Timbre Perception of Sounds from Impacted Materials: Behavioral, Electrophysiological and Acoustic Approaches

Mitsuko Aramaki[1,2], Mireille Besson[1,2], Richard Kronland-Martinet[3], and Sølvi Ystad[3]

[1] CNRS - Institut de Neurosciences Cognitives de la Méditerranée,
31, chemin Joseph Aiguier 13402 Marseille Cedex 20, France
{aramaki,besson}@incm.cnrs-mrs.fr
[2] Aix-Marseille - Université,
58, Bd Charles Livon 13284 Marseille Cedex 07, France
[3] CNRS - Laboratoire de Mécanique et d'Acoustique,
31, chemin Joseph Aiguier 13402 Marseille Cedex 20, France
{kronland,ystad}@lma.cnrs-mrs.fr

Abstract. In this paper, timbre perception of sounds from 3 different impacted materials (Wood, Metal and Glass) was examined using a categorization task. Natural sounds were recorded, analyzed and resynthesized and a sound morphing process was applied to construct sound continua between different materials. Participants were asked to categorize the sounds as Wood, Metal or Glass. Typical sounds for each category were defined on the basis of the behavioral data. The temporal dynamics of the neural processes involved in the categorization task were then examined for typical sounds by measuring the changes in brain electrical activity (Event-Related brain Potentials, ERPs). Analysis of the ERP data revealed that the processing of Metal sounds differed significantly from Glass and Wood sounds as early as 150 ms and up to 700 ms. The association of behavioral, electrophysiological and acoustic data allowed us to investigate material categorization: the importance of damping was confirmed and additionally, the relevancy of spectral content of sounds was discussed.

1 Introduction

Natural sounds carry a large amount of acoustic information that can be extracted by classical signal analysis techniques such as spectral analysis or time-frequency decompositions. Nevertheless, such techniques do not aim at explaining how the brain captures the information that allows us to identify different aspects of the sound source, such as its size, shape or material. The current study focuses on the perception of sounds produced by impacted materials, i.e., Wood, Metal and Glass. We aim at identifying the perceptual properties of these sounds (which are related to the concept of "invariants"

S. Ystad, R. Kronland-Martinet, and K. Jensen (Eds.): CMMR 2008, LNCS 5493, pp. 1–17, 2009.

in ecological acoustics [1]), that characterize material categories, without considering the mechanical correlates of the sound source[1].

Results of acoustic analysis are considered together with the results of behavioral and electrophysiological analyses. In particular, we examine the agreement between the acoustic relevancy of sound descriptors with respect to sound categorization as revealed by statistical analysis and their perceptual/cognitive relevancy as revealed by the objective measures of the electrical brain activity. As a first approach to the identification of these invariants for material perception, we investigate the sound descriptors that are known to be relevant for timbre perception and for material identification.

Previous studies on the perception of sound categories (of musical instruments or human voices) have mainly been based on the notion of timbre. Timbre is a perceptual and subjective attribute of sounds that is not well understood and is often defined by what it is not: "an indispensable key feature for the appreciation of sound quality other than pitch, loudness, duration and spatial location" (American National Standards Institute, 1973). Timbre is a complex feature of sounds that requires a multidimensional representation. Several authors [2,3,4,5] have used dissimilarity ratings to characterize the sounds of musical instruments and to identify timbre spaces based on the perceptual distances between sounds. The dimensions of these timbre spaces are correlated with various descriptors such as attack time (the way the energy rises at the sound onset), spectral bandwidth (spectrum spread), spectral center of gravity or spectral centroid (correlated to the brightness of sound) and spectral flux (the way the spectrum varies with time). Another important dimension of timbre, introduced by [6], is the roughness (distribution of interacting frequency components within the limits of a critical band). In a musical context, the roughness is closely linked to the concept of consonance/dissonance [7,8].

Regarding the perception of impact sounds, previous acoustic studies have examined the perception of the physical characteristics of the sound source, that are the properties of the actions and objects. Regarding perception of objects, [9] has shown that the perceived hardness of a mallet striking a metallic object is predictable from the characteristics of attack time. Further, strong correlations have been found between the perceived size of objects and the pitch of the generated sounds [10,11,12,13,14] and between the perceived shape of objects and the distribution of spectral components [15,16]. Finally, perception of material seems mainly to correlate with the damping of spectral components [17,18,19]. Damping also remains a robust acoustic descriptor to identify macro-categories (i.e., wood-plexiglass and steel-glass categories) across variations in the size of objects [20]. From a physical point of view, the damping is due to various mechanisms of loss and differs as a function of materials [21,22]. The quality of percussive sounds is highly correlated to the frequency-dependency of the damping: high frequency components being generally more heavily damped than low frequency ones. Consequently, different damping for high and for low frequency components involve

[1] Had the aim been to study the perceptual influence of the mechanical properties of the object (material, geometry), we would have studied structures and material independently. From a mechanical point of view, the respective contribution of object and material properties to sound perception is difficult to determine since identical sounds can be obtained by manipulating the characteristics of one or of the other.

intrinsic timbre modifications. Nevertheless, a global characterization of the damping can be given by the sound decay which measures the decrease of the sound energy as a function of time. Since damping is known to be important for material identification, we investigated the relevancy of sound decay for sound categorization in addition to previously presented timbre descriptors.

We used the flexibility of synthesis processes to construct the stimuli by extending the concept of analysis by synthesis [23]. Thus, we recorded, analyzed and resynthesized sounds from everyday life objects made of different materials (i.e., impacted wooden beams, metallic plates and various glass bowls). Further, we generated realistic sounds unambiguously evoking each material category. In practice, we used an additive synthesis model so that sounds were reconstructed as the sum of a number of frequency components controlled separately in amplitude, frequency and temporal evolution. Then, we created continuous transitions between sounds from different materials (i.e., {Wood-Metal}, {Wood-Glass} and {Glass-Metal}) through a morphing process based upon an interpolation between the values of the parameters of the sounds at the extremes of a continuum. Sounds from these continua were used in a categorization task so as to be sorted as Wood, Metal or Glass. In this way, the limits of the categories on each continuum and consequently, a set of unambiguous, "typical" sounds for each category could be identified.

We analyzed both the percentage of responses in each category and Reaction Times (RTs), as RTs are known to provide a good measure of the chronometry of mental processes [24,25]. Then, we examined the temporal dynamics of the brain processes associated with the perception and categorization of these sounds. In practice, we recorded Event Related Potentials (ERPs) time-locked to the sounds to analyze the different stages of information processing as they unfold in real time. 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 not only by their polarity but also by 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 first positive-negative-positive components following stimulus presentation, the P100, N100 and P200, reflect the sensory and perceptual stages of information processing, and are obligatory responses to the stimulation [26,27]. Depending upon the characteristics of the stimulation, this N100/P200 complex presents differences. Then, depending of the experimental design and of the task at hand, different late ERP components are elicited (N200, P300, N400 ...).

Results from studies based on the analysis of the ERPs have provided interesting information regarding the time course of timbre processing. For instance, the Mismatch Negativity (MMN; [28]) is considered as a good index of pre-attentive discrimination between tones with different acoustic, perceptual, cognitive and/or emotional attributes (see [29] and [30] for a review). [31] have shown that a MMN is elicited by infrequent pure tones presented among frequent harmonically rich tones. Moreover, recent results have shown increased MMN amplitude and latency with changes in spectral centroid [32,33], roughness [34], attack time and spectrum fine structure (even harmonic attenuation; [33]. These results thus suggest that such timbre descriptors are processed pre-attentively. Otherwise, in a discrimination task, [35] found increased N100 and P200 amplitude as a function of spectro-temporal complexity (see also [36,37]). Moreover, the use of a source estimation method (LORETA, [38]) allowed them to uncover possible generators of the N100 and P200 components: N100 resulting from activity in the right posterior Sylvian and inferior parietal regions and P200 from left and right regions of the posterior Sylvian fissure in the vicinity of the secondary auditory cortex. Finally, the late positive component, with maximum amplitude around 300 ms, the P300, is thought to reflect the categorization of task relevant events [39,40]. For instance, [29] showed that P300 latency was longest for emotion deviants, intermediate for timbre deviants and shortest for pitch deviants.

In the present experiment, regarding the ERPs, since Metal sounds show richer spectra than Wood or Glass sounds, we predicted that Metal sounds would elicit larger P200 amplitude [37]. Indeed, the spectrum of sounds generated from impacted structures with a given geometry is generally richer for Metal than for Wood or Glass. This can be explained by physical and signal processing considerations because the spectral density of energy integrates the energy of each frequency component over the entire duration of the sound. Due to differences in mechanisms of loss between materials, Metal is associated to less damped vibrations than Wood and Glass [21]. Consequently, the spectral energy of each mode is higher for Metal sounds, particularly in the high frequency range where modes are rapidly damped for Wood and Glass sounds. This general spectral behavior is in line with the results of the acoustic analyses of the sounds used in the present experiment. Sounds also differed from each other on several acoustic parameters: the number of spectral components (higher for Metal than for Wood or Glass), the damping (lower for Metal than for Wood and Glass) and the spectral distribution (quasi-harmonic for Glass, non harmonic for Metal and Wood) and these differences may also influence the ERP components. Note that since we used an explicit categorization task, we also expected N200-P300 components to be generated [41] and their amplitude and/or latency should reflect the characteristics of the categorization processes (i.e., shorter latency and larger amplitude for the sounds that are easier to categorize).

2 Methods

2.1 Participants and Stimuli

Twenty-two participants (11 women and 11 men), 19 to 35 years-old, were tested in this experiment. They were all right-handed, non-musicians (no formal musical training), had normal audition and no known neurological disorders. They all gave written consent to participate in the experiment and they were paid for their participation.

To design the stimuli, we first recorded impact sounds from everyday life objects made of different materials (i.e. impacted wooden beams, metallic plates and various glass bowls) to insure the generation of realistic sounds. Then, we used a simplified version of the analysis-synthesis model described in [42], based on an additive synthesis technique to resynthesize these recorded sounds (at 44.1 kHz sampling frequency):

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

where $\theta(t)$ is the Heaviside function, M the number of components, and the parameters A_m, α_m and ω_m are respectively the amplitude, damping coefficient and frequency of the mth component.

To minimize timbre variations induced by pitch changes, all resynthesized sounds were tuned to the same chroma (note C): Wood sounds to the pitch C4, Metal sounds to C4 and C5 and Glass sounds to C6 and C7. The specific properties of the different materials did not, however, allow us to tune all sounds to the same octave. For instance, Glass sounds can not be transposed to low pitches as they will no longer be recognized as Glass sounds. Nevertheless, based on [43], we considered that the octave differences between sounds should not influence sound categorization. When the pitch was changed, the new value of the damping coefficient of each tuned frequency component was evaluated according to a damping law measured on the original sound [44]. Sounds were also equalized in loudness by gain adjustments. These sounds were called *reference* sounds of each material categories.

Finally, we built 15 continua (5 for each transition: $\{$Wood-Metal$\}$, $\{$Wood-Glass$\}$ and $\{$Glass-Metal$\}$) composed of 20 hybrid sounds that simulate a progressive transition between different materials. By using a morphing process, the interpolation was computed both on the amplitudes of the spectral components and on their damping parameters. Sound examples are available at http://www.lma.cnrs-mrs.fr/$\sim$kronland/Categorization/sounds.html.

2.2 Procedure and Recording ERPs

Participants were asked to categorize sounds pseudo-randomly presented through one loudspeaker, as Wood, Metal or Glass, as fast as possible, by pressing one response button out of three. The association between response buttons and material categories was balanced across participants. Brain electrical activity (Electroencephalogram, EEG) was recorded continuously from 32 Biosemi Pin-type active electrodes (Amsterdam University) mounted on an elastic headcap and located at standard left and right hemisphere positions over frontal, central, parietal, occipital and temporal areas (International extended 10/20 system sites [45]: Fz, Cz, Pz, Oz, Fp1, Fp2, AF3, AF4, F7, F8, F3, F4, Fc5, Fc6, Fc1, Fc2, T7, T8, C3, C4, Cp1, Cp2, Cp5, Cp6, P3, P4, PO3, PO4, P7, P8, O1, O2). Moreover, to detect horizontal eye movements and blinks, the electrooculogram (EOG) was recorded from Flat-type active electrodes placed 1 cm to the left and right of the external canthi, and from an electrode beneath the right eye. Two additional electrodes were placed on the left and right mastoids. EEG was recorded at a sampling rate of 512 Hz using Biosemi amplifier. The EEG was re-referenced off-line to the algebric average of the left and right mastoids and filtered with a lowpass of 40 Hz. Data were analyzed using routines of Matlab EEGLAB toolbox [46]. Data were segmented in single trials of 2500 ms, starting 200 ms before the onset of the sound.

3 Sound Descriptors

To characterize sounds from an acoustic point of view, we considered the following sound descriptors: attack time AT, spectral centroid SC, spectral bandwidth SB, spectral flux SF, roughness R and normalized sound decay α.

The attack time AT is a temporal timbre descriptor which characterizes signal onset. It is defined by the time necessary for the signal energy to raise from 10% to 90% of the maximum amplitude of the signal.

The spectral centroid SCG is a spectral timbre descriptor which is commonly associated to the brightness of the sound and is defined by [47]:

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

while the spectral bandwidth SB, commonly associated to the spectrum spread is defined by [48]:

$$SB = \frac{1}{2\pi} \sqrt{\frac{\sum_k \mid \hat{s}(k) \mid (\omega(k) - 2\pi \times SC)^2}{\sum_k \mid \hat{s}(k) \mid}} \tag{3}$$

where ω and $\hat{s}$ represent the frequency and the Fourier transform of the signal, respectively.

The spectral flux SF is a spectro-temporal timbre descriptor quantifying the time evolution of the spectrum and is defined by [49]:

$$SF = \frac{1}{C} \sum_{c=1}^{C} \mid p_{n,n-1} \mid \tag{4}$$

where C represents the number of local spectra (frames) and $p_{n,n-1}$ the Pearson product moment correlation coefficient between the local spectra at the discrete times n and $n-1$.

Based on the [50] and [51] models, the roughness R of a sound, commonly associated to the presence of several frequency components within the limits of a critical band, can be computed by summing up the partial roughness r_{ij} for all pairs of frequency components (i,j) contained in the sound [52]:

$$r_{ij} = 0.5 \times (A_i A_j)^{0.1} \times \left(\frac{2 \min(A_i, A_j)}{A_i + A_j} \right)^{3.11} \times (e^{-3.5s|\omega_i - \omega_j|} - e^{-5.75s|\omega_i - \omega_j|}) \tag{5}$$

where

$$s = \frac{0.24}{0.0207 \times \min(\omega_i, \omega_j) + 2\pi \times 18.96} \tag{6}$$

and where A_i and A_j, ω_i and ω_j are respectively the amplitudes and frequencies of a pair of frequency components (i,j).

Finally, sound decay quantifies the global amplitude decrease of the temporal signal. Since the sounds consist of the sum of exponentially damped sine waves (Equation (1)), the sound decay is directly estimated by the slope of the logarithm of the envelope of the temporal signal. Nevertheless, since the damping is frequency dependent, this decrease depends on the spectral content of the sound. Consequently, we chose to consider a normalized sound decay α with respect to a reference that takes into account the spectral localization of the energy. Then, we defined α as the ratio of the sound decay to the SC value.

4 Results

4.1 Behavioral Data

Participants' responses and RTs were collected for each sound. Sounds were considered as *typical* if they were classified in one category (i.e., Wood, Metal or Glass) by more than 70% of the participants. This threshold value was determined using a statistical approach based on hierarchical cluster analysis (see [53] for details).

Data were averaged as a function of participants' responses in each category (Table 1). Results were analyzed using repeated-measures Analysis of Variance (ANOVAs) including Material (Wood, Metal and Glass) as a within-subject factor. For this and following analyses, effects were considered significant if the p-value was less than .05. When interactions between 2 or more factors were significant, post-hoc comparisons (Tukey tests) were computed. Clearly, participants more often categorized sounds as Metal than as Wood or Glass [$F(2,42)=25.31$, $p<.001$] and RTs were slower for Glass sounds [$F(2,42)=29.1$, $p<.001$] than for Wood and Metal sounds.

Table 1. Percentages of responses (N) in each material category (W=Wood, M=Metal, G=Glass), of responses N_{typ} associated to typical sounds of each category (i.e., sounds classified in the category by more than 70% of the participants), and of responses N_{nontyp} associated to non typical sounds (i.e., sounds classified by less than 70% of the participants). Mean RTs for each category are also indicated in milliseconds. The data are reported with their respective standard deviation (±SD) across participants.

	N	RT	N_{typ}	RT	N_{nontyp}	RT
W	31(±7)	936(±195)	17(±2)	829(±185)	14(±5)	1080(±250)
M	42(±5)	1028(±275)	29(±1)	964(±254)	13(±3)	1170(±323)
G	27(±5)	1152(±247)	9(±1)	1010(±211)	18(±3)	1203(±263)

4.2 Electrophysiological Data

To examine the time course of the perception of typical sounds, as defined from the results of the behavioral data, as well as potential differences between the 3 sound categories, ERPs data were averaged separately for the typical sounds of each material category (Wood, Metal and Glass). Based upon visual inspection of the ERP traces, the following time windows were chosen for statistical analysis to focus on specific ERP components: 0-50 ms, 50-80 ms (P100), 80-150 ms (N100), 150-230 ms (P200), 230-330 ms (N280), 400-550 ms and 550-700 ms (Negative Slow Wave, NSW and P550).

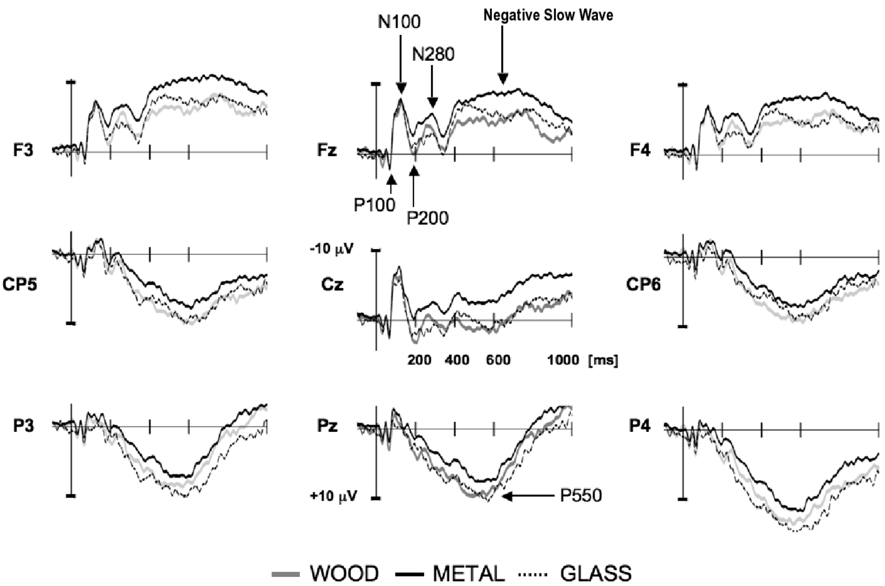

Fig. 1. Event-Related Potentials (ERPs) to typical sounds of Wood (grey solid trace), Metal (black solid trace) and Glass (dotted trace) at midline (Fz, Cz, Pz) and at lateral electrodes ({F3, CP5, P3} / {F4, CP6, P4}). The amplitude (in microvolt μV) is represented on the ordinate and negativity is up. The time from sound onset is on the abscissa (in millisecond, ms).

Separate repeated-measures ANOVAs were conducted for midline and lateral electrodes. Factors were Material (Wood, Metal and Glass) and Electrodes (Fz, Cz, Pz) for midline analyses. Factors were Material, Hemispheres (left vs. right), Regions of Interest (3 ROIs: fronto-central, centro-temporal and centro-parietal) and Electrodes (3 for each ROI: {AF3, F3, FC5} / {AF4, F4, FC6}; {T7, C3, CP5} / {T8, C4, CP6}; {CP1, P7, P3} / {CP2, P8, P4}) for lateral analyses. Effects were considered significant if the p-value was less than 0.05. Results are reported in Table 2.

Figure 1 shows the ERPs recorded at midline electrodes and at one lateral electrode per ROI (selected as the most representative of the ROI). Typical sounds from each category elicited small positive components, with maximum amplitude around 65 ms post-sound onset (P100), large negative components peaking around 100 ms (N100) followed by relatively positive (P200) and negative (N280 and NSW) components at fronto-central sites or large positive components (P550) at parietal sites. Results of ANOVAs revealed no main effect of Material on P100 amplitude (Midlines, p=.76; Lateral, p=.69) or on N100 amplitude (Midlines, p=.07; Lateral, p=.22) either at midline or at lateral electrodes. By contrast, at both midline and lateral electrodes, Metal sounds elicited smaller P200 and larger N280 and NSW components than Wood and Glass sounds that did not differ from each other (Table 3). Moreover, Metal sounds elicited smaller P550 component than Wood sounds (Table 2). This effect was largest over fronto-central and centro-temporal regions in the 400-550 ms latency range and over fronto-central region in the 550-700 ms latency range (see Table 4).

Table 2. F statistics and mean amplitude values (in μV) for the main effect of the Material factor (W = Wood, M = Metal, G = Glass) in the latency windows in which the effects were significant. Amplitude differences (W-M, M-G and W-G in μV) are indicated only when significant. For this and following tables, the level of significance is given by * (p<.05), ** (p<.01), *** (p<.001). $^{(a)}$ indicates a marginally significant effect at p=.059.

	150-230 ms		**230-330 ms**		**400-550 ms**		**550-700 ms**	
	P200 comp.		N280 comp.		NSW and P550 comp.			
	Mid.	Lat.	Mid.	Lat.	Mid.	Lat.	Mid.	Lat.
F(2,42)	5.26^{**}	4.11^{*}	4.99^{*}	4.88^{*}	4.00^{*}	4.08^{*}	4.00^{*}	3.18^{*}
W	0.92	0.46	0.47	0.96	1.63	3.65	0.99	4.00
M	-1.57	-1.35	-1.30	-0.78	-1.39	1.05	-1.97	1.56
G	0.48	0.13	1.35	1.32	0.51	2.25	0.78	3.13
W-M	2.49^{*}	1.81^{*}	–	1.74^{*}	3.02^{*}	2.6^{*}	2.96^{*}	2.44^{*}
M-G	-2.05^{*}	–	-2.65^{**}	-2.10^{*}	–	–	$-2.75^{(a)}$	–
W-G	–	–	–	–	–	–	–	–

Table 3. F statistics, mean amplitude values (in μV) for Material (W = Wood, M = Metal, G = Glass) by Electrode (Fz, Cz, Pz) interaction in the latency windows in which the effects were significant. Amplitude differences (W-M, M-G and W-G in μV) are indicated only when significant.

	150-230 ms			**230-330 ms**			**400-550 ms**			**550-700 ms**		
	P200 component			N280 component			NSW and P550 components					
	$F(4,84)=3.63^{**}$			$F(4,84)=2.83^{*}$			$F(4,84)=3.99^{**}$			$F(4,84)=7.06^{***}$		
	Fz	Cz	Pz	Fz	Cz	Pz	Fz	Cz	Pz	Fz	Cz	Pz
W	-1.25	1.57	2.44	-2.63	0.06	3.98	-4.21	0.99	8.12	-4.45	0.75	6.67
M	-3.93	-1.43	0.65	-4.77	-2.00	2.86	-7.75	-2.82	6.39	-8.71	-3.20	6.00
G	-1.49	1.09	1.83	-1.56	1.20	4.42	-5.05	-0.15	6.74	-4.61	0.32	6.65
W-M	2.68^{***}	3.00^{***}	1.79^{***}	2.14^{***}	2.06^{***}	–	3.54^{***}	3.81^{***}	1.73^{*}	4.26^{***}	2.45^{***}	–
M-G	-2.44^{***}	-2.52^{***}	-1.18^{**}	-3.21^{***}	-3.20^{***}	-1.56^{**}	-2.7^{***}	-2.67^{***}	–	-4.1^{***}	-3.52^{***}	–
W-G	–	–	–	–	–	–	–	–	–	–	–	–

4.3 Acoustic Data

A set of sound descriptors (as defined in Section 3) were calculated for the typical sounds from each material category: attack time AT, spectral centroid SCG, spectral bandwidth SB, spectral flux SF, roughness R and normalized sound decay α. An acoustic analysis was conducted to determine whether each sound descriptor explains sound categorization.

Table 4. F statistics, mean amplitude values (in μV) for Material (W = Wood, M = Metal, G = Glass) by Region (R1 = Fronto-central, R2 = Centro-temporal, R3 = Centro-parietal) interaction in the latency windows in which the effects were significant. Amplitude differences (W-M, M-G and W-G in μV) are indicated only when significant.

	400-550 ms F(4,84)=3.9**			**550-700 ms** F(4,84)=4.59**		
	R1	R2	R3	R1	R2	R3
W	-2.18	4.84	8.28	-1.76	5.55	8.23
M	-6.28	2.53	6.91	-6.23	3.53	7.39
G	-3.68	3.03	7.41	-2.87	4.17	8.10
W-M	4.1***	2.31***	1.37*	4.47***	–	–
M-G	-2.6***	-0.5***	–	-3.36***	–	–
W-G	–	–	–	–	–	–

Table 5. Statistics (χ^2) and mean values (reported with their respective standard deviation, $\pm$SD) for the Material factor (Wood, Metal, Glass) corresponding to AT (in ms), SCG (in Hz), SF, R (in asper) and α (in dB). The level of significance of post-hoc comparisons are indicated next to the mean values.

	AT	SCG	SF	R	α
χ^2(2,198)=	20.08***	33.55***	97.21***	71.71***	159.81***
Wood	0.24($\pm$0.47)	2573*($\pm$597)	105*($\pm$69)	0.4($\pm$0.32)	-0.088*($\pm$0.033)
Metal	0.11*($\pm$0.63)	3536($\pm$1154)	655($\pm$719)	1.71*($\pm$1.24)	-0.013*($\pm$0.006)
Glass	0.37($\pm$0.48)	3426($\pm$891)	550($\pm$403)	0.58($\pm$0.45)	-0.037*($\pm$0.016)

A Kruskal-Wallis test was computed for each sound descriptor including Material (Wood, Metal and Glass) as a factor. Results are presented in Table 5. Except for SB, all descriptors explained the differentiation of at least one category from the other 2. Specifically, AT and R were relevant to discriminate Metal from both Wood and Glass; SCG and SF to discriminate Wood from both Metal and Glass and α to discriminate the 3 categories.

5 Discussion

Behavioral data allowed us to define which of the 20 sounds from each continuum were most typical of each material category (i.e., the sounds that were categorized as Wood, Metal or Glass by more than 70% of the participants). The ERPs to these typical sounds were then analyzed to examine the time course of sound perception. Acoustic analysis of these typical sounds revealed the relevancy of sound descriptors to explain sound categorization. Results of these analyses allowed us to discuss sound descriptors that are most relevant for material perception.

Results of the acoustic analysis showed that the normalized sound decay α alone allowed the discrimination of the 3 material categories. Since in our case, the normalized sound decay characterizes the damping, the relevancy of α is in line with previous results showing that damping is an important cue in the perception of sounds from impacted materials [17,18,19,20]. However, while damping may be of particular importance for sound perception, results of several studies have shown that it is not sufficient for a complete discrimination between different materials. For instance, [19] concluded that the length of bars influenced the perception of material with glass and wood being associated with shorter bars (i.e., higher frequencies) than rubber and steel. Similarly, [15] also found frequency effects on material perception: glass and rubber were associated with higher pitches than wood and steel. Finally, [20] concluded that within a macro-category, material identification is based on plate size (i.e., signal frequency) so that steel and plastic were identified more often from larger plates than wood and glass.

Consequently, our perception of material seems to be guided by additional cues other than damping that most likely are linked to the spectral content of sounds. To directly test for the influence of damping and spectral complexity on the perception of the sounds used here, we constructed 9 stimuli by independently combining the spectra of typical sounds (of Wood, Metal or Glass) with the damping characteristics of these sounds. The experimental design is detailed in[2]. The percentage of responses obtained for the 9 hybrid sounds is presented in Table 6. Results confirmed that typical sounds (with spectrum and damping within the same category) were associated with the highest percentage of responses in their respective category. Moreover, results revealed that sounds with Metal damping and spectrum of Wood or Glass, were categorized as Metal, thereby showing that damping remains a robust descriptor to characterize the Metal category. For sounds with Wood damping, the one with Metal spectrum was still categorized as Wood while the one with Glass spectrum was categorized as Glass so that damping did not entirely characterize the Wood category. Finally, sounds with Glass

[2] **Experimental protocol: Participants:** 10 participants (3 women, 7 men) were tested in this experiment. They were all non-musicians and had normal audition. **Stimuli:** one reference sound (Section 2.1) of each material category (Wood, Metal and Glass) was selected. From the 3 reference sounds, 9 hybrid sounds were created using the additive synthesis model defined in Equation (1). The sounds were constructed by independently combining each spectrum (A_m and ω_m values) with the damping law characteristic of each material categorie (measured on the original sound). Thus, these sounds correspond to the 9 possible combinations between the 3 spectra and the 3 damping law characteristics of each material category. The sounds are available at http://www.lma.cnrs-mrs.fr/∼kronland/Categorization/sounds.html. **Procedure:** sounds were presented through headphones (Stax SR-202 with amplifier SRM-310). As a training, participants first listened passively to 2 typical sounds of each material category (different to the typical sounds used for the construction of the stimuli). Then, they listened passively to the 9 hybrid sounds presented randomly. Finally, they were asked to categorize these 9 sounds as Wood, Metal or Glass by using a computer keyboard. Participants could listen to the sounds as often as they wanted before answering. The responses for the 9 sounds were collected.

damping combined with either Wood or Metal spectrum lost their Glass specificity: they were mostly categorized as Metal. Thus, damping is not sufficient to characterize the Wood and Glass categories. Clearly, this experiment further showed that both damping and the spectral content of sounds are relevant cues for material categorization. Therefore, at least, one descriptor other than the normalized sound decay α (that robustly characterizes the damping of each material category since the 3 values of α for a given material damping have close values across different spectra, see Table 6) needs to be considered for an accurate sound categorization of materials.

Table 6. Percentage of responses in each material category (W, M and G) and normalized sound decay values (α) for the 9 sounds constructed by combination of Wood, Metal and Glass spectrum with Wood, Metal and Glass damping. The highest percentage obtained by each sound among the 3 categories is highlighted in bold.

	Wood damp.				Metal damp.				Glass damp.			
	W	M	G	α	W	M	G	α	W	M	G	α
Wood spec.	**80**	20	0	-0.033	0	**90**	10	-0.001	10	**70**	20	-0.012
Metal spec.	**60**	20	20	-0.047	0	**100**	0	-0.002	20	**70**	10	-0.020
Glass spec.	40	0	**60**	-0.047	0	**60**	40	-0.001	0	20	**80**	-0.013

Regarding the ERPs, the analysis revealed no significant differences on the amplitude of the N100 component that is notably influenced by variations in sound onset parameters [54]. This is not surprising insofar as the AT differences between sound categories were approximately 0.1 ms (see values in Table 5) and were therefore below the temporal resolution of hearing [55]. Then, typical Metal sounds elicited brain activity that differed from the other 2 typical sound categories starting around 150 ms after sound onset and lasting until the end of the recording period at some scalp sites. In particular, Metal sounds elicited smaller P200 and P550 components, and larger N280 and NSW components than Wood and Glass sounds. Results from previous experiments have shown that the P200 component, that typically reflects sensory and perceptual processing, is sensitive to spectral complexity [37,56]. In particular, [35] found larger P200 amplitude, and an additional effect on N100 amplitude, for instrumental tones (complex spectra) relative to sine wave tones. It should be noted, however, that while increased spectral complexity enhanced the amplitude of the N100 and/or P200 components in previous studies [37,35], the present results showed decreased P200 amplitude and increased N280 amplitude to Metal sounds. Two factors may account for these differences. First, the present task and experimental design differed from previous experiments because we used an explicit sound categorization task. Such tasks are known to elicit N200-P300 components that are thought to reflect the categorization and decision processes [41,39]. Thus, the occurrence of the N280 and P550 (P300-like component) is likely to be task-dependent. Note that since N280 amplitude was larger for Metal sounds, the reduction in P200 amplitude may result from an overlap effect. Second, the present sounds differed from those used in previous studies. In the experiments by [35] and [37], stimuli differed in spectral complexity but were equated in pitch (F4/B4 and C4, respectively) and in duration (400 and 500 ms, respectively). By contrast, our stimuli differed both in pitch (the chroma was the same, i.e., C, but the octaves sometimes differed) and in duration (mean values of normalized sound decay[3] for each category in Table 5). However, differences in pitch are not likely to explain the large ERPs differences between Metal sounds and the other 2 sound categories because the largest pitch differences are found between Glass sounds (notes C6 and C7) and the other 2 sound categories (Wood: note C4 and Metal: notes C4 and C5). By contrast, [57] have shown that increasing sound duration from 200 to 300 ms significantly increased N200 amplitude. Consequently, the N280 enhancement for Metal sounds can reflect a sound duration effect because of the longer duration of Metal sounds. Finally, a long-lasting NSW larger for Metal sounds than the other 2 sound categories developed over frontal regions. Based on previous studies, NSW may reflect processes related to the maintenance of stimuli in working memory, expectancy since a row of XXXX followed sound offset [58], attention [59] and also sound duration (as the "sustained potential" reported by [60]).

To summarize, the analysis of the electrophysiological data revealed that spectral complexity and sound duration seem to be relevant cues to explain the differentiation of the ERPs to Metal sounds compared to the ERPs to Glass and Wood sounds. Note that the occurrence of an N280 component is most likely driven by the categorization task but the increased amplitude of this component to Metal sounds is most likely linked to the longer duration of Metal compared to Wood and Glass sounds. The relevance of sound descriptors investigated in Section 4.3 can be compared with results from ERPs. In particular, we can support that descriptors reflecting the spectral content of sounds (i.e., SB, SCG, SF and R) and sound duration (which is related to α) may be relevant from a perceptual point of view. Note that the roughness R seems to be the most relevant descriptor to account for electrophysiological data since it allowed the distinction of Metal sounds from the other 2 categories. Moreover, from a perceptual point of view, the roughness R most accurately describes the typical dissonance of Metal sounds. By contrast, AT (characterizing sound onset) revealed as relevant descriptor by acoustic analysis to discriminate Metal sounds from both Wood and Glass sounds was not considered as a relevant descriptor from a perceptual point of view since no differences on N100 components were found in the ERPs results.

6 Conclusion

The aim of the current experiment was to investigate the perception of sounds from different impacted materials. For that purpose, we collected behavioral and electrophysiological data from a categorization experiment and we conducted acoustic analysis of typical sounds by investigating sound descriptors known to be relevant for timbre

[3] In our case, the normalized sound decay is related to sound duration. We chose to investigate sound decay instead of sound duration since the perceptual correlate of sound duration is not well defined. Actually, its estimation still is an open issue for rapidly damped signals since the perceived duration differs from the actual duration of the signals.

and material perception. Synthesis allowed the generation of perfectly controlled stimuli and the construction of sound continua through a morphing process. From the behavioral data, we were able to define typical sounds for each material category. ERPs results allowing the investigation of the brain dynamics processes supported the assumption that our perception of impact sounds from material categories is based on the damping (characterized by normalized sound decay) together with spectral content (characterized by spectral descriptors). Further, these relevant descriptors should allow us to address the problem of identifying acoustic invariants of sound categories. Future studies will aim at defining these invariants, assumed to be based on a combination of the relevant descriptors.

Acknowledgments

This research was supported by a grant from the Human Frontier Science Program (HSFP #RGP0053) to Mireille Besson and a grant from the French National Research Agency (ANR, JC05-41996, “senSons”) to Sølvi Ystad. Mitsuko Aramaki was supported by a post-doctoral grant from the HFSP and the ANR.

References

1. Michaels, C.F., Carello, C.: Direct perception. Prentice-Hall, Englewood Cliffs (1981)
2. Grey, J.M.: Multidimensional perceptual scaling of musical timbres. Journal of the Acoustical Society of America 61(5), 1270–1277 (1977)
3. Krumhansl, C.L.: Why is musical timbre so hard to understand. In: Structure and perception of electroacoustic sound and music. Elsevier, Amsterdam (1989)
4. Krimphoff, J., McAdams, S., Winsberg, S.: Caractérisation du timbre des sons complexes. ii: Analyses acoustiques et quantification psychophysique. Journal de Physique 4(C5), 625–628 (1994)
5. McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., Krimphoff, J.: Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. Psychological Research 58, 177–192 (1995)
6. von Helmholtz, H.L.F.: On the sensations of tone as the physiological basis for the theory of music, 2nd edn (1877)
7. Sethares, W.A.: Local consonance and the relationship between timbre and scale. Journal of the Acoustical Society of America 94(3), 1218–1228 (1993)
8. Vassilakis, P.N.: Selected Reports in Ethnomusicology (Perspectives in Systematic Musicology). In: Auditory roughness as a means of musical expression, Department of Ethnomusicology, University of California, vol. 12, pp. 119–144 (2005)
9. Freed, D.J.: Auditory correlates of perceived mallet hardness for a set of recorded percussive events. Journal of the Acoustical Society America 87(1), 311–322 (1990)
10. Tucker, S., Brown, G.J.: Investigating the perception of the size, shape and material of damped and free vibrating plates. Technical Report CS-02-10, Université de Sheffield, Department of Computer Science (2002)
11. van den Doel, K., Pai, D.K.: The sounds of physical shapes. Presence 7(4), 382–395 (1998)
12. Kunkler-Peck, A.J., Turvey, M.T.: Hearing shape. J. of Experimental Psychology: Human Perception and Performance 26(1), 279–294 (2000)

13. Carello, C., Anderson, K.L., Kunkler-Peck, A.J.: Perception of object length by sound. Psychological Science 9(3), 211–214 (1998)
14. Lutfi, R.A., Oh, E.L.: Auditory discrimination of material changes in a struck-clamped bar. Journal of the Acoustical Society of America 102(6), 3647–3656 (1997)
15. Rocchesso, D., Fontana, F.: The sounding object (2003)
16. Lakatos, S., McAdams, S., Caussé, R.: The representation of auditory source characteristics: simple geometric form. Perception & Psychophysics 59, 1180–1190 (1997)
17. Wildes, R.P., Richards, W.A.: Recovering material properties from sound. In: Richards, W.A. (ed.), ch. 25, pp. 356–363. MIT Press, Cambridge (1988)
18. Gaver, W.W.: How do we hear in the world? explorations of ecological acoustics. Ecological Psychology 5(4), 285–313 (1993)
19. Klatzky, R.L., Pai, D.K., Krotkov, E.P.: Perception of material from contact sounds. Presence: Teleoperators and Virtual Environments 9(4), 399–410 (2000)
20. Giordano, B.L., McAdams, S.: Material identification of real impact sounds: Effects of size variation in steel, wood, and plexiglass plates. Journal of the Acoustical Society of America 119(2), 1171–1181 (2006)
21. Chaigne, A., Lambourg, C.: Time-domain simulation of damped impacted plates: I. theory and experiments. Journal of the Acoustical Society of America 109(4), 1422–1432 (2001)
22. Valette, C., Cuesta, C.: Mécanique de la corde vibrante (Mechanics of vibrating string). Traité des Nouvelles Technologies, série Mécanique. Hermès (1993)
23. Risset, J.-C., Wessel, D.L.: Exploration of timbre by analysis and synthesis. In: The psychology of music, 2nd edn. Cognition and Perception, pp. 113–169. Academic Press, London (1999)
24. Donders, F.C.: On the speed of mental processes. Acta Psychol. 30, 412–431 (1969)
25. Sternberg, S.: Memory-scanning: mental processes revealed by reaction-time experiments. American Scientist 57(4), 421–457 (1969)
26. Rugg, M.D., Coles, M.G.H.: The ERP and Cognitive Psychology: Conceptual issues. In: Electrophysiology of mind. Event-related brain potentials and cognition. Oxford Psychology, vol. 25, pp. 27–39. Oxford University Press, Oxford (1995)
27. Eggermont, J.J., Ponton, C.W.: The neurophysiology of auditory perception: from single-units to evoked potentials. Audiology & Neuro-Otology 7, 71–99 (2002)
28. Näätänen, R.: Attention and brain function. Erlbaum, Hillsdale, NJ (1992)
29. Goydke, K.N., Altenmuller, E., Moller, J., Munte, J.: Changes in emotional tones and instrumental timbre are reflected by the mismatch negativity. Cognitive Brain Research 21, 351–359 (2004)
30. Tervaniemi, M.: Musical sound processing in the human brain: evidence from electric and magnetic recordings. Annals of the New York Academy of Sciences 930, 259–272 (2001)
31. Tervaniemi, M., Winkler, I., Näätänen, R.: Pre-attentive categorization of sounds by timbre as revealed by event-related potentials. NeuroReport 8, 2571–2574 (1997)
32. Toiviainen, P., Tervaniemi, M., Louhivuori, J., Saher, M., Huotilainen, M., Näätänen, R.: Timbre similarity: convergence of neural, behavioral, and computational approaches. Music Perception 16, 223–241 (1998)
33. Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M.-H., McAdams, S.: Separate neural processing of timbre dimensions in auditory sensory memory. Journal of Cognitive Neuroscience 18(12), 1959–1972 (2006)
34. De Baene, W., Vandierendonck, A., Leman, M., Widmann, A., Tervaniemi, M.: Roughness perception in sounds: behavioral and erp evidence. Biological Psychology 67, 319–330 (2004)
35. Meyer, M., Baumann, S., Jancke, L.: Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. NeuroImage 32, 1510–1523 (2006)

36. Shahin, A., Bosnyak, D.J., Trainor, L.J., Roberts, L.E.: Enhancement of neuroplastic p2 and n1c auditory evoked potentials in musicians. The Journal of Neuroscience 23(12), 5545–5552 (2003)
37. Shahin, A., Roberts, L.E., Pantev, C., Trainor, L.J., Ross, B.: Modulation of p2 auditory-evoked responses by the spectral complexity of musical sounds. NeuroReport 16(16), 1781–1785 (2005)
38. Pascual-Marqui, R.D., Michel, C.M., Lehmann, D.: Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. International Journal of Psychophysiology 18, 49–65 (1994)
39. Sutton, S., Braren, M., Zubin, J., John, E.R.: Evoked potential correlates of stimulus uncertainty. Science 150, 1187–1188 (1965)
40. Donchin, E., Coles, M.G.H.: Is the p300 component a manifestation of context updating? Behavioral and Brain Sciences 11, 357–374 (1988)
41. Ritter, W., Simson, R., Vaughan, H.G.: Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. Psychophysiology 20, 168–179 (1983)
42. Aramaki, M., Kronland-Martinet, R.: Analysis-synthesis of impact sounds by real-time dynamic filtering. IEEE Transactions on Audio, Speech, and Language Processing 14(2), 695–705 (2006)
43. Parncutt, R.: Harmony - A Psychoacoustical Approach. Springer, Heidelberg (1989)
44. Aramaki, M., Baillères, H., Brancheriau, L., Kronland-Martinet, R., Ystad, S.: Sound quality assessment of wood for xylophone bars. Journal of the Acoustical Society of America 121(4), 2407–2420 (2007)
45. Jasper, H.H.: The ten-twenty electrode system of the international federation. Electroencephalography and Clinical Neurophysiology 10, 371–375 (1958)
46. Delorme, A., Makeig, S.: Eeglab: an open source toolbox for analysis of single-trial eeg dynamics. Journal of Neuroscience Methods 134, 9–21 (2004)
47. Beauchamp, J.W.: Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones. Journal of the Audio Engineering Society 30(6), 396–406 (1982)
48. Marozeau, J., de Cheveigné, A., McAdams, S., Winsberg, S.: The dependency of timbre on fundamental frequency. Journal of the Acoustical Society of America 114, 2946–2957 (2003)
49. McAdams, S.: Perspectives on the contribution of timbre to musical structure. Computer Music Journal 23(3), 85–102 (1999)
50. Plomp, R., Levelt, W.J.M.: Tonal consonance and critical bandwidth. Journal of the Acoustical Society of America 38, 548–560 (1965)
51. Terhardt, E.: On the perception of periodic sound fluctuations (roughness). Acustica 30(4), 201–213 (1974)
52. Vassilakis, P.N.: Sra: A web-based research tool for spectral and roughness analysis of sound signals. In: Proceedings of the 4th Sound and Music Computing (SMC) Conference, pp. 319–325 (2007)
53. Aramaki, M., Brancheriau, L., Kronland-Martinet, R., Ystad, S.: Computer Music Modeling and Retrieval - Genesis of Meaning of Sound and Music. In: Perception of impacted materials: sound retrieval and synthesis control perspectives. LNCS, vol. 5493, pp. 138–151. Springer, Heidelberg (2009)
54. Hyde, M.: The n1 response and its applications. Audiology & Neuro-Otology 2, 281–307 (1997)
55. Gordon, J.W.: The perceptual attack time of musical tones. Journal of Acoustical Society of America 82(1), 88–105 (1987)
56. Kuriki, S., Kanda, S., Hirata, Y.: Effects of musical experience on different components of meg responses elicited by sequential piano-tones and chords. The Journal of Neuroscience 26(15), 4046–4053 (2006)

57. Kushnerenko, E., Ceponiene, R., Fellman, V., Huotilainen, M., Winkler, I.: Event-related potential correlates of sound duration: similar pattern from birth to adulthood. NeuroReport 12(17), 3777–3781 (2001)
58. Walter, W.G., Cooper, R., Aldridge, V.J., McCallum, W.C., Winter, A.L.: Contingent negative variation: an electrical sign of sensorimotor association and expectancy in the human brain. Nature 230, 380–384 (1964)
59. King, J., Kutas, M.: Who did what and when? using word- and clause-level erps to monitor working memory usage in reading. Journal of Cognitive Neuroscience 7(3), 376–395 (1995)
60. Alain, C., Schuler, B.M., McDonald, K.L.: Neural activity associated with distinguishing concurrent auditory objects. Journal of Acoustical Society of America 111(2), 990–995 (2002)