ANALYSIS-SYNTHESIS OF IMPACT SOUNDS

Mitsuko Aramaki and Richard Kronland-Martinet

CNRS-Laboratoire de Mécanique et d'Acoustique,
31, chemin Joseph Aiguier, 13402 Marseille Cedex 20, France
phone: +33 491 16 42 50, fax: +33 491 16 40 12, email: {aramaki,kronland}@lma.cnrs-mrs.fr
web: www.lma.cnrs-mrs.fr/mscssm.htm

ABSTRACT

This article presents a sound synthesis model reproducing impact sounds from a perceptual point of view. We mainly address here the synthesis of sounds generated by impacting complex structures for which classical techniques are not well adapted. In the context of analysis-synthesis, the synthesis parameters are directly estimated from the analysis of real sounds.

1. INTRODUCTION

Impact sounds and more generally transient sounds, still pose problems for what the synthesis is concerned. They are important for both musical synthesis (percussive instruments) and audio virtual reality applications (in animation movies to illustrate the bang on a door or impact sounds on objects made with different materials, ...). Recent works have shown the existence of perceptual clues allowing the recognition of impact sounds only by the listening. The synthesis model presented in this study is based on the reproduction of the two main clues, corresponding to the material and the geometry of the structure. In particular, the sound damping is simulated by a time varying filtering process. In the context of analysis-synthesis, the synthesis parameters, which are correlated to the physical parameters such as the damping law and the eigen modes of the structure, are directly extracted from the analysis of real sounds. This leads to an analysis-synthesis model that will be illustrated by two examples: sounds of impacted thin plates and piano soundboard response.

2. WHAT IS AN IMPACT SOUND?

The term *impact sound* can be used to describe many situations. We adopt here a conceptual definition close to the one given in [17]. In the time domain, an impact sound is a short duration sound not lasting more than a few seconds and characterized by an abrupt onset and a short decay. Typically, it corresponds to the vibratory response of a given structure under free oscillations after been excited by an impact, or to the sound produced by the collision of objects. As a consequence, the spectral content of such a sound generally is broadband.

Synthesis methods based on physical modeling are very efficient to simulate the sound produced by simple vibrating structures for which the classical theory gives analytical solutions. Nevertheless, for complex structures, these methods are not adapted since they require a precise description of the structure itself and lead to complicated physical models. To go beyond this limitation, we look for a sound synthesis model the aim of which is to reproduce the main perceptual features of impact sounds.

Many hearing tests have shown the existence of perceptual clues allowing the recognition of impact sound only by the listening [5] [6] [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):

Material: The perception of the material is mainly correlated to the damping coefficient of the eigen modes of the vibrating object [18]. This coefficient generally is frequency dependent and can be related to the internal friction coefficient which is eventually considered as the most important characteristic of the material.

Dimensions of the structure: For what the geometry is concerned, the size is mainly perceived by the detection of the emergent spectral components associated to the modes of the structure [3]. The frequencies of these components correspond to the so-called eigen frequencies of the structure and are deduced for simple cases from the movement equation. Especially, for multidimensional structures, the modal density increases with respect to the frequency so that the modes overlap and become indiscernible at high frequency. We shall then consider that the spectral content is composed of some emergent modes related to the geometry at low frequency and an overlapping modal density at high frequency.

3. SOUND SYNTHESIS MODEL

In this section, we propose a sound synthesis model aiming at simulating impact sounds from a perceptual point of view. This model is based on the reproduction of the two main contributions, in particular those corresponding to the material and the dimensions of the structure. We propose to model separately these two contributions (material and structure dimensions) even if they can't be totally disconnected from a physical point of view.

3.1 Material contribution

As mentioned in the previous section, 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 proposed by Van Duyne and Smith [13] [14] consisting in simulating the damping by a time-varying filtering process acting on the input signal. The time varying filter, noted F, is based on an IIR structure

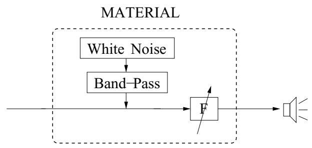

Figure 1: Synthesis model reproducing the main characteristics of the material.

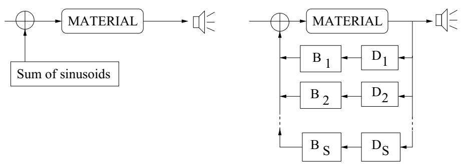

Figure 2: **On the left:** Sound synthesis model using an additive approach. **On the right:** Sound synthesis model using a digital waveguide model approach.

(Infinite Impulse Response). We call *dynamic filtering* the extension of the definition of such a filter in the case where the coefficients vary with respect to the time. Nevertheless, it is assumed that this variation is small enough so that the filter can be considered as non evolutive under a small time interval. To control the bandwidth of the generated spectrum, we used an additional band-pass filter. The response of this filter is strongly related to both the strenght of the impact and the characteristics of the excitator. The model is represented in Figure 1.

3.2 From the material to the object

The simplest approach to generate emergent modes consists in using an additive method, by simply adding a sum of sinusoids to the white noise (figure 2, left). The sinusoids are characterized by their amplitudes and frequencies, leading to an easy implementation and control of the model. Nevertheless, it suffers of a lack of correlation between the parameters and the physical description of the vibrating structure.

We addressed this problem by proposing another approach based on physical modelling using digital waveguide concept [12] and especially the banded digital waveguide model [4] (figure 2, right). The idea consists in favouring some of the propagative waves by using loops simulating the creation of stationary waves in the structure. Each resonance is reproduced by a feedback loop composed of a delay line D_s and a band pass filter B_s (a second order filter has shown to be sufficient in most of the cases). We call these filters the *modal filters*. The band pass filter is centered on the desired frequency to select one resonance in the harmonic spectrum generated by the delay line. Generally, we generate as many resonances as feedback loops. But several harmonic resonances can be produced with only one feedback loop by increasing the bandwith of the band pass filter.

4. FROM REAL SOUNDS TO SYNTHESIS

In the context of analysis-synthesis, synthesis model parameters can be estimated from the analysis of natural sounds. In this section, we present an analysis method to extract both the damping law (characteristic of the material) and the modal parameters (characteristic of the structure geometry). We shall then establish a relationship between the synthesis model parameters and the estimated physical parameters.

4.1 Analysis of transient sounds

The transient feature of the sounds we consider leads at using joint representations to precisely 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 term of contributions which are well localized in both the time and the scale domain, giving a representation coherent with the hearing [7]. We chose a slightly modified version of the wavelet transform where the wavelets are constructed to mimic the response bandwidths of the basilar membrane at various locations [10]. This leads to what we call the time-scale Bark representation. This representation is obtained by decomposing the signal by a filter bank, the responses of which are given by gaussian functions located at central frequency corresponding to the Bark scale.

Estimation of the damping law: The damping law characterizing the material is estimated from the time-scale Bark representation. In each sub-band, we first calculate the associated analytical signal. This signal gives a complex representation of the sub-band signal, leading to an easy estimation of both the instantaneous frequency and amplitude law. 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 time signal exponentially decreases. Thus, we fit each sub-band amplitude law with an exponential function $e^{-\alpha t}$. The damping law is then characterized by the α for each Bark sub-band.

Estimation of the modal parameters: We have to determine the amplitudes and the eigen frequencies of emergent modes. For that purpose, many analysis methods are available. Among the so-called parametric methods, the Prony method [8] (or Steiglitz-Mac Bride method [15]) consists in identifying a given signal to a sum of exponentially damped sinusoids. 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 non-parametric methods, we can use the Fourier transform method to determine the amplitudes and eigen frequencies of emergent modes. We used either method, but it is worth noticing that parametric techniques are better adapted to the separation of close components.

4.2 Synthesis parameter estimation

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

4.2.1 *Estimation of the dynamic filter coefficients*

The dynamic filter F responsible of the sound damping is based on an IIR filter structure. In practice, we used a first order filter, the response of which is given by:

$$F(z) = \frac{a_0(1+z^{-1})}{1+b_1 z^{-1}} \quad \text{with} \quad z = e^{i\omega} \tag{1}$$

The expression is simplified so that only the two coefficients $\{a_0,b_1\}$ are needed. This is due to the fact that we assume the modulus of F to be zero at the Nyquist frequency. We can consider higher order filters (for example to simulate complex and not monotonous evolution of the signal) but a first order filter has shown to be sufficient for most of our applications in impact sound synthesis. The coefficient values are supposed to be time varying and must be calculated at regular time intervals. As the filter F reproduces the damping, we chose to estimate the set of unknown parameters $\{a_0,b_1\}$ by keeping invariant the damping coefficient in three specific Bark intervals. A first interval corresponds to the one containing the spectral centroid G. The two others correspond to the bands that contain the energetic centroids of the two domains on either side of G. This division is closely linked to the concept of tristimulus [11].

4.2.2 *Estimation of modal filter parameters*

In this section, we describe the *modal* filter parameter estimation for additive and digital waveguide approaches. In both cases, it is necessary to define precisely the emergence ratio of the modes with respect to the “smooth” part of the spectrum. This leads in selecting the modes of greater amplitudes and considering the rest of the spectrum as the “noise contribution” (took into account by the material model).

For what the additive approach is concerned, the model parameters are directly given by the spectral representation of the signal (see section 4.1). In the digital waveguide approach case, the estimation process is different. To avoid the formalism to be non-stationary, we use the concept of average filter. This filter, noted $\tilde{F}(\omega)$, corresponds to a filter the frequency response of which coincides with the Power Spectral Density of a white noise filtered by the dynamic filter. Its determination is of great importance since it allows us in further considering the system as a time invariant one. This assumption has shown to be valid for most of the sounds we worked on, since the short duration of the sounds together with the regular behavior of the damping makes the noise part well characterized by its global energy. Thus, one reduce the model to a time invariant one which is now fully determined by a transfer function $H(\omega)$ given by:

$$H(\omega) = \frac{\tilde{F}(\omega)}{1-\tilde{F}(\omega)\sum_{s=1}^{S}\left(B_s(\omega)D_s(\omega)\right)} \qquad (2)$$

with S the number of feedback loops. The parameters characterizing the modal filters are then obtained by minimizing the difference between the Fourier transform of the measured signal and the transfer function of the synthesis model.

4.3 Examples of analysis-synthesis

4.3.1 *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 made with different materials. Figure 3 shows the Bark representations of the signals obtained from wood and steel material. The right part of figure 3 shows the damping laws respectively corresponding to the wood, the glass and the steel as a function of the frequency. As we assumed, they

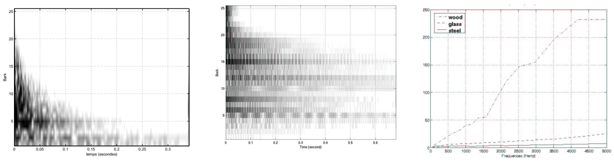

Figure 3: Bark representations corresponding to signal from wood thin plate (on the left) and signal from steel thin plate (on the middle). Frequency dependent damping laws for the wood, the glass and the steel as function of frequency (on the right). They are estimated from Bark representations.

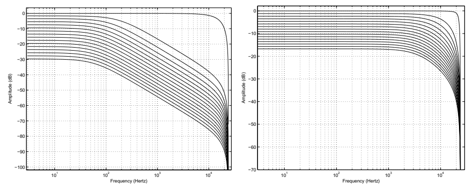

Figure 4: Evolution of the dynamical filter (dB scale) as function of frequency (logarithmic scale) for the wood material (on the left) and for the steel (on the right).

are frequency-dependent. The damping laws have been estimated on the relevant part of the sound spectrum for which the energy is sufficient. Nevertheless, the regularity of the curves obtained this way allows a pertinent extrapolation along the frequency axis. Figure 4 represents the evolution of the frequency response of the *dynamic filter* at various time intervals. One can clearly see that the filter becomes more and more low pass, meaning that high frequency components faster decrease than low frequency ones. These results present a good correlation with the physics of material which states that high frequency modes are generally more damped than low frequency ones. From the perceptual point of view, this is coherent with the fact that wood is perceived as a less resonant material than steel. As a consequence, a sound generated by a wood structure is generally shorter than a sound generated by a metallic structure.

For what the influence of the emergent modes is concerned, we have checked their relevance in the perception of the structure dimensions. According to the physics, our perception is sensitive to the frequency of the emergent modes: the higher, the smaller the structure is mentally visualized.

4.3.2 *Impulse response of a piano soundboard*

An important problem in piano sound synthesis is the reproduction of 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 using our synthesis model, the parameters of which have been estimated from the analysis of the sound obtained experimentally. This model will be further linked to an existing synthesis model of hammer-string interaction [1] [2] using the Commuted Synthesis concept [13] [14].

Figure 5 shows the extracted damping law and the time evolution of the modulus of the *dynamic filter* simulating this damping law. The dynamic filter presents a lowpass filter behavior. For what the contribution to the emergent modes

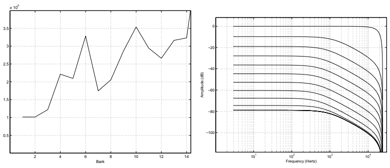

Figure 5: **On the left:** estimation of the damping law from the experimental signal of an impulse response of a piano soundboard. **On the right:** temporal evolution of the dynamic filter.

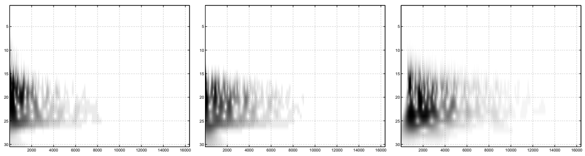

Figure 6: Time-scale representations of the measured signal (on the left), the synthesis signal obtained by the additive approach (on the middle), the synthesis signal obtained by the waveguide approach (on the right) as function of time (x-axis) and scale parameter (y-axis).

is concerned, we only took into account the ten modes of greater amplitudes. The rest of the modes is generated by the initial white noise. Consequently, in the additive approach, we have considered a sum of ten sinusoids and in the waveguide approach, ten feedback loops (one loop by mode).

From a perceptual point of view, the global behavior of the original signal is conserved and the synthesis sounds similar. Figure 6 respectively shows the time-scale representations of the measured signal, the synthesis signal obtained by the additive approach and the synthesis signal obtained by the digital waveguide approach.

5. CONCLUSION

We have presented a method to efficiently simulate impact sounds by reproducing the main physical features which play an important role from a perceptual point of view: the frequency-dependent damping law and the emergent spectral content. Using specific analysis methods adapted to transient sounds, the synthesis parameters are estimated from the analysis of natural sounds. Moreover, new sounds can be created using these models leading to an efficient tool for the generation of sounds in the context of virtual reality. In prospect, we expect generalizing the concept so that we can take into account the type of actions (rubbing, scratching, . . .).

We recently implemented the synthesis model based on the additive approach in real-time using the Max-MSP software. We then piloted the model using the Radio Baton interface [9], composed of two batons and a flat box into which captors are placed. This experiment has shown the efficiency of the synthesis model, allowing many sound effects such as continuous morphings between distinct materials or sizes.

REFERENCES

[1] M. Aramaki, J. Bensa, L. Daudet, Ph. Guillemain, R. Kronland-Martinet (2002). *Resynthesis of coupled piano string vibrations based on physical modeling*, Journal of New Music Research, Vol. 30, Issue 3, pp. 213-226, Ed. Swets & Zeitlinger.

[2] J. Bensa, K. Jensen, R. Kronland-Martinet. *A Hybrid Resynthesis Model for Hammer-Strings Interaction of Piano Tones*, to be published in EURASIP Journal on Applied Signal Processing, fall 2004.

[3] C. Carello, K. L. Anderson, A. J. Kunkler-Peck (1998). *Perception of object lenght by sound*, Psychological Science, vol. 9(3), pp. 211-214.

[4] G. Essl and P. Cook (1999). *Banded waveguides: towards physical modeling of bowed bar percussion instruments*, Actes du International Computer Music Conference, Beijing, Chine, pp. 321-324.

[5] W. W. Gaver (1988). *Everyday listening and auditory icons*, PhD thesis, University of California (San Diego).

[6] R. L. Klatzky, D. K. Pai, E. P. Krotkov (2000). *Perception of material from contact sounds*, Presence, MIT Press, vol. 9(4), pp. 399-410.

[7] R. Kronland-Martinet (1988). *The wavelet transform for analysis, synthesis, and processing of speech and music sounds*, Computer Music Journal, vol. 12(4).

[8] J. Laroche (1989). *Etude d'un système d'analyse et de synthèse utilisant la méthode de Prony - Application aux instruments de musique de type percussif*, PhD thesis, Ecole Nationale Supérieure des Télécommunications Paris.

[9] M. V. Mathews (1991). *The Radio Baton and conductor program or: pitch, the most important and least expressive part of music*, Computer Music Journal, vol. 15(4).

[10] Brian C. J. Moore (1982). *An introduction to the psychology of hearing*, Academic Press Inc., London.

[11] H. F. Pollard, E. V. Jansson (1982). *A Tristimulus Method for the Specification of Musical Timbre*, Acustica, vol. 51, pp. 162-171.

[12] J. O. Smith (1992). *Physical modeling using digital waveguides*, Computer Music Journal, vol. 16(4), pp. 74-87.

[13] J. O. Smith, Scott A. Van Duyne (1995). *Commuted Piano Synthesis*, Proceedings of the International Computer Music Conference, Banff, Canada, pp. 319-326.

[14] J. O. Smith, Scott A. Van Duyne (1995). *Developments for the Commuted Piano*, Proceedings of the International Computer Music Conference, Banff, Canada, Companion paper.

[15] K. Steiglitz, L. E. McBride (1965). *A Technique for the Identification of Linear Systems*, IEEE Trans. Automatic Control, vol. AC-10, pp. 461-464.

[16] S. Tucker and G. J. Brown (2002). *Investigating the perception of the size, shape and material of damped and free vibrating plates*, Technical Report CS-02-10, Université de Sheffield, Department of Computer Science.

[17] B. Truax (1999). *Handbook for acoustic ecology*, Cambridge Street Publishing, originally published by the World Soundscape Project, Simon Fraser University, and ARC Publications (1978).

[18] R. P. Wildes, W. A. Richards (1988). *Recovering material properties from sounds*, Natural computation, MIT Press, pp. 356-363.