeller er brukt til dette formålet.

Denne fremstillingen starter med en presentasjon av de viktigste syntesemetodene. Videre søkes en sammenheng mellom synteseparametre og analysedata. De ikke-stasjonære egenskapene til lydsignalene nødvendiggjør vurdering og tilpasning av analysemetoder så som tids-frekvens representasjoner. Parametrene fra analysen kan direkte brukes på additive syntesemodeller. Denne type modellering er anvendt på fløytetoner.

Modeller som simulerer bølgeforplantingen i mediet er videre konstruert for i større grad å ta hensyn til fysiske karakteristikker av lydproduksjonssystemet. Modellene er anvendt på spente strenger og rør. En estimeringsmetode av synteseparametrene basert på en sammenligning av bølgelikningenes løsninger og responsen til bølgeledersystemet er utviklet. Denne modellen tar hensyn til dispersjon og dissipasjon som oppstår ved bølgeforplanting i mediet.

For vedlikeholdte lyder er kilde og resonator adskildt ved dekonvolusjon. Ved hjelp av en adaptiv filtreringsmetode deles kildesignalet videre i et deterministisk og et stokastisk bidrag. Modelleringen av den deterministiske komponenten, som vanligvis har en sterkt ulineær oppførsel, nødvendiggjør bruken av globale syntesemetoder så som « waveshaping » samt lyttemessige kriterier så som Tristimulus kriteriet. Ved modellering av den stokastiske komponenten er det tatt hensyn til sannsynlighetstettheten og effektspektraltettheten til prosessen.

Et eksempel på sann-tids kontroll av en fløytemodell er til slutt presentert. En klassisk tverrfløyte utstyrt med sensorer er brukt til å styre den foreslåtte modellen. Manipuleringsmuligheter ved endring av parametrene som inngår i modellen er diskutert.

Résumé

La modélisation sonore consiste à construire des modèles de synthèses susceptibles de reproduire et de manipuler des sons naturels. L'objectif de ce travail est de définir des modèles sonores prenant en compte aussi bien les aspects physiques liés à la source sonore que leur conséquences perceptives. Pour cela, on s'intéresse à des modèles constitués par une combinaison de modèles physiques et de modèles de signaux.

Après avoir discuté les principales méthodes de synthèse sonore, on aborde la problématique de la mise en correspondance entre les paramètres de synthèse et les données de l'analyse. La nature non-stationnaire des signaux sonores conduit alors à considérer et à adapter des méthodes d'analyse de type temps-fréquence. Les paramètres issus de l'analyse permettent directement de caler des modèles de synthèse additive. Une application à ce type de modélisation est donnée dans le cadre des sons de flûte.

Dans le but d'une meilleure prise en compte des caractéristiques physiques du système de production sonore, on s'intéresse ensuite à la construction de modèles basés sur le comportement des ondes se propageant dans une structure résonante. Les cas des cordes tendues et des tuyaux ont été considérés. La comparaison entre les solutions des équations du mouvement et la réponse du système « guide d'onde » a permis la construction d'une méthode d'estimation des paramètres de synthèse prenant en compte les effets dispersifs et dissipatifs liés au milieu de propagation.

Dans le cas de sons entretenus, la séparation source-résonateur a été réalisée par déconvolution. Le signal source est alors décomposé, par filtrage adaptatif, sous forme de deux contributions: une composante déterministe et une composante stochastique. La modélisation de la composante déterministe, dont le comportement est généralement fortement non-linéaire, a nécessité l'utilisation de méthodes de synthèses globales telle que la distorsion non-linéaire, et de critères perceptifs tel que le Tristimulus. La composante stochastique est quand à elle modélisée en prenant en compte la densité de probabilité et la densité spectrale de puissance du processus.

On aborde enfin le problème du contrôle temps-réel en prenant l'exemple d'un modèle de flûte. La description d'une interface calquée sur l'instrument réel est donnée, et quelques possibilités de manipulation intime des sons par action sur les paramètres du modèle sont discutées.

Mots clefs

Modèles sonores, Analyse-synthèse, Analyse temps-fréquence, Modèles physiques, Modèles de signaux, Distorsion non-linéaire, Tristimulus, Temps-réel, Interface, Flûte, Informatique musicale.