Thinking the sounds: an intuitive control of an impact sound synthesizer

Mitsuko Aramaki, Charles Gondre, Richard Kronland-Martinet, Thierry Voinier, Sølvi Ystad

THINKING THE SOUNDS : AN INTUITIVE CONTROL OF AN IMPACT SOUND SYNTHESIZER

Mitsuko Aramaki

Institut de Neurosciences Cognitives
de la Méditerranée
Centre National de la Recherche
Scientifique (CNRS)
31 Chemin Joseph Aiguier
13402 Marseille Cedex, France
aramaki@incm.cnrs-mrs.fr

Charles Gondre, Richard Kronland-Martinet,
Thierry Voinier, Sølvi Ystad

Laboratoire de Mécanique et d'Acoustique
Centre National de la Recherche
Scientifique (CNRS)
31 Chemin Joseph Aiguier
13402 Marseille Cedex, France
{gondre, kronland, voinier,
ystad}@lma.cnrs-mrs.fr

ABSTRACT

In this paper we describe a synthesizer to be used both for virtual reality and musical purposes, based on an additive synthesis model, and 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 through verbal descriptions, the middle layer to a control of perceptually relevant signal descriptors, while the bottom layer is directly linked to the parameters of the 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 the sounds in real-time through the control of the different parameters.

1. INTRODUCTION

The aim of the current study is to be able to intuitively control an additive synthesis model simulating impact sounds [1]. This is of importance within several domains, like for instance sound design and virtual reality, where sounds are created 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 have been developed in different contexts. In spite of the high quality of such models, the control issue, 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 signal parameters and then to find out how these parameters can be combined to propose a high-level evocative control of the synthesizer. The timbre descriptors that are found to be relevant are further considered for the control strategy. In addition to the timbre descriptors, physical wave propagation phenomena such as dispersion and dissipation are also considered.

Based on these findings, we propose a complete mapping strategy that links three control layers: top layer (verbal description of the mental representation of the 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 properties of the sound producing object, like for instance the material category, size and shape. The middle layer is composed of sound descriptors that characterize impact sounds from a perceptual point of view, identified thanks to a categorization test (described in section 5.1.1). The bottom layer directly depends on the parameters of the synthesis process (e.g., oscillators' frequencies and amplitudes). Finally, the mapping between the top and middle layers is based on results from previous studies on the perception of the 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 [3].

This paper is organized as follows: first we present the general background of this study by describing the theoretical model of impact sounds based on physical considerations and by defining the sound descriptors that are known to be perceptually relevant in the case of impact sounds. Then, we describe the architecture of the synthesizer and we detail the mapping between the different layers for the control of the synthesizer. We finally present some additional functionalities allowing real-time piloting or specific controls in a musical context.

2. SYNTHESIS 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:

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

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 [4] 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:

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

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 α_k, generally frequency-dependent, is linked to the mechanical characteristics of the material and particularly to the internal friction coefficient [5]. 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 the 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 synthesis technique that consists in decomposing the signal $s(t)$ into deterministic $d(t)$ and stochastic $s(t)$ contributions :

$$s(t) = d(t) + b(t) \quad (3)$$

where $d(t)$ is defined in Equ. 2 and $b(t)$ is a filtered noise. Interestingly, we present in Section 5 some control strategies issued from physical considerations, in particular the control of the inharmonicity.

3. PERCEPTUALLY RELEVANT SOUND DESCRIPTORS

The perceptual relevance of signal parameters must be investigated before the design of a mapping strategy. The investigation of timbre properties is therefore a natural starting point. The descriptors that are found to be of interest, will be used in the middle control layer of the synthesizer (described in the introduction). These descriptors will be chosen with the top control layer in mind.
Previous timbre studies revealed that timbre is a complex feature that requires a multidimensional representation [6][7]. The timbre descriptors that are most commonly used in the literature to characterize impact sounds are attack time AT, spectral centroid SC, spectral bandwidth SB, spectral flux SF and roughness R. The attack time is a measure for the way the energy rises at sound onset. The spectral centroid SC is a measure for the center of gravity of the spectrum, while 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 [8]. The roughness R 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 [9][10]. 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 [9][11]. 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 [12][13], and for our investigations we used a model [14] 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 will govern the roughness control implementation (section 5.2.3).

Previous acoustic studies on the links between perception and physical characteristics of sound sources have revealed several properties that are important to identify the action on the object and the object itself. Regarding perception of excitation, [15] has shown that the perceived hardness of a mallet striking a metallic object is predictable from the characteristics of the attack time. 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 [16][17][18][19][20][21][22]. Finally, the perception of material seems to be mainly correlated with the damping of spectral components [5][17][23][24][25] and seems in addition to be a robust acoustic descriptor to identify macro-categories (i.e., wood-plexiglass and steel-glass categories) [26].

These results give important cues about the mapping strategy to be used in the synthesizer. Some parameters, such as the attack time (related to the excitation) and the pitch (related to the size), can be directly used as control parameters, while the distribution of spectral components to control the perceived shape is a more intricate task. By returning to 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. In addition to the inharmonicity, the roughness has shown to be important to distinguish metal from glass and wood [27]. Therefore, both the inharmonicity and the roughness have been chosen to control the perceived shape. Finally, as the material perception is closely linked to the damping of the spectral components, the SF could be an interesting candidate to control this aspect. However, since the damping is frequency dependent (high frequency components being more rapidly damped than low-frequency components), it necessitates a fine control of both the global and frequency dependent behavior, that is proposed in the current interface.

4. IMPLEMENTATION OF THE SYNTHESIZER

Our real-time implementation using Max/MSP is based on the theoretical synthesis model (section 2). The input signal of the model consists of a stochastic contribution (limited here to a gaussian noise generator) providing the broadband spectrum and a tonal contribution simulating the emergent modes. The tonal contribution is obtained by combining a sum of sinusoids (96 oscillators) and a narrow-band filtered noise (96 resonant filters), whose respective output levels can be adjusted (Precise/blur control), enabling the creation of more or less 'fuzzy' pitches. The output levels of stochastic and tonal contributions may also be adjusted by a fader (Noisy/tonal control). The resulting signal is then formatted by an envelope generator which role is to shape the spectral and temporal envelopes of the signal. This real time implementation is shown in Fig. 3. On this figure, different modules corresponding to the mapping have been added (grey boxes) which functionalities are specified in section 5.2.

5. CONTROL STRATEGY OF THE SYNTHESIZER

The large number of synthesis parameters of our model necessitates 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 musi-

cal context [28]. For that reason, different mapping strategies can be proposed with respect to the context of use. We here describe one of those strategies that provides an intuitive control of the synthesis parameters based on the evocations of sound sources for the class of impact sounds.

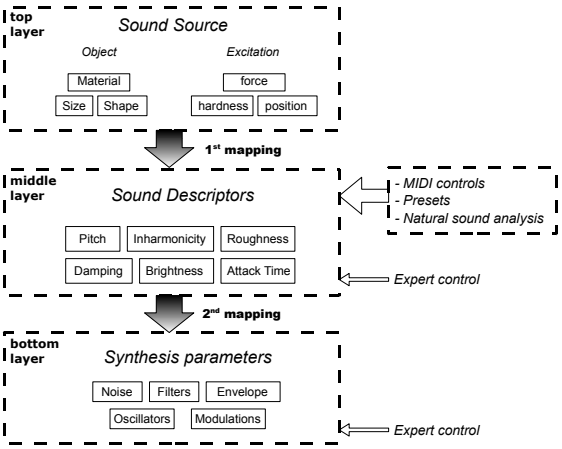

Figure 1: *From intuitive control to signal processing, a three layer implementation*

We propose a control strategy based on three hierarchical layers allowing us to route and dispatch the control parameters from an intuitive level to the signal and algorithmic level (see Fig. 1). The top layer represents the controls that the user will have to manipulate in an intuitive manner. Those controls are based on verbal descriptions of the physical sound source that are the characteristics of the object (nature of material, size and shape) and of 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 and that we described in Section 3. Finally, the bottom layer is composed of synthesis parameters described in Section 4. Note that by default, the user only has access to the top layer. Nevertheless, we give the possibility to an expert user to directly access the middle or bottom layers. Between these three layers, there are two mappings to implement (represented as black arrows in figure 1). 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 mappings in the following sections.

5.1. First mapping : from sound source to sound descriptors

The first mapping links the verbal description of the sound source to the sound descriptors. Figure 2 shows the main connections between these two layers that are described in the following sections.

5.1.1. 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

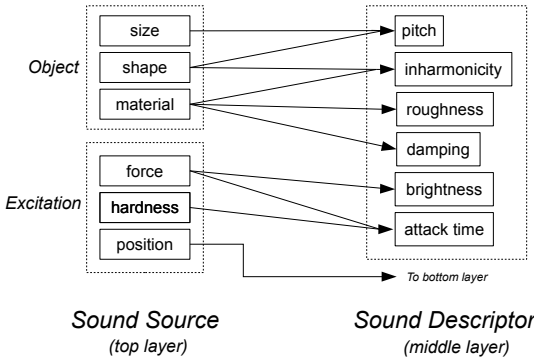

Figure 2: *First mapping - from sound source to sound descriptors*

of sounds. Consequently, the control of the perceived material involves the control of Damping but also of spectral sound descriptors such as the Inharmonicity or the Roughness (and consequently the Brightness). 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 [27]: 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 allowed identifying the variation range of Damping parameter values (i.e., the global damping a_g and the relative damping a_r; see Section 5.2.1) 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).

The perception of the size of the object correlated mainly 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 [29]. 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 in which the ear is particularly pitch-sensitive. Otherwise, for impacted objects presenting a cavity (e.g., empty bottle), physical considerations (Helmholtz resonance) led us to predict a resonant frequency value with respect to the air volume inside the cavity [30].

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.

5.1.2. Excitation (impact force, hardness and position)

The excitation is characterized by the force and hardness of the impact as well as by the excitation point. The force of the impact corresponds to the amount of energy transmitted to the impacted object so that the harder the impact, the brighter the sound and the sharper the sound attack. Consequently, the force is linked to the Brightness and to the Attack time. Based on studies described in Section 3, the hardness is linked to the Attack Time. The impact position, which strongly influences the amplitudes of the components by causing envelope modulations in the spectrum, may also be taken into account. In our implementation, the impact point position is directly mapped to the bottom layer and is simulated by a comb filter that shapes the spectrum.

5.2. Second mapping : from sound descriptors to synthesis parameters

The second mapping (connection between middle and bottom layers) is intended to act upon the synthesis parameters according to the variations of the sound descriptors. Figure 3 shows the connections between each sound descriptor and the signal processing chain that are described in the following sections.

5.2.1. Pitch

As mentioned previously in section 4, the input signal of the model consists in a stochastic part (gaussian noise) and a tonal part simulating the emergent modes. The perceived pitch is controlled by acting on the amplitudes and frequencies of the tonal part. To control the high number of parameters (96 frequency-amplitude pairs), a tuning control based on standard western tonal definitions has been implemented, which allows one to define 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 both to provide an intuitive control to musicians, as it is to facilitate the complex task of structuring rich spectra.

5.2.2. Inharmonicity

As already mentioned in section 5.1.1, the inharmonicity is an important parameter, as it may change one's perception of the size, shape, and material of an object. The inharmonicity control allows the user to alter the spectral relationship between all the 96 components of the tonal contribution using three parameters a, b and c of the inharmonicity law defined by:

$$\widetilde{f_k} = a f_k \left(1 + b \left(\frac{f_k}{f_0}\right)^2\right)^c \tag{4}$$

where f_k is the frequency of the k^{th} partial, and f_0 is 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. For example, setting $a > 1$ and $b > 0$ leads to spectral dilations (i.e., frequencies will be deviated to higher values) providing a way to get piano or bell-like inharmonicity profiles, while setting $a < 1$ and $b < 0$ leads to spectral contractions (deviation to lower values) such the membranes or plates inharmonicity profiles. The parameter c allows both to get a piano-like inharmonicity formula when $c = 0.5$ (after [31]) and to strongly increase the frequency deviation for bigger values. Some pre-defined presets allow a direct access to typical inharmonity profiles.

5.2.3. Roughness

As described in section 3, 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. 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 (Fig. 3).

– Amplitude modulation :

$$d_k(t) = [1 + m\cos(2\pi f_m t)] \times A_k \cos(2\pi f_k t) \tag{5}$$

$$\begin{aligned} d_k(t) = A_k \cos(2\pi f_k t) + \frac{A_k m}{2} \cos((2\pi f_k + 2\pi f_m)t) \\ + \frac{A_k m}{2} \cos((2\pi f_k - 2\pi f_m)t) \end{aligned} \tag{6}$$

where $m \in [0, 1]$ 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 :

$$d_k(t) = A_k \cos(2\pi f_k t + m\cos(2\pi f_m t)) \tag{7}$$

$$d_k(t) = A_k \sum_{n=-\infty}^{\infty} J_n(m) cos((2\pi f_k + n2\pi f_m)t) \tag{8}$$

where $n \in N$, 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 m is restraint 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 has to define the modulating frequency and the modulation indices. The modulating frequency is common to every modulation, while the modulation indices are controlled through 24 frequency bands, corresponding to the critical bands of hearing, which provides a better control over the resulting spectrum.
Note that the control of roughness partly depends on the control of the inharmonicity (section 5.2.2) since the spectrum dilation modifies the number of spectral components within a bark band. The roughness has the advantage of being controlled locally for each partial component. Consequently, roughness may be defined as a local inharmonicity.

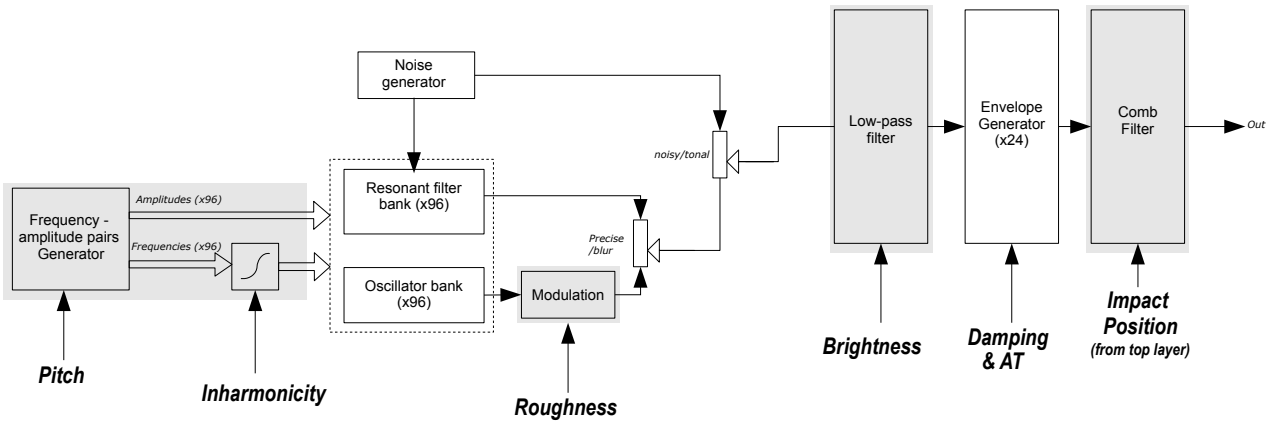

Figure 3: *Second mapping - from sound descriptors to synthesis parameters*

5.2.4. Brightness

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

5.2.5. Damping

The frequency-dependency of the Damping is modeled by an exponential function that is called the damping law :

$$\alpha(\omega) = e^{a_g + a_r \omega} \tag{9}$$

where a_g is defined as a global damping and a_r is defined as a relative damping. To take into account perceptual considerations, the damping law is defined on Bark scale, corresponding to the critical bands of hearing [32]. Moreover, we assume that the damping is constant in each Bark band, so that the damping is defined by 24 values. In practice, the damping is controlled by the two parameters a_g and a_r of the exponential function. 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 law of a parabolic form where parameters depend on the wood species [33]. Note also that the damping can be controlled independently in each Bark band by acting on the 24 values of damping. In our implementation, the damping is simulated by 24 envelope generators (one per bark band). The generated envelopes are exponentially shaped envelopes ($e^{-\alpha(\omega)t}$), according to equation 1.

5.2.6. Attack time

The Attack time, which strongly characterizes the excitation (Section 5.1.2) is simulated by the creation of a dB-linear fade-in whose duration is set-up as the attack-time duration.

5.3. further functionalities

5.3.1. Extracting synthesis parameters from natural sounds

An analysis module has been implemented, which enables the extraction of synthesis parameters from natural percussive sounds [34]. Although some parameters, such as the damping or the noise level, may need to be re-adjusted in order to obtain a perceptually satisfying resynthesis of the sound, this analysis module provide an efficient way to get global synthesis parameter settings for a given impact sound, that can then be refined by the user through the interface's controls.

5.3.2. MIDI controls

In order to enhance the playing expressivity when the synthesizer is used in a musical context, sound descriptors (middle layer level) can be controlled by using the MIDI protocol. In particular, MIDI velocity may be mapped to the impact force on a drum interface, as well as MIDI note messages may be mapped to the pitch. In addition, it is possible to map any parameter accessible from the graphical interface (e.g., presets, attack time, etc.) to any MIDI control-change channel. This functionality allows one to create singular or useful mappings when using MIDI sensors (impact position may be changed in real time for example).

5.3.3. Expert controls

Besides the damping and inharmonicity models, the possibility is given to freely design their respective behaviors : the inharmonicity law can be drawn by the user and the damping can be defined for each bark band. In this case, damping is defined as the duration for the components to be attenuated by 60dB. This representation, as it gives an indication about the dynamic profile of the damping, provides a more intuitive control than a direct representation of the $\alpha(\omega)$ values. This situation allows to define some singular damping and inharmonicity profiles and thus generate interesting sound effects. In addition, an 'expert mode' gives the possibility to accurately define the partials' frequencies and amplitudes (any value of the 96 frequency-amplitude pairs may be changed). Such features are as useful to sound design and musical experimentation as they are to study the influence of those parameters on the perception.

6. CONCLUSION & 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 (mental representation of the source), a middle layer (sound descriptors) and a bottom layer (synthesis parameters). The top layer is characterized by the characteristics of the sound source (object and excitation). At the middle layer, the sound descriptors have been partly chosen on the basis of listening tests, partly on the basis of the physical behavior of wave propagation. This mapping strategy offers various possibilities of sound creation and of 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 using MIDI protocol. Actually, 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 a complete intuitive control of sounds coherent with our mental representations.

7. REFERENCES

[1] M. Aramaki, R. Kronland-Martinet, Th. Voinier and S. Ystad, "A Percussive Sound Synthetizer Based on Physical and Perceptual Attributes," *Computer Music Journal*, vol. 30, n. 2, pp. 32–41, 2006.

[2] K. van den Doel, P.G. Kry and D.K. Pai, "FoleyAutomatic: Physically-based Sound Effects for Interactive Simulation and Animation," in *Proceedings SIGGRAPH 2001*, pp. 537–544, 2001.

[3] J.O. Smith and S.A. van Duyne, "Developments for the Commuted Piano," in *Proceedings of the 1995 International Computer Music Conference*, San Francisco, California : International Computer Music Association, pp 335–343, 1995.

[4] C. Valette and C. Cuesta, "Mécanique de la corde vibrante," Lyon, France: Hermès, 1993.

[5] R.P. Wildes, W.A. Richards, "Recovering material properties from sound," in *Natural Computation*, MIT Press, Cambridge, pp. 356–363, 1988.

[6] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff, "Perceptual scaling of synthesized musical timbres: common dimensions, specicities, and latent subject classes," *Psychological Research*, vol. 58, pp. 177–192, 1995.

[7] J.M. Grey, "Multidimensional perceptual scaling of musical timbres," *Journal of the Acoustical Society of America*, vol. 61, n. 5, pp. 1270–1277, 1977.

[8] S. McAdams, "Perspectives on the contribution of timbre to musical structure," *Computer Music Journal*, vol. 23, n. 3, pp. 85–102, 1999.

[9] W.A. Sethares, "Local consonance and the relationship between timbre and scale," *Journal of the Acoustical Society of America*, vol. 93, n. 3, pp. 1218–1228, 1993.

[10] P.N. Vassilakis, "Auditory roughness as a means of musical expression," *Selected Reports in Ethnomusicology*, vol. 12 (Perspectives in Systematic Musicology), pp. 119-144, 2005.

[11] P.N. Vassilakis, "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*, Lefkada, Greece, July 2007, pp. 319–325.

[12] P. Daniel and D. Weber, "Psychoacoustical roughness implementation of an optimized model," *Acustica*, vol. 83, pp. 113–123, 1997.

[13] D. Pressnitzer, "Perception de rugosité psychoacoustique : D'un attribut élémentaire de l'audition à l'écoute musicale," PhD thesis, Université Paris 6, 1998.

[14] M. Leman, "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*, 2000.

[15] D.J. Freed, "Auditory correlates of perceived mallet hardness for a set of recorded percussive events," *Journal of the Acoustical Society of America*, vol. 87, n. 1, pp. 311–322, 1990.

[16] C. Carello, K.L. Anderson, and A.J. KunklerPeck, "Perception of object length by sound," *Psychological Science*, vol. 9, no. 3, pp. 211–214, 1998.

[17] S. Tucker and G.J. Brown, "Investigating the Perception of te Size, Shape, and Material of Damped and Free Vibrating Plates," University of Sheffield, Departement of Computer Science, Technical Report CS-0210, 2002.

[18] K. van den Doel and D.K. Pai,"The Sounds of Physical Shapes," *Presence*, vol. 7, no. 4, pp. 382–395, 1998

[19] A.J. Kunkler-Peck and M.T. Turvey, "Hearing shape," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 26, n. 1, pp. 279–294, 2000.

[20] F. Avanzini, D. Rocchesso, "Controlling material properties in physical models of sounding objects," in *Proceedings of the International Computer Music Conference 2001*, La Habana, Cuba, September 2001, pp. 91–94.

[21] D. Rocchesso and F. Fontana, "The Sounding Object," 2003.

[22] S. Lakatos, S. McAdams, and R. Caussé, "Perception of auditory source characteristics: Simple geometric form," in *Perception & Psychophysics*, vol. 59, 1180–1190, 1997.

[23] S. McAdams, A. Chaigne and V. Roussarie, "The psychomechanics of simulated sound sources: material properties of impacted bars," *Journal of the Acoustical Society of America*, vol. 115, n. 3, pp 1306–1320, 2004.

[24] 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.

[25] W.W. Gaver, "How do we hear in the world ? Explorations of ecological acoustics," *Ecological Psychology*, vol. 5, n. 4, pp. 285–313, 1993.

[26] B.L. Giordano and S. McAdams, "Material identification of real impact sounds: Effects of size variation in steel, wood, and plexiglass plates," *Journal of the Acoustical Society of America*, vol. 119, n. 2, pp 1171–1181, 2006.

[27] M. Aramaki, M. Besson, R. Kronland-Martinet and S. Ystad, "Timbre perception of sounds from impacted materials: behavioral, electrophysiological and acoustic approaches," in *Computer Music Modeling and Retrieval – Genesis of Meaning of Sound and Music*, Springer Berlin, Heidelberg, 2009.

[28] P. Gobin, R. Kronland-Martinet, G.-A. Lagesse, T. Voinier, and S. Ystad, "Designing Musical Interfaces with Composition in Mind," in *Lecture Notes in Computer Science LNCS 2771*, Springer, pp. 225-246, 2004.

[29] E. Terhardt, G. Stoll and M. Seewann, "Pitch of Complex Signals According to Virtual-Pitch Theory: Tests, Examples, and Predictions," *Journal of Acoustical Society of America*, vol. 71, pp. 671–678, 1982.

[30] P.R. Cook, "Real Sound Synthesis for Interactive Applications," A. K Peters Ltd., Natick, Massachusetts, 2002.

[31] H. Fletcher, "Normal Vibration Frequencies of a Stiff Piano String," *Journal of the Acoustical Society of America*, vol. 36, n. 1, pp. 203–209, 1964.

[32] B.C.J. Moore, "Introduction to the Psychology of Hearing," 2nd Ed., New York: Academic Press, 1982

[33] M. Aramaki, H. Baillères, L. Brancheriau, R. Kronland-Martinet and S. Ystad, "Sound quality assessment of wood for xylophone bars," *Journal of the Acoustical Society of America*, vol. 121, n. 4, pp. 2407–2420, 2007.

[34] M. Aramaki and R. Kronland-Martinet, "Analysis-Synthesis of Impact Sounds by Real-Time Dynamic Filtering," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, n. 2, pp 695–705, 2006.