A Percussive Sound Synthesizer Based on Physical and Perceptual Attributes

Author(s): Mitsuko Aramaki, Richard Kronland-Martinet, Thierry Voinier and Sølvi Ystad

Source: *Computer Music Journal*, Summer, 2006, Vol. 30, No. 2 (Summer, 2006), pp. 32-41

Published by: The MIT Press

Stable URL: https://www.jstor.org/stable/3682002

Mitsuko Aramaki, Richard Kronland-Martinet, Thierry Voinier, and Sølvi Ystad
Laboratoire de Mécanique et d'Acoustique
Centre National de la Recherche Scientifique (CNRS)
31, Chemin Joseph Aiguier
13402 Marseille Cedex 20, France
{aramaki,kronland,voinier,ystad}@lma.cnrs-mrs.fr

A Percussive Sound Synthesizer Based on Physical and Perceptual Attributes

Synthesis of impact sounds is far from a trivial task owing to the high density of modes generally contained in such signals. Several authors have addressed this problem and proposed different approaches to model such sounds. The majority of these models are based on the physics of vibrating structures, as with for instance modal synthesis (Adrien 1991; Pai et al. 2001; van den Doel, Kry, and Pai 2001; Cook 2002; Rocchesso, Bresin, and Fernström 2003). Nevertheless, modal synthesis is not always suitable for complex sounds, such as those with a high density of mixed modes. Other approaches have also been proposed using algorithmic techniques based on digital signal processing. Cook (2002), for example, proposed a granular-synthesis approach based on a wavelet decomposition of sounds.

The sound-synthesis model proposed in this article takes into account both physical and perceptual aspects related to sounds. Many subjective tests have shown the existence of perceptual clues allowing the source of the impact sound (its material, size, etc.) to be identified merely by listening (Klatzky, Pai, and Krotkov 2000; Tucker and Brown 2002). Moreover, these tests have brought to the fore some correlations between physical attributes (the nature of the material and dimensions of the structure) and perceptual attributes (perceived material and perceived dimensions). Hence, it has been shown that the perception of the material mainly correlates with the damping coefficient of the spectral components contained in the sound. This damping is frequency-dependent, and high-frequency modes are generally more heavily damped than low-frequency modes. Actually, the dissipation of vibrating energy owing to the coupling between the structure and the air increases with frequency (see, for example, Caracciolo and Valette 1995).

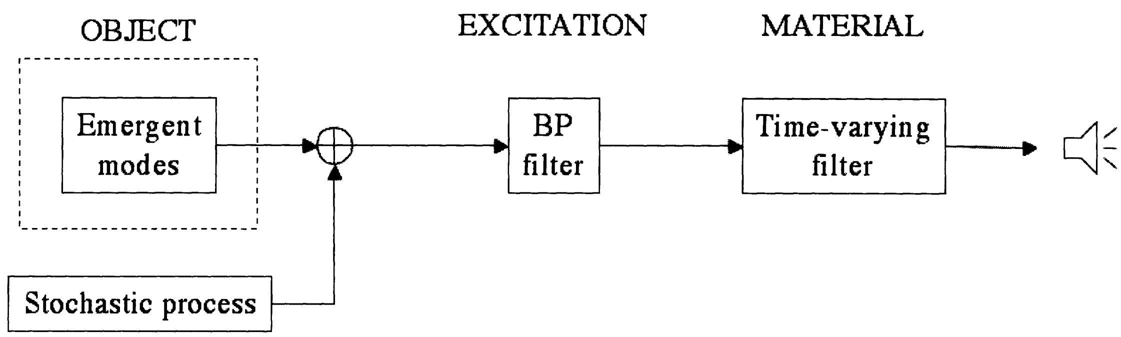
To take into account this fundamental sound behavior from a synthesis point of view, a time-varying filtering technique has been chosen. It is well known that the size and shape of an object's attributes are mainly perceived by the pitch of the generated sound and its spectral richness. The perception of the pitch primarily correlates with the vibrating modes (Carello, Anderson, and Kunkler-Peck 1998). For complex structures, the modal density generally increases with the frequency, so that high frequency modes overlap and become indiscernible. This phenomenon is well known and is described for example in previous works on room acoustics (Kuttruff 1991).

Under such a condition, the human ear determines the pitch of the sound from emergent spectral components with consistent frequency ratios. When a complex percussive sound contains several harmonic or inharmonic series (i.e., spectral components that are not exact multiples of the fundamental frequency), different pitches can generally be heard. The dominant pitch then mainly depends on the frequencies and the amplitudes of the spectral components belonging to a so-called dominant frequency region (Terhardt, Stoll, and Seewann 1982) in which the ear is pitch sensitive. (We will discuss this further in the Tuning section of this article.) With all these aspects in mind, and wishing to propose an easy and intuitive control of the model, we have divided it into three parts represented by an excitation element, a material element, and an object element.

The large number of parameters available through such a model necessitates a control strategy. This strategy (generally called a 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 (Gobin et al. 2004).

Computer Music Journal, 30:2, pp. 32–41, Summer 2006

Figure 1. Impact sound synthesis model. This model is divided in three parts representing the object, the excitation, and the material contributions.



In this article, we mention some examples of possible strategies, like an original tuning approach based on the theory of harmony. This approach makes it possible to construct complex sounds like musical chords, in which the root of the chord, its type (major or minor), and its inversions can be chosen by the performer. However, owing to the strong influence between mapping and composition, the choice of the strategy should, as far as possible, be available to the composer.

Theoretical Synthesis Model

The synthesis model we propose is shown in Figure 1. It is an extension of that proposed by Smith and Van Duyne (1995, Van Duyne and Smith 1995), developed to simulate the soundboard's influence on piano tones. This model is based on a time-varying subtractive synthesis process that acts on a noisy input signal. This sound-synthesis model reproduces two main contributions characterizing the perceived material (determined by the damping factors) and the perceived dimensions of the impacted object (determined by pitch and modal density). We decided to model these two contributions separately, even if they cannot be totally disconnected from a physical point of view. Actually, we believe this separation yields an easier and more intuitive control of the sounds.

Another important aspect of the model is its ability to resynthesize natural sounds, meaning that one can also reproduce a given impact sound that is perceptually identical to the original. Nevertheless, this aspect is not described here, and we refer the reader to a more theoretical article (Aramaki and Kronland-Martinet 2006). In what follows, we give a more precise description of the three main elements contained in the model.

Material Element

From the literature, it is well known that damping is frequency-dependent, implying that high-frequency modes generally are more heavily damped than low-frequency modes (Caracciolo and Valette 1995). This is important from a perceptual point of view, because damping is a characteristic of the object's material and allows us to distinguish, for example, wood from steel. The damping of the modes is here simulated by a digital infinite-impulse-response (IIR) filter structure in which coefficients vary with time (here called a time-varying filter). Nevertheless, from a theoretical point of view, it is assumed that this variation is small enough for the filter to be considered stationary in a small time interval (Mourjopoulos, Kyriakis-Bitzaros, and Goutis 1990).

The filter used for the model generally is a low pass with a gain and cutoff frequency that decrease with time. In this way, by simply acting on the damping coefficients, we can reproduce the main perceptual features of an impacted material. In particular, if we strongly damp an initial white noise input, the sound will have wooden characteristics, whereas for the same initial white noise, it will have metallic characteristics when the damping is weak (Sound Examples 1 and 2, available online at www.lma.cnrs-mrs.fr/~kronland/CMJ/sounds.html). At this stage, this "material" model is adequate to reproduce perceptual effects reflecting the main characteristics of the impacted materials, even though the technique does not simulate modes.

Object Element

To provide a subjective notion of the size and shape of the sounding object, a few spectral components are added to the initial white noise. If, for example, we add one or a few low-frequency components, the sound will evoke a sense that the impacted structure is relatively big. Conversely, if we add high-frequency components, the sound will evoke a sense that the impacted structure is relatively small (Sound Examples 3 and 4, www.lma.cnrs-mrs.fr/~kronland/CMJ/sounds.html). From a physical point of view, these spectral components mainly correspond to the eigenmodes of the structures. These modes can be deduced for simple cases from the movement equation and can be generated simply by adding a sum of sinusoids to the white-noise input signal.

Nevertheless, the approach suffers from a lack of correlation between the stochastic part of the sound and the deterministic part, making the sounds unrealistic. To overcome this drawback, we generated the deterministic part from narrow bands of the initial white noise, improving the correlation between the two parts (Sound Examples 5 and 6, www.lma.cnrs-mrs.fr/~kronland/CMJ/sounds.html). Note that another method to generate resonances based on physical modeling, namely the banded digital waveguide approach (Essl et al. 2004), has been proposed in Aramaki and Kronland-Martinet (2006). Even though this method gives very satisfactory sounds, we have not used it in real-time applications, because it can lead to instability problems and increase the calculation time. By perceptually reproducing the most pertinent spectral components related to these eigenmodes, we can simulate sounds that evoke various structures like strings, plates, bells, etc. (Sound Examples 7–11, www.lma.cnrs-mrs.fr/~kronland/CMJ/sounds.html).

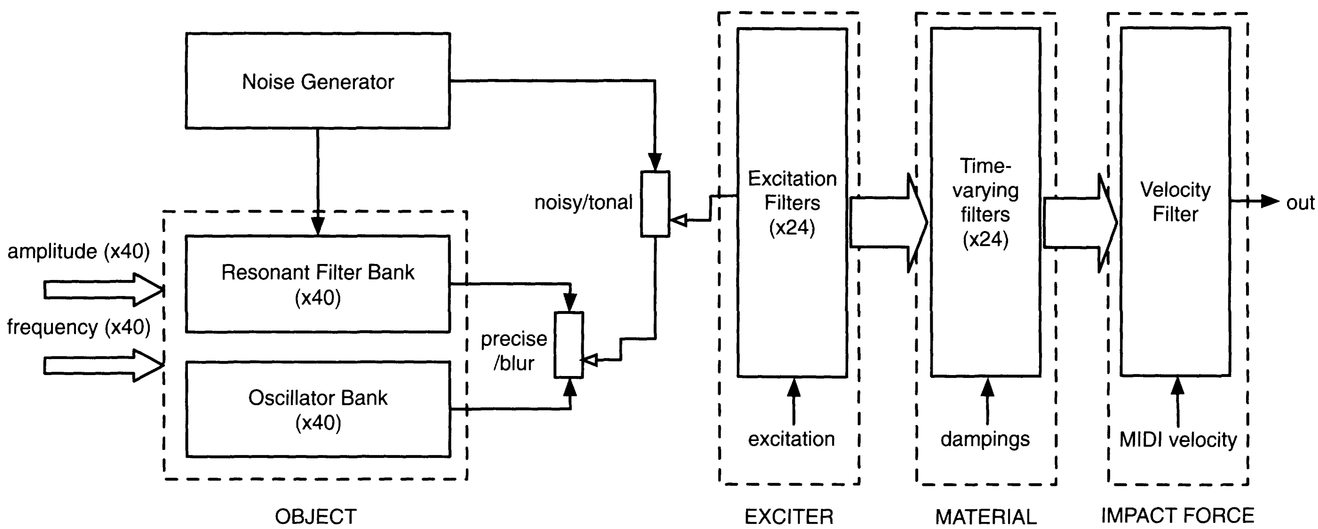
Excitation Element

Finally, to model the excitation, a band-pass filter is used to control the bandwidth of the generated spectrum. From a physical point of view, the response of this filter is strongly related to the strength of the impact, that is, the bandwidth increases as a function of the impact velocity. We can also add a time envelope that controls the attack time of the sound, thereby characterizing the collision between the exciter and the object. This possibility has been added in the real-time implementation of the model. (The slower the attack time, the smoother the excitation, as illustrated in Sound Example 12, www.lma.cnrs-mrs.fr/~kronland/CMJ/sounds.html.)

Tuning

In this section, we discuss the problem of tuning the pitch of the impact sounds. Even though we aim at designing an intuitive tool for musicians rather than a complete impact-sound tuning system, pitch tuning is not a trivial task. Actually, complex sounds often evoke several spectral pitches, because our hearing system tends to associate spectral components having consistent frequency ratios. Moreover, the perceived pitch of a series of spectral components, either harmonic or inharmonic, is not necessarily given by the frequency of the first component of the series. As Terhardt, Stoll, and Seewann (1982) explain, complex tones elicit both spectral and virtual pitches. Spectral pitches correspond to the frequency of spectral peaks contained in the sound spectrum, whereas virtual pitches are deduced by the auditory system from the upper partials in the Fourier spectrum, leading to pitches that may not correspond to any peak contained in the sound spectrum. A well-known example is the auditory generation of the missing fundamental of a harmonic series of pure tones. In addition, owing to the presence of a dominant frequency region situated around 700 Hz in which the ear is particularly pitch-sensitive, the perceived pitch depends on both the frequencies and the amplitudes of the spectral components. Hence, the pitch of complex tones with low fundamental frequencies (under 500 Hz)

Figure 2. Real-time implementation of the three main elements of the theoretical model (related to the object, material, and excitation). A low-pass filter is added to take into account the impact force on the drum interface (controlled by MIDI velocity).



depends on higher partials, while the pitch of tones with high fundamental frequencies is rather determined by the fundamental frequency, because it lies in the dominant region.

As a consequence, when a complex tone contains inharmonic partials, the perceived pitch is determined by the frequencies in the dominant region and might differ from the fundamental frequency. If, for example, a complex inharmonic tone has partials of 204, 408, 612, 800, 1,000, and 1,200 Hz, all with similar amplitudes, the first three partials yield a perceived pitch equal to the fundamental frequency (204 Hz). Nevertheless, the six partials together give a pitch of 200 Hz, because the higher partials determine the pitch, given that they lie in the dominant region (Terhardt, Stoll, and Seewann 1982).

Another aspect that can modify pitch perception is the masking effect between partials, because mental reconstruction of the fundamental frequency of a residue tone might be difficult or impossible if partials in the dominant region are masked by noise or other frequency components. Finally, in most musical situations, tones in context are less ambiguous, because the context normally suggests the pitch register in which the tone is most likely to be heard (Parncutt 1989). This might be of importance in future studies. For the time being, we only focus on the tuning of isolated complex tones.

We shall now see how the real-time model is implemented and adapted to give the user access to these parameters. As this synthesis model simulates percussive sounds, a drum interface is a natural choice for piloting the model.

Real-Time Implementation

Our real-time implementation using Max/MSP is based on the structure of the theoretical synthesis model: the "object" element, devoted to the simulation of the emergent modes; the "material" element, simulating the damping of the sounds; and the "excitation" element (see Figure 2). A low-pass filter is added to take into account the impact force on the drum interface (controlled by MIDI velocity).

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. As mentioned earlier, the modes can be simulated by a sum of sinusoids, but the lack of correlation between the stochastic and the deterministic parts makes the sound unrealistic. The spectral peaks are therefore obtained by combining a sum of sinusoids (40 oscillator banks) and a narrow-band filtered white noise (40 resonant filter banks), enabling the creation of more or less "fuzzy" pitches. Indeed, fuzzy pitches are useful for adding reverberative effects to the sound when the stochastic part is weak.

The material element simulating the damping controls the evolution of the spectrum through 24 frequency bands, corresponding to the critical bands of hearing, known as the Bark bands (Zwicker and Fastl 1990). This configuration allows the reproduction of the frequency dependence of the damping, where the damping coefficients are taken as constant in each Bark band. Consequently, the damping is simulated by a time-varying gain for each Bark band.

The excitation part is reproduced by two contributions: the "exciter" element and the "impact force" element (see Figure 2). The "exciter" element controls the spectral repartition of the initial energy given to the system and conveys the mechanical characteristics of the excitation element. For common cases, a band-pass filter generally is sufficient. Hence, this contribution is simulated by a static gain adjustment in each Bark band. In comparison to the theoretical model, a supplementary filter ("velocity filter") is added to take into account the player's gesture (MIDI velocity input) and is composed of a one-pole low-pass filter.

Control of the Synthesis Model

As seen in the previous sections, the synthesis model contains a large number of elements that make it possible to control different aspects of the sounds. Although the model has been divided into three parts for more intuitive control, its complexity necessitates the development of strategies to control the parameters. In this section, we first provide an overview of the basic control parameters of the model, and then we discuss how mapping strategies can be developed for musical purposes. Figure 3 shows the actual user interface and reveals the possibilities of control of the synthesis model.

Control of the Input Signal Parameters

As previously mentioned, the input signal is composed of tonal and noisy components. The user can define the relative gain of each contribution (the "tonal/noisy" slider in Figure 3). Concerning the tonal contribution, which consists of the sinusoids and narrow-band filtered white noise (the "object" element in Figure 2), the user can also define the ratio between these two parts (the "precise/blur" slider in Figure 3).

The object element contains 80 parameters (40 frequency and 40 amplitude values) that are associated with the control of the oscillator banks and the resonant filter banks. To minimize the high number of frequency parameters, a proposed tuning preset ("Chord Generator" in Figure 3) based on standard Western tonal definitions is constructed. Players can here choose whether they wish to construct the complex sound with a single pitch, with several pitches forming a specific four-note chord, or with arbitrary combinations of pitches.

When the unison case is chosen, the four notes generated by the chord generator have the same pitch values. When the chord case is chosen, four different pitches are generated. In this case, the player selects the root of the chord, the type (major or minor), the harmonization (4th, 7th, 9th, diminished, etc.) and the inversion (four possible). (This is illustrated in Sound Example 13, www.lma.cnrs-mrs.fr/~kronland/CMJ/sounds.html.)

The inversions indicate the lowest note of the chord so that, for example, the first inversion ("inversion +1") corresponds to a chord built on the second note of the main chord, and the second inversion ("inversion +2") corresponds to a chord built on the third note of the main chord. Negative inversions are also possible. "Inversion –1" and "inversion –2" indicate that the lowest note of a four-note chord is taken as the fourth and third notes, respectively, the non-inverted chord. When the main chord contains four notes (e.g., "C7"), then "inversion +2" and "inversion –2" are identical (differing by one octave). This would not have been the case if the chords had contained more than four notes (e.g., "C4,7," "D7,9", etc.). As an example, Figure 4 shows some possible inversions of "Cminor7."

It is well known, for example, that a triad in root position played on a piano is more "stable" than its inversions, and that the relative stability of two chords is determined by the relative proximity to

Figure 3. User interface for control of the synthesis model.

Figure 4. Example of a C-minor seventh chord with some possible inversions.

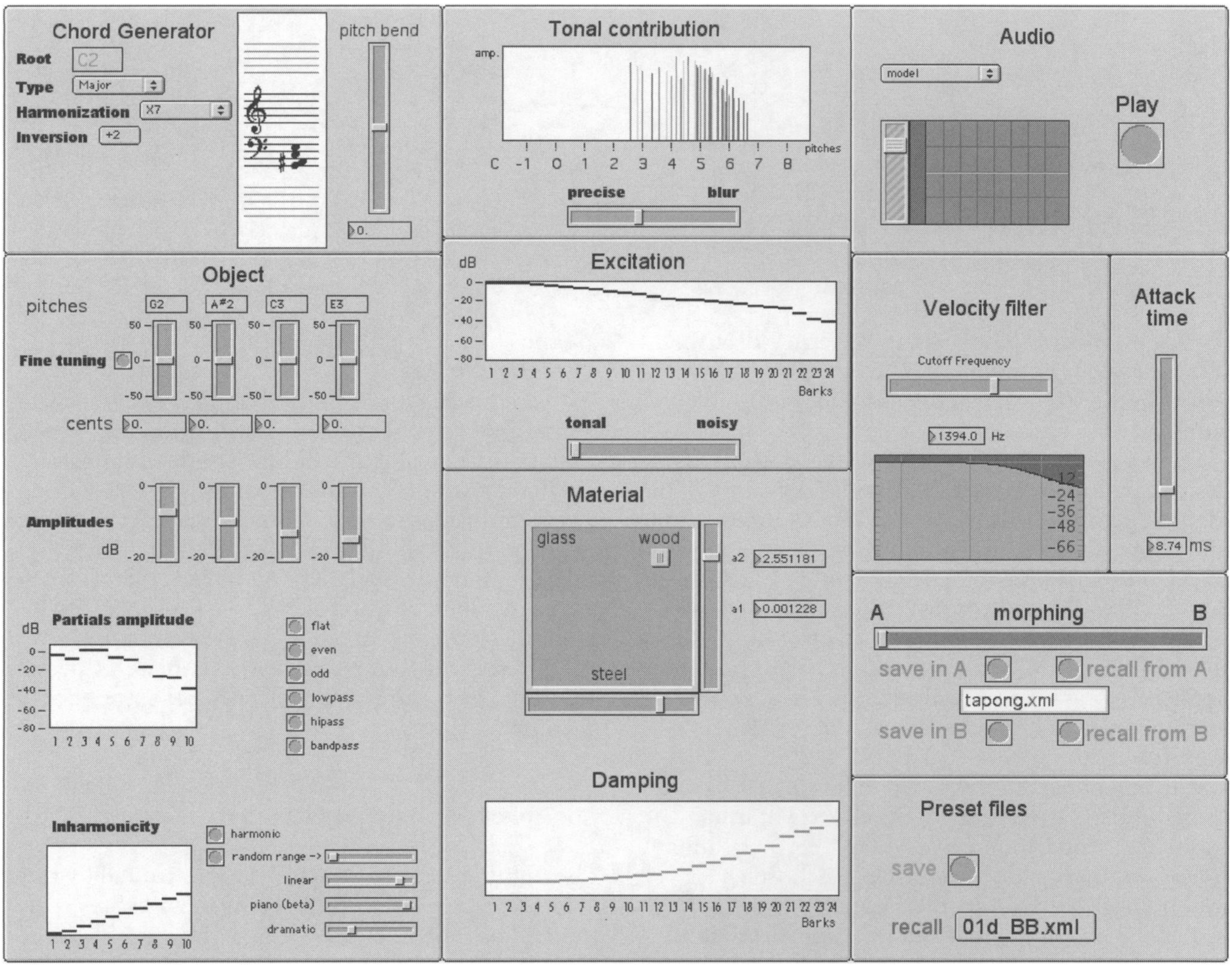


Figure 3

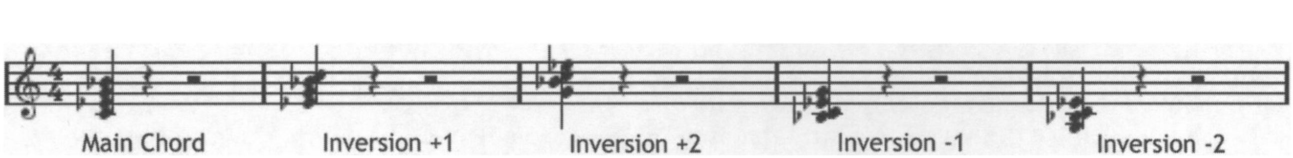
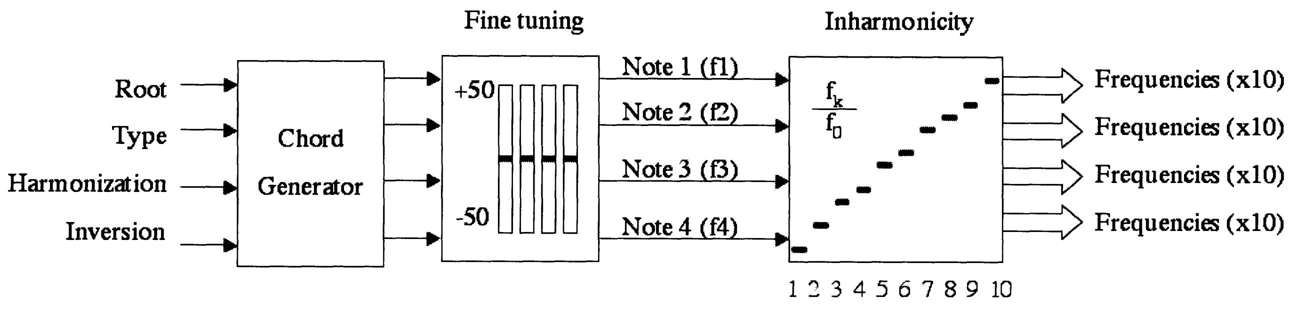


Figure 4

Figure 5. Control of the object element (i.e., of the 40 frequency values) with the chord generator, fine tuning, and inharmonicity.



the local tonic of their roots in the circle of fifths (Lerdahl and Jackendoff 1977). We will later investigate whether this is also the case for our impulsive sounds.

From a given chord preset, pitch deviations can also be created by moving sliders ("Fine tuning" in Figure 3). For each note, a set of ten oscillator banks and ten resonant filter banks are associated, meaning that a spectrum composed of ten partials is generated. The user can control the amplitude values globally by acting on each note's gain ("Amplitudes" in Figure 3) or more precisely by acting independently on the amplitudes of the ten spectral components associated with each note ("Partials amplitude" in Figure 3).

The player can also alter the relationship between the partials of each note by the control of inharmonicity ("Inharmonicity" in Figure 3). The inharmonicity relationship can either be chosen individually by defining the frequency ratio f_k/f_0 of each partial in the series as a function of the fundamental frequency, or by adjusting parameters of different presets. We have chosen the following presets for this purpose: harmonic ($f_k = k \times f_0$), linear ($f_k = a \times k \times f_0$) or piano-like ($f_k = k \times f_0 \times \sqrt{1 + \beta k^2}$, after Fletcher 1964 and Valette and Cuesta 1993). In each relationship, f_0 is the fundamental frequency, and $k \in$ N+, $a \in \mathbf{R}^*+$, and $\beta \in \mathbf{R}+$ are the control parameters. For example, when the signal is harmonic, the spectral components are integer multiples of the fundamental frequency, and the inharmonicity curve is given by a straight line, where the frequency of the kth component equals k times the frequency of the fundamental component. Figure 5 provides a more detailed view of the control of the "object" element (i.e., the 40 frequency values) using the chord generator, fine-tuning, and inharmonicity settings.

A schematic representation of the tonal contribution is given on the top of the user interface ("Tonal contribution" in Figure 3). The degree of inharmonicity of the spectral components together with their amplitudes makes it possible to alter the perceived pitch as a function of the dominant frequency region, as explained previously. This tool represents both an interesting musical interface to tune the synthesis model in a musical context and an important research tool to study the relationship between pitch perception and spectral components of complex sounds.

Control of the Material Element

The material part of the model is controlled by the damping parameters, with access to 24 values corresponding to the frequency bands on which the user can act independently ("Damping" in Figure 3). In addition, we have chosen to parameterize the set of damping values by an exponential function, defining a damping law $\alpha(\omega)$ that can be written

$$\alpha(\omega) = e^{a_1\omega + a_2} \quad (1)$$

Hence, we reduce the damping control to only two parameters, a_1 and a_2. This damping law is a function of the frequency and can be directly estimated from physical considerations or from the analysis of natural sounds. Thus, a bi-dimensional space defined by $\{a_1; a_2\}$ in which the player can move a cursor is proposed ("Material" in Figure 3). As the damping values are strongly characteristic of the nature of the perceived material, this space can be considered as a "material space" where specific zones can be representative of different materials. Analysis of natural sounds of different materials (in particular wood, glass, and steel) allowed us to calibrate this bi-dimensional space and roughly determine the domains of each material, as shown in Figure 3. Hence, the player can move from one material to another, for example, from a wooden to a metallic sound.

Control of the Excitation Element

The excitation filter is controlled with 24 gains that can be modified graphically ("Excitation" in Figure 3). These gains act on 24 Bark-scale band-pass filters. We can also take into account the excitation point, which from a physical point of view causes envelope modulations in the corresponding spectrum. The player can change the attack time to vary the perceived excitation ("Attack time" in Figure 3). For instance, a slow attack simulates the perception of a rubbed object. More generally, one could propose an excitation space where the type of excitation could be chosen (e.g. plucking, striking, rubbing). Finally, the player's gesture is taken into account by a low-pass filter ("velocity filter" in Figure 3) which cutoff frequency depends on the force sensed by the trigger control interface. In this way, we can imitate the well-known non-linear effect that leads to an increase in the spectral width as a function of the force of the impact.

Other Mapping Strategies

We now propose possible mapping strategies for more intuitive control of the real-time synthesis model. These examples are intended to give a hint of many possible useful strategies. In addition to the material space proposed in the previous section, it would be of interest to define a space related to the size and the shape of the object. As already seen, the parameters related to such an object space would mainly be related to the pitch and the inharmonicity of the input signal. Actually, small objects are generally envisioned by the listener when high pitches are perceived, whereas big objects are envisioned when low pitches are perceived. Furthermore, a one-dimensional structure (e.g., a string) is perceived when a unison preset is chosen for the chord generator, whereas a multi-dimensional structure (e.g., a plate or a bell) is perceived when several pitches are chosen.

In this way, according to our mental representations of sounds, a more intuitive control can be proposed based on verbal input parameters such as "string-like" or "plate-like" that could be linked to a geometry dimension, and parameters such as "big" and "small" that could be linked to a size dimension in the object space. By proposing chord presets when multiple pitches are chosen, we believe the musician will have access to an interesting tool to control the combination of spectral components. The spectral content of complex sounds is often very dense and hence difficult to control intuitively. Being able to construct spectra from a musical approach (i.e., basic chord theory) attracts musicians and facilitates the complex task of structuring rich spectra.

Another possibility is to act directly on the sound quality, namely, on the timbre itself. In this case, we focus on the perceptual effects of the sound without taking into account physical aspects of the source. Thus, we can act directly on the timbre descriptors, such as the attack time, the spectral centroid, and the spectral flux (McAdams et al. 1995). The attack time is a measure for how quickly the energy envelope attains its maximum value, the spectral centroid is a measure of the spectral center of gravity and is directly related to the brightness of the sound, and the spectral flux is a measure for the variation of the spectral envelope over the duration of the note. Aspects of timbre can then be controlled by acting on a timbre space with two or three dimensions represented by different timbre descriptors. Such a control is not available currently, but it will be available in future versions.

In addition to the different control spaces linked to the material, size, shape, and timbre of the sounds, we have given the user the ability to morph between two different sounds ("morphing" in Figure 3). For this purpose, interpolations (linear or logarithmic) between the parameters of the two reference sounds are employed. This control possibility gives access to the creation of hybrid sounds, which for example makes it possible to simulate continuous transitions between different materials (Sound Examples 14 and 15, www.lma.cnrs-mrs.fr/~kronland/CMJ/sounds.html) or between different structures.

Conclusion

We have presented an efficient, hybrid synthesis technique for percussive sounds. This sound-synthesis model reproduces two main contributions characterizing the perceived material and the perceptual dimensions of the structure. A real-time implementation of the model showed its accuracy, allowing the generation of a wide variety of impact sounds. The system has been used in a musical context with a drum-like MIDI interface. As the drum interface itself offers limited controls, we have employed additional controllers (sliders, pedals, etc.) to act on different parameters, such as pitch and damping coefficients. To avoid additional controllers, the system has also been piloted by a MIDI keyboard that allows a direct control of the pitch and velocity and offers other control possibilities (e.g., aftertouch and pitch bend).

The adjustment of the model's parameters, however, is often difficult and necessitates the development of a mapping strategy. We have presented some mapping strategies, such as a morphing control, a material space, and an original approach to tune the complex sounds based on the standard-practice Western theory of harmony. This approach also makes the synthesis model an interesting research tool to investigate pitch perception of complex sounds. However, the choice of these strategies is left open to the composer, because it strongly influences the music that is to be written for this instrument.

This study is a first step toward a better understanding of the nature of percussive sounds, and especially toward a description of their most pertinent parameters from a perceptual and cognitive point of view. For this purpose, a larger, interdisciplinary project related to the semiotics of sounds associating sound modeling and neurosciences has been initiated.

In addition to the generation of synthesis sounds, we are also attempting to construct an analysis-synthesis platform. Actually, analysis-synthesis techniques that allow a given impact sound to be resynthesized have already been designed (Aramaki and Kronland-Martinet 2006). The association of nonlinear analysis-synthesis processes can allow the resynthesis of sounds generated by source-resonance systems, while perceptual and cognitive approaches will be proposed to study the influence of each synthesis parameter on listeners.

References

Adrien, J. M. 1991. "The Missing Link: Modal Synthesis." In G. De Poli, A. Piccialli, and C. Roads, eds. *Representations of Musical Signals.* Cambridge, Massachusetts: MIT Press, pp. 269–297.

Aramaki, M., and R. Kronland-Martinet. 2006. "Analysis-Synthesis of Impact Sounds by Real-Time Dynamic Filtering." *IEEE Transactions on Audio, Speech and Language Processing* 14(2):695–705.

Caracciolo, A., and C. Valette. 1995. "Damping Mechanisms Governing Plate Vibration." *Acta Acustica* 3:393–404.

Carello, C., K. L. Anderson, and A. J. Kunkler-Peck. 1998. "Perception of Object Length by Sound." *Psychological Science* 9(3):211–214.

Cook, P. R. 2002. *Real Sound Synthesis for Interactive Applications.* Natick, Massachusetts: AK Peters.

Essl, G., et al. 2004. "Theory of Banded Waveguides." *Computer Music Journal* 28(1):37–50.

Fletcher, H. 1964. "Normal Vibration Frequencies of a Stiff Piano String." *Journal of the Acoustical Society of America* 36(1):203–209.

Gobin, P., et al. 2004. "Designing Musical Interfaces with Composition in Mind." *Lecture Notes in Computer Science LNCS 2771.* Vienna: Springer, pp. 225–246.

Klatzky, R. L., D. K. Pai, and E. P. Krotkov. 2000. "Perception of Material from Contact Sounds." *Presence* 9(4): 399–410.

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

Lerdahl, F., and R. Jackendoff. 1977. "Toward a Formal Theory of Tonal Music." *Journal of Music Theory* 21:110–171.

McAdams, S., et al. 1995. "Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities, and Latent Subject Classes." *Psychological Research* 58:177–192.
Mourjopoulos, J. N., E. D. Kyriakis-Bitzaros, and C. E. Goutis. 1990. "Theory and Real-Time Implementation of Time-Varying Digital Audio Filters." *Journal of the Audio Engineering Society* 38(7/8):523–536.
Pai, D. K., et al. 2001. "Scanning Physical Interaction Behavior of 3D Objects." *Proceedings of SIGGRAPH 2001.* New York: Association for Computing Machinery, pp. 87–96.
Parncutt R. 1989. *Harmony: A Psychoacoustical Approach.* Vienna: Springer.
Rocchesso, D., R. Bresin, and M. Fernström. 2003. "Sounding Objects." *IEEE Multimedia* 10(2):42–52.
Smith, J. O., and S. A. Van Duyne. 1995. "Commuted Piano Synthesis." *Proceedings of the 1995 International Computer Music Conference.* San Francisco, California: International Computer Music Association, pp. 319–326.
Terhardt, E., G. Stoll, and M. Seewann. 1982. "Pitch of Complex Signals According to Virtual-Pitch Theory: Tests, Examples, and Predictions." *Journal of the Acoustical Society of America* 71:671–678.
Tucker, S., and G. J. Brown. 2002. "Investigating the Perception of the Size, Shape, and Material of Damped and Free Vibrating Plates." University of Sheffield, Department of Computer Science Technical Report CS-02-10.
Valette, C., and C. Cuesta. 1993. *Mécanique de la corde vibrante.* Lyon, France: Hermès.
van den Doel, K., P. G. Kry, and D. K. Pai. 2001. "FoleyAutomatic: Physically-Based Sound Effects for Interactive Simulation and Animation." *Proceedings of SIGGRAPH 2001.* New York: Association for Computing Machinery, pp. 537–544.
Van Duyne, S. A., and J. O. Smith. 1995. "Developments for the Commuted Piano." *Proceedings of the 1995 International Computer Music Conference.* San Francisco, California: International Computer Music Association, pp. 335–343.
Zwicker, E., and H. Fastl. 1990. *Psychoacoustics: Facts and Models.* Vienna: Springer.