Imagine the Sounds: An Intuitive Control of an Impact Sound Synthesizer

Mitsuko Aramaki[1], Charles Gondre[2], Richard Kronland-Martinet[2], Thierry Voinier[2], and Sølvi Ystad[2]

[1] CNRS - Institut de Neurosciences Cognitives de la Méditerranée, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex, France
aramaki@incm.cnrs-mrs.fr

[2] CNRS - Laboratoire de Mécanique et d'Acoustique, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex, France
{gondre,kronland,voinier,ystad}@lma.cnrs-mrs.fr

Abstract. In this paper we present a synthesizer developed for musical and Virtual Reality purposes that offers an intuitive control of impact sounds. A three layer control strategy is proposed for this purpose, where the top layer gives access to a control of the sound source through verbal descriptions, the middle layer to a control of perceptually relevant sound descriptors, while the bottom layer is directly linked to the parameters of the additive synthesis model. The mapping strategies between the parameters of the different layers are described. The synthesizer has been implemented using Max/MSP, offering the possibility to manipulate intrinsic characteristics of sounds in real-time through the control of few parameters.

1 Introduction

The aim of the current study is to propose an intuitive control of an additive synthesis model simulating impact sounds [1]. This is of importance within several domains, like sound design and virtual reality, where sounds are created from high-level verbal descriptions of the sound source and are to be coherent with a visual scene [2]. In this context, the challenge consists in being able to synthesize sounds that we have in mind. Efficient synthesis models that enable perfect resynthesis of natural sounds were developed in different contexts. In spite of the high quality of such models, the control, and the so-called mapping strategy, is an important aspect that has to be taken into account when constructing a synthesizer. To propose an intuitive control of sounds, it is in the first place necessary to understand the perceptual relevance of the sound attributes and then to find out how they can be combined to propose a high-level evocative control of the synthesizer. The sound attributes can be of different types and can either be directly linked to the physical behavior of the source [3], to the signal parameters [4] or to timbre descriptors obtained from perceptual tests [5][6][7]. In this particular study, perceptually relevant descriptors together with physical parameters linked to wave propagation phenomena such as dispersion and dissipation are considered.

S. Ystad et al. (Eds.): CMMR/ICAD 2009, LNCS 5954, pp. 408–421, 2010.

Based on these findings, we propose a complete mapping strategy that links three control layers: top layer (verbal description of the imagined sound source), middle layer (descriptors related to the characteristics of the signal) and bottom layer (parameters related to the synthesis model). The top layer offers the most intuitive way for a non-expert user to create impact sounds by specifying the perceived properties of the impacted object (like the material category, size and shape) and of the nature of the action (force, hardness, excitation point). The middle layer is composed of sound descriptors that characterize impact sounds from a perceptual point of view. The bottom layer directly depends on the parameters of the synthesis process. Finally, the mapping between the top and middle layers is based on results from previous studies on the perception of physical characteristics of the sound source (i.e., perception of material, object and action). The mapping between middle and bottom layers is defined based on results from synthesis experiments.

The paper is organized as follows: we first describe the theoretical model of impact sounds based on physical considerations and the real-time implementation of the synthesizer. Then, we define sound descriptors that are known to be relevant from a perceptual point of view in the case of impact sounds. The three-layer control strategy based on these descriptors is presented and the mappings between the different layers are detailed. We finally present some additional uses allowing analysis of natural impact sounds or real-time control in a musical context.

2 Signal Model of Impact Sounds

From a physical point of view, impact sounds are typically generated by an object under free oscillations that has been excited by an impact, or by the collision between solid objects. For simple cases, the vibratory response of such vibrating system (viewed as a mass-spring-damper system) can be described by a linear PDE:

$$\frac{\partial^2 x}{\partial t^2} = \frac{E}{\rho} Lx \tag{1}$$

where x represents the displacement, E the Young modulus and ρ the mass density of the material. L represents the differential operator describing the local deformation and corresponds to the Laplacian operator for strings (in 1D) or membranes (in 2D) and to the Bi-Laplacian for bars (in 1D) or thin plates (in 2D). To take into account loss mechanisms, the Young modulus generally is defined as complex valued [8] so that the solution $d(t)$ of the movement equation can be expressed by a sum of eigen modes $d_k(t)$, each of them decreasing exponentially:

$$d(t) = \sum_{k=1}^{K} d_k(t) = \sum_{k=1}^{K} A_k e^{2i\pi f_k t} e^{-\alpha_k t} \tag{2}$$

where A_k is the amplitude, f_k the eigen frequency, α_k the damping coefficient of the k^{th} mode, and K the number of components. The damping coefficient

α_k, generally frequency-dependent, is linked to the mechanical characteristics of the material and particularly to the internal friction coefficient [9]. The eigen frequencies f_k are deduced from the eigen values of the operator L with respect to the boundary conditions. Note that for multidimensional structures, the modal density increases with frequency so that the modes may overlap in the high frequency domain.

Consequently, we consider that from a signal point of view, an impact sound is accurately modeled by an additive model that consists in decomposing the signal $s(t)$ into deterministic $d(t)$ and stochastic $b(t)$ contributions :

$$s(t) = d(t) + b(t) \tag{3}$$

where $d(t)$ is defined in (2) and $b(t)$ is an exponentially damped noise defined by :

$$b(t) = \sum_{n=1}^{N} b_n(t) e^{-\alpha_n t} \tag{4}$$

where N is the number of frequency subbands. To take into account perceptual considerations, the subbands are defined on the Bark scale corresponding to the critical bands of human hearing [10]. We assume that the damping coefficient α_n is constant in each Bark band so that the damping of the noise signal is defined by 24 values.

3 Implementation of the Synthesizer

The real-time implementation of the theoretical model (defined in (3) and (4)) was made with MaxMSP [11]. The whole architecture is shown in Figure 1. The input signal consists of a stochastic contribution providing the noisy broadband

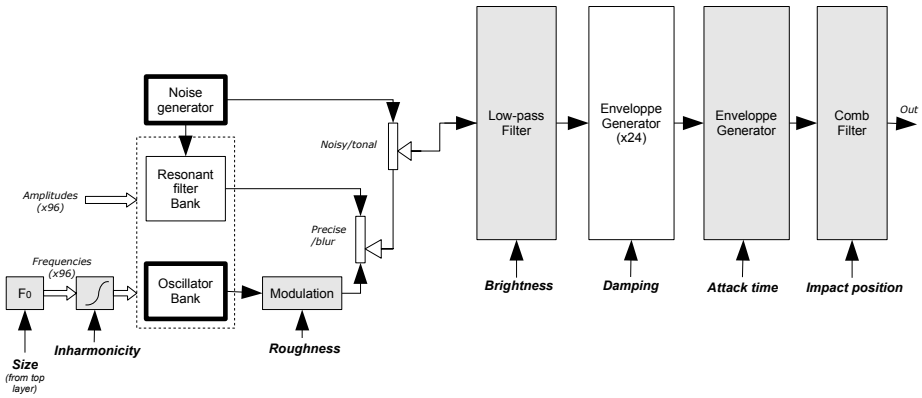

Fig. 1. Implementation of the impact sound synthesizer based on the additive signal model defined in Equation 3. The boxes represented in grey correspond to the modules added for the mapping toward higher levels (Section 5.2).

spectrum and a tonal contribution simulating the modes (boxes in bold). The stochastic contribution is produced by a noise generator which statistics can be defined by the user (we here used a gaussian noise). The tonal contribution is obtained by combining a sum of 96 sinusoids (oscillators) and 96 narrow-band filtered noises (obtained by filtering the output of the noise generator with resonant filters). The respective output level of sinusoids and filtered noises can be adjusted by a fader (*precise/blur* control), enabling the creation of interesting sound effects such as 'fuzzy' pitches. The output levels of stochastic and tonal contributions may also be adjusted by a fader (*tonal/noisy* control).

Then, the signal is damped by an envelope generator providing exponentially decaying envelopes in the form of $e^{-\alpha t}$. These envelopes differed with respect to frequency to take into account the frequency-dependency of the damping. Based on equation (4), we considered 24 envelopes, i.e., one per Bark band, characterized by the damping coefficient α_n. In each frequency subband, the same envelope is applied on both stochastic and deterministic parts of the signal to increase the merging between them. Nevertheless, for a pure deterministic signal, a damping coefficient can be defined for each partial of the tonal contribution.
At signal level, the sound generation necessitated the manipulation of hundreds of parameters and consequently, was only intended for experts. Thus, the large number of signal parameters necessitates the design of a control strategy. This strategy (generally called mapping) is of great importance for the expressive capabilities of the instrument, and it inevitably influences the way it can be used in a musical context [12]. For that reason, different mapping strategies can be proposed with respect to the context of use.

4 Perceptually Relevant Sound Descriptors

In this paper, we aim at proposing a mapping providing an intuitive control of the synthesizer from verbal descriptions of the sound source to the acoustic parameters of the signal. For that, we focused on previous psychoacoustic studies that investigated the links between the perception of the physical properties of the source and the acoustic attributes of the resulting sound. They revealed important sound features that uncover some characteristics of the object itself and the action. In particular, the perceived object size is found to be strongly correlated with the pitch of the generated sounds while the perceived shape of objects is correlated with the distribution of spectral components [13][14][15][16][17][18][19]. The perception of material seems to be mainly correlated with the damping of spectral components [9][14][3][20][21] and seems in addition to be a robust acoustic descriptor to identify macro-categories (i.e., wood-plexiglass and steel-glass categories) [22]. Regarding perception of excitation, [23] has shown that the perceived hardness of a mallet striking a metallic object is predictable from the characteristics of the attack time (a measure for the energy rise at sound onset). In addition, the perceived force of the impact is related to the brightness of the sound commonly associated with the spectral centroid, i.e, a measure for the center of gravity of the spectrum.

The attack time and spectral centroid were also identified as relevant descriptors in studies investigating the timbre perception for other types of sounds (e.g., sounds from musical instruments [5][6]). These studies revealed that timbre is a complex feature that requires a multidimensional representation characterized by several timbre descriptors. The most commonly used descriptors in the literature, in addition to attack time and spectral centroid, are spectral bandwidth, spectral flux and roughness. The spectral bandwidth is a measure for the spectrum spread. The spectral flux is a spectro-temporal descriptor that quantifies the time evolution of the spectrum. Its definition is given in [7]. The roughness is closely linked to the presence of several frequency components within the limits of a critical band and is closely linked to the notion of consonance/dissonance [24][25].

The control proposed in the synthesizer will be based on results from these psycho-acoustical studies since they give important cues about the intuitive aspect of the mapping strategy.

5 Control Strategy of the Synthesizer

We propose a control strategy based on three hierarchical layers allowing us to route and dispatch the control parameters from an evocative level to the signal level (see Figure 2). The top layer represents the control parameters that the user manipulates in an intuitive manner. Those control parameters are based on verbal descriptions of the physical source that characterize the object (nature of material, size and shape) and the excitation (impact force, hardness and position). The middle layer is based on sound descriptors that are known to be relevant from a perceptual point of view as described in Section 4. Finally,

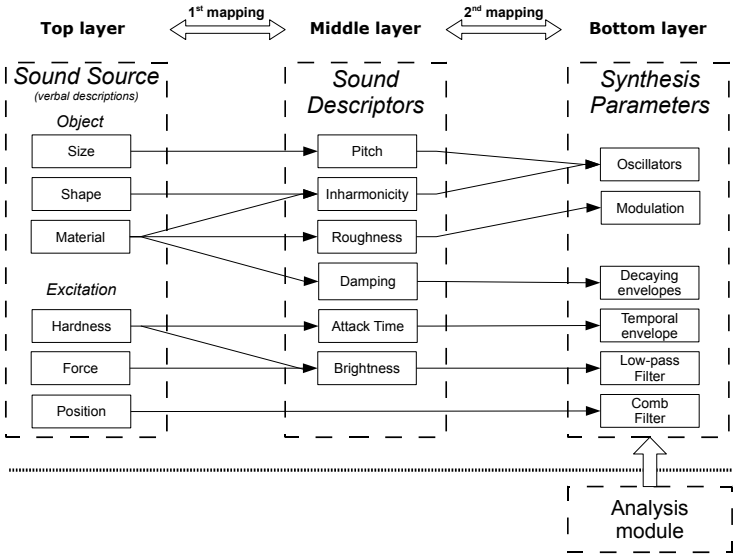

Fig. 2. Overview of the control strategy designed for the impact sound synthesizer

the bottom layer is composed of synthesis parameters as described in Section 3. Note that by default, the user only has access to the top layer. Nevertheless, we give the possibility for an expert user to directly access the middle or bottom layers. Such features are in some cases useful for sound design and musical experimentation to study the perceptual influence of specific parameters.

Between these three layers, two different mappings are to be implemented (represented as black arrows in Figure 2). As the parameters that allow intuitive controls are not independent and might be linked to several signal characteristics at a time, the mappings are far from being straight-forward. We describe these two mapping strategies in the following sections.

5.1 First Mapping: From Verbal Descriptions of Sound Source to Sound Descriptors

The first mapping links verbal descriptions characterizing the sound source to the perceptually relevant sound descriptors.

Object (material, size and shape). The characteristics of the object are defined by its perceived material, shape and size. As described in Section 3, previous studies have shown that the perception of material is related to the damping but also to additional cues mostly linked to the spectral content of sounds. In particular, the roughness has shown to be important to distinguish metal from glass and wood [26]. Consequently, the control of the perceived material involves the control of Damping but also of spectral sound descriptors such as Inharmonicity or Roughness.

The perception of the size of the object is mainly correlated with the pitch. Indeed, based on the physics, the pitch is related to the dimension of the object: actually, a big object is generally vibrating at lower eigenfrequencies than a small one. For quasi-harmonic sounds, we assume the pitch to be related to the frequency of the first spectral component. By contrast, complex sounds (i.e., numerous and overlapping modes), may elicit both spectral and virtual pitches [27]. Spectral pitches correspond to existing spectral peaks contained in the sound, whereas virtual pitches are deduced by the auditory system from upper partials of the spectrum. The virtual pitches may not correspond to any existing peak owing to the presence of a dominant frequency region situated around 700 Hz for which the ear is particularly pitch-sensitive. Thus, the pitch of complex sounds is still an open issue. In the present case, the perceived size of the object is directly linked to the fundamental frequency of the sound.

Furthermore, for impacted objects presenting a cavity (e.g., empty bottle), physical considerations (Helmholtz resonance) led to the prediction of a resonant frequency value with respect to the air volume inside the cavity [28]. In practice, the size of the cavity is directly mapped to the bottom layer and is simulated by adjusting the gain of a second-order peak filter which center frequency is set to the fundamental frequency, and which quality factor is fixed to $\frac{\sqrt{2}}{2}$.

Finally, the shape of the impacted object determines the spectral content of the generated impact sound from a physical point of view. As described in

Section 2, the frequencies of the spectral components correspond to the so-called eigenfrequencies that are characteristic of the modes of the vibrating object. Consequently, the perceived shape of the object is linked to the control of the Inharmonicity together with the pitch.

Excitation (impact force, hardness and position). The control of the excitation is based on the hardness of the mallet, the force of the impact as well as the excitation point. The excitation characterizes the nature of the interaction between the excitator and the vibrating object. From a physical point of view, this interaction can be described by a contact model such as the model proposed by [29] based on the contact theory of Hertz. The author found that the contact time τ can be defined by:

$$\tau = \pi \sqrt{\frac{\delta H}{E} \frac{LS*}{S}} \tag{5}$$

where δ is the density, H the height and $S*$ the section of the impacting object. E is the modulus of elasticity, S the section and L the length of the impacted object. This expression allows us to notably deduce that the harder the mallet (i.e., higher the modulus of elasticity), the shorter the contact time (the $S*/S$ ratio also acts on the contact time). In addition, the frequency range sollicited by the impact is governed by the contact time. From the theoretical spectrum $S_C(\omega)$ of the contact expressed by [30]:

$$S_C(\omega) \propto \frac{\sqrt{2}}{\left|\omega_0 \left(1 - (\omega/\omega_0)^2\right)\right|} \sqrt{1 - \cos\left(\frac{\pi\omega}{\omega_0}\right)} \tag{6}$$

where $\omega_0 = 1/\tau$, we can conclude that the shorter the contact time (the limit case would be the dirac), the larger the spectrum spread and the brighter the resulting sound. Based on these considerations, we linked the control of the hardness to the Attack time and the Brightness.

Concerning the force of the impact, its maximum amplitude is determined by the velocity of the impacting object. [31] showed that the contact time is weakly influenced by the force. Thus, the force is linked to the Brighness so that the heavier the force, the brighter the sound.

The excitation point, which strongly influences the amplitudes of the components by causing envelope modulations in the spectrum, is also taken into account. In practice, the impact point position is directly mapped to the bottom layer and is simulated by shaping the spectrum with a feedforward comb filter, defined by the transfer function :

$$H_\beta(z) = 1 - z^{-\lfloor \beta P \rfloor} \tag{7}$$

where P is the period in samples, $\beta \in (0, 1)$ denotes a normalized position, and $\lfloor . \rfloor$ the floor function [32].

5.2 Second Mapping: From Sound Descriptors to Signal Parameters

The second mapping (connection between middle and bottom layers) is intended to act upon the signal parameters according to the variations of the sound descriptors.

Inharmonicity. As already mentioned in Section 5.1, the distribution of the spectral components is an important parameter, as it may change one's perception of the size, shape and material of the impacted object. Its control is an intricate task since many strategies are possible. Based on the physical considerations described in Section 2, the inharmonicity induced by dispersion phenomena produces changes in the distribution of spectral components, and has shown to be an efficient parameter to control different shapes. Thus, we propose a control of the inharmonicity that allows the user to alter the spectral relationship between all the 96 initial harmonic components of the tonal contribution using three parameters a, b and c of the inharmonicity law defined by:

$$\widetilde{f_k} = a f_k \left(1 + b \left(\frac{f_k}{f_0} \right)^2 \right)^c \tag{8}$$

where $\widetilde{f_k}$ is the modified frequency, f_k the frequency of the k^{th} harmonic, and f_0 the fundamental frequency.

Thus the inharmonicity control changes the frequency ratio f_k/f_0 of each spectral component and provides an efficient way to get different types of inharmonicity profiles. Setting $a \geq 1$ and $b > 0$ leads to spectral dilations (i.e., frequencies will be deviated to higher values) providing a way to get stiff string or bell-like inharmonicity profiles, while setting $a < 1$ and $b < 0$ leads to spectral contractions (deviation to lower values) such as membrane or plate inharmonicity profiles. For example, a piano string inharmonicity is obtained for $a = 1$, $c = 0.5$ and b between 5.10^{-5} and 70.10^{-5} in the lower half of the instrument compass [33]. Large values of parameter c allows to strongly increase the frequency deviation. Some pre-defined presets offer a direct access to typical inharmonity profiles. Besides the proposed inharmonicity law, the possibility is given to the user to freely design desired behaviors by defining an arbitrary frequency ratio, independent for each component.

Roughness. Roughness is strongly linked to the presence of several spectral components within a bark band. Thus the control of roughness involves the generation of additional spectral components associated to the original ones. Based on this concept of presence of several components within a critical band, several methods have been proposed for the estimation of roughness for stationary tonal sounds [24][34]. A roughness estimation is obtained from the frequencies and amplitudes of the components. It is more difficult to evaluate the roughness of noisy and/or rapidly time-varying sounds. A computation model based on the auditory system has to be used. Several models have been developed [35][36], and for our investigations we used a model [37] that leads to a 'time-frequency representation' of the roughness. This representation reveals, for a given sound, the

critical bands that contain roughness, and how the roughness varies with respect to time. These investigations show that roughness is not equally distributed on the whole sound spectrum. For many impact sounds roughness exists in some frequency regions or 'roughness formants'.

This observation governed the roughness control implementation. For that, we implemented a way to increase the roughness independently for each bark band by means of amplitude and frequency modulations. Both methods are applied on each component at the oscillator bank level (Figure 1):

– Amplitude modulation :

$$d_k(t) = [1 + I\cos(2\pi f_{\mathrm{m}} t)] \times A_k \cos(2\pi f_k t) \tag{9}$$

$$\begin{aligned} d_k(t) = A_k \cos(2\pi f_k t) &+ \frac{A_k I}{2} \cos((2\pi f_k + 2\pi f_{\mathrm{m}})\, t) \\ &+ \frac{A_k I}{2} \cos((2\pi f_k - 2\pi f_{\mathrm{m}})\, t) \end{aligned} \tag{10}$$

where $I \in [0, 1]$ is the modulation index, f_{m} the modulating frequency, and A_k and f_k the k^{th} partial's amplitude and frequency respectively. Thus, for each partial, the amplitude modulation creates two additional components on both sides of the original partial, that consequently increases locally the roughness.

– Frequency modulation :

$$d_k(t) = A_k \cos(2\pi f_k t + I\cos(2\pi f_{\mathrm{m}} t)) \tag{11}$$

$$d_k(t) = A_k \sum_{n=-\infty}^{\infty} J_n(I) cos((2\pi f_k + n2\pi f_{\mathrm{m}})\, t) \tag{12}$$

where $n \in \mathbb{N}$, and J_n is the Bessel function of order n. Thus, for each partial, the frequency modulation creates an infinite number of additional components whose amplitudes are given by the partial's amplitude and the value of the Bessel function of order n for the given modulation index. In practice, the modulation index I is confined between 0 and 1, so that only a limited number of those additional components will be perceived.

In both the amplitude and frequency modulations, the user only defines the modulating frequency and the modulation indices. The modulating frequency is defined as a percentage of bark bandwidth. Based on [38], the percentage was fixed at 30%, while the modulation indices are controlled through 24 frequency bands, corresponding to the bark scale.

Note that the control of roughness can be considered as a local control of inharmonicity. Indeed, both controls modify the modal density (by creating additional components or by dilating the original spectrum) but the control of roughness has the advantage of being controlled locally for each component.

Brightness. The brightness, that is linked to the impact force, is controlled by acting on the amount of energy in the signal. In practice, the signal is filtered with a second order low pass filter of cut-off frequency f_c. Thus brightness perception decreases by progressively removing the highest frequencies of the broadband spectrum.

Damping. The material perception is closely linked to the damping of the spectral components. Since the damping is frequency dependent (high frequency components being more rapidly damped than low-frequency components), it necessitates a fine control of its frequency dependent behavior. Based on equation (4), the damping is controlled independently in each Bark band by acting on 24 values. To provide a more meaningful indication of the dynamic profile, the damping coefficient values were converted to duration values, i.e., the time necessary for the signal amplitude to be attenuated by 60dB. In addition, we defined a damping law expressed as an exponential function:

$$\alpha(\omega) = e^{a_{\mathrm{g}}+a_{\mathrm{r}}\omega} \tag{13}$$

so that the control of damping was reduced to two parameters: a_{g} is defined as a global damping and a_{r} is defined as a frequency-relative damping. The choice of an exponential function enables us to efficiently simulate various damping profiles characteristic of different materials by acting on few control parameters. For instance, it is accepted that in case of wooden bars, the damping coefficients increase with frequency following an empirical parabolic law which parameters depend on the wood species [39]. The calibration of the Damping was effectuated based on behavioral results from our previous study investigating the perception of sounds from different material categories based on a categorization task [26]: sounds from 3 impacted materials (i.e., Glass, Metal and Wood) were analyzed and synthesized, and continuous transitions between these different materials were further synthesized by a morphing technique. Sounds from these continua were then presented randomly to participants who were asked to categorize them as Glass, Metal or Wood. The perceptual limits between different categories were defined based on participants' responses and a set of unambiguous 'typical' sounds were determined. The acoustic analysis of these typical sounds determined the variation range of Damping parameter values (i.e., the global damping a_{g} and the relative damping a_{r}) for each category. Thus, the control of these two parameters provided an easy way to get different damping profiles directly from the label of the perceived material (Wood, Metal or Glass).

Attack time. The Attack time, which characterizes the excitation, is applied by multiplying the signal with a temporal envelope defined as a dB-linear fade-in function. The fade-in duration is set-up as the attack time duration.

5.3 Further Functionalities

Extracting synthesis parameters from natural sounds. An analysis module providing the extraction of the signal parameters (i.e., amplitudes, frequencies, damping coefficients, PSD of the noisy contribution) from natural percussive

sounds was implemented in Matlab [40] (see also [4]). This module provided a set of signal parameters for a given impact sound and was linked to the synthesis engine at the bottom level. From these settings, the controls offered at different layers allowed the user to manipulate characteristics of the resynthesized sound and to modify its intrinsic timbre attributes. Then, the modified sound could be stored in a wave file. Note that if the initial spectrum was non-harmonic, the control of inharmonicity was still valid: in that case, f_k corresponded to the frequency of the component of rank k.

MIDI controls. The synthesizer can be also used in a musical context. In order to enhance the playing expressivity, parameters that are accessible from the graphical interface (e.g., presets, attack time, size, material, impact position...) can be controlled by using the MIDI protocol. In practice, parameters are mapped to any MIDI channel, and can be controlled using either "control change" or "note on" messages. For instance, if an electronic drum set is used to control the synthesizer, MIDI velocity provided by the drum pad can be mapped to the impact force and the pitch value can be mapped to the size of the object. This functionality enables the creation of singular or useful mappings when using MIDI sensors.

In addition, to control the high number of parameters (96 frequency-amplitude pairs), a tuning control based on standard western tonal definitions was implemented, which enables the definition of chords composed of four notes [1]. Each note is defined by a fundamental frequency and is then associated with 24 harmonics, so that the 96 frequencies are defined 'automatically' by only four note pitches. In this chord configuration, the controls of sound descriptors related to spectral manipulation is effectuated on the 24 spectral components associated with each note and replicated on all the notes of the chord. Such a feature is thus useful to provide an intuitive control to musicians, as it is to facilitate the complex task of structuring rich spectra.

6 Conclusion and Perspectives

In this study, we have developed an intuitive control of a synthesizer dedicated to impact sounds based on a three level mapping strategy: a top layer (verbal descriptions of the source), a middle layer (sound descriptors) and a bottom layer (signal parameters). The top layer is defined by the characteristics of the sound source (object and excitation). At the middle layer, the sound descriptors were partly chosen on the basis of perceptual considerations, partly on the basis of the physical behavior of wave propagation. The bottom layer corresponded to the parameters of the additive signal model. This mapping strategy offers various possibilities to intuitively create realistic sounds and sound effects based on few control parameters. Further functionalities were also added such as an analysis module allowing the extraction of synthesis parameters directly from natural sounds or a control via the MIDI protocol. The mapping design is still in progress and some improvements are considered. In particular, although the

sound descriptors chosen for the control are perceptually relevant, the link between top and middle layers is far from being evident, since several middle layer parameters interact and cannot be manipulated independently. Additional tests will therefore be needed to choose the optimal parameter combinations that allow for an accurate control of sounds coherent with timbre variations.

Acknowledgment

This work was supported by the French National Research Agency (ANR, JC05-41996, "senSons").

References

1. Aramaki, M., Kronland-Martinet, R., Voinier, T., Ystad, S.: A percussive sound synthesizer based on physical and perceptual attributes. Computer Music Journal 30(2), 32–41 (2006)
2. van den Doel, K., Kry, P.G., Pai, D.K.: FoleyAutomatic: physically-based sound effects for interactive simulation and animation. In: Proceedings SIGGRAPH 2001, pp. 537–544 (2001)
3. McAdams, S., Chaigne, A., Roussarie, V.: The psychomechanics of simulated sound sources: material properties of impacted bars. Journal of the Acoustical Society of America 115(3), 1306–1320 (2004)
4. Kronland-Martinet, R., Guillemain, P., Ystad, S.: Modelling of Natural Sounds Using Time-Frequency and Wavelet Representations. Organised Sound 2(3), 179–191 (1997)
5. McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., Krimphoff, J.: Perceptual scaling of synthesized musical timbres: common dimensions, specicities, and latent subject classes. Psychological Research 58, 177–192 (1995)
6. Grey, J.M.: Multidimensional perceptual scaling of musical timbres. Journal of the Acoustical Society of America 61(5), 1270–1277 (1977)
7. McAdams, S.: Perspectives on the contribution of timbre to musical structure. Computer Music Journal 23(3), 85–102 (1999)
8. Valette, C., Cuesta, C.: Mécanique de la corde vibrante. Hermès, Lyon (1993)
9. Wildes, R.P., Richards, W.A.: Recovering material properties from sound. In: Richard, W.A. (ed.) ch. 25, pp. 356–363. MIT Press, Cambridge (1988)
10. Moore, B.C.J.: Introduction to the Psychology of Hearing, 2nd edn. Academic Press, New York (1982)
11. MaxMSP, Cycling 1974, `http://www.cycling74.com/downloads/max5`
12. Gobin, P., Kronland-Martinet, R., Lagesse, G.A., Voinier, T., Ystad, S.: Designing Musical Interfaces with Composition in Mind. In: Wiil, U.K. (ed.) CMMR 2003. LNCS, vol. 2771, pp. 225–246. Springer, Heidelberg (2004)
13. Carello, C., Anderson, K.L., KunklerPeck, A.J.: Perception of object length by sound. Psychological Science 9(3), 211–214 (1998)
14. Tucker, S., Brown, G.J.: Investigating the Perception of te Size, Shape, and Material of Damped and Free Vibrating Plates. Technical Report CS-02-10, University of Sheffield, Departement of Computer Science (2002)
15. van den Doel, K., Pai, D.K.: The Sounds of Physical Shapes. Presence 7(4), 382–395 (1998)

16. Kunkler-Peck, A.J., Turvey, M.T.: Hearing shape. Journal of Experimental Psychology: Human Perception and Performance 26(1), 279–294 (2000)
17. Avanzini, F., Rocchesso, D.: Controlling material properties in physical models of sounding objects. In: Proceedings of the International Computer Music Conference 2001, pp. 91–94 (2001)
18. Rocchesso, D., Fontana, F.: The Sounding Object (2003)
19. Lakatos, S., McAdams, S., Caussé, R.: The representation of auditory source characteristics: simple geometric form. Perception & Psychophysics 59, 1180–1190 (1997)
20. Klatzky, R.L., Pai, D.K., Krotkov, E.P.: Perception of material from contact sounds. Presence 9(4), 399–410 (2000)
21. Gaver, W.W.: How do we hear in the world? Explorations of ecological acoustics. Ecological Psychology 5(4), 285–313 (1993)
22. Giordano, B.L., McAdams, S.: Material identification of real impact sounds: Effects of size variation in steel, wood, and plexiglass plates. Journal of the Acoustical Society of America 119(2), 1171–1181 (2006)
23. Freed, D.J.: Auditory correlates of perceived mallet hardness for a set of recorded percussive events. Journal of the Acoustical Society of America 87(1), 311–322 (1990)
24. Sethares, W.A.: Local consonance and the relationship between timbre and scale. Journal of the Acoustical Society of America 93(3), 1218–1228 (1993)
25. Vassilakis, P.N.: Selected Reports in Ethnomusicology (Perspectives in Systematic Musicology). Auditory roughness as a means of musical expression, Department of Ethnomusicology, University of California, vol. 12, pp. 119–144 (2005)
26. Aramaki, M., Besson, M., Kronland-Martinet, R., Ystad, S.: Computer Music Modeling and Retrieval – Genesis of Meaning of Sound and Music. In: Timbre perception of sounds from impacted materials: behavioral, electrophysiological and acoustic approaches. LNCS, vol. 5493, pp. 1–17. Springer, Heidelberg (2009)
27. Terhardt, E., Stoll, G., Seewann, M.: Pitch of Complex Signals According to Virtual-Pitch Theory: Tests, Examples, and Predictions. Journal of Acoustical Society of America 71, 671–678 (1982)
28. Cook, P.R.: Real Sound Synthesis for Interactive Applications. A. K. Peters Ltd., Natick (2002)
29. Graff, K.F.: Wave motion in elastic solids. Ohio State University Press (1975)
30. Broch, J.T.: Mechanical vibration and shock measurements. In: Brel, Kjaer (eds.) (1984)
31. Sansalone, M.J., Streett, W.B.: Impact-Echo: Nondestructive Testing of Concrete and Masonry. Bullbrier Press (1997)
32. Jaffe, D.A., Smith, J.O.: Extensions of the Karplus-Strong plucked string algorithm. Computer Music Journal 7(2), 56–69 (1983)
33. Fletcher, H.: Normal Vibration Frequencies of a Stiff Piano String. Journal of the Acoustical Society of America 36(1), 203–209 (1964)
34. Vassilakis, P.N.: SRA: A web-based research tool for spectral and roughness analysis of sound signals. In: Proceedings of the 4th Sound and Music Computing (SMC) Conference, pp. 319–325 (2007)
35. Daniel, P., Weber, D.: Psychoacoustical roughness implementation of an optimized model. Acustica 83, 113–123 (1997)
36. Pressnitzer, D.: Perception de rugosité psychoacoustique: D'un attribut élémentaire de l'audition à l'écoute musicale. PhD thesis, Université Paris 6 (1998)

37. Leman, M.: Visualization and calculation of the roughness of acoustical musical signals using the synchronisation index model (sim). In: Proceedings of the COST-G6 Conference on Digital Audio Effects, DAFX 2000 (2000)
38. Plomp, R., Levelt, W.J.M.: Tonal Consonance and Critical Bandwidth. Journal of the Acoustical Society of America 38(4), 548–560 (1965)
39. Aramaki, M., Baillères, H., Brancheriau, L., Kronland-Martinet, R., Ystad, S.: Sound quality assessment of wood for xylophone bars. Journal of the Acoustical Society of America 121(4), 2407–2420 (2007)
40. Aramaki, M., Kronland-Martinet, R.: Analysis-Synthesis of Impact Sounds by Real-Time Dynamic Filtering. IEEE Transactions on Audio, Speech, and Language Processing 14(2), 695–705 (2006)