Analysis-Synthesis of Impact Sounds by Real-Time Dynamic Filtering

Mitsuko Aramaki and Richard Kronland-Martinet

***Abstract*—This paper presents a sound synthesis model that reproduces impact sounds by taking into account both the perceptual and the physical aspects of the sound. For that, we used a subtractive method based on dynamic filtering of noisy input signals that simulates the damping of spectral components. The resulting sound contains the perceptual characteristics of an impact on a given material. Further, the addition of very few modal contributions—using additive or banded digital waveguide synthesis—together with a bandpass filtering taking into account the interaction with the excitator, allows realistic impact sounds to be synthesized. The synthesis parameters can be linked to a perceptual notion of material and geometry of the sounding object. To determine the synthesis parameters, we further address the problem of analysis-synthesis aiming at reconstructing a given impact sound. The physical parameters are extracted through a time-scale analysis of natural sounds. Examples are presented for sounds generated by impacting plates made of different materials and a piano soundboard.**

***Index Terms*—Analysis-synthesis, audio virtual reality, banded digital waveguide, impact sounds, timbre, time-varying filtering.**

I. Introduction

SOUND synthesis processes dedicated to impact sounds are of great interest in audio sound design and musical applications. For virtual reality, impact sounds are needed for sound animation movies or video games. They can be colliding objects, gunshots, banging doors, or falling objects. For music, impact sounds are characteristic of percussive instruments and are widely used in computer music compositions. This paper aims at designing a timbre synthesizer dedicated to impact sounds so that both perceptual and physical aspects related to the sounds are taken into account. In addition, the system should satisfy the two following constraints:

- be simple enough so that the system can be used in real-time;
- be able to extract the parameters of the model from the analysis of real sounds, allowing their resynthesis.

Synthesis of impact sounds has already been addressed by several authors. Most of the proposed models are based on the physics of vibrating structures, leading to a modal approach of the synthesis process [1]–[7]. Modal representation of vibrating systems is a generic way of describing both the sound and its relations with the physical characteristics of the structure. Yet, modal representation is not always suitable for complex sounds for example, sounds containing a high density of mixed and/or nonlinear modes. These sounds include cymbals or gongs [8] or strongly struck strings, for which the nonlinear coupling of transverse and longitudinal modes generate additional partials [9], [10]. The synthesis of transient sounds has also been addressed using algorithmic techniques based on digital signal processing. Cook [2] proposed a granular synthesis approach based on a wavelet decomposition of sounds. This approach provides realistic perceptual effects (like walking sounds [11]) but the relation between the synthesis parameters and the physics is not always established. Smith and Van Duyne [12], [13] used a clever way of simulating the effect produced on piano tones by the soundboard of the instrument. They assumed the modes overlapped and the "modal response" was flat to further simulate the damping of the energy through a time-varying filter acting on a white noise.

Following Smith and Van Duyne [12], [13], we present a sound analysis-synthesis model based on a time-varying subtractive synthesis process that acts on a noisy input signal. An additional bandpass filter serves to calibrate both the excitator and the frequency response of the system. This model is adequate to reproduce perceptual effects corresponding to impacted materials, even though the technique does not simulate modes. Further, adding just a few modes that correspond to the most prominent modes of the impacted structure allows sounds to be improved by providing a subjective notion of the size of the sounding object.

In accordance with previous studies on the perception of sounds generated by vibrating structure [14]–[16], we then show how the synthesis parameters can be linked to the physical parameters. This linkage makes it possible to generate realistic sound timbres or blurred ones if morphing techniques are applied.

Another important aspect of the paper is the resynthesis of natural sounds, meaning that we also attempt to reproduce a given impact sound from a perceptual point of view. For that purpose, we show how the parameters of our synthesis model can be estimated from the analysis of natural sounds, using perceptually relevant criteria based on time-scale decomposition. The results are discussed for two sounds: the response of steel and wooden rectangular plates and the response of a wooden piano soundboard (complex geometry).

II. What Is an Impact Sound?

The term *impact sound* is generally used to describe a subjective concept associated with a huge class of sounds. To better quantify this concept, we adopt here a definition close to the one in [17]: in the time domain, an impact sound lasts only a few seconds and is characterized by an abrupt onset and a rapid decay. Typically, the sound corresponds to the vibratory response of

Manuscript received March 4, 2004; revised December 12, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gerald Schuller.

The authors are with the CNRS-Laboratoire de Mécanique et d'Acoustique, 13402 Marseille, France (e-mail: aramaki@lma.cnrs-mrs.fr; kronland@lma.cnrs-mrs.fr).

Digital Object Identifier 10.1109/TSA.2005.855831

a structure under free oscillations that has been excited by an impact, or to the sound produced by the collision of objects. Consequently, the spectral content of such a sound generally is broadband. Impact sounds belong to the class of *nonstationary signals*, and both the time and the frequency behaviors of the corresponding signal have to be taken into account for synthesis purposes. Moreover, despite their short duration, natural impact sounds can easily be related to the vibrating material. This operation shows they contain much relevant perceptual information, making it difficult to model and to parameterize especially when the sounds contain a high density of mixed modes.

Synthesis methods based on physical modeling are very efficient for simulating the sound produced by simple vibrating structures for which the classical theory gives analytical solutions. For complex structures, however, these methods rapidly become unusable; the structure itself must be precisely described, making the physical models complicated. From a perceptual point of view, their refinement is not always successful. To overcome these drawbacks, we sought a sound synthesis model that reproduces the main perceptual features of impact sounds from a reference exclusively given by the original sound signal.

Many subjective tests have shown that perceptual clues allow the source of the impact sound to be identified merely by listening [14]–[16]. These tests have brought to the fore some correlations between physical attributes (the nature of the material, dimensions of the structure) and perceptual attributes (perceived material, perceived dimensions). We present below how the material and the dimensions of the structure are perceived.

A. Perception of the Material

The perception of the material correlated mainly with the damping coefficient $\alpha(f)$ of the spectral components contained in the sound [18]. By considering that the signal is a solution of a linear PDE representing a simple mechanical system (mass-spring-damper), one can assume that the temporal evolution of each mode $s_k(t)$ decreases exponentially

$$s_k(t) = A_k e^{2i\pi f_k t} e^{-\alpha(f_k)t} \tag{1}$$

where A_k is the amplitude and f_k is the eigenfrequency. The spectral representation of the modes is given by

$$\hat{s}_k(f) = \frac{A_k}{\alpha(f_k) + 2i\pi(f - f_k)} \tag{2}$$

which corresponds to a peak localized around the frequency f_k. The damping α generally is frequency-dependent and can be related to the internal friction coefficient, usually considered the most important characteristic of the material. Several expressions of this coefficient can be found [14], [19]–[21].

B. Physical and Perceptual Dimensions of the Structure

1) Physical Dimensions of the Structure: The dimensions of the impacted structure determine the spectral content of the generated sound. The frequencies of the spectral components correspond to the so-called eigenfrequencies of the structure and are deduced for simple cases from the movement equation. In general, the movement equation governing the displacement u of a given structure can be written [22]

$$\frac{\partial^2 u}{\partial t^2} = \frac{E}{\rho} L u \tag{3}$$

where E represents the Young modulus, ρ the mass density of the material and L the differential operator expressing the local deformation. For example, L corresponds to the Laplacian operator in the case of a string (in one dimension) or membranes (in two dimensions) and to the Laplacian squared in the case of a bar (in one dimension) and a thin plate (in two dimensions). The geometry of the structure together with the boundary conditions determine the eigenvalues λ_n of the operator L. The eigenfrequencies ω_n are then defined by

$$\omega_n^2 = \lambda_n c^2 \quad \text{with} \quad c^2 = \frac{E}{\rho}. \tag{4}$$

Thus, the eigenfrequencies are characteristic of the modes of the structure. For multidimensional structures, the modal density increases with the frequency so that the modes overlap at high frequency and become indiscernible. This phenomenon is well known and is described for example in previous works on room-acoustics [23].

2) Perceptual Dimensions of the Structure: The perception of the size of the structure correlated mainly with the modes [24], [25]. In particular, if the modes are numerous and overlap, the human ear can perceive only the most prominent components, defined as the spectral components of greater amplitudes. The rest of the spectrum is composed of overlapping modes of smaller amplitudes. From a perceptual point of view, we shall then consider the spectral content to be composed of the most prominent modes (generally located at low frequency) and of a density of overlapping modes (generally located at high frequency).

III. Sound Synthesis Model

In this section, we propose a sound synthesis model simulating impact sounds from a perceptual point of view. This model is based on the reproduction of the two main contributions previously described and characterizing the material and the perceptual dimensions of the structure. From a synthesis point of view, the most prominent modes are reproduced using modal synthesis. The modal density at high frequency composed of overlapping modes is reproduced by a white noise that generates a broadband spectrum. We propose to model these two contributions separately (material and structure dimensions) even if they cannot be totally disconnected from a physical point of view. Indeed, by the relation (4), one notes that the modes characterizing the dimensions of the structure also depend on the material inasmuch as the eigenfrequencies ω_n are defined from the parameter c, characteristic of the material.

A. Material Contribution

As already mentioned, one of the characteristics of impact sounds is a broadband spectrum. This spectral behavior is due to the possibly high density of modes together with their fast damping. To simulate such a spectrum behavior, we used a white noise to generate a broadband spectrum (from an energetic point of view). As a basis of the synthesis model, we chose the one

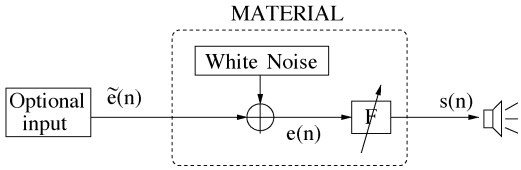


Fig. 1. Synthesis model reproducing the main characteristics of the material.

proposed by Smith and Van Duyne [12], [13] which consists in simulating the damping by a time-varying filtering process that acts on the input signal e (Fig. 1). Here, the optional input $\tilde{e}$ is not used, but will be later devoted to taking into account the geometry of the structure (see the next Section III-B). The time-varying filter, noted F, is based on an IIR structure (Infinite Impulse Response) with a *frozen-time approximation transfer function* defined at the sampling time n by [26]

$$F(z,n) = \frac{\sum_{\ell=0}^{L} b_\ell(n) z^{-\ell}}{1 + \sum_{k=1}^{K} a_k(n) z^{-k}} \quad \text{where} \quad z = e^{i\omega}. \tag{5}$$

We shall call the filter F a *dynamic filter*. The coefficients $\{a_k, b_\ell\}$ vary with time. Nevertheless, it is assumed that this variation is small enough for the filter to be considered stationary under a small time interval [27]. The difference equation giving the output signal $s(n)$ of the dynamic filter F at nth sample is then expressed by

$$s(n) = \sum_{\ell=0}^{L} b_\ell(n) e(n-\ell) - \sum_{k=1}^{K} a_k(n) s(n-k) \tag{6}$$

where $e(n)$ is the input signal at nth sample.

From a perceptual point of view, the resulting sounds already present the main characteristics of an impacted material. In particular, in the case of strong damping, the sounds present some wooden characteristics and in the case of weak damping, the sounds present some metallic characteristics [28].

B. From the Material to the Object

As previously mentioned, the modal density at high frequency composed of overlapping modes of small amplitudes is reproduced using a filtered white noise. To take into account the perceptual notion of the structure geometry, we need to add only a few modes to the input noisy signal. Concerning the most prominent modes, the simplest approach to generate them consists of using an additive method, by simply adding to the input noise a signal $M(t)$ that consists of a sum of S sinusoids

$$M(t) = \sum_{s=1}^{S} A_s \cos(2\pi f_s t + \Phi_s). \tag{7}$$

The sinusoids are characterized by their amplitudes A_s, frequencies f_s and phases Φ_s, making the model easy to implement and control [Fig. 2(a)]. Nevertheless, this approach suffers from a lack of correlation between the parameters and the physical description of the vibrating structure.

We overcame this lack of correlation by another approach based on physical modeling that uses the digital waveguide concept [29] and especially the banded digital waveguide model [30]–[32] [Fig. 2(b)]. The idea consists in favoring some of the propagative waves by using loops that create stationary waves in the structure. Each resonance is reproduced by a feedback loop composed of a delay line D_s and a bandpass filter B_s (a second-order filter is sufficient in most cases). We call these filters the *modal loop filters*. The delay line filter noted D_s is defined by the following expression:

$$D_s(\omega) = e^{-i\omega d_s} \tag{8}$$

with d_s corresponding to the number of delayed samples (not necessary an integer). The bandpass filter B_s is defined by its transfer function. We chose it as

$$B_s(z) = \gamma_s \frac{c_0 + c_1 z^{-1} + c_2 z^{-2}}{1 + c_3 z^{-1} + c_4 z^{-2}} \tag{9}$$

where the coefficients $\{c_0, c_1, c_2, c_3, c_4\}$ are given by [33]

$$\begin{aligned} c_0 &= \frac{\tau}{1+\tau} \quad c_1 = 0 \quad c_2 = \frac{-\tau}{1+\tau} \\ c_3 &= \frac{-2\cos(2\pi f_c)}{1+\tau} \quad c_4 = \frac{1-\tau}{1+\tau}. \end{aligned} \tag{10}$$

The τ parameter is defined by

$$\tau = \frac{\sin(2\pi f_c)}{2Q} \quad \text{with} \quad Q = \frac{\Delta f}{f_c} \tag{11}$$

where f_c is the central frequency, Δf the frequency bandwidth at -3 dB, and γ_s the maximal amplitude of the filter modulus at frequency f_c. The bandpass filter is centered on the desired frequency in the harmonic spectrum generated by the loop and the delay line. Generally, as many resonances as feedback loops are generated. But by increasing the bandwidth of the filter, several resonances can be produced with only one feedback loop. In this case, the frequency of these resonances depends on the phase of the bandpass filter B_s. With our choice of B_s, the phase is zero only at the central frequency (for the influence of the filter phase on the harmonicity of the generated spectrum, see for example [34] which relates to synthesis of piano string vibrations). The output signal $s(n)$ of the model is written

$$\begin{aligned} s(n) = \sum_{\ell=0}^{L} b_\ell(n) \left(e(n-\ell) + \tilde{e}(n-\ell) \right) \\ - \sum_{k=1}^{K} a_k(n) s(n-k) \end{aligned} \tag{12}$$

with

$$\begin{cases} \tilde{e}(n) = \sum_{s=1}^{S} y_s(n) \\ y_s(n) = \gamma_s \left(c_0 s(n-d_s) + c_1 s(n-d_s-1) \right. \\ \qquad \left. + c_2 s(n-d_s-2) \right) - c_3 y_s(n-1) \\ \qquad - c_4 y_s(n-2) \end{cases} \tag{13}$$

where S is the number of feedback loops.

C. Taking Into Account the Excitator

To control the bandwidth of the generated spectrum, we used an additional bandpass filter (Fig. 2). The response of this filter

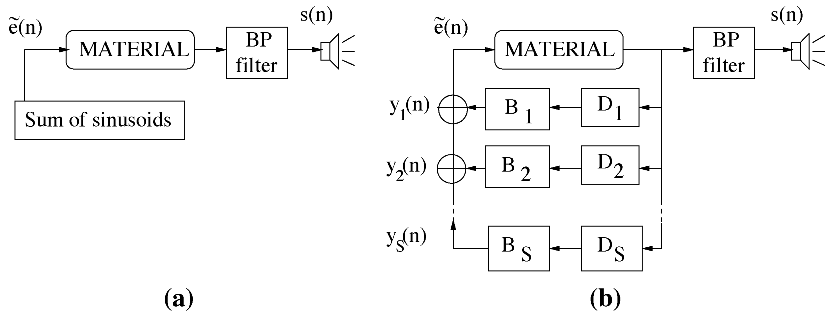


Fig. 2. (a) Sound synthesis model using an additive approach. (b) Sound synthesis model using a banded digital waveguide model approach. An additional bandpass filter is shown to take into account the excitator.

is strongly related to both the strength of the impact and the characteristics of the collision between the excitator and the resonator. In some cases, the response of this filter can be obtained analytically from the physical description of the phenomena. But this aspect is not in the scope of this paper.

D. Simulation Results

We present some simulation results from the two implementations of the synthesis model (using the additive and the banded digital waveguide approaches), for a given set of parameters. The dynamic filter F is chosen as a one pole-one zero filter [which corresponds to the case of $L = 1$ and $K = 1$ in expression (5)]. The dynamic filter is then characterized at each time step by a set of three coefficients $\{a_1, b_0, b_1\}$. Because the dynamic filter F is responsible for the damping of the signal, the temporal evolution of the coefficients was considered an exponentially decreasing function (Fig. 3).

For the additive approach, we chose to add a sum of five sinusoids to the initial input noise. The frequency values and the amplitudes are arbitrarily chosen. For the banded digital waveguide approach, we considered five feedback loops simulating the same modes. In practice, we chose a fourth-order Butterworth filter for the additional bandpass filter associated to the excitator.

From a perceptual point of view, the sounds generated by the two approaches are similar [28]. Fig. 4 shows the temporal synthesis signals and Fig. 5 the corresponding spectra. An important difference between the two approaches is that the additive model always generated the same perceived sound when provided with the same set of parameters. In contrast, the banded digital waveguide model did not generate exactly the same sounds because modes are created by amplifying a narrow band of a stochastic signal. This approach leads to subtle differences between sounds, making the synthesis of repetitive impacts more realistic.

By acting on the evolution curves of the coefficients, one can modify the damping. In addition, by changing the frequency values of the most prominent modes, one can modify the perception of the structure dimensions. In particular, the higher the frequencies, the smaller the structure is perceived to be.

IV. From Real Sounds to Synthesis

In the context of analysis-synthesis, synthesis parameters can be estimated from the analysis of natural sounds. In this section, we present an analysis method for extracting the damping law (characteristic of the material) and the parameters of the most prominent modes, which we call *modal parameters* (characteristic of the structure geometry). We then establish relations between the synthesis parameters and the estimated physical parameters.

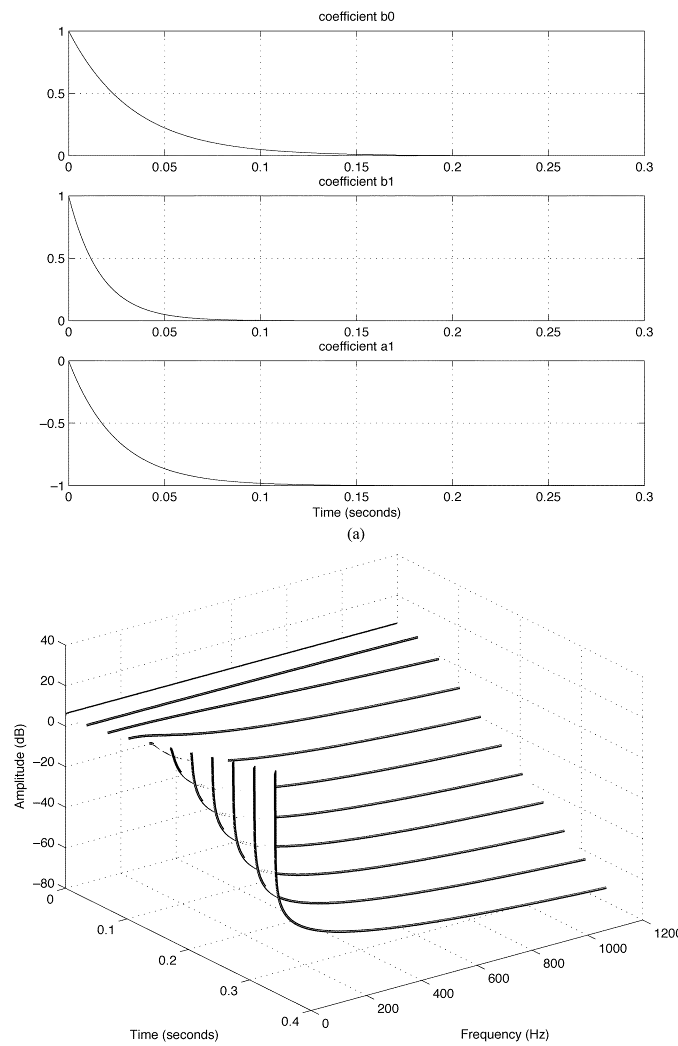


Fig. 3. Time evolution of the coefficients of the dynamic filter (a): evolution of the coefficients b_0 (on the top), b_1 (on the middle), and a_1 (on the bottom). They are arbitrarily considered exponentially decreasing functions. Time evolution of the dynamic filter (b) corresponding to this set of coefficients.

A. Analysis of Transient Sounds

1) Time-Scale Representation: The transient feature of the sounds we consider calls for using joint representations to pre-

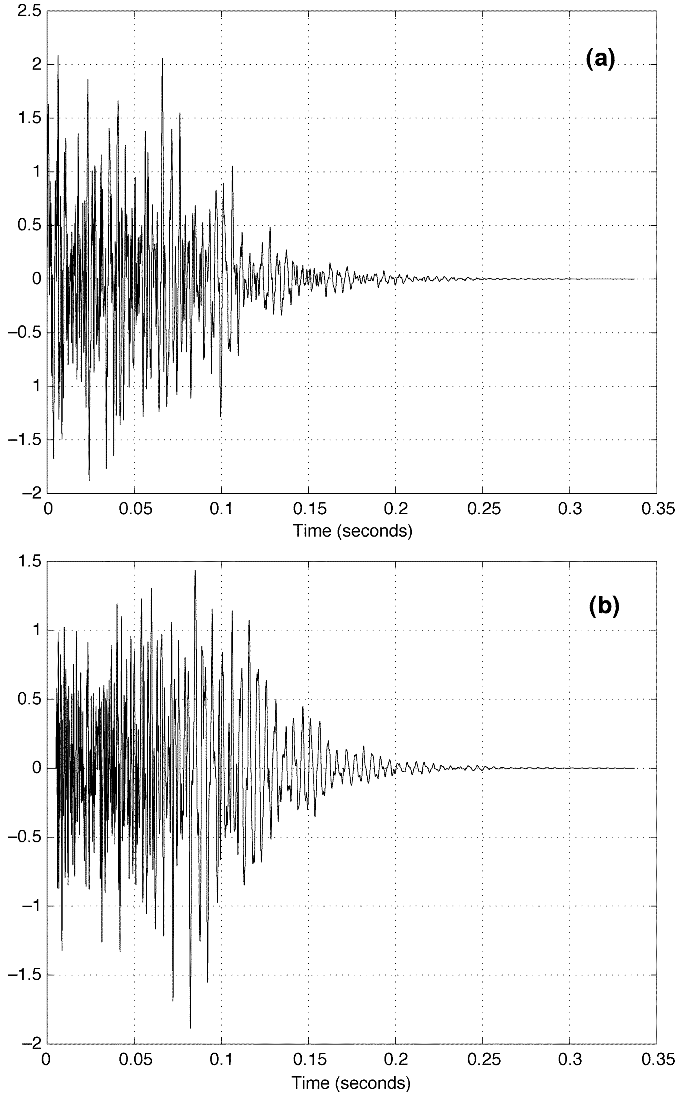


Fig. 4. One temporal realization of the synthesis process obtained by (a) the model based on the additive approach and (b) the model based on the banded digital waveguide approach.

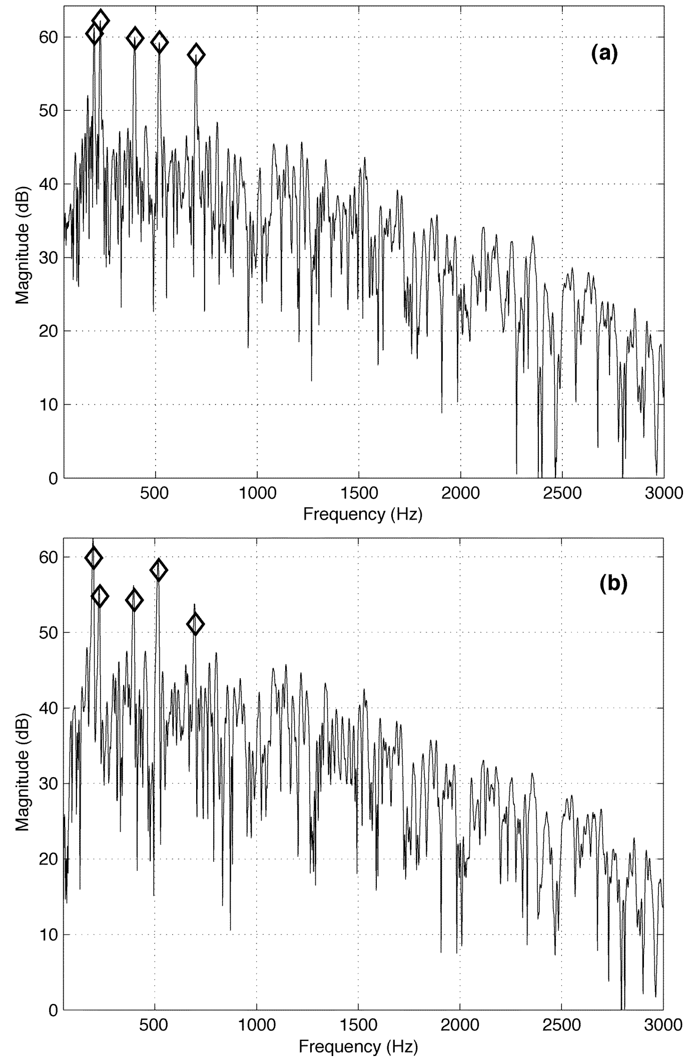


Fig. 5. Spectra of synthesis signals represented on Fig. 4 obtained by (a) the model based on the additive approach and (b) the model based on the banded digital waveguide approach as a function of frequency. The most prominent modes added are marked by an $\diamondsuit$.

cisely describe their most relevant perceptual characteristics. For that, we propose analysis methods based on time-scale decomposition (wavelets). Such methods consist in decomposing a signal in terms of contributions that are localized in the time and the scale domains. Because of the constant relative bandwidths, this approach provides a representation that is consistent with what one hears [35]. The coefficients $\widehat{S}(b,a)$ of such a representation are given by

$$\widehat{S}(b,a) = \int_{-\infty}^{+\infty} s(t)\bar{g}_{b,a}(t)dt \tag{14}$$

where $s(t)$ is the temporal signal, and $g_{b,a}(t)$ is called the analysis wavelet. Here $\bar{g}$ denotes to the complex conjugate of g. It is of finite energy, of zero mean, and is obtained from a so-called *mother wavelet* $g(t)$ by the expression

$$g_{b,a}(t) = \frac{1}{\sqrt{a}}g\left(\frac{t-b}{a}\right) \text{ with } a \in \mathbb{R}^{*+} \text{ and } b \in \mathbb{R} \tag{15}$$

where $\mathbb{R}$ represents the set of real numbers and $\mathbb{R}^{*+}$ the set of strictly positive real numbers. We chose a slightly modified version of the wavelet transform in which the wavelets are constructed to mimic the response bandwidths of the basilar membrane at various locations [36]. This approach leads to what we call the time-scale Bark representation. Among the approximations of the Bark-scale function, we chose the correspondence between frequency and Bark-scale given by the relation [37]

$$B = 13\arctan(0.76f) + 3.5\arctan\left(\frac{f}{7.5}\right)^2 \tag{16}$$

where f is the frequency in kilohertz and B the Bark value. The time-scale Bark representation is obtained by decomposing the signal by a filter bank (Fig. 6). We chose the responses $R_k(\omega)$ to be given by Gaussian functions located at the central frequency f_k corresponding to the Bark scale (integer values of B). The bandwidth of each filter is designed to minimize information loss.

2) Estimation of the Damping Law: The damping law characterizing the material is estimated from the time-scale Bark representation. In each subband, we first calculate the associated analytical signal. This signal gives a complex representation of the subband signal, leading to easy estimation of both the instantaneous frequency and the amplitude law. As already

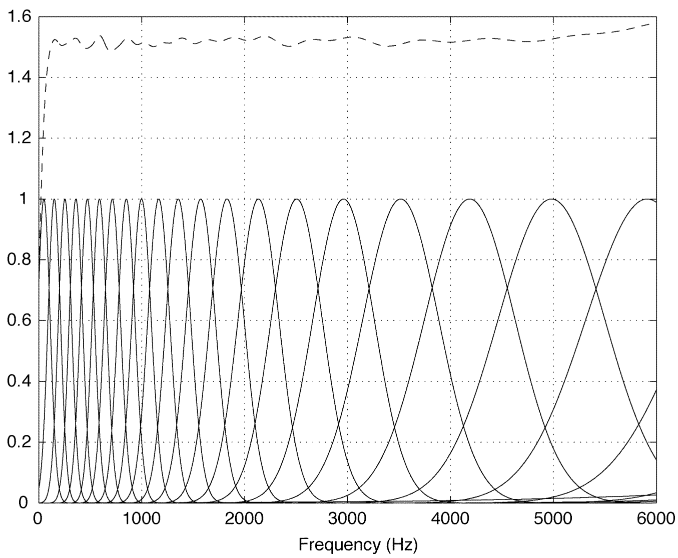


Fig. 6. Filter bank composed of Gaussian functions located at the central frequency corresponding to the Bark scale. The bandwidth of each filter is designed to minimize information loss. The dashed curve represents the sum of the modulus of the responses of the filter bank.

mentioned, one can assume that the time evolution of the modes decreases exponentially. Hence, we fit each subband amplitude law with an exponential function $\Lambda e^{-\alpha t}$. The damping law is then characterized by Λ and α in each Bark subband.

3) Estimation of the Modal Parameters: To determine the amplitudes and the eigenfrequencies of most prominent modes, many analysis methods are available. Among the so-called parametric methods, the Prony method [38] (or Steiglitz-Mac Bride method [39]) consists of considering a given signal as a sum of exponentially damped sinusoids. In the case of many closely spaced modes, these methods need high order filters. To solve this problem, Karjalainen *et al.* have proposed a subband method [40]. As one needs to estimate only the frequency and amplitude values (the dynamic filter ensures the damping of the modes), the estimation process is simplified and very high order filters do not need to be considered. Starting from initial values, the optimal amplitude and frequency values are obtained by minimizing the quadratic error between the measured and the theoretical signals. Otherwise, among the so-called nonparametric methods, the Fourier transform can be used to determine the amplitudes and eigenfrequencies of most prominent modes. We used both methods, but note that parametric techniques are better adapted to the separation of close components.

B. Synthesis Parameter Estimation

In this section, we describe how the parameters of the synthesis model can be estimated from the physical ones that were extracted by the analysis methods previously described. In particular, we propose to estimate the dynamic filter coefficients from the damping law and the modal loop filter coefficients from the modal parameters.

1) Estimation of the Dynamic Filter Coefficients: The dynamic filter F is based on an IIR filter structure. In practice, we used a first-order filter, the response of which is given by

$$F(z,n) = \frac{b_0(n)(1+z^{-1})}{1+a_1(n)z^{-1}} \quad \text{with} \quad z = e^{i\omega}. \tag{17}$$

The expression is simplified. Only the two coefficients $\{a_1, b_0\}$ are needed because the modulus of F is assumed to be zero at the Nyquist frequency. Higher order filters can be considered (for example to simulate complex and not monotonous evolution of the signal), but a first-order filter was sufficient for most of our applications in impact sound synthesis. The coefficient values are expected to be time-varying and must be calculated at regular time intervals. As the filter F reproduces the damping, we chose to estimate the filter coefficients so that the damping is kept invariant in three specific Bark intervals. A first interval corresponds to the one containing the energy spectral centroid G_1 defined by

$$G_1 = \frac{\int_0^{\omega_{\max}} \omega |\hat{s}(\omega)|^2 d\omega}{\int_0^{\omega_{\max}} |\hat{s}(\omega)|^2 d\omega} \tag{18}$$

where $\hat{s}(\omega)$ corresponds to the Fourier transform of the signal. The upper bound of the integral $\omega_{\max}$ corresponds to the higher frequency of the relevant part of the spectrum. In practice, we define the frequency $\omega_{\max}$ as the frequency beyond which the spectrum envelope decreases by 60 dB from the maxima. $\hat{s}(\omega)$ corresponds to the Fourier transform of the signal. The two other specific Bark intervals in which the damping coefficient is kept invariant correspond to the bands that contain the spectral centroids G_2 and G_3 of the domains on either side of G_1. This division is closely linked to the concept of tristimulus [41], [42].

At each time step Δt, the unknown coefficients $\{a_1, b_0\}$ are estimated by minimizing the criterion $\mathcal{C}_1$ defined by the following expression:

$$\mathcal{C}_1 = \frac{1}{2}\sum_{k=1}^{3}\left(|\delta_k| - |\tilde{\delta}_k|\right)^2 \tag{19}$$

where

$$\begin{cases} \delta_k = |F(z_k, t_0)| - |F(z_k, t_1)| \\ \tilde{\delta}_k = \Lambda(G_k)\left(e^{-\alpha(G_k)t_0} - e^{-\alpha(G_k)t_1}\right) \end{cases} \tag{20}$$

where $z_k = e^{iG_k}$, $\Delta t = |t_0 - t_1|$, and Λ the "gain" of the damping law (see Section IV-A2). After the minimization process, one obtains the optimal set of coefficients characterizing the dynamic filter. To ensure the filter is stable, a bound constraint is added in the minimization process so that the values of the coefficient a_1 satisfy the condition [43]

$$|a_1| < 1. \tag{21}$$

2) Estimation of Modal Loop Filter Parameters: In this section, we describe how the parameters of the modal loop filters are estimated for additive and banded digital waveguide approaches. In both cases, one must precisely define the emergence ratio of the modes with respect to the "smooth" part of the spectrum. This definition leads to selecting the modes of greater amplitudes and considering the rest of the spectrum as the "noise contribution" (taken into account by the material model). One could also take into account masking phenomena occurring between modes and exclude the ones that are inaudible [44], but this aspect is not in the scope of this paper.

For the additive approach, the modal parameters are directly obtained from the spectral representation of the signal (see Section IV-A) by taking into account the amplitude correction due to the damping. For the banded digital waveguide approach, the estimation process is different. For simplicity's sake and to avoid the formalism dealing with the nonstationary behavior of the signal, we use the concept of *average filter*. This filter, noted $\widetilde{F}(\omega)$, corresponds to a filter with a frequency response that coincides with the power spectral density (PSD) of a white noise filtered by the dynamic filter. Determining the average filter is of great importance since it allows us to assume the system is time invariant. This assumption was valid for most sounds on which we worked, since the short duration of the sounds, along with the regular behavior of the damping, allowed the noisy part of the signal to be well characterized by its PSD. One thus reduces the model to one that is time invariant, now fully characterized by a transfer function.

For the filter $D_s(\omega)$, the parameter d_s is determined from the value of the frequency ω_s of the mode

$$d_s = \frac{\Phi_{\widetilde{F}}(\omega_s) - 2p\pi}{\omega_s} \quad \text{with} \quad p \in \mathbb{Z} \tag{22}$$

where $\Phi_{\widetilde{F}}$ is the phase of the filter $\widetilde{F}(\omega)$. $\mathbb{Z}$ represents the set of relative integers. The integer p is defined so that the parameter d_s is positive and corresponds to the smaller delay.

The filter $B_s(\omega)$ is defined from the expression (9). One then estimates the central frequency f_s, the frequency bandwidth Δf, and the amplitude γ_s. The central frequencies of all the S filters are given by the frequency mode values. The frequency bandwidths are arbitrarily set between 1 and 5 Hz to ensure only one component is generated. Nevertheless, a bigger bandwidth could be used if more components are needed. The set of optimal amplitudes $\gamma_{s=1,\ldots,S}$ of all the filters $B_s(\omega)$ is obtained by minimizing the criterion $\mathcal{C}_2$

$$\mathcal{C}_2 = \frac{1}{2}\sum_{s=1}^{S}\left(\left|H(\gamma_1,\ldots,\gamma_S,\omega_s)\right| - \left|\widetilde{H}(\omega_s)\right|\right)^2. \tag{23}$$

Here, $\widetilde{H}(\omega)$ is the Fourier transform of the measured signal and $H(\gamma_1,\ldots,\gamma_S,\omega)$ is the transfer function of the synthesis model reduced to one that is time invariant

$$H(\gamma_1,\ldots,\gamma_S,\omega) = \frac{\widetilde{F}(\omega)}{1 - \widetilde{F}(\omega)\sum\limits_{s=1}^{S}\left(B_s(\gamma_s,\omega)D_s(\omega)\right)}. \tag{24}$$

C. Examples of Analysis-Synthesis

1) Impact Sounds on Thin Plates: We address the problem of resynthesizing sounds produced by impacted thin plates. These sounds were experimentally obtained by recording the vibrations of rectangular thin plates of different materials [28].

Fig. 7 shows the time-scale Bark representations of the signals obtained with wood and steel materials. Fig. 8 shows the damping laws corresponding to the wood, the glass, and the steel plates as a function of the frequency. As expected, they were frequency-dependent. The damping laws were estimated on the relevant part of the sound spectrum for which the energy is significant. Yet, note that the regularity of the curves thereby obtained allows pertinent extrapolation along the frequency axis.

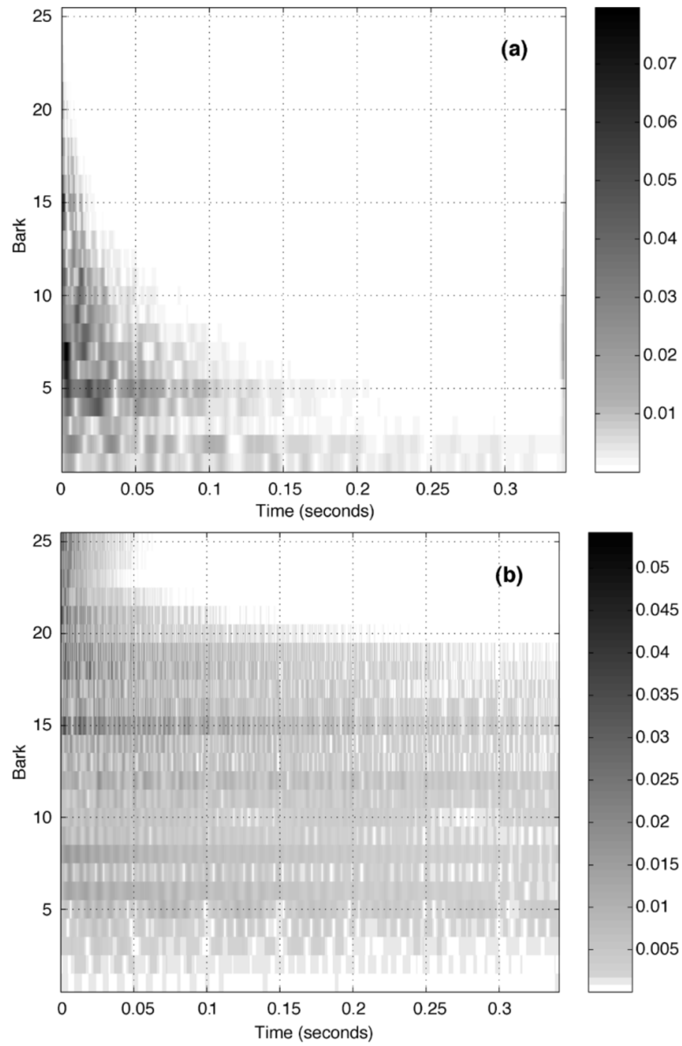


Fig. 7. Time-scale Bark representations corresponding to the beginning of signals from thin wood plate (a) and a thin steel plate (b).

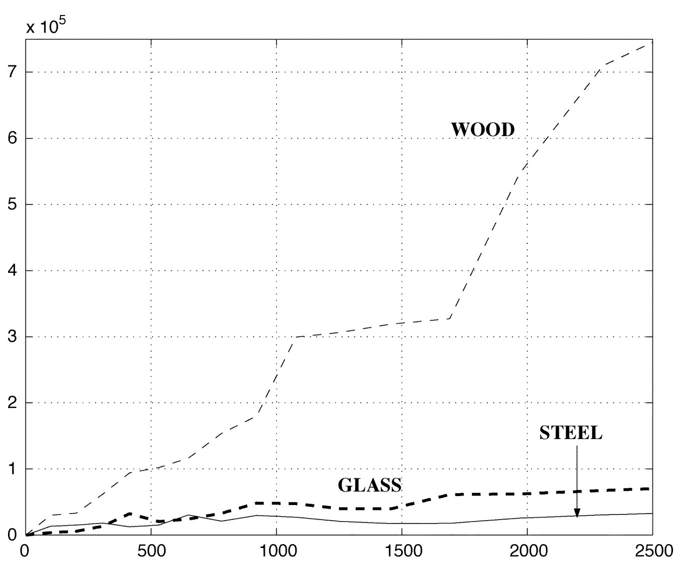


Fig. 8. Frequency-dependent damping laws $\alpha(\mathrm{f})$ for the wood, glass, and steel as a function of frequency. They are estimated from time-scale Bark representations.

Fig. 9 represents the evolution of the coefficients $\{a_1, b_0\}$ of the dynamic filter corresponding to the wood, the glass, and the

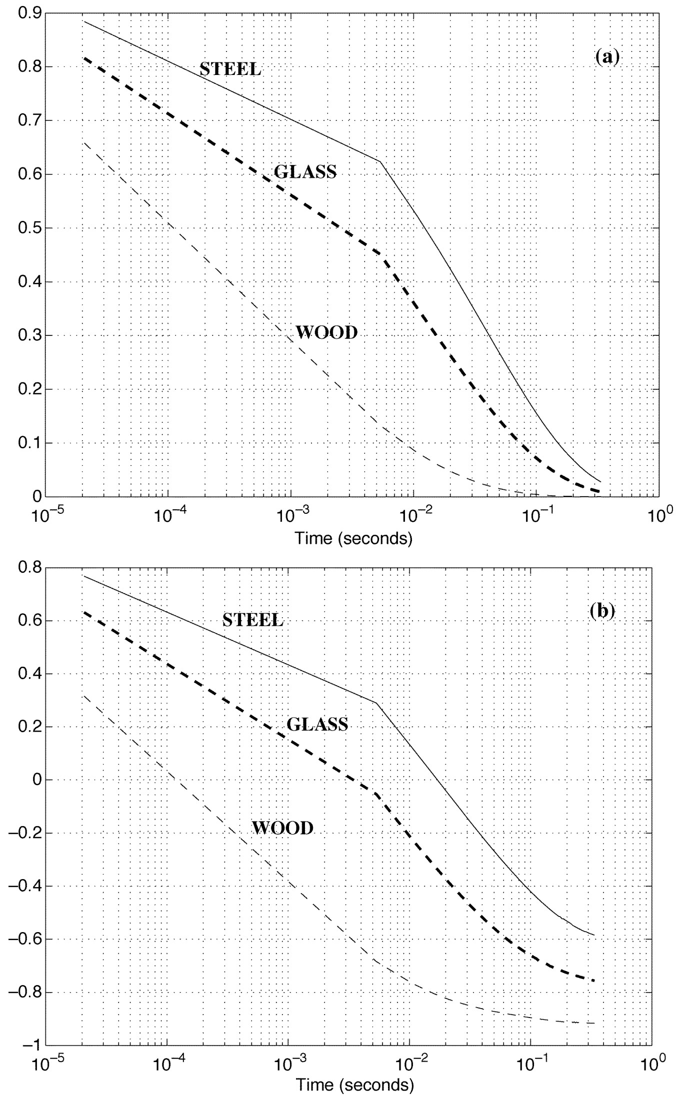


Fig. 9. Evolution of the coefficients b_0 (a) and a_1 (b) of the dynamic filter for the wood, steel, and glass material. Zoom on the first milliseconds of the signal.

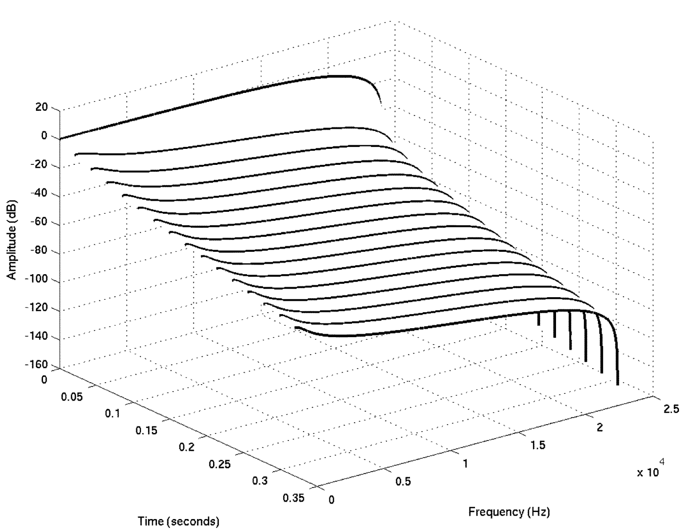


Fig. 10. Time evolution of the dynamic filter (decibel scale) as a function of frequency for the wood material.

steel as a function of time. Figs. 10 and 11 represent the time evolution of the frequency response of the dynamic filter for

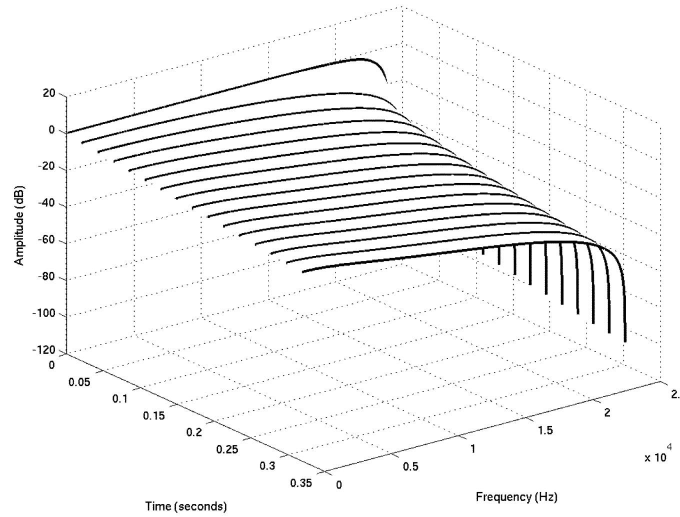


Fig. 11. Time evolution of the dynamic filter (decibel scale) as a function of frequency for the steel material.

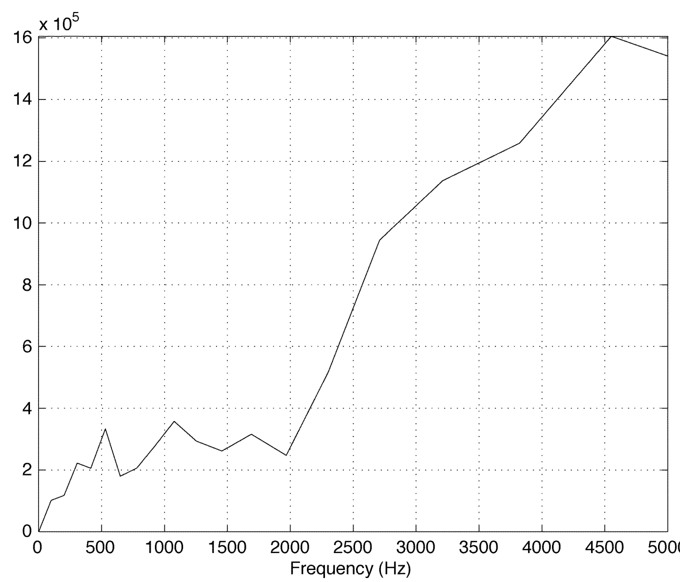


Fig. 12. Estimation of the damping law $\alpha(f)$ as a function of frequency from the experimental signal of an impulse response of a piano soundboard.

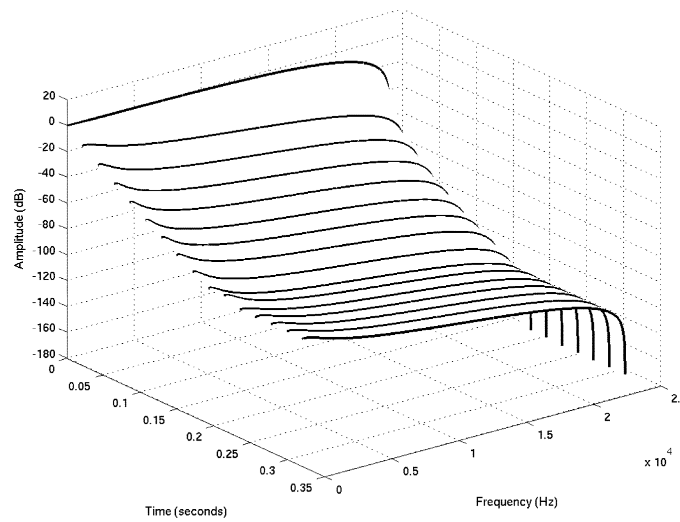


Fig. 13. Time evolution of the dynamic filter as a function of frequency, simulating the damping law shown in Fig. 12, which corresponds to the impulse response of a piano soundboard.

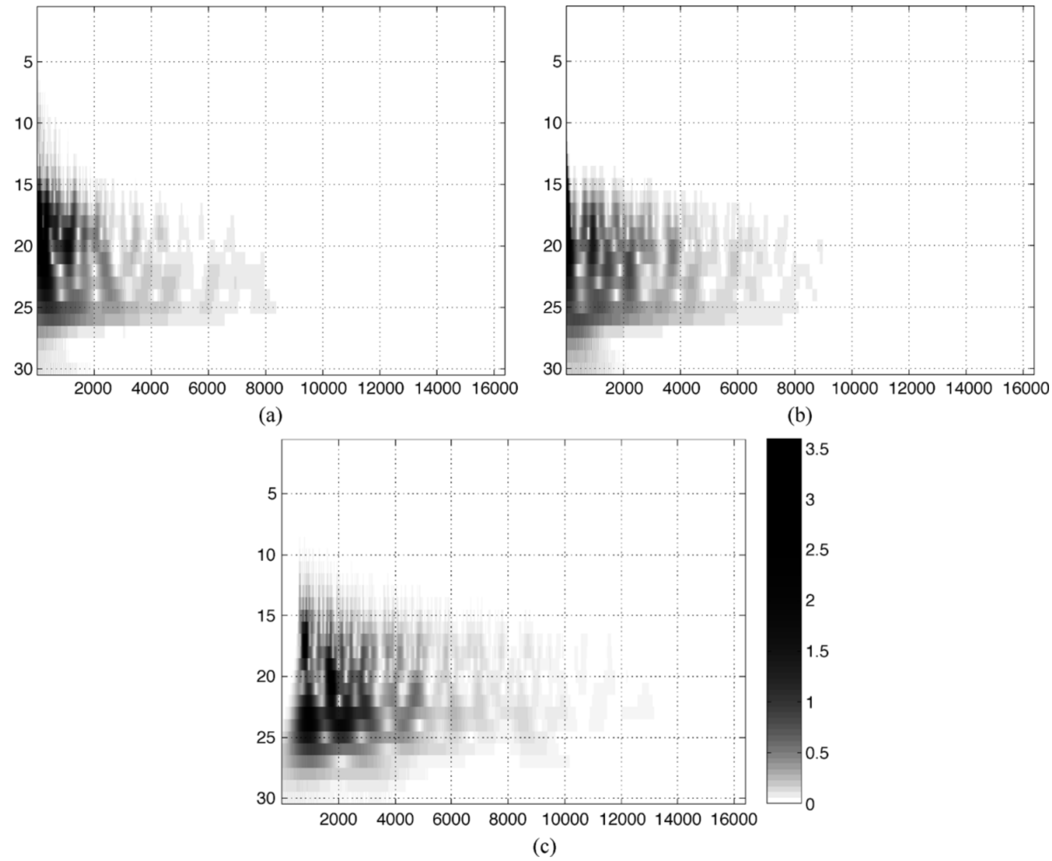


Fig. 14. Time-scale representations of the measured signal (a), the synthesis signal obtained by the additive approach (b), and the synthesis signal obtained by the banded digital waveguide approach (c) as a function of samples (x axis) and scale parameter (y axis).

wood and steel materials. The filters become increasingly low-pass, meaning that high frequency components decrease faster than low frequencies. These results agree well with the physics of materials, which states that high frequency modes are generally more heavily damped than low frequencies. Actually, the dissipation of vibrating energy due to the coupling of the structure and the air increases with frequency (see for example [45]). From a perceptual point of view, the results are consistent with the fact that wood is perceived as a less resonant material than steel. Consequently, a sound generated by a wooden structure is generally shorter than a sound generated by a steel structure.

Concerning the influence of the most prominent modes, we checked their relevance in the perception of the structure dimensions. According to the physics, our perception is sensitive to the frequency of the most prominent modes: the higher they are, the smaller the structure is thought to be.

2) Impulse Response of a Piano Soundboard: A problem in piano sound synthesis is to reproduce the radiation of the instrument. This radiation is produced by the soundboard, which is difficult to model from a physical point of view. We have resynthesized the impulse response of a piano soundboard through our synthesis model, whose parameters were estimated from the analysis of the sound obtained experimentally. This model will be further linked to an existing synthesis model of hammer-string interaction [34], [46] using the Commuted Synthesis concept [12], [13].

Fig. 12 shows the extracted damping law and Fig. 13 shows the time evolution of the modulus of the dynamic filter simulating this damping law. The dynamic filter behaves like a low-pass filter. Concerning the contribution of the most prominent modes, we took into account only the ten modes of greater amplitudes. The rest of the modes are generated by the initial white noise. Consequently, in the additive approach, we considered a sum of ten sinusoids and in the banded digital waveguide approach, ten feedback loops (one loop per mode).

From a perceptual point of view, the global behavior of the original signal is conserved and the synthesis sounds are similar [28]. Fig. 14 shows the time-scale representations of the measured signal, one realization of the synthesis signal obtained by the additive approach, and one realization of the synthesis signal obtained by the banded digital waveguide approach.

V. Conclusion—Current Works

The synthesis of impact sounds plays an important role in virtual reality and musical applications. Rather than a precise sound signal, one seeks a well-perceived sound. The perception of sound correlates with the nature of the material, the geometry of the structure, and the type of impact. Consequently, impact sounds have specific relations among their components: the damping of the components is mainly linked to the material, the most prominent components are mainly linked to the geometry of the structure, and the type of impact influences the spectral bandwidth. Constructing such relations between the components is analogous to making a tapestry from odd bits of yarn. Actually, this metaphor fits the synthesis model we presented since it mainly consists in "reorganizing" the relations between incoherent components provided by a white noise generator. For that purpose, a dynamic filter proved well adapted. We have then shown that only few modal components are needed to perceptually simulate an impact on a given structure: the global material information is contained in the dynamic filter output that takes into account the damping of the sound components. The model was very efficient for synthesis as well as for resynthesis purposes. Actually, through perceptually based analysis, the synthesis parameters can be estimated from natural sounds, making it easy to build a database containing the model parameters for various materials. Even though this paper focuses on the design and the efficiency of the synthesis model, note that it is real-time compatible. An implementation using the MAX/MSP [1] software allowed us to widely explore the field of impact sound synthesis and transformations and also opened new directions of research. We currently use the model to better understand the relations between physical parameters (damping, most prominent modes) and sound categorization, as well as the way the brain processes the information (using Event Related brain Potentials measurements). From a synthesis point of view, we expect the results to help in determining the variation of the synthesis parameters for sounds within a given material category. Sound transformations such as continuous morphing between different materials [28] lead to valuable stimuli for such psychophysiological experiments. We further expect the model to be extended to the case of complex excitations such as stacking, rubbing, and scratching.

REFERENCES

[1] J. M. Adrien, "The missing link: Modal synthesis," in *Representations of Musical Signals*, G. D. Poli, A. Piccialh, and C. Roads, Eds. Cambridge, MA: MIT Press, 1991, ch. 8, pp. 269–297.

[2] P. R. Cook, *Real Sound Synthesis for Interactive Applications*. Natick, MA: Peters, 2002.

[3] D. Rocchesso, R. Bresin, and M. Femström. Sounding objects. [Online]. Available: http://www.soundobject.org/papers/sob.pdf.

[4] ——, Sounding objects. [Online]. Available: http://www.soundobject.org/SObBook/SObBook_JUL03.pdf.

[5] ——, "Sounding objects," *IEEE Multimedia*, vol. 10, no. 2, pp. 42–52, 2003.

[6] K. van den Doel, P. G. Kry, and D. K. Pai, "Foleyautomatic: Physically based sound effects for interactive simulation and animation," in *Proc. SIGGRAPH 2001*, Aug. 2001, pp. 537–544.

[7] D. K. Pai, K. van den Doel, D. L. James, J. Lang, J. E. Lloyd, J. L. Richmond, and S. M. Yau, "Scanning physical interaction behavior of 3-D objects," in *Proc. SIGGRAPH 2001*, Aug. 2001, pp. 87–96.

[8] A. Chaigne, C. Touzé, and O. Thomas, "Nonlinear axisymmetric vibrations of gongs," in *Proc. Int. Symp. Musical Acoustics*, Perugia, Italy, 2001.

[9] B. Bank and L. Sujbert, "Modeling the longitudinal vibration of piano strings," in *Proc. SMAC Conf.*, Stockholm, Sweden, 2003.

[10] J. Bensa and L. Daudet, "Efficient modeling of "phantom" partials in piano tones," in *Proc. Int. Symp. Musical Acoustics (ISMA2004)*, Nara, Japan, 2004.

[11] P. R. Cook, "Modeling Bill's Gait: Analysis and parametric synthesis of walking sounds," in *Proc. Audio Engineering Soc. 22nd Int. Conf. Virtual, Synthetic and Entertainment Audio*, Helsinki, Finland, 2002.

[12] J. O. Smith and S. A. van Duyne, "Commuted piano synthesis," in *Proc. Int. Computer Music Conf.*, Banff, Canada, 1995, pp. 319–326.

[13] ——, "Developments for the commuted piano," in *Proc. Int. Computer Music Conf.*, Banff, Canada, 1995.

[14] W. W. Gaver, "EveryDay Listening and Auditory Icons," Ph.D. dissertation, Univ. California, San Diego, 1988.

[15] R. L. Klatzky, D. K. Pai, and E. P. Krotkov, "Perception of material from contact sounds," *Presence*, vol. 9, no. 4, pp. 399–410, 2000.

[16] S. Tucker and G. J. Brown, "Investigating the Perception of the Size, Shape and Material of Damped and Free Vibrating Plates," Dept. Comput. Sci., Univ. Sheffield, Sheffield, U.K., Tech. Rep. CS-02-10, 2002.

[1] see http://www.cycling74.com/index.html.

[17] B. Truax, *Handbook for Acoustic Ecology*. Burnaby, BC, Canada: Simon Fraser Univ., 1999.

[18] R. P. Wildes and W. A. Richards, "Recovering material properties from sounds," *Nat. Comput.*, pp. 356–363, 1988.

[19] C. A. Wert, "Internal friction in solids," *J. Appl. Phys.*, vol. 60, no. 6, pp. 1888–1895, 1986.

[20] A. Chaigne and C. Lambourg, "Time-domain simulation of damped impacted plates: I. Theory and experiments," *J. Acoust. Soc. Amer.*, vol. 109, no. 4, pp. 1422–1432, 2001.

[21] A. Chaigne, C. Lambourg, and D. Matignon, "Time-domain simulation of damped impacted plates: II. Numerical models and results," *J. Acoust. Soc. Amer.*, vol. 109, no. 4, pp. 1433–1447, 2001.

[22] P. Djoharian, "Material design in physical modeling sound synthesis," *J. New Music Res.*, vol. 30, no. 3, pp. 227–241, 2001.

[23] H. Kuttruff, *Room Acoustics*. New York: Elsevier, 1991.

[24] C. Carello, K. L. Anderson, and A. J. Kunkler-Peck, "Perception of object length by sound," *Psychol. Sci.*, vol. 9, no. 3, pp. 211–214, 1998.

[25] P. R. Cook, *Real Sound Synthesis for Interactive Applications*. Natick, MA: Peters, Jul. 2002, pp. 124–127.

[26] T. Y. Leou and J. K. Aggarwall, "Recursive implementation of LTV filters frozen-time transfer function versus generalized transfer function," *Proc. IEEE*, vol. 72, no. 7, Jul. 1984.

[27] J. N. Mourjopoulos, E. D. Kyriakis-Bitzaros, and C. E. Goutis, "Theory and real-time implementation of time-varying digital audio filters," *J. Audio Eng. Soc.*, vol. 38, no. 7/8, pp. 523–536, 1990.

[28] M. Aramaki and R. Kronland-Martinet. Sound Examples Illustrating the Article. [Online]. Available: http://www.1ma.cnrs-mrs.fr/~kronland/IEEE_SAP/sounds.html.

[29] J. O. Smith, "Physical modeling using digital waveguides," *Comput. Music J.*, vol. 16, no. 4, pp. 74–87, 1992.

[30] G. Essl and P. R. Cook, "Banded waveguides: Toward physical modeling of bowed bar percussion instruments," in *Proc. Int. Comput. Music Conf.*, Beijing, China, 1999, pp. 321–324.

[31] S. Serafin, P. Huang, S. Ystad, C. Chafe, and J. O. Smith, "Analysis and synthesis of unusual friction driven musical instruments," in *Proc. Int. Comput. Music Conf.*, Göteborg, Sweden, 2002, pp. 57–60.

[32] G. Essl, S. Serafin, P. R. Cook, and J. O. Smith, "Theory of banded waveguides," *Comput. Music J.*, vol. 28, no. 1, pp. 37–50, 2004.

[33] R. Bristow-Johnson. Cookbook formulae for audio EQ biquad filter coefficients. [Online]. Available: http://www.harmony-central.com/Computer/Programming/Audio-EQ-Cookbook.txt.

[34] M. Aramaki, J. Bensa, L. Daudet, P. Guillemain, and R. Kronland-Martinet, "Resynthesis of coupled piano string vibrations based on physical modeling," *J. New Music Res.*, vol. 30, no. 3, pp. 213–226, 2002.

[35] R. Kronland-Martinet, "The wavelet transform for analysis, synthesis, and processing of speech and music sounds," *Comput. Music J.*, vol. 12, no. 4, pp. 11–20, 1988.

[36] B. C. J. Moore, *An Introduction to the Psychology of Hearing*. London: Academic Press Inc., 1982.

[37] E. Zwicker and H. Fastl, *Psychoacoustics, Facts and Models*. New York: Springer-Verlag, 1990.

[38] J. Laroche, "Etude d'un système d'analyse et de synthèse utilisant la méthode de Prony—Application aux instruments de musique de type percussif," Ph.D. dissertation, Ecole Nationale Supérieure des Télécommunications, Paris, France, 1989.

[39] K. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-10, pp. 461–464, 1965.

[40] M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Makiviria, and V. Välimäki, "Frequency-zooming ARMA modeling of resonant and reverberant systems," *J. Audio Eng. Soc.*, vol. 50, no. 12, pp. 1012–1029, 2002.

[41] H. F. Pollard and E. V. Jansson, "A tristimulus method for the specification of musical timbre," *Acustica*, vol. 51, pp. 162–171, 1982.

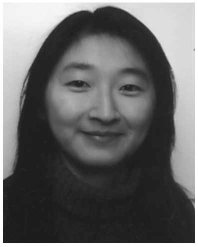
[42] S. Ystad, "Sound modeling applied to flute sounds," *J. Audio Eng. Soc.*, vol. 48, no. 9, pp. 810–825, 2000.

[43] J. O. Smith, "An introduction to digital filter theory," in *Digital Audio Signal Processing, An Anthology*, J. Strawn, Ed., 1983, ch. 2, pp. 69–135.

[44] K. van den Doel, D. K. Pai, T. Adam, L. Kortchmar, and K. Pichora-Fuller, "Measurements of perceptual quality of contact sound models," in *Proc. 2002 Int. Conf. Auditory Display*, Kyoto, Japan, 2002.

[45] A. Caracciolo and C. Valette, "Damping mechanisms governing plate vibration," *Acta Acustica*, vol. 3, pp. 393–404, 1995.

[46] J. Bensa, K. Jensen, and R. Kronland-Martinet, "A hybrid resynthesis model for hammer-strings interaction of piano tones," *EURASIP J. Appl. Signal Process.*, vol. 2004, no. 7, pp. 1021–1035, 2004.



Mitsuko Aramaki received the master degree in mechanics (speciality in acoustic and dynamics of vibrations) in 1999 from the University of Aix-Marseille II, Marseille, France, and the Ph.D. degree in 2003 for her work at the Laboratoire de Mécanique et d'Acoustique, Marseille, France, on analysis and synthesis of impact sounds using physical and perceptual approaches.

She currently is a Postdoctoral Researcher at the Institut de Neurosciences Cognitives de la Méditerranée, Marseille, where she works on a pluridisciplinary project combining cognitive aspects of the timbre perception, physical modeling of sounds, and neurosciences methods.

Richard Kronland-Martinet received the Ph.D. degree in acoustics from the University of Aix-Marseille II, Marseille, France, in 1983, and the Doctorat d'Etat es Sciences degree in 1989 for his work on analysis and synthesis of sounds using time-frequency and time-scale representations.

He is currently Director of Research at the National Center for Scientific Research (CNRS), Laboratoire de Mécanique et d'Acoustique, Marseille, where he is the Head of the group Modeling, Synthesis and Control of Sound and Musical Signals. His primary research interests are in analysis and synthesis of sounds with a particular emphasis on musical sounds. He has recently been involved in a multidisciplinary research project associating sound synthesis processes and brain imaging techniques to better understand the way the brain is processing sounds and music.