Controlling the perceived material in an impact sound synthesizer

Mitsuko Aramaki, Mireille R Besson, Richard Kronland-Martinet, Sølvi Ystad

Controlling the Perceived Material in an Impact Sound Synthesizer

Mitsuko Aramaki, *Member, IEEE*, Mireille Besson, Richard Kronland-Martinet, *Senior Member, IEEE*, and Sølvi Ystad

***Abstract*—In this paper, we focused on the identification of the perceptual properties of impacted materials to provide an intuitive control of an impact sound synthesizer. To investigate such properties, impact sounds from everyday life objects, made of different materials (wood, metal and glass), were recorded and analyzed. These sounds were synthesized using an analysis–synthesis technique and tuned to the same chroma. Sound continua were created to simulate progressive transitions between materials. Sounds from these continua were then used in a categorization experiment to determine sound categories representative of each material (called *typical* sounds). We also examined changes in electrical brain activity (using event related potentials (ERPs) method) associated with the categorization of these typical sounds. Moreover, acoustic analysis was conducted to investigate the relevance of acoustic descriptors known to be relevant for both timbre perception and material identification. Both acoustic and electrophysiological data confirmed the importance of damping and highlighted the relevance of spectral content for material perception. Based on these findings, controls for damping and spectral shaping were tested in synthesis applications. A global control strategy, with a three-layer architecture, was proposed for the synthesizer allowing the user to intuitively navigate in a "material space" and defining impact sounds directly from the material label. A formal perceptual evaluation was finally conducted to validate the proposed control strategy.**

***Index Terms*—Analysis–synthesis, control, event related potentials, impact sounds, mapping, material, sound categorization, timbre.**

I. Introduction

THE current study describes the construction of a synthesizer dedicated to impact sounds that can be piloted using high-level verbal descriptors referring to material categories (i.e., wood, metal and glass). This issue is essential for sound design and virtual reality where sounds coherent with visual scenes are to be constructed. Control strategies for synthesis (also called mapping) is an important issue that has interested the computer music community ever since it became possible to produce music with computers [1]. A large number of interfaces and control strategies have been proposed by several authors [2]–[10]. Most of these interfaces were designed for musical purposes and are generally not adapted to build environmental sounds used in sound design and virtual reality. As opposed to music-oriented interfaces that generally focus on the control of acoustic factors such as pitch, loudness, or rhythmic deviations, a more intuitive control based on verbal descriptors that can be used by non-experts is needed in these new domains. This issue requires knowledge on acoustical properties of sounds and how they are perceived. As a first approach towards the design of such an environmental sound synthesizer, we focus on the class of impact sounds and on the control of the perceived material. In particular, our aim is to develop efficient mapping strategies between words referring to certain material categories (i.e., wood, metal and glass) and signal parameters to allow for an intuitive sound synthesis based on a smaller number of control parameters.

To point out perceptual properties that characterize the categories, a listening test was conducted. Stimuli were created first by recording impact sounds from everyday life objects made of different materials. Then, these recorded sounds were synthesized by analysis–synthesis techniques and tuned to the same chroma. Finally, we created continua from the tuned sounds to simulate progressive transitions between the categories by interpolating signal parameters. The use of sound continua was of interest to closely investigate transitions and limits between material categories. Sounds from these continua were used in a categorization task so as to be classified by participants as Wood, Metal, or Glass. From the percentage of responses, we determined sound categories representative of each material (called sets of *typical* sounds).

Then, we examined the acoustic characteristics that differ across typical Wood, Metal, and Glass sounds. For this purpose, we considered acoustic descriptors known to be relevant both for timbre perception and for material identification. Previous studies on the perception of sound categories have mainly been based on the notion of timbre. Several authors have used dissimilarity ratings to identify timbre spaces in which sounds from different musical instruments can be distinguished [11]–[14]. They found correlations between dimensions of these timbre spaces and acoustic descriptors such as attack time (the way the energy rises at the sound onset), spectral bandwidth (spectrum spread), or spectral centroid (center of gravity of the spectrum). More recently, roughness (distribution of interacting frequency components within the limits of a critical band) was considered

Manuscript received April 27, 2009; revised October 08, 2009; accepted March 18, 2010. Date of publication April 08, 2010; date of current version October 27, 2010. This work was supported by the Human Frontier Science Program under Grant HFSP #RGP0053 to M. Besson and a grant from the French National Research Agency (ANR, JC05-41996, "senSons") to S. Ystad. The work of M. Aramaki was supported by a postdoctoral grant, first from the HFSP and then from the ANR. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gaël Richard.

M. Aramaki and M. Besson are with CNRS-Institut de Neurosciences Cognitives de la Méditerranée, 13402 Marseille Cedex 20 France, and also with Aix-Marseille-Université, 13284 Marseille Cedex 07 France (e-mail: aramaki@incm.cnrs-mrs.fr; besson@incm.cnrs-mrs.fr).

R. Kronland-Martinet and S. Ystad are with the CNRS-Laboratoire de Mécanique et d'Acoustique, 13402 Marseille Cedex 20 France (e-mail: kronland@lma.cnrs-mrs.fr; ystad@lma.cnrs-mrs.fr).

Digital Object Identifier 10.1109/TASL.2010.2047755

as a relevant dimension of timbre since it is closely linked to the concept of consonance in a musical context [15], [16]. In the case of impact sounds, the perception of material seems mainly to correlate with the frequency-dependent damping of spectral components [17], [18] ([19], [20] in the case of struck bars), due to various loss mechanisms. Interestingly, damping remains a robust acoustic descriptor to identify macro-categories (i.e., between wood–Plexiglas and steel–glass categories) across variations in the size of objects [21]. From an acoustical point of view, a global characterization of the damping can be given by the sound decay measuring the decrease in sound energy as a function of time.

The above-mentioned timbre descriptors, namely attack time, spectral bandwidth, roughness, and normalized sound decay, were considered as potentially relevant signal features for the discrimination between sound categories. An acoustic analysis was conducted on these descriptors to investigate their relevance. At this stage, it is worth mentioning that signal descriptors that are found to be significant in traditional timbre studies may not be directly useful in the case of sound synthesis and control. Some descriptors might not give access to a sufficiently fine control of the perceived material. It might be necessary to act on a combination of descriptors. To more deeply investigate perceptual/cognitive aspects linked to the sound categorization, we exploited electrophysiological measurements for synthesis purposes since they provide complementary information regarding the nature of sound characteristics that contribute to the differentiation of material categories from a perceptual/cognitive point of view. In particular, we examined changes in brain electrical activity [using event related potentials (ERPs)] associated with the perception and categorization of typical sounds (we refer the reader to a related article for more details [22]).

Based on acoustic and electrophysiological results, sound characteristics relevant for an accurate evocation of material were determined and control strategies related to physical and perceptual considerations were proposed. The relevance of these strategies in terms of an intuitive manipulation of parameters was further tested in synthesis applications. High-level control was achieved through a calibration process to determine the range values of the damping parameters specific to each material category. In particular, the use of sound continua in the categorization experiment highlighted transition zones between categories that allowed for continuous control between different materials.

The paper is organized as follows: first the sound categorization experiment with stimuli construction and results is presented. Statistical analyses are further carried out on the set of sounds defined as *typical* to determine the acoustic descriptors that best discriminate sound categories. Then, sound characteristics that are relevant for material perception are obtained from physical considerations, timbre investigations and electrophysiological measurements. Control strategies allowing for an intuitive manipulation of these sound characteristics, based on these findings and on our previous works [23]–[25], are proposed in a three-layer control architecture providing the synthesis of impact sounds directly from the material label. A formal perceptual evaluation of the proposed control strategy is finally presented.

II. Sound Categorization Experiment

A. Participants

Twenty-five participants (13 women and 12 men, 19 to 35 years old, mean age $= 22.5$) were tested in this experiment that lasted for about one hour. They were all right-handed, nonmusicians (no formal musical training), had normal audition and no known neurological disorders. They all gave written consent and were paid to participate in the experiment.

B. Stimuli

We first recorded 15 sounds by impacting everyday life objects made of three different materials (wooden beams, metallic plates, glass bowls) that are five sounds per material. Synthetic versions of these recorded sounds were generated by an analysis–synthesis process and tuned to the same chroma. Then, we created J-step sound continua that simulate progressive transitions between two sounds of different materials by acting on amplitudes and damping parameters. The different stages of the stimuli construction are detailed below.

1) Analysis–Synthesis of Natural Sounds: Recordings of natural sounds were made in an acoustically treated studio of the laboratory using a microphone placed 1 m from the source. The objects from different materials were impacted by hand. We tried to control the impact on the object by using the same drumstick and the same impact force. The impact position on the different objects was chosen so that most modes were excited (near the center of the object for wooden beams and metallic plates; near the rim for glass bowls). Sounds were digitally recorded at 44.1-kHz sampling frequency.

From a physical point of view, the vibrations of an impacted object (under free oscillations) can generally be modeled as a sum of M exponentially damped sinusoids:

$$s(t) = \theta(t) \sum_{m=1}^{M} A_m \sin(\omega_m t + \Phi_m) e^{-\alpha_m t} \qquad (1)$$

where $\theta(t)$ is the Heaviside function and the parameters A_m, α_m, ω_m, and Φ_m, the amplitude, damping coefficient, frequency, and phase of the mth component, respectively. Based on the signal model corresponding to (1), we synthesized the recorded sounds at the same sampling frequency. Several different techniques allow precise estimating the signal parameters $\{A_m, \alpha_m, \omega_m\}_{m=1,\ldots,M}$ based on high-resolution analysis such as the Steiglitz–McBride technique [26] or more recently Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT), MUltiple SIgnal Classification (MUSIC), Least Squares or Maximum-Likehood techniques [27]–[30] (see also [31], [32]). These latter methods provide an accurate estimation and can be used to conduct spectral analysis. We here used a simplified analysis technique based on discrete Fourier transform (DFT) since we aimed at reproducing the main characteristics of the original sounds in terms of perceived material rather than achieving a perfect resynthesis.

The number of components M to synthesize was estimated from the modulus of the spectral representation of the signal.

Only the most prominent components, which amplitudes were larger than a threshold value fixed at 30 dB below the maximum amplitude of the spectrum, were synthesized. In addition, to keep the broadness of the original spectrum, we made sure that at least the most prominent component in each critical bandwidth was synthesized. Since Wood and Glass sounds had relatively poor spectra (i.e., few components), most of the components were synthesized. By contrast, Metal sounds had rich and broadband spectra. Some components were due to the nonlinear vibrations of the impacted object (favored by a low dissipation for Metal) and could not be reproduced by the signal model that only considers linear vibrations. Thus, the number of components for synthetic Metal sounds were generally inferior to the number of components of the original sound.

The frequency values ω_m were directly inferred from the abscissa of the local maxima corresponding to the prominent components. Since the spectrum was obtained by computing a fast Fourier transform (FFT) over 2^{16} samples, the frequency precision of each component was equal to 0.76 Hz ($= 44\,100/2^{16}$). Each component m was isolated using a gaussian window centered on the frequency ω_m. The frequency bandwidth of the gaussian window was adapted to numerically minimize the smoothing effects and to avoid the overlap of two successive components which causes interference effects. The gaussian window presents the advantage of preserving the exponential damping when convolved with an exponentially damped sine wave. Then, the analytic signal $\hat{s}_m(t)$ of the windowed signal was calculated using the Hilbert transform and the modulus of $\hat{s}_m(t)$ was modeled by an exponentially decaying function

$$|\hat{s}_m(t)| = A_m e^{-\alpha_m t} \tag{2}$$

Thus, by fitting the logarithm of $|\hat{s}_m(t)|$ with a polynomial function of degree 1 at best in a least-squares sense, the amplitude A_m was inferred from the ordinate at the origin while the damping coefficient α_m was inferred from the slope. Finally, the phases Φ_m were set to 0 for all components. This choice is commonly adopted in synthesis processes since it avoids undesirable clicks at sound onset. It is worth noticing that this phase adjustment does not affect the perception of the material because phase relationships between components mainly reflect the position of the microphone relative to the impacted object.

2) Tuning: The pitches of the 15 synthetic sounds (five per material category) differed since they resulted from impacts on various objects. Consequently, sounds were tuned to the same chroma to minimize pitch variations. Tuning is needed to build homogeneous sound continua with respect to pitch (Section II-B4) and to accurately investigate acoustic descriptors (Section III). In particular, the relationships between descriptors will be better interpreted if they are computed on a set of tuned sounds with normalized pitches rather than on a set of sounds with various pitch values.

We first defined the initial pitch of the sounds from informal listening tests: four participants (different from those who participated in the categorization experiment) listened to each sound and were asked to evaluate the pitch by playing the matching note on a piano keyboard. For each sound, the pitch was defined by the note that was most often associated with the sound. Thus, we defined the pitches $G\sharp 3$ (fundamental frequency of 415.30 Hz), $C\sharp 3$ (277.18 Hz), $F\sharp 3$ (369.99 Hz), $C\sharp 3$, and $C\sharp 3$ for the 5 Wood sounds; $A3$ (440.00 Hz), $F\sharp 3$, $D5$ (1174.65 Hz), $E4$ (659.25 Hz), and $E3$ (329.62 Hz) for the 5 Metal sounds, and $C5$ (1046.50 Hz), $E6$ (2637.02 Hz), $C\sharp 6$ (2217.46 Hz), $D5$, and $F5$ (1396.91 Hz) for the 5 Glass sounds. Then, we tuned the sounds to the closest note C with respect to the initial pitch to minimize signal transformations applied on the sounds: Wood sounds were tuned to the pitch C3, Metal sounds to C3 and C4 and Glass sounds to C5 and C6. Therefore, sounds differed by 1, 2, or 3 octaves depending upon the material. Based upon previous results showing high similarity ratings for tone pairs that differed by octaves [33], an effect known as the octave equivalence, we assume that the octave differences between sounds belonging to a same category should have little influence on sound categorization.

In practice, tuned sounds were generated using the previous synthesis technique [(1)]. The amplitudes and phases of components were kept unchanged but the frequencies (noted $\tilde{\omega}_m$ for tuned sounds) and damping coefficients (noted $\tilde{\alpha}_m$) were recalculated as follows. The tuned frequencies $\tilde{\omega}_m$ were obtained by transposing original ones $\{\omega_m\}_{m=1,\ldots,M}$ with a dilation factor η defined from the fundamental frequency values (in Hz), noted F and $\tilde{F}$, of the sound pitches before and after tuning, respectively,

$$\tilde{\omega}_m = \eta\omega_m \quad \text{with} \quad \eta = \frac{\tilde{F}}{F}. \tag{3}$$

The damping coefficient $\tilde{\alpha}_m$ of each tuned component was recalculated by taking into account the frequency-dependency of the damping. For instance, it is known that in case of wooden bars, the damping coefficients increase with frequency following an empirical expression of a parabolic form where parameters depend on the wood species [34]–[36]. To achieve our objectives, we defined a general expression of a damping law $\alpha(\omega)$ chosen as an exponential function

$$\alpha(\omega) = e^{(\alpha_G + \alpha_R \omega)}. \tag{4}$$

The exponential expression presents the advantage of easily fitting various and realistic damping profiles with a reduced number of parameters. $\alpha(\omega)$ is defined by two parameters α_G and α_R characteristic of the intrinsic properties of the material. The parameter α_G reflects global damping and the parameter α_R reflects frequency-relative damping (i.e., difference between high-frequency component damping and low-frequency component damping). Thus, a damping law $\alpha(\omega)$ was estimated on the original sound by fitting the damping coefficients $\{\alpha_m\}_{m=1,\ldots,M}$ with the (4) at best in a least-squares sense. Then, the damping coefficient $\tilde{\alpha}_m$ of the mth tuned component was recalculated according to this damping law (see also [37])

$$\tilde{\alpha}_m = \alpha(\tilde{\omega}_m). \tag{5}$$

3) Gain Adjustment: Sounds were equalized by gain adjustments to avoid the influence of loudness in the categorization judgments. The gain adjustments were determined on the basis of a pretest with four participants (different from those who participated in the categorization experiment). They were asked to balance the loudness level of the tuned sounds. These

tuned sounds were previously normalized by a gain of reference $\Gamma_0 = 1.5 \times A$ with A corresponding to the largest value of the maxima of the signal modulus among the 15 tuned sounds. The coefficient 1.5 is a safety coefficient commonly used in gain adjustment tests to avoid the saturation of the signals after the adjustment. The gain values Γ to be applied on the 5 Wood sounds were equal to [70, 20, 30, 15, 30], on the 5 Metal sounds were equal to [3.5, 1.1, 1, 1.5, 1.3] and on the five Glass sounds were equal to [35, 15, 15, 30, 10].

Finally, the four participants were asked to evaluate the final sounds in terms of perceived material. Results showed that sounds were categorized in the same material category as the original sounds by all participants thereby showing that the main characteristics of the material were preserved.

4) Sound Continua: To closely investigate transitions between material categories, we created 15 J-step sound continua noted Ω_i with five continua for each material transition. The five Wood-Metal continua were indexed from Ω_1 to Ω_5, the five Wood–Glass continua from Ω_6 to Ω_{10} and finally, the five Glass-Metal continua from Ω_{11} to Ω_{15}. Each continuum was composed of 22 hybrid sounds ($J = 22$) that were obtained by mixing the spectra and by interpolating the damping laws of the two extreme sounds. We chose to mix spectra to fix the values of the frequency components which allows minimizing pitch variations across sounds within a continuum (it is known that shifting components modifies pitch). We chose to interpolate damping laws to gradually modify the damping that conveys fundamental information on material perception. Thus, the sound H_j(t) at step j of the continuum is expressed by

$$H_j(t) = \gamma_1(j)\frac{\Gamma_1}{\Gamma_0}\sum_{m=1}^{M} A_m \sin(\omega_m t)e^{-\alpha^j(\omega_m)t} + \gamma_2(j)\frac{\Gamma_2}{\Gamma_0}\sum_{n=1}^{N} A_n \sin(\omega_n t)e^{-\alpha^j(\omega_n)t} \quad (6)$$

where $\{A_m, \omega_m\}_{m=1,\ldots,M}$ and $\{A_n, \omega_n\}_{n=1,\ldots,N}$ correspond to the sets of amplitudes and frequencies of the two extreme sounds and j varies from 1 to 22. The gains Γ_1 and Γ_2 correspond to the gains of the extreme sounds defined from the gain adjustment test according to a gain of reference Γ_0 (see Section II-B2). The gains $\gamma_1(j)$ and $\gamma_2(j)$ vary at each step j on a logarithmic scale, according to the dB scale

$$\gamma_1(j) = 1 - \frac{\log(j)}{\log(J)}$$
$$\gamma_2(j) = 1 - \frac{\log(J-j+1)}{\log(J)}. \quad (7)$$

The damping variation along the continua is computed by interpolating the damping parameters α_G and α_R of the damping law [defined in (4)] estimated on the two extreme sounds (located at step $j = 1$ and $j = 22$, respectively), leading to the determination of a hybrid damping law $\alpha^j(\omega)$ that progressively varies at each step j of the continuum (see Fig. 1)

$$\alpha^j(\omega) = e^{(\alpha_G^j + \alpha_R^j \omega)} \quad (8)$$

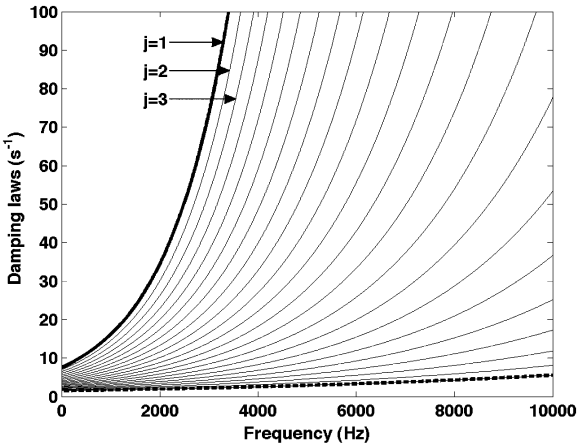


Fig. 1. Damping laws $\alpha^j(\omega)$ for $j = 1, \ldots, 22$ as a function of frequency corresponding to a Wood–Metal continuum. Bold curves correspond to damping laws of Wood (in bold plain) and Metal (in bold dashed) sounds at the extreme positions.

with

$$\alpha_G^j = \left(\alpha_G^{22} - \alpha_G^1\right)\frac{j-1}{J-1} + \alpha_G^1$$
$$\alpha_R^j = \left(\alpha_R^{22} - \alpha_R^1\right)\frac{j-1}{J-1} + \alpha_R^1. \quad (9)$$

The use of an interpolation process on the damping allowed for a better merging between the extreme sounds since the spectral components of the two spectra are damped following the same damping law $\alpha^j(\omega)$. As a consequence, hybrid sounds (in particular, at centered positions of the continua) differed from sounds obtained by only mixing the extreme sounds.

The obtained sounds had different signal lengths (Metal sounds are longer than Wood or Glass sounds). To restrain the lengths to a maximum of 2 seconds, sound amplitudes were smoothly dropped off by multiplying the temporal signal with the half decreasing part of a Hann window.

A total of 330 sounds were created. The whole set of sounds are available at [38]. The averaged sound duration was 861 ms for all sounds and 1053 ms in the Wood–Metal continua, 449 ms in the Wood-Glass continua, and 1081 ms in the Glass–Metal continua.

C. Procedure

The experiment was conducted in a quiet Faradized (electrically shielded) room. Sounds were presented once (i.e., no repetition of the same sound) in random order through one loudspeaker (Tannoy S800) located 1 m in front of the participant. Participants were asked to categorize sounds as Wood, Metal, or Glass, as fast as possible, by pressing one response button out of three on a three-buttons response box[1] (right, middle, and left buttons; one button per material category label). The association between response buttons and material categories was balanced across participants to avoid any bias linked with the

[1]Since participants were not given the option to choose that sounds did not belong to either one of these three categories, results may be biased, but this potential ambiguity would be raised only for intermediate sounds.

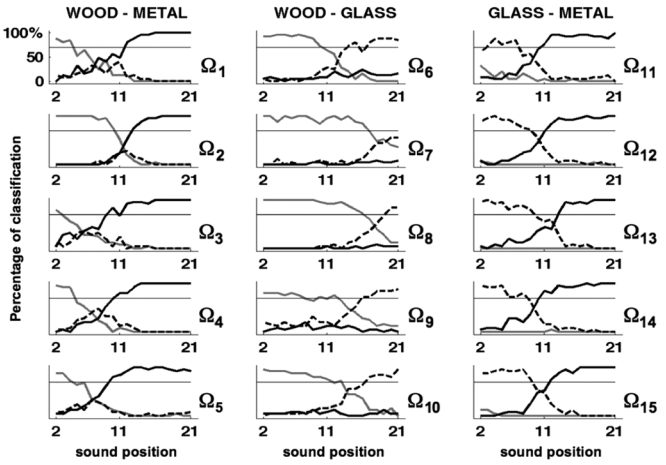


Fig. 2. Percentage of classification as Wood (gray curves), Metal (black curves), or Glass (dashed curves) for each sound as a function of its position j on the continuum for the 15 continua Ω_i. Sounds were considered as typical if they were classified in one category by more than 70% of participants (threshold represented by an horizontal line). No data were collected for extreme sounds ($j = 1$ and $j = 22$) since they were used in the training session.

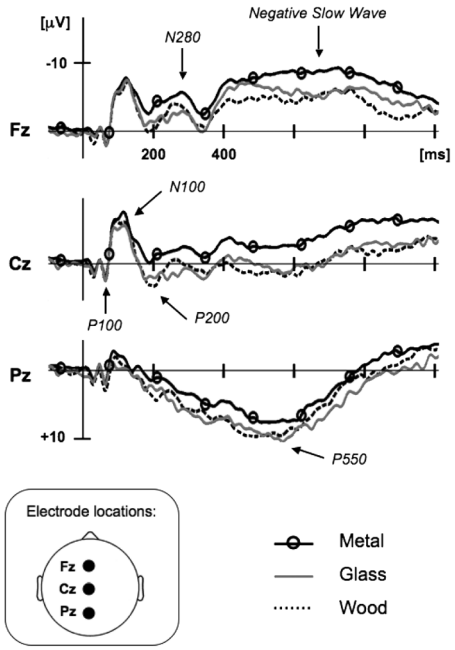


Fig. 3. ERPs to typical sounds of Wood (dotted), Metal (circle marker), and Glass (gray) at electrodes (Fz, Cz, Pz) along the midline of the scalp (see [22] for lateral electrodes). The amplitude (in microvolts) is represented on the ordinate and negativity is up. The time from sound onset is on the abscissa (in milliseconds).

order of the buttons. A row of 4 crosses, i.e., "XXXX," was presented on the screen 2500 ms after sound onset during 1500 ms to give participants time to blink. The next sound was presented after a 500-ms silence. Participants' responses were collected for all sounds except for the extreme sounds since they were used in the training session. The electrical brain activity (Electroencephalogram, EEG) was recorded continuously for each participant during the categorization task.

D. Results

1) Behavioral Data: Percentages of categorization as Wood, Metal, or Glass were obtained for each sound by averaging responses across participants. Fig. 2 allows visualization of these data as a function of sound position along the continua. From these data, we determined a set of *typical* sounds for each material category: sounds were considered as *typical* if they were categorized as Wood, Metal, or Glass by more than 70% of participants (we refer the reader to [39] for more details on the determination of this threshold of percentage value). In addition, sound positions delimiting categories along the continua can be defined from the position ranges of typical sounds. Note that due to the different acoustic characteristics of sounds, the category limits are not located at the same position for all continua (see Fig. 2).

2) Electrophysiological Data: We examined ERPs time-locked to sound onset to analyze the different stages of information processing as they unfold in real time.[2] ERP data were averaged separately for typical sounds of each material category (Fig. 3). We summarize here the main findings (we refer the reader to a related article for more details [22]).

Typical sounds from each category elicited small P100 components, with maximum amplitude around 65-ms post-sound onset, large N100, and P200 components followed by N280 components and Negative Slow Wave (NSW) at fronto-central sites or large P550 components at parietal sites. Statistical analyses revealed no significant differences on the P100 and N100 components as a function of material categories. By contrast, they showed that typical Metal sounds elicited smaller P200 and P550 components, and larger N280 and NSW components than typical sounds from Wood and Glass. From an acoustic point of view, Metal sounds have richer spectra and longer durations (i.e., lower damping) than Wood and Glass sounds. The early differences on the P200 components most likely reflect the processing of spectral complexity (see [42] and [43]) while the later differences on the N280 and on the NSW are likely to reflect differences in sound duration (i.e., differences in damping; see [44] and [45]).

III. Acoustic Analysis

The typical sounds as defined based upon behavioral data (Section II-D1) form a set of sounds representative of each material category. To characterize these typical sounds from an acoustical point of view, we investigated the following descriptors: attack time (AT), spectral bandwidth (SB), roughness (R), and normalized sound decay (α) that are defined below. Then, we examined the relationships between acoustic descriptors and their relevance to discriminate material categories.

A. Definition of Acoustic Descriptors

Attack time is a temporal timbre descriptor which characterizes signal onset. It is defined by the time (in second) necessary

[2]The ERPs elicited by a stimulus (a sound, a light, etc.) are characterized by a succession of positive (P) and negative (N) deflections relative to a baseline (usually measured within the 100 ms or 200 ms that precedes stimulus onset). These deflections (called components) are characterized by their polarity, their latency of maximum amplitude (relative to stimulus onset), their distribution across different electrodes located at standard positions on the scalp and by their functional significance. Typically, the P100, N100, and P200 components reflect the sensory and perceptual stages of information processing, and are obligatory responses to the stimulation [40], [41]. Then, depending on the experimental design and on the task at hand, different late ERP components are elicited (N200, P300, N400, etc.).

for the signal energy to raise from a threshold level to the maximum energy in the temporal envelope (for percussive sound) or to the sustained part (for a sustained sound with no decay part) [46], [47]. Different values have been proposed in the literature for both minimum and maximum thresholds. For our concern, we chose to compute the attack time from 10% to 90% of the maximum amplitude of the temporal envelope as in [48]. This descriptor is known to be relevant to distinguish different classes of instrumental sounds. For instance, sounds from percussive and woodwind instruments have respectively short and long AT.

Spectral bandwidth (in Hz), commonly associated with the spectrum spread, is defined by [49]

$$SB = \frac{1}{2\pi}\sqrt{\frac{\sum_k |\hat{s}(k)|\left(\omega(k) - 2\pi \times SC\right)^2}{\sum_k |\hat{s}(k)|}} \tag{10}$$

where SC is the spectral centroid (in Hz) defined by [50]

$$SC = \frac{1}{2\pi}\frac{\sum_k \omega(k)\,|\hat{s}(k)|}{\sum_k |\hat{s}(k)|} \tag{11}$$

and where ω represents frequency, $\hat{s}$ the Fourier transform of the signal estimated using the FFT algorithm and k the FFT bin index. The FFT was calculated on 2^{16} samples.

Roughness (in asper) is commonly associated with the presence of several frequency components within the limits of a critical band. From a perceptual point of view, roughness is correlated with tonal consonance based on results from experiments on consonance judgments conducted by [51]. From a signal point of view, [52] have shown that roughness and fluctuation strength are proportional to the square of the modulation factor of an amplitude modulated pure tone. We computed roughness based on Vassilakis's model by summing up the partial roughness r_{mn} for all pairs of frequency components contained in the sound [53]

$$r_{mn} = 0.5(A_m A_n)^{0.1} \times \left(\frac{2\min(A_m, A_n)}{A_m + A_n}\right)^{3.11} \times \left(e^{-3.5v|\omega_m - \omega_n|} - e^{-5.75v|\omega_m - \omega_n|}\right) \tag{12}$$

with

$$v = \frac{0.24}{0.0207 \times \min(\omega_m, \omega_n) + 2\pi \times 18.96}$$

and where A_m and A_n are amplitudes and ω_m and ω_n are the frequencies of components m and n, respectively.

Finally, the sound decay D (in s^{-1}) quantifies the amplitude decrease of the whole temporal signal and globally characterizes the damping in the case of impact sounds. In particular, D approximately corresponds to the decay of the spectral component with the longest duration (i.e., generally the lowest frequency one). The sound decay is directly estimated by the slope of the logarithm of the temporal signal envelope. This envelope is given by calculating the analytic signal using the Hilbert transform and by filtering the modulus of this analytic signal using a second-order low-pass Butterworth filter with cutoff frequency of 50 Hz [36]. Since damping is frequency dependent (Section II-B1), sound decay depends on the spectral content of the sound. Consequently, we considered a normalized sound decay denoted α with respect to the spectral localization of the energy and we defined the dimensional descriptor α as the ratio of the sound decay D to the SC value

$$\alpha = \frac{D}{SC}. \tag{13}$$

TABLE I
COEFFICIENTS OF DETERMINATION BETWEEN THE ATTACK TIME AT, THE SPECTRAL BANDWIDTH SB, THE ROUGHNESS R, AND THE NORMALIZED SOUND DECAY α. SINCE THE MATRIX IS SYMMETRIC, ONLY THE UPPER PART IS REPORTED. THE P-VALUES ARE ALSO REPORTED BY *** ($p < .001$) WHEN COEFFICIENTS ARE SIGNIFICANT (WITH BONFERRONI ADJUSTMENT)

	AT	SB	R	α
AT	1	0.05^{***}	0.07^{***}	0
SB	–	1	0.25^{***}	0.02
R	–	–	1	0.23^{***}
α	–	–	–	1

B. Relationships Between Acoustic Descriptors

As a first step, we examined the relationships between the acoustic descriptors estimated on typical sounds. Table I shows the coefficients of determination that are the square of the Bravais–Pearson coefficients between pairs of descriptors. We found highest significant correlation (although not high in terms of absolute value) between the two spectral descriptors SB and R. Lowest correlations were found between AT and the other ones, reflecting the fact that sound onset has little influence on the spectral characteristics and does not depend on the decaying part of the sound (described by α).

Second, a principal component analysis (PCA) was conducted on standardized values of acoustic descriptors (i.e., values centered on the mean value and scaled by the standard deviation value). Results showed that the first two principal components explained about 72% of the total variance (the first component alone explained about 48%). As shown in Fig. 4, the first component was mainly correlated to the spectral descriptors (SB and R) and the second component to the temporal descriptors (AT and α). Thus, PCA revealed that sounds could reliably be represented in a reduced bi-dimensional space which orthogonal axes are mainly correlated to spectral (Component I) and temporal descriptors (Component II), respectively. This result confirmed that spectral and temporal descriptors bring complementary information on the sound characterization from an acoustic point of view.

C. Discrimination Between Material Categories

We examined the relevance of acoustic descriptors to discriminate material categories using a discriminant canonical analysis. This analysis was conducted using Materials (Wood, Metal, and Glass) as groups and standardized values of acoustic descriptors $\{\mathrm{AT}, \mathrm{SB}, \mathrm{R}, \alpha\}$ as independent variables. Since three sound categories were considered, two discriminant functions that allow for the clearest separation between sound categories were computed (the number of discriminant functions is equal to the number of groups minus one). These

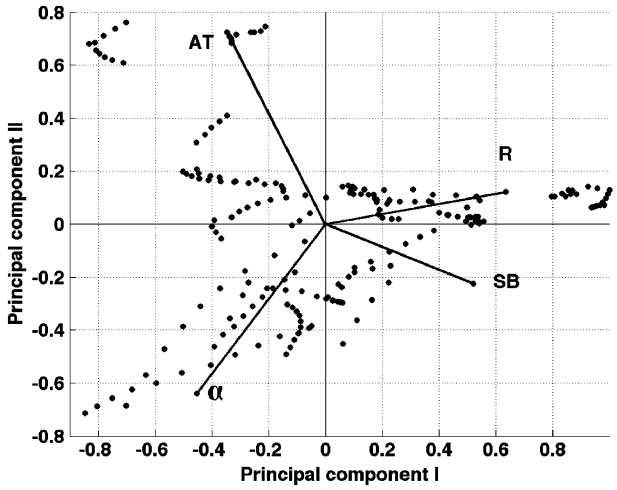


Fig. 4. Biplot visualization: the observations corresponding to typical sounds are represented by dots and the acoustic descriptors by vectors. The contribution of descriptors to each Principal Component (PC) can be quantified by the R^2 statistics given by a regression analysis: Attack time AT ($R^2 = .23$ for PC I and $R^2 = .52$ for PC II), Spectral bandwidth SB ($R^2 = .52$ for PC I and $R^2 = .05$ for PC II), Roughness R ($R^2 = .77$ for PC I and $R^2 = .01$ for PC II) and Normalized sound decay α ($R^2 = .39$ for PC I and $R^2 = .40$ for PC II).

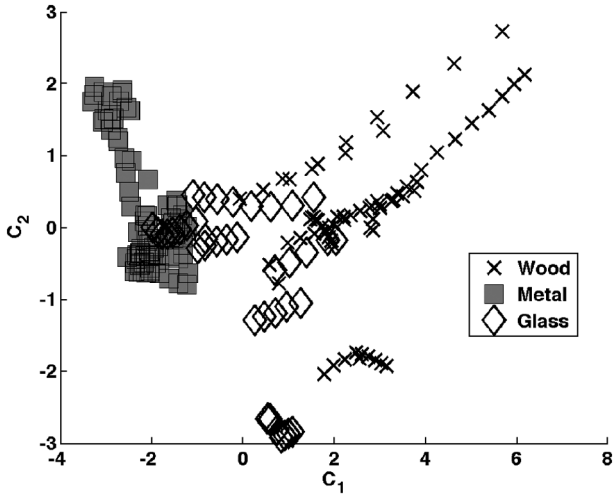


Fig. 5. Scatter plot of the canonical variables allowing the clearest separation between typical Wood ($\times$), Metal ($\square$), and Glass ($\diamond$) sound categories.

functions C_1 and C_2 were expressed as a combination of the independent variables

$$\begin{aligned} C_1 &= 1.04\alpha + 0.76SB - 0.58R + 0.47AT \\ C_2 &= 0.70R - 0.15SB - 0.56AT + 0.38\alpha. \end{aligned} \tag{14}$$

The Wilks's Lambda show that both functions C_1 (Wilks's $\Lambda = .15$; $\chi^2 = 366.65$; $p < .001$) and C_2 (Wilks's $\Lambda = .87$; $\chi^2 = 28.02$; $p < .001$) are significant. The first function C_1 explains 96% of the variance (coefficient of determination $= 0.82$) while the second function C_2 explains the remaining variance (coefficient of determination $= 0.13$). The coefficient associated with each descriptor indicates its relative contribution to the discriminating function. In particular, the first function C_1 is mainly related to α and allows clear distinction particularly between typical Wood and Metal sounds as shown in Fig. 5. This result is in line with previous studies showing that damping is a fundamental cue in the perception of sounds from impacted materials (see the Introduction). The second axis C_2 is mainly related to the spectral descriptor R and allows for a distinction of Glass sounds.

IV. Control Strategy for the Synthesizer

Results from acoustic and electrophysiological data are now discussed in the perspective of designing an intuitive control of the perceived material in an impact sound synthesizer. In particular, we aim at determining relevant sound characteristics for an accurate evocation of different materials and at proposing intuitive control strategies associated with these characteristics. In practice, the synthesis engine and the control strategies were implemented using Max/MSP [54] thereby allowing for the manipulation of parameters in real-time and consequently, providing an easy way to evaluate the proposed controls. The observations and conclusions from these synthesis applications are also reported.

A. Determination of Sound Characteristics

As a starting point, results from acoustic analysis revealed that α (characterizing the damping) was the most relevant descriptor to discriminate material categories, therefore confirming several findings in the literature on the relevance of damping for material identification (see the Introduction). Thus, damping was kept as a relevant sound characteristic to control in the synthesizer.

Furthermore, acoustic analysis showed that in addition to the damping, a second dimension related to spectral characteristics of sounds was significant, in particular for the distinction between Glass and Metal sounds. Interestingly, this result is supported by electrophysiological data that revealed ERP differences between Metal on one side and both Glass and Wood sounds on the other side. Since these differences were interpreted as reflecting processing of sound duration (related to damping) and spectral content, ERP data showed the relevance of both these aspects for material perception. Thus, from a general point of view, it is relevant to assume that material perception seems to be guided by additional cues (other than damping) that are most likely linked to the spectral content of sounds. This assumption is in line with several studies showing the material categorization can be affected by spectral characteristics, in particular, Glass being associated with higher frequencies than Metal sounds [20], [55].

In line with this assumption, synthesis applications confirmed that damping was relevant but was not in some cases sufficient to achieve material categories. For instance, it was not possible to transform a given Wood or Glass sound into a Metal sound by only applying a Metal damping on a Wood or Glass spectrum. The resulting sound did not sound metallic enough. It was therefore necessary to also modify the spectral content, and in particular the spectral distribution, to emphasize the metallic aspect of the sounds (examples in [38]). We found another limitation of damping in the case of Glass synthesis. Indeed, Glass sounds match a wide range of damping values (from highly damped

jar sounds to highly resonant crystal glass sounds) and are most often characterized by a sparse distribution of spectral components (i.e., few distinct components). These observations indicated that the perceptual distinction between Glass and Metal may be due to the typical dissonant aspect of Metal and can be accurately reflected by the roughness that was highlighted as the most relevant descriptor in the acoustic analysis after the damping. Thus, we concluded on the necessity to take into account a control of the spectral shape, in addition to the control of damping, for a more accurate evocation of the perceived material.

Besides, electrophysiological data provided complementary information regarding the temporal dynamics of the brain processes associated with the perception and categorization of typical sounds. First, it is known that P100 and N100 components are influenced by variations in sound onset parameters [56]. The lack of differences on these ERP components is taken to indicate similar brain processes for all typical sounds, showing that the information of the perceived material does not lie in the sound onset. As a synthesis outcome, it means that a modification of sound onset does not affect the nature of the perceived material and consequently, that AT is not a relevant parameter for the control of the perceived material. Second, it is known that N100 component is also influenced by pitch variations [40]. While octave differences were largest between Glass and the other two categories, the lack of differences on N100 component is taken to indicate that pitch may not affect sound categorization. Thus, electrophysiological data support our previous assumption concerning the weak influence linked to the octave differences on sound categorization (Section II-B2).

Based on these considerations, damping and spectral shaping were determined as relevant sound characteristics for material perception. Control strategies associated with these characteristics are proposed and detailed in the following sections.

B. Control of damping

The control of damping was designed by acting on parameters α_G and α_R of the damping law (4). This control gave an accurate manipulation of sound dynamics (time evolution) with a reduced number of control parameters. In particular, the parameter α_G governed the global sound decay (quantified by the descriptor D, Section III) and the parameter α_R allowed controlling the damping differently for high- and low-frequency components. This control made it possible to synthesize a wide variety of realistic sounds. Since from a physical point of view, high frequency components generally are more heavily damped than low frequency ones, we expected both parameters α_G and α_R to be positive in the case of natural sounds.

C. Control of Spectral Shaping

We here propose two control strategies. The first one relied on spectral dilation and was based on physical considerations, in particular on the dispersion phenomena. The second control was based on amplitude and frequency modulations and relied on adding components to the original spectrum. This latter control had a smaller influence on pitch compared with the first control since the original spectrum is not distorted. It also has interesting perceptual consequences. For instance, by creating components within a specific frequency band (i.e., critical band of hearing), this control specifically influenced the perception of roughness that was highlighted as a relevant acoustic descriptor in the acoustic analysis (Section III). These two control strategies are detailed in the following sections.

1) Control by Spectral Dilation: From the analysis of natural sounds, and from physical models describing wave propagation in various media (i.e., various physical materials), two important phenomena can be observed: dispersion and dissipation [32], [57]. Dissipation is due to various loss mechanisms and is directly linked to the damping parameters (α_G and α_R) as described above. Dispersion is linked to the fact that the wave propagation speed varies with respect to frequency. This phenomenon occurs when the phase velocity of a wave is not constant and introduces inharmonicity in the spectrum of the corresponding sound. An example of dispersive medium is the stiff string for which the mth partial is not located at $m\omega_1$ but at $m\omega_1\sqrt{1+\beta m^2}$ where ω_1 is the fundamental frequency and β the coefficient of inharmonicity depending on the physical parameters of the string [58]. We based our first spectral shaping strategy on the spectral dilation defined by

$$\tilde{\omega}_m = W(\breve{\omega}_{\min}, \breve{\omega}_{\max}, \omega)\breve{\omega}_m + (1 - W(\omega_{\min}, \omega_{\max}, \omega))\,\omega_m \tag{15}$$

where W is a window function (defined later in the text) and

$$\breve{\omega}_m = S_G \omega_m \sqrt{1 + S_R \left(\frac{\omega_m}{\omega_1}\right)^2} \tag{16}$$

with ω_m and $\tilde{\omega}_m$ that correspond to the frequency of the initial and shifted component of rank m, respectively. Equation (16) is a generalization of the inharmonicity law previously defined for stiff strings so that the expression is not limited to harmonic sounds but can be applied to any set of frequencies. S_G and S_R are defined as the global and relative shaping parameters, respectively. Ranges of S_G and S_R are constrained so that $\breve{\omega}_m$ are real-valued and $\breve{\omega}_m > \omega_1$ for all $m = 1, \ldots, M$ with M the number of components. Thus, S_R should be lower bounded

$$\min S_R = -\frac{1}{M^2} \tag{17}$$

and S_G should satisfy

$$S_G \frac{\omega_m}{\omega_1} \sqrt{1 + S_R \left(\frac{\omega_m}{\omega_1}\right)^2} > 1 \quad \text{for all } m = 1, \ldots, M. \tag{18}$$

A window function $W(\omega_{\min}, \omega_{\max}, \omega)$ provided a local control of spectral shaping within a given frequency range $[\omega_{\min}; \omega_{\max}]$. In particular, it was of interest to keep the first components unchanged during spectral control to reduce pitch variations. For instance, a window function $W(\omega_3, F_s/2, \omega)$ where F_s is the sampling frequency can be applied to only act on frequencies higher than ω_2. In practice, we chose a Tukey (tapered cosine) window defined between $\omega_{\min}$ and $\omega_{\max}$. The window is parameterized by a ratio ρ (between 0 and 1) allowing the user to choose intermediate profiles from rectangular $(\rho = 0)$ to Hann $(\rho = 1)$ windows. Consequently, the user is able to act on the weight of the local control.

From an acoustic point of view, the control acts on the spectral descriptors SB and R in a global way. For example, a decrease of the S_G value leads to a decrease of the SB value and at the same time to an increase of the R value.

2) Control by Amplitude and Frequency Modulations: Amplitude modulation creates two components on both sides of the original one and the modulated output waveform is expressed by

$$\begin{aligned} d_m^{\mathrm{AM}}(t) &= A_m\left(1 + I\cos(\omega_n t)\right)\cos(\omega_m t) \\ &= A_m\cos(\omega_m t) + \frac{A_m I}{2}\cos\left((\omega_m + \omega_n)t\right) \\ &\quad + \frac{A_m I}{2}\cos\left((\omega_m - \omega_n)t\right) \end{aligned}$$

where $I \in [0,1]$ is the modulation index, ω_n the modulating frequency, A_m the amplitude, and ω_m the frequency of the mth component.

Frequency modulation creates a set of components on both sides of the original one and the modulated output waveform is expressed by

$$\begin{aligned} d_m^{\mathrm{FM}}(t) &= A_m\cos\left(\omega_m t + I\sin(\omega_n t)\right) \\ &= A_m\sum_{k=-\infty}^{\infty} J_k(I)\cos\left((\omega_m + k\omega_n)t\right) \end{aligned}$$

where $k \in N$ and $J_k(I)$ is the Bessel function of order k. The amplitude of these additional components are given by the amplitude of the original partial A_m and the values of $J_k(I)$ for a given modulation index I.

For both amplitude and frequency modulations, synthesis applications showed that applying the same value of the modulating frequency ω_n to all components led to synthetic sounds perceived as too artificial. To avoid this effect, we proposed a definition of the modulating frequency $\omega_{n,m}$ for each spectral component m based on perceptual considerations. Thus, $\omega_{n,m}$ was expressed as a percentage of the critical bandwidth Δf_m associated with each component m [59]

$$\Delta f_m = 25 + 75\left(1 + 1.4 f_m^2\right)^{0.69} \tag{19}$$

where f_m is expressed in kHz. Since Δf_m increases with respect to frequency, components created at high frequencies are more distant (in frequency) on both sides of the central component than components created at low frequencies. This provided an efficient way to control roughness since the addition of components within a critical bandwidth increases the perception of roughness. In particular, it is known that the maximum sensory dissonance corresponds to an interval between spectral components of about 25% of the critical bandwidth [51], [60].

Synthesis applications showed that both spectral shaping controls allowed for morphing particularly between Glass and Metal sounds while keeping the damping unchanged. In this case, the damping coefficients of the modified frequencies were recalculated according to the damping law. Both controls provided a local control since modifications can be applied on each original component independently. The control based on amplitude and frequency modulations allowed subtle spectral modifications compared with the control based on spectral dilation and in particular, led to interesting fine timbre effects such as cracked glass sounds (sound examples can be found in [38]).

D. Control of the Perceived Material

A global control strategy of the perceived material that integrates the previous damping and spectral shaping controls is proposed in this section. This strategy is hierarchically built on three layers: the "Material space" (accessible to the user), the "Damping and Spectral shaping parameters" and the "Signal parameters" (related to the signal model). Note that the mapping strategy does not depend on the synthesis technique. As a consequence, the proposed control can be applied to any sound generation process.[3] Note also that the proposed strategy is not unique and represents one among several other possibilities [24], [25].

Fig. 6 illustrates the mapping between these three layers based on the first spectral shaping control (using S_G and S_R). The Material space is designed as a unit disk of center C with three fixed points corresponding to the three reference sounds (Wood, Metal, and Glass) equally distributed along the external circle. The Glass sound position is arbitrarily considered as the angle's origin ($\theta = 0$) and consequently, the Metal sound is positioned at $\theta = 2\pi/3$ and the Wood sound at $\theta = 4\pi/3$. The three reference sounds were synthesized from the same initial set of harmonic components (fundamental frequency of 500 Hz and 40 components) so that Wood, Metal, and Glass sounds were obtained by only modifying the damping and spectral shaping parameters (values given in Table II and sound positions shown in Fig. 6). These parameters were chosen on the basis of the sound quality of the evoked material. Note that the reference sounds could be replaced by other sounds.

The user navigates in the Material space between Wood, Metal, and Glass sounds by moving a cursor and can synthesize the sound corresponding to any position. When moving along the circumference of the Material space circle, the corresponding sound $S_h(\theta)$ characterized by its angle θ is generated with Damping and Spectral shaping parameters defined by

$$\mathbf{P}_{S_h}(\theta) = T(\theta)\mathbf{P}_G + T\left(\theta - \frac{2\pi}{3}\right)\mathbf{P}_M + T\left(\theta - \frac{4\pi}{3}\right)\mathbf{P}_W \tag{20}$$

where $\mathbf{P}$ represent the parameter vector $\{\alpha_G, \alpha_R, S_G, S_R\}$ of the sound S_h and of the reference sound of Glass (G), Metal (M), and Wood (W). The function $T(\theta)$ was defined so that the interpolation process was exclusively made between two reference sounds at a time

$$T(\theta)\begin{cases} = -\frac{3}{2\pi}\theta + 1, & \text{for } \theta \in \left[0; \frac{2\pi}{3}\right[\\ = 0, & \text{for } \theta \in \left[\frac{2\pi}{3}; \frac{4\pi}{3}\right] \\ = \frac{3}{2\pi}\theta - 2, & \text{for } \theta \in \left]\frac{4\pi}{3}; 2\pi\right[\end{cases}. \tag{21}$$

Inside the circle, a sound $S'_h(r,\theta)$ characterized by its angle θ and its radius r is generated with parameters defined by

$$\mathbf{P}_{S'_h}(r,\theta) = (1-r)\mathbf{P}_C + r\mathbf{P}_{S_h}(\theta) \tag{22}$$

[3]In practice, we implemented an additive synthesis technique (sinusoids plus noise) in the synthesizer previously developed [23] since it was the most natural one according to the signal model. Other techniques could have been considered as well such as frequency modulation (FM) synthesis, subtractive synthesis, etc.

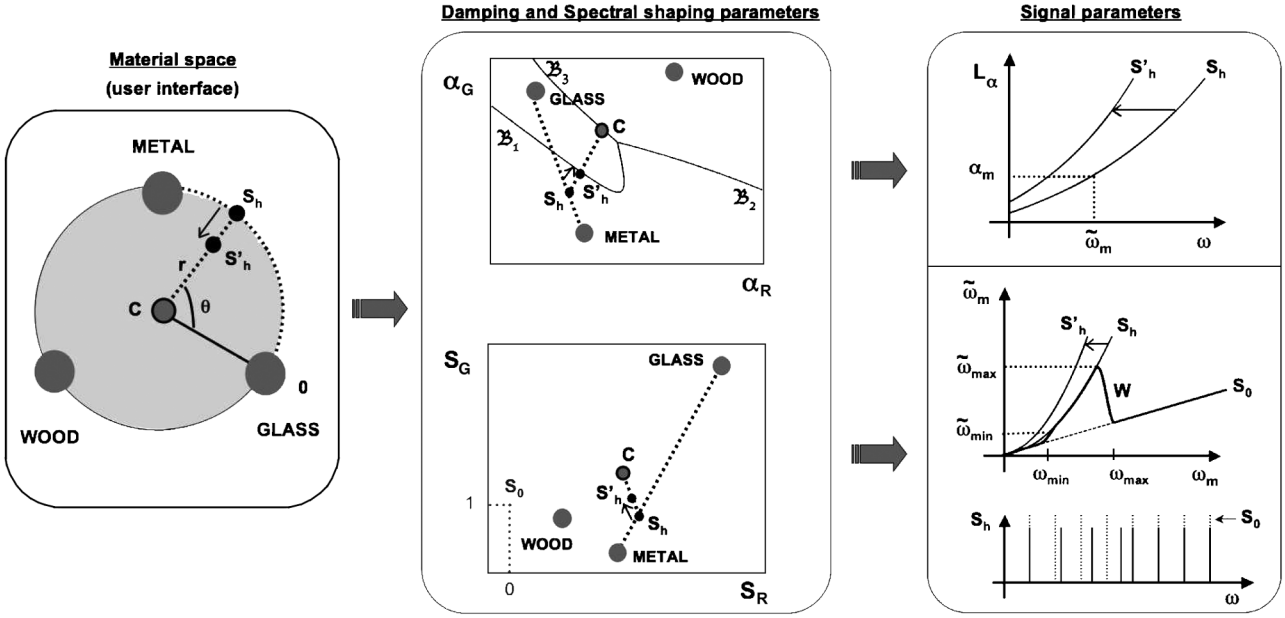


Fig. 6. Control of the perceived material based on three-layer architecture: the "Material space" (user interface), the "Damping and spectral shaping parameters" and "Signal parameters." The navigation in the Material space (e.g., from sound S_h to sound S'_h) involves modifications of both Damping and Spectral shaping parameters. The $\{\alpha_G, \alpha_R\}$ space was calibrated with specific domain for each material category (borders defined in Fig. 7). In this example, we represented the Spectral shaping parameters $\{S_G, S_R\}$ corresponding to the first spectral shaping control. Finally, at signal level, the damping coefficients α_m are computed from (4) with values of $\{\alpha_G, \alpha_R\}$ and the frequencies $\tilde{\omega}_m$ were computed from (15) with values of $\{S_G, S_R\}$. The Metal, Wood, and Glass reference sounds are constructed from the same initial harmonic sound S_0 located at point (0,1) in the $\{S_G, S_R\}$ space. The role of spectral shaping with the window function W is illustrated at the bottom. S_0 is represented in dotted and S_h in bold. The amplitude of the spectrum S_h was arbitrarily reduced for a sake of clarity.

TABLE II
VALUES OF DAMPING (α_G AND α_R) AND SPECTRAL SHAPING (S_G AND S_R) PARAMETERS CORRESPONDING TO THE REFERENCE SOUNDS OF METAL, WOOD, AND GLASS CATEGORY IN THE MATERIAL SPACE

	α_G	α_R ($\times 10^{-4}$)	S_G	S_R
Metal	0.6	2	0.5	0.1
Wood	3	4	0.85	0.05
Glass	2.5	1.5	2.4	0.2

where $\mathbf{P}_C$ represents the parameter vector of the sound C defined by $\{\bar{\alpha}_G, \bar{\alpha}_R, \bar{S}_G, \bar{S}_R\}$ with bar symbol denoting the average of the three values (corresponding to Wood, Metal, and Glass reference sounds) for each parameter and where $\mathbf{P}_{S_h}(\theta)$ is defined in (20). A similar strategy was designed for the mapping based on the second spectral shaping control (amplitude and frequency modulations). In that case, the parameter vector $\mathbf{P}$ corresponded to $\{\alpha_G, \alpha_R, I, \omega_n\}$.

The second layer concerns the controls of Damping and Spectral shaping parameters. For each control, the parameters are represented in two-dimensions to propose an intuitive configuration, called damping (α_G, α_R) and spectral shaping (S_G, S_R) spaces, respectively. The intuitive manipulation of α_G and α_R was achieved by a calibration process that consisted in determining a specific domain for each material category in the (α_G, α_R) space. Borders between material domains were determined based on results from predictive discrimination analysis. In practice, we calibrated the (α_G, α_R) space delimited by extreme values of α_G and α_R for typical sounds (range of $\alpha_G = [0.25; 3.34]$ and range of $\alpha_R = [0.5; 6.64] \times 10^{-4}$; see Fig. 7). This space was sampled in 2500 evenly spaced points

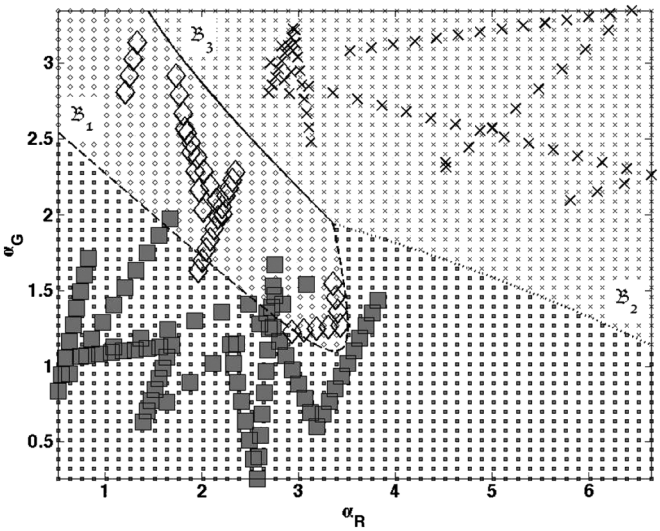


Fig. 7. Calibration of the $\{\alpha_G, \alpha_R\}$ space from border sections $\mathfrak{B}_1$ between Metal and Glass (dashed line), $\mathfrak{B}_2$ between Metal and Wood (dotted line) and $\mathfrak{B}_3$ between Glass and Wood (solid line). The positions of typical sounds for Wood ($\times$), Metal ($\square$), and Glass ($\diamond$) used for the classification process are also represented.

and each point was associated with a posterior probability of belonging to a material category. This probability was computed from a Bayesian rule based on the knowledge of the positions of typical sounds in the (α_G, α_R) space. Classification functions were determined between pairs of material categories and were expressed as quadratic combination of α_G and α_R ($\times 10^{-4}$). Boundary curves were materialized from the set of points that have similar classification probabilities δ for both categories

$$\{\mathbf{x} : \delta_{G_1}(\mathbf{x}) = \delta_{G_2}(\mathbf{x})\} \tag{23}$$

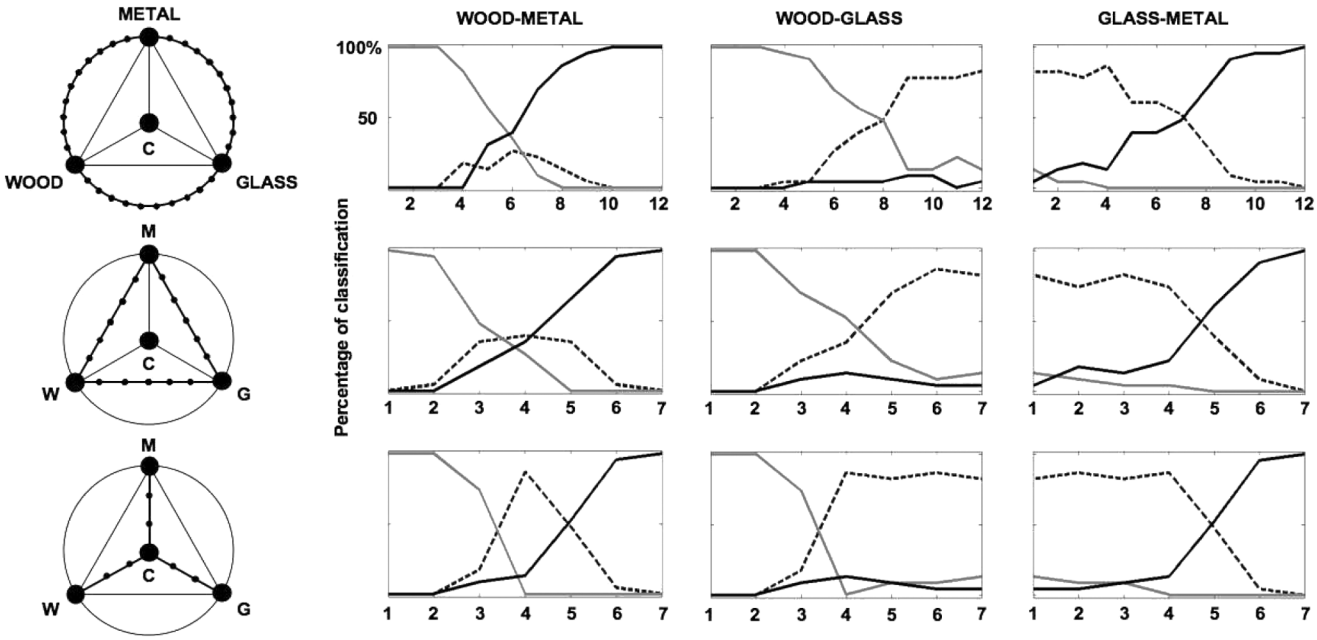


Fig. 8. Perceptual evaluation test of the control strategy. Left: position of sounds (black markers) used for the test in the Material Space as a function of the trajectory: along the external circle (first row), along the chords whose endpoints are the reference sounds (second row) and along the radii of the circle from the center C to the reference sounds (third row). Right: percentage of classification as Wood (gray curves), Metal (black curves), or Glass (dashed curves) corresponding to the three types of trajectory for each sound as a function of its position of the continuum.

or equivalently

$$\{\mathbf{x} : 0 = \delta_{G_1}(\mathbf{x}) - \delta_{G_2}(\mathbf{x})\} \tag{24}$$

where $\mathbf{x} = (\alpha_R, \alpha_G)$ and G_1 and G_2 the categories.

For our concern, the border noted $\mathfrak{B}_1$ between Metal and Glass categories was defined by

$$\mathfrak{B}_1 : 0 = 109.69 - 48.44\alpha_R - 48.30\alpha_G + 5.33(\alpha_R)^2 + 11.33\alpha_R\alpha_G + 3.37(\alpha_G)^2 \tag{25}$$

and all the points for which this function is negative were classified into the Metal category. The border $\mathfrak{B}_2$ between Wood and Metal was defined by

$$\mathfrak{B}_2 : 0 = -41.18 + 1.50\alpha_R + 17.75\alpha_G + 0.27(\alpha_R)^2 - 0.09\alpha_R\alpha_G - 0.20(\alpha_G)^2 \tag{26}$$

and all the points for which this function is negative were classified into the Wood category. Finally, the border $\mathfrak{B}_3$ between Wood and Glass regions was defined by

$$\mathfrak{B}_3 : 0 = 68.51 - 46.94\alpha_R - 30.55\alpha_G + 5.60(\alpha_R)^2 + 11.24\alpha_R\alpha_G + 3.17(\alpha_G)^2 \tag{27}$$

and all the points for which this function is negative were classified into the Wood category. The calibration of the (α_G, α_R) space was completed by keeping the section of the borders that directly separate the two sound categories: as shown in Fig. 7, the border section that was kept for $\mathfrak{B}_1$ is represented by a dashed line, the section kept for $\mathfrak{B}_2$ is represented by a dotted line and the section kept for $\mathfrak{B}_3$ is represented by a solid line. These borders were reported in the Middle layer (Fig. 6) allowing an intuitive manipulation of damping parameters. Note that these borders do not represent a strict delimitation between sound categories and a narrow transition zone may be taken into account on both sides of the borders. In particular, sounds belonging to this transition zone may be perceived as ambiguous sounds such as sounds created at intermediate positions of the continua.

Finally, the bottom layer concerns the signal parameters determined as follows: the damping coefficients α_m are computed from (4) with $\{\alpha_G, \alpha_R\}$ values and frequencies $\tilde{\omega}_m$ from (15) with $\{S_G, S_R\}$ values. The amplitudes A_m are assumed to be equal to one.

V. Perceptual Evaluation of the Control Strategy

The proposed control strategy for the perceived material was evaluated with a formal perceptual test. Twenty-three participants (9 women, 14 men) participated in the experiment. Sounds were selected in the Material Space as shown in Fig. 8 (left). Three types of trajectory between two reference sounds were investigated: along the external circle (by a 12-step continuum), along chords whose endpoints are the reference sounds (7-step continuum) and along the radii of the circle from the center C to the reference sounds (7-step continuum). Sounds were presented once randomly through headphones. The whole set of sounds are available at [38]. Participants were asked to categorize each sound as Wood, Metal, or Glass, as fast as possible, by selecting with a mouse on a computer screen the corresponding label. The order of labels displayed on the screen was balanced across participants. The next sound was presented after a 2-seconds silence. Participants' responses were collected and averaged for each category (Wood, Metal, and Glass) and for each sound.

Fig. 8 (right) shows results as a function of sound position along the continua. Sounds at extreme positions were classified by more than 70% of participants in the correct category, leading

to the validation of the reference sounds as typical exemplars of their respective material category. In Wood–Metal transition, sounds at intermediate positions were classified as Glass with highest percentages for those along the trajectory via the center C. By contrast, in both Wood–Glass and Glass–Metal transitions, intermediate sounds were most often classified in one of the two categories corresponding to the extreme sounds. From an acoustic point of view, this reflects the fact that, the interpolation of Damping parameters between Metal and Wood sounds crosses the Glass category while this is not the case for the other two transitions (see Fig. 6).

These results were in line with the ones obtained from the behavioral data in the first categorization experiment (Fig. 2) and consequently, allowed us to validate the proposed control strategy as an efficient way to navigate in the Material space. Note that the interpolation process was computed on a linear scale between parameters of the reference sounds [cf. (21) and (22)]. The next step will consist in taking into account these results and modify the interpolation rules so that the metric distance between a given sound and a reference sound in the Material space closely reflects perceptual distance.

VI. Conclusion

In this paper, we proposed a control strategy for the perceived material in an impact sound synthesizer. To this end, we investigated the sound characteristics relevant for an accurate evocation of material by conducting a sound categorization experiment. To design the stimuli, sounds produced by impacting three different materials, i.e., wood, metal, and glass, were recorded and synthesized by using analysis–synthesis techniques. After tuning, sound continua simulating progressive transitions between material categories were built and used in the categorization experiment. Both behavioral data and electrical brain activity were collected. From behavioral data, a set of typical sounds for each material category was determined and an acoustic analysis including descriptors known to be relevant for timbre perception and material identification was conducted. The most relevant descriptors that allow discrimination between material categories were identified: the normalized sound decay (related to the damping) and the roughness. Electrophysiological data provided complementary information regarding the perceptual/cognitive aspects related to the sound categorization and were discussed in the context of synthesis. Based on acoustic and ERP data, results confirmed the importance of damping and highlighted the relevance of spectral descriptors for material perception. Control strategies for damping and spectral shaping were proposed and tested in synthesis applications. These strategies were further integrated in a three-layer control architecture allowing the user to navigate in a "Material Space." A formal perceptual evaluation confirmed the validity of the proposed control strategy. Such a control offered an intuitive manipulation of parameters and allowed defining realistic impact sounds directly from the material label (i.e., Wood, Metal, or Glass).

References

[1] M. V. Mathews, "The digital computer as a musical instrument," *Science*, vol. 142, no. 3592, pp. 553–557, 1963.

[2] R. Moog, "Position and force sensors and their application to keyboards and related controllers," in *Proc. AES 5th Int. Conf.: Music Digital Technol.*, 1987, pp. 179–181, A. E. S. New York, Ed..

[3] M. Battier, "L'approche gestuelle dans l'histoire de la lutherie électronique. Etude d'un cas: Le theremin," *Proc. Colloque International* , ser. Collection Eupalinos, Editions Parenthèses, 1999, Les nouveaux gestes de la musique.

[4] J. Tenney, "Sound-generation by means of a digital computer," *J. Music Theory*, vol. 7, no. 1, Spring, 1963.

[5] A. Camurri, M. Ricchetti, M. Di Stefano, and A. Stroscio, "Eyesweb—Toward gesture and affect recognition in dance/music interactive systems," in *Proc. Colloquio di Informatica Musicale*, 1998.

[6] P. Gobin, R. Kronland-Martinet, G. A. Lagesse, T. Voinier, and S. Ystad, *From Sounds to Music: Different Approaches to Event Piloted Instruments*, ser. Lecture Notes in Computer Science. : Springer-Verlag, 2003, vol. 2771, pp. 225–246.

[7] M. Wanderley and M. Battier, "Trends in gestural control of music," IRCAM-Centre Pompidou, 2000.

[8] J.-C. Risset and D. L. Wessel, "Exploration of timbre by analysis and synthesis," in *The Psychology of Music*, ser. Cognition and Perception, 2nd ed. New York: Academic, 1999, pp. 113–169.

[9] D. L. Wessel, "Timbre space as a musical control structure," *Comput. Music J.*, vol. 3, no. 2, pp. 45–52, 1979.

[10] S. Ystad and T. Voinier, "A virtually-real flute," *Comput. Music J.*, vol. 25, no. 2, pp. 13–24, Summer, 2001.

[11] J. M. Grey, "Multidimensional perceptual scaling of musical timbres," *J. Acoust. Soc. Amer.*, vol. 61, no. 5, pp. 1270–1277, 1977.

[12] C. L. Krumhansl, "Why is musical timbre so hard to understand," in *Structure and Perception of Electroacoustic Sound and Music*. Amsterdam, The Netherlands: Elsevier, 1989.

[13] J. Krimphoff, S. McAdams, and S. Winsberg, "Caractérisation du timbre des sons complexes. II: Analyses acoustiques et quantification psychophysique [characterization of timbre of complex sounds. II: Acoustical analyses and psychophysical quantification]," *J. Phys.*, vol. 4, no. C5, pp. 625–628, 1994.

[14] S. McAdams, S. Winsberg, S. Donnadieu, G. D. Soete, and J. Krimphoff, "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," *Psychol. Res.*, vol. 58, pp. 177–192, 1995.

[15] W. A. Sethares, "Local consonance and the relationship between timbre and scale," *J. Acoust. Soc. Amer.* , vol. 94, no. 3, pp. 1218–1228, 1993.

[16] P. N. Vassilakis, "Auditory roughness as a means of musical expression," Dept. of Ethnomusicology, Univ. of California, Selected reports in ethnomusicology (perspectives in systematic musicology), 2005, vol. 12, pp. 119–144.

[17] R. P. Wildes and W. A. Richards, *Recovering Material Properties From Sound*, W. A. Richards, Ed. Cambridge, MA: MIT Press, 1988, ch. 25, pp. 356–363.

[18] W. W. Gaver, "How do we hear in the world ? explorations of ecological acoustics," *Ecol. Psychol.*, vol. 5, no. 4, pp. 285–313, 1993.

[19] R. Lutfi and E. Oh, "Auditory discrimination of material changes in a struck-clamped bar," *J. Acoust. Soc. Amer.*, vol. 102, no. 6, pp. 3647–3656, 1997.

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

[21] B. L. Giordano and S. McAdams, "Material identification of real impact sounds: Effects of size variation in steel, wood, and Plexiglas plates," *J. Acoust. Soc. Amer.*, vol. 119, no. 2, pp. 1171–1181, 2006.

[22] 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*, ser. LNCS, S. Ystad, R. Kronland-Martinet, and K. Jensen, Eds. Berlin, Heidelberg, Germany: Springer-Verlag, 2009, vol. 5493, pp. 1–17.

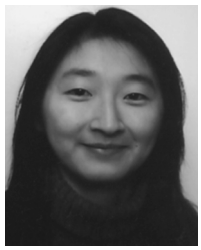
[23] M. Aramaki, R. Kronland-Martinet, T. Voinier, and S. Ystad, "A percussive sound synthesizer based on physical and perceptual attributes," *Comput. Music J.*, vol. 30, no. 2, pp. 32–41, 2006.

[24] M. Aramaki, R. Kronland-Martinet, T. Voinier, and S. Ystad, "Timbre control of a real-time percussive synthesizer," in *Proc. 19th Int. Congr. Acoust. (CD-ROM)*, 2007, 84-87985-12-2.

[25] M. Aramaki, C. Gondre, R. Kronland-Martinet, T. Voinier, and S. Ystad, "Thinking the sounds: An intuitive control of an impact sound synthesizer," in *Proc. 15th Int. Conf. Auditory Display (ICAD 2009)*, 2009.

[26] K. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," *IEEE Trans. Autom. Control*, vol. AC-10, no. 10, pp. 461–464, Oct. 1965.

[27] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
[28] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.
[29] R. Badeau, B. David, and G. Richard, "High-resolution spectral analysis of mixtures of complex exponentials modulated by polynomials," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1341–1350, Apr. 2006.
[30] L.-M. Reissell and D. K. Pai, "High resolution analysis of impact sounds and forces," in *Proc. WHC '07: 2nd Joint EuroHaptics Conf. Symp. Haptic Interfaces for Virtual Environment and Teleoperator Syst.*, Washington, DC, 2007, pp. 255–260, IEEE Computer Society.
[31] M. Aramaki and R. Kronland-Martinet, "Analysis-synthesis of impact sounds by real-time dynamic filtering," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 695–705, Mar. 2006.
[32] R. Kronland-Martinet, P. Guillemain, and S. Ystad, "Modelling of natural sounds by time-frequency and wavelet representations," *Organised Sound*, vol. 2, no. 3, pp. 179–191, 1997.
[33] R. Parncutt, *Harmony—A Psychoacoustical Approach*. Berlin/Heidelberg, Germany: Springer, 1989.
[34] A. Chaigne and C. Lambourg, "Time-domain simulation of damped impacted plates: I. Theory and experiments," *J. Acoust. Soc. Amer.*, vol. 109, no. 4, pp. 1422–1432, 2001.
[35] T. Ono and M. Norimoto, "Anisotropy of dynamic young's modulus and internal friction in wood," *Jpn. J. Appl. Phys.*, vol. 24, no. 8, pp. 960–964, 1985.
[36] S. McAdams, A. Chaigne, and V. Roussarie, "The psychomechanics of simulated sound sources: Material properties of impacted bars," *J. Acoust. Soc. Amer.*, vol. 115, no. 3, pp. 1306–1320, 2004.
[37] M. Aramaki, H. Baillères, L. Brancheriau, R. Kronland-Martinet, and S. Ystad, "Sound quality assessment of wood for xylophone bars," *J. Acoust. Soc. Amer.*, vol. 121, no. 4, pp. 2407–2420, 2007.
[38] 2010 [Online]. Available: http://www.lma.cnrs-mrs.fr/~kronland/Categorization/sounds.html, last checked: Oct. 2009
[39] M. Aramaki, L. Brancheriau, R. Kronland-Martinet, and S. Ystad, "Perception of impacted materials: Sound retrieval and synthesis control perspectives," in *Computer Music Modeling and Retrieval—Genesis of Meaning of Sound and Music*, ser. LNCS, S. Ystad, R. Kronland-Martinet, and K. Jensen, Eds. Berlin, Heidelberg, Germany: Springer-Verlag, 2009, vol. 5493, pp. 134–146.
[40] M. D. Rugg and M. G. H. Coles, "The ERP and cognitive psychology: Conceptual Issues," in *Electrophysiology of Mind. Event-Related Brain Potentials and Cognition*, ser. Oxford Psychology. New York: Oxford Univ. Press, 1995, pp. 27–39, no. 25.
[41] J. Eggermont and C. Ponton, "The neurophysiology of auditory perception: From single-units to evoked potentials," *Audiol. Neuro-Otol.*, vol. 7, pp. 71–99, 2002.
[42] A. Shahin, L. E. Roberts, C. Pantev, L. J. Trainor, and B. Ross, "Modulation of p2 auditory-evoked responses by the spectral complexity of musical sounds," *NeuroReport*, vol. 16, no. 16, pp. 1781–1785, 2005.
[43] S. Kuriki, S. Kanda, and Y. Hirata, "Effects of musical experience on different components of meg responses elicited by sequential piano-tones and chords," *J. Neurosci.*, vol. 26, no. 15, pp. 4046–4053, 2006.
[44] E. Kushnerenko, R. Ceponiene, V. Fellman, M. Huotilainen, and I. Winkler, "Event-related potential correlates of sound duration: Similar pattern from birth to adulthood," *NeuroReport*, vol. 12, no. 17, pp. 3777–2781, 2001.
[45] C. Alain, B. M. Schuler, and K. L. McDonald, "Neural activity associated with distinguishing concurrent auditory objects," *J. Acoust. Soc. Amer.*, vol. 111, no. 2, pp. 990–995, 2002.
[46] S. McAdams, "Perspectives on the contribution of timbre to musical structure," *Comput. Music J.*, vol. 23, no. 3, pp. 85–102, 1999.
[47] H.-G. Kim, N. Moreau, and T. Sikora, *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. New York: Wiley, 2005.
[48] G. Peeters, "A Large set of audio features for sound description (similarity and description) in the Cuidado Project," IRCAM, Paris, France, 2004, Tech. Rep..
[49] J. Marozeau, A. de Cheveigné, S. McAdams, and S. Winsberg, "The dependency of timbre on fundamental frequency," *J. Acoust. Soc. Amer.*, vol. 114, pp. 2946–2957, 2003.
[50] J. W. Beauchamp, "Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones," *J. Audio Eng. Soc.*, vol. 30, no. 6, pp. 396–406, 1982.
[51] R. Plomp and W. J. M. Levelt, "Tonal consonance and critical bandwidth," *J. Acoust. Soc. Amer.*, vol. 38, pp. 548–560, 1965.
[52] E. Terhardt, "On the perception of periodic sound fluctuations (roughness)," *Acustica*, vol. 30, no. 4, pp. 201–213, 1974.
[53] P. N. Vassilakis, "SRA: A web-based research tool for spectral and roughness analysis of sound signals," in *Proc. 4th Sound Music Comput. (SMC) Conf.*, 2007, pp. 319–325.
[54] 2009 [Online]. Available: http://www.cycling74.com/, last checked: Octob. 2009
[55] D. Rocchesso and F. Fontana, 2003, "The Sounding Object," [Online]. Available: http://www.soundobject.org/SobBook/SobBook_JUL03.pdf last checked: Oct. 2009
[56] M. Hyde, "The N1 response and its applications," *Audiol. Neuro-Otol.*, vol. 2, pp. 281–307, 1997.
[57] C. Valette and C. Cuesta, *Mécanique de la corde vibrante (Mechanics of vibrating string)*, ser. Traité des Nouvelles Technologies, série Mécanique. London, U.K.: Hermès, 1993.
[58] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments (Second Edition)*. Berlin, Germany: Springer-Verlag, 1998.
[59] E. Zwicker and H. Fastl, *Psychoacoustics, Facts and Models*. Berlin, Germany: Springer-Verlag, 1990.
[60] H. L. F. von Helmholtz, *On the Sensations of Tone as the Physiological Basis for the Theory of Music*, 2nd ed. New York: Dover, 1877, reprinted 1954.



Mitsuko Aramaki (M'09) received the M.S. degree in mechanics (speciality in acoustic and dynamics of vibrations) from the University of Aix-Marseille II, Marseille, France, and the Ph.D. degree for her work at the Laboratoire de Mécanique et d'Acoustique, Marseille, France, in 2003, on analysis and synthesis of impact sounds using physical and perceptual approaches.

She is currently a Researcher at the Institut de Neurosciences Cognitives de la Méditerranée, Marseille, where she works on a pluridisciplinary project combining sound modeling, perceptual and cognitive aspects of timbre, and neuroscience methods in the context of virtual reality.

Mireille Besson received the Ph.D. degree in neurosciences from the University of Aix-Marseille II, Marseille, France, in 1984.

After four years of post-doctorate studies at the Department of Cognitive Science, University of California at San Diego, La Jolla, working with Prof. M. Kutas and at the Department of Psychology, University of Florida, Gainesville, working with Prof. I. Fischler, she obtained a permanent position at the National Center for Scientific Research (CNRS), Marseille, France. She is currently Director of Research at the CNRS, Institut de Neurosciences Cognitives de la Méditerranée (INCM), where she is the head of the "Language, Music, and Motor" team. Her primary research interests are centered on brain imaging of linguistic and non linguistic sound perception and on brain plasticity mainly using event-related brain potentials. She is currently conducting a large research project on the influence of musical training on linguistic sound perception in normal reading and dyslexic children.

Richard Kronland-Martinet (M'09–SM'10) received the M.S. degree in theoretical physics in 1980, the Ph.D. degree in acoustics from the University of Aix-Marseille II, Marseille, France, in 1983, and the "Doctorat d'Etat es Sciences" degree from the University of Aix-Marseille II in 1989 for his work on analysis and synthesis of sounds using time–frequency and time–scale (wavelets) representations.

He is currently Director of Research at the National Center for Scientific Research (CNRS), Laboratoire de Mécanique et d'Acoustique, Marseille, where he is the Head of the group "Modeling, Synthesis and Control of Sound and Musical Signals." His primary research interests are in analysis and synthesis of sounds with a particular emphasis on high-level control of synthesis processes. He recently addressed applications linked to musical interpretation and semantic description of sounds using a pluridisciplinary approach associating signal processing, physics, perception, and cognition.

Sølvi Ystad received the Ph.D. degree in acoustics from the University of Aix-Marseille II, Marseille, France, in 1998.

She is currently a Researcher at the National French Research Center (CNRS) in the research team S2M—Synthesis and Control of Sounds and Musical Signals—in Marseille, France. Her research activities are related to sound modeling with a special emphasis on the identification of perceptually relevant sound structures to develop efficient synthesis models. She was in charge of the research project "Towards the sense of sounds," financed by the French National Agency (ANR—http://www.sensons.cnrs-mrs.fr) from 2006–2009.