Perception of Impacted Materials: Sound Retrieval and Synthesis Control Perspectives

Mitsuko Aramaki[1,2], Loïc Brancheriau[3], Richard Kronland-Martinet[4], and Sølvi Ystad[4]

[1] CNRS - Institut de Neurosciences Cognitives de la Méditerranée,
31, chemin Joseph Aiguier 13402 Marseille Cedex 20, France
aramaki@incm.cnrs-mrs.fr
[2] Aix-Marseille - Université,
58, Bd Charles Livon 13284 Marseille Cedex 07, France
[3] CIRAD - PERSYST Department,
TA B-40/16, 73 Rue Jean-Franois Breton, 34398 Montpellier Cedex 5, France
loic.brancheriau@cirad.fr
[4] 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 study, we aimed at determining statistical models that allowed for the classification of impact sounds according to the perceived material (Wood, Metal and Glass). For that purpose, everyday life sounds were recorded, analyzed and resynthesized to insure the generation of realistic sounds. Listening tests were conducted to define sets of typical sounds of each material category by using a statistical approach. For the construction of statistical models, acoustic descriptors known to be relevant for timbre perception and for material identification were investigated. These models were calibrated and validated using a binary logistic regression method. A discussion about the applications of these results in the context of sound synthesis concludes the article.

1 Introduction

Sound classification systems are based on the calculation of acoustic descriptors that are extracted by classical signal analysis techniques such as spectral analysis or time-frequency decompositions. In this context, many sound descriptors depending on the specificities of sound categories were defined in the literature, in particular in the framework of MPEG 7 [1]. Indeed, descriptors were proposed for speech recognition [2], audio indexing [3,4], music classification [5] or for psycho-acoustical studies related to timbre [6,7]. Nevertheless, these classification processes would be significantly improved if perceptually relevant information conveyed in acoustic signals could be identified to enable fewer and more relevant descriptors to characterize the signal.

In this current study, we aim at determining statistical models that allow categorization of impact sounds as a function of the perceived material based on few acoustic descriptors. For that purpose, we investigated the acoustic descriptors that are known to be relevant for timbre perception and material identification. In practice, a sound

S. Ystad, R. Kronland-Martinet, and K. Jensen (Eds.): CMMR 2008, LNCS 5493, pp. 134–146, 2009.

data bank constituted of realistic impact sounds from different materials (Wood, Metal, Glass) was generated using analysis-synthesis techniques. Then, a morphing process allowed us to build sound continua that simulate continuous transitions between sounds corresponding to different materials. Listening tests were conducted and we used a statistical approach to determine the sets of sounds that were judged as typical and non typical for each material category. The use of sound continua allowed for the determination of perceptual borders between these typical and non typical sounds as a function of the position along the continua to identify the variation range of parameters for each material category. A statistical model of binary logistic regression was calibrated and validated based on the calculation of selected descriptors for each sound category. We finally address some perspectives of this study in particular, in the domain of sound synthesis.

2 Categorization Experiment

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

Sounds were pseudo-randomly presented through one loudspeaker (Tannoy S800) located 1 m in front of the participants who were asked to categorize sounds as from impacted Wood, Metal or Glass materials, as fast as possible, by pressing one response button out of three. The association between response buttons and material categories was balanced across participants. Participants' responses were collected and were averaged for each sound.

To design stimuli, we first recorded (at 44.1 kHz sampling frequency) impact sounds from everyday life objects of various materials (i.e., impacted wooden beams, metallic plates and various glass bowls) that unambiguously evoked each material category. Then, based on the analysis-synthesis model described in [8], we resynthesized these recorded sounds. To minimize timbre variations induced by pitch changes, all sounds were tuned to the same chroma (note C), but not to the same octave, due to the specific properties of the different materials. For instance, Glass sounds could not be transposed to low pitches as they would no longer be recognized as Glass sounds. The synthesized sounds therefore differed by 1, 2 or 3 octaves depending upon the material. The new pitches of the tuned sounds were obtained by transposition (dilation of the original spectra). In practice,Wood sounds were tuned to the pitch C4, Metal sounds to the pitches C4 and C5 and Glass sounds to the pitches C6 and C7. Based upon previous results showing high similarity ratings for tone pairs that differed by octaves [9], an effect known as the octave equivalence, we presumed that the octave differences between sounds should not influence categorization. Each time the pitch was modified, the new value of the damping coefficient of each tuned frequency component was recalculated according to a damping law measured on the original sound [10], since the frequency-dependency of the damping is fundamental for material perception [11]. Sounds were finally equalized by gain adjustments to avoid an eventual influence of loudness in the categorization judgments.

The resynthesized sounds were further used to create 15 sound continua that simulate a progressive transition between the different materials. In particular, we built 5 continua for each of the following 3 transitions: {Wood-Metal}, {Wood-Glass} and {Glass-Metal}. Each continuum, composed of 20 hybrid sounds, is built using a morphing process. The interpolation on the amplitudes of the spectral components was computed by a crossfade technique. Concerning the damping, the coefficients were estimated according to a hybrid damping law calculated at each step of the morphing process. This hybrid damping law was computed from an effective linear interpolation between the 2 extreme damping laws. In practice, 15 continua (5 for each of the 3 transitions) were built. Sound examples are available at http://www.lma.cnrs-mrs.fr/∼kronland/Categorization/sounds.html.

3 Determination of Sound Categories

Based on participants' responses, we aim at determining sounds considered as *typical* of each material category (Wood, Metal or Glass) by using a statistical approach. Actually, the sounds were produced to form progressive transitions from one category to another. The limits between categories along the continua are thus not known. This causes an intermediate zone with sounds that represent a mixture of the two extreme sounds. From a perceptual point of view, the sounds contained in this zone are perceived as ambiguous.

The limits of each category along continua have to be estimated to define sets of sounds that are typical of Wood, Metal and Glass categories. The specific perceptive space used to address this problem is defined by the three material axes {Wood, Metal, Glass}. Since the experimental protocol was based on a categorization task with three-choice answers, the percentages of responses obtained for material categories are correlated as shown in Table 1. Thus, these axes are not orthogonal and a principal component analysis is therefore carried out (Table 2).

Table 1. Pearson correlation coefficients of percentages of responses obtained in Wood, Metal and Glass material category (N= 300). The correlations are significant at the .01 level (2-tailed) and are represented by **.

	Wood	**Metal**
Metal	-0.68**	1
Glass	-0.33**	-0.48**

Two principal components are extracted; the third component can be neglected (Table 2). The perceptive space is thus totally defined in a plane and sounds are grouped inside a triangle as shown in Figure 1. This data representation shows the repartition of responses between the three material categories. In a case of a clear discrimination between categories, all sounds should be distributed in three distinct groups located

Table 2. Total variance explained by principal component analysis on percentages of responses obtained for Wood, Metal and Glass categories (N = 300)

Component	Eigenvalue	% of variance	Cumulative (%)
1	1.7	57	57
2	1.3	43	100
3	0	0	100

at the apexes of the triangle and the gaps between groups would mark the limits between categories. Figure 1 does not show three distinct groups and in this sense, reveals difficulties encountered by participants to classify sounds at intermediate position of the continua. Moreover, sounds located on the Metal-Glass side of the triangle are well-aligned, meaning that these sounds were classified as Metal or Glass but never as Wood. By contrast, sounds located on the Wood-Metal side are not well-aligned, meaning that these sounds were classified as Wood or Metal but also as Glass. Consequently, it is not possible to determine the limits of the categories directly from the perceptive space because of the high number of sounds scattered inside the triangle (ambiguous zone).

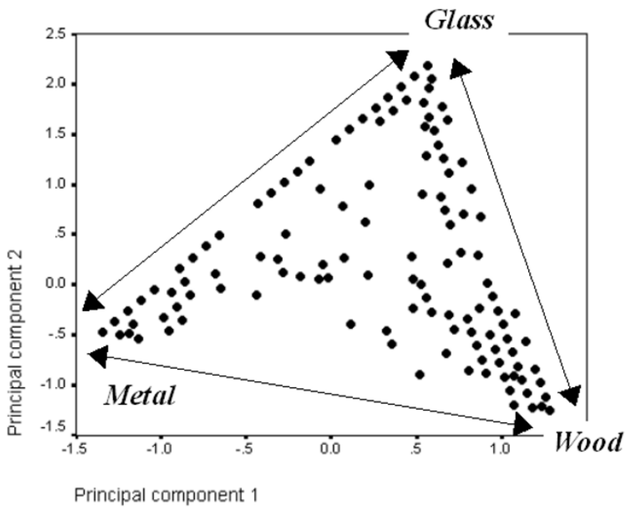

Fig. 1. Perceptive space defined by the two principal components (factor score plot, N = 300)

In order to determine the limits of the categories, a hierarchical cluster analysis is carried out on the principal components in a second stage. As presented previously, the two principal components completely define the perceptive space and are orthogonal. The distance measurement used is then the Euclidian distance. The method used for combining clusters is the furthest neighbor (complete linkage). This method is adapted to the research of globular groups with unequal variances and unequal sizes. Table 3

Table 3. Simplified agglomeration schedule of the hierarchical clustering procedure (N = 292; 8 sounds were missing because their distance does not allow an attribution to one of the three groups A, B or C). At each agglomerating step, the type of continua, the Min and Max indices, the number of samples in the agglomerate (S), the identified group (A, B or C) and the associated category (Wood, Metal or Glass) are indicated.

Agglomerating step	Continua	Min index	Max index	S	Identified group	Associated category
1	W-G	14	21	20	A	Glass
2	G-M	2	9	35	A	Glass
3	W-G	13	21	9	A	Glass
4	G-M	2	12	11	A	Glass
5	W-M	9	12	8	B	Metal
6	G-M	10	13	4	B	Metal
7	W-M	10	21	52	B	Metal
8	G-M	11	21	50	B	Metal
9	W-M	2	10	23	C	Wood
10	W-G	2	17	62	C	Wood
11	W-M	4	11	12	C	Wood
12	W-G	9	19	6	C	Wood

presents the simplified diagram of agglomeration obtained by hierarchical cluster analysis. The agglomeration procedure first identifies a group denoted A (from step 1 to 4) to which we associated the Glass category, because Glass is the common element of the first 4 agglomerating steps. From the relative position between sounds in the perceptive space, Glass thus constitutes the most easily identifiable category. The procedure identifies a second group denoted B (from step 5 to 8) to which we associated the Metal category. Finally, the third group C (from step 9 to 12) was associated with the Wood category. Wood thus constitutes the material category which is most difficult to identify. Interestingly, for most agglomerating steps (except for steps 5 and 6), the clustering procedure identified groups of sounds (A, B or C) located at the extreme of continua since the Min or Max index corresponds to an extreme position of the continua (i.e., position 2 or 21) or to a position close to an extreme (i.e., 4 or 17). This procedure is relevant from a perceptual point of view since it means that sets of typical sounds were located at the extreme of the continua.

Then, the values of “Min” and “Max” indices in Table 3 were used as far as possible to define sound indices delimiting categories. As an example, the step 2 revealed an agglomerate from index 2 (beginning of the Glass agglomerate) to 9 (end of the Glass agglomerate approaching Metal) associated to Glass category. In this case, the retained index delimiting the Glass category along Glass-Metal continua is 9. This process is

Table 4. Indices delimiting categories for each type of continua. The indices from steps linked with symbol - mark the membership to an agglomerate in an intermediate zone. The mark X indicates the inconsistency in the agglomerating step 12.

Transition	**Indices from steps**	**Indices delimiting categories**
Glass-Metal	9, 12, 10-13, 11	{9-13}
Wood-Metal	9-12, 10, 10, 11	{9-12}
Wood-Glass	14, 13, 17, X	{13-17}

Table 5. Percentages of responses in each category (Wood, Metal or Glass) for sounds located at the indices delimiting the categories for the 5 different continua. Sounds are denoted G_i-M_i-K, W_i-M_i-K and W_i-G_i-K for the i^{th} continua with K the index along the continua (09 or 13 for Glass-Metal, 09 or 12 for Wood-Metal and 13 or 17 for Wood-Glass transition). Averaged percentage values (Mean) across the 5 continua of a given transition in the associated categories (e.g., mean values of G and M columns for Glass-Metal transition) and bilateral errors (Error) associated to the mean estimation with an unknown variance (significance level of .05).

GLASS-METAL				**WOOD-METAL**				**WOOD-GLASS**			
	W	M	**G**		**W**	M	G		**W**	M	G
G_1-M_1-09	8	36	**56**	W_1-M_1-09	**44**	40	16	W_1-G_1-13	**28**	12	60
G_2-M_2-09	0	28	**72**	W_2-M_2-09	**96**	4	0	W_2-G_2-13	**92**	0	8
G_3-M_3-09	0	20	**80**	W_3-M_3-09	**12**	68	20	W_3-G_3-13	**96**	0	0
G_4-M_4-09	0	36	**64**	W_4-M_4-09	**16**	52	32	W_4-G_4-13	**64**	8	28
G_5-M_5-09	0	24	**76**	W_5-M_5-09	**20**	60	20	W_5-G_5-13	**72**	4	24
Mean			**70**		**38**				**70**		
Error			12		43				34		
	W	**M**	G		W	**M**	G		W	M	**G**
G_1-M_1-13	0	**96**	4	W_1-M_1-12	12	**80**	8	W_1-G_1-17	4	16	**80**
G_2-M_2-13	0	**96**	4	W_2-M_2-12	20	**52**	28	W_2-G_2-17	68	4	**28**
G_3-M_3-13	0	**72**	28	W_3-M_3-12	8	**92**	0	W_3-G_3-17	60	4	**36**
G_4-M_4-13	0	**96**	4	W_4-M_4-12	4	**80**	16	W_4-G_4-17	24	4	**72**
G_5-M_5-13	0	**88**	12	W_5-M_5-12	4	**92**	0	W_5-G_5-17	12	4	**80**
Mean		**90**				**79**					**59**
Error		13				20					31

repeated for all the agglomerates of each step. The agglomerating step 12 is withdrawn from the analysis since the indices of the agglomerate (i.e., 9-19 on Wood-Glass continua) delimit a zone on the Glass side, that is incoherent with the identified group (C)

Table 6. Criterion of category membership according for each transition

Transition	Criterion	
Glass-Metal	Glass if	$G \geq 70\%$
	Metal if	$M \geq 90\%$
Wood-Metal	Wood if	$W \geq 38\%$
	Metal if	$M \geq 79\%$
Wood-Glass	Wood if	$W \geq 70\%$
	Glass if	$G \geq 59\%$

and associated category (Wood). The indices delimiting categories (and consequently delimiting intermediate zones) are defined for each type of continua in Table 4. For instance, along the Glass-Metal continua, the Glass category goes until index 9 and the Metal category starts from index 13.

Then, from these indices delimiting categories, we define the criteria of category membership as a threshold of percentage of responses in a category for each type of transition. In practice, we define these criteria as equal to the average of the percentages of responses obtained for sounds at the indices defined in Table 4 across the 5 different continua of a given transition. The calculation of these average values is shown in Table 5. The membership criteria for each transition are summarized in Table 6. Note that all criteria are superior to 50% except for the Wood membership criterion in the Wood-Metal continua (i.e., $38\% \pm 43\%$). This value is not accurate from a perceptual point of view since it would lead to define the sounds (issued from Wood-Metal continua) that were categorized in the Wood category by only 38% of participants as typical Wood sounds (actually, these sounds were more often classified in the Metal category, see Table 5). This incoherency can be explained by the fact that the average calculation is based on a small number of values (average on 5 values). In order to minimize the uncertainty associated with these membership criteria (revealed by large intervals of confidence, see error values in Table 5) that directly influences the construction and the validation of predictive models, we determine an unique category membership criterion equal to the average value of all the membership criteria, i.e., $67.6\% \pm 9.9\%$ (N=30). For the sake of simplicity, we defined a criterion of category membership equal to 70% (value that belongs to the interval of confidence) for this current study.

4 Relationship between Acoustic Descriptors and Sound Categories

4.1 Acoustic Descriptors

To characterize sounds from an acoustic point of view, we considered the following sound descriptors known to be relevant for timbre perception and material identification: attack time AT, spectral centroid CGS, spectral bandwidth SB, spectral flux SF, roughness R and normalized sound decay α.

The attack time AT is defined as the time necessary for the signal energy to raise from 10% to 90% of the maximum amplitude of the signal. The spectral centroid CGS and the spectral bandwidth SB were defined by:

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

where ω and $\hat{s}$ respectively represent the frequency and the Fourier transform of the signal. The spectral flux SF is a spectro-temporal timbre descriptor quantifying the time evolution of the spectrum. The definition presented in [12] was chosen. The roughness R of a sound is commonly associated to the presence of several frequency components within the limits of a critical band. In particular, R is closely linked to the concept of consonance/dissonance [13]. The definition presented in [14] was chosen.

Finally, the sound decay quantifies the global amplitude decrease of the temporal signal and is directly correlated to the damping. Since the damping is a fundamental cue for material perception [11,15,16,17], the sound decay is assumed to be a relevant acoustic descriptor for our sounds. In practice, the sound decay is estimated from 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. In our case, typical sounds present a high variability of spectral content across material categories. Consequently, to allow comparisons between sound decay values, we further chose to consider a sound decay that was normalized with respect to a reference that takes into account the spectral localization of the energy, i.e., the CGS value.

4.2 Binary Logistic Regression Analysis

From classifications obtained from perceptual tests, we aim at estimating the membership of a sound in a material category starting from the calculation of the 6 acoustic descriptors presented in the previous section: $\{$AT, CGS, SB, SF, R, $\alpha\}$.

In our case, the dependent variables are qualitative; they represent the membership (True) or the non membership (False) of the category. In order to build statistical models to estimate the membership to a category, the binary logistic regression method is used. The associated method of multinomial logistic regression is not adapted to the problem because the best estimators can be different from one category to another.

One statistical model of binary logistic regression for each material category (Wood, Metal and Glass) are then built based on the acoustic descriptors. The problem of collinear parameters is overcome by a forward stepwise regression. Logistic regression allows one to predict a discrete outcome, such as group membership, from a set of variables that may be continuous, discrete, dichotomous, or a mixture of any of these. The dependent variable in logistic regression is usually dichotomous, that is, the dependent variable can take the value 1 (True) with a probability of success π, or the value 0 (False). This type of variable is called a Bernoulli (or binary) variable. Logistic regression makes no assumption about the distribution of the independent variables. They do not have to be normally distributed, linearly related or of equal variance within each group. The relationship between the predictors and response variables is not a linear function in logistic regression. The logistic regression function which is the logit transformation of π is used:

$$\pi(x) = P(Y = 1/X = x) = \frac{e^{L_{Cat}(x)}}{1 + e^{L_{Cat}(x)}} \tag{2}$$

with

$$L_{Cat}(x) = \beta_0 + \sum_i \beta_i x_i \tag{3}$$

The function $L_{Cat}(x)$ was calibrated and validated for each category (Cat={Wood, Metal, Glass}). Because we are dealing with sound continua, a segmented cross validation procedure was used. The validation set was built by selecting 1 of 3 sounds (corresponding to a set of 67 sounds). The calibration set was composed by the remaining sounds (corresponding to a set of 133 sounds). Note that for a given material model, the validation and calibration sets were built by excluding sound continua that did not contain the material. For instance, the Metal category model is not concerned by the 5 sound continua of the transition {Wood-Glass}. For each category, the membership of sounds corresponded to the set of "typical" sounds that were defined from the results of the listening test and from a statistical approach (cf. sections 2 and 3). A stepwise selection method was used and the statistical analysis was conducted with SPSS software (Release 11.0.0, LEAD Technologies).

4.3 Results and Discussion

For each category model, the step summary is given in Table 7. The statistics Cox & Snell R^2 and Nagelkerke adjusted R^2 try to simulate determination coefficients which, when used in linear regression, give the percentage variation of the dependent variable explained by the model. Because a binary logistical model is used, the interpretation of R^2 is not quite the same. In this case, the statistics give an idea of the strength of the association between the dependent and independent variables (a pseudo-R^2 measure).

The results for calibration and validation processes are given in Table 8. Thus, the predictive models are expressed by the function $\pi(x)$ in Eq. (2) and $L_{Cat}(x)$ for each category {Wood, Metal, Glass} is respectively given by:

$$\begin{aligned} L_{Wood}(\alpha, CGS, SB) &= -38.5 - 196\alpha - 0.00864CGS + 0.0161SB \\ L_{Metal}(\alpha, SB) &= 14.7 + 322\alpha - 0.00253SB \\ L_{Glass}(SB, CGS, R, \alpha, SF) &= 14.33 - 0.006SB + 0.002CGS - 3.22R \\ &\quad +52.69\alpha - 0.001SF \end{aligned} \tag{4}$$

Table 7. Step summary (Nagelkerke R^2 adjusted value and the variable entered) of the logistic regression method for each material category (Wood, Metal, Glass)

Category	Step	Nagelkerke R^2 adjusted	Variable entered
WOOD	1	.616	α
	2	.644	CGS
	3	.854	SB
METAL	1	.637	α
	2	.718	SB
GLASS	1	.086	SB
	2	.164	CGS
	3	.300	R
	4	.377	α
	5	.410	SF

Table 8. Classification table for each category (Wood, Metal, Glass) calculated on the calibration (N=133) and validation (N=67) populations. The cut value is .5.

		Calibration			Validation		
	Observed \ Predicted	False	True	% correct	False	True	% correct
WOOD	False	78	4	95.1	40	1	97.5
	True	8	43	84.3	4	22	84.6
	Overall %	90.7	91.5	91	91	95.6	92.5
	Observed \ Predicted	False	True	% correct	False	True	% correct
METAL	False	50	12	80.6	27	7	79.4
	True	6	65	91.5	1	32	97
	Overall %	89.3	84.4	86.5	96.4	82	88
	Observed \ Predicted	False	True	% correct	False	True	% correct
GLASS	False	91	7	92.9	43	3	93.4
	True	17	19	52.8	10	10	50
	Overall %	84.2	73	82.1	81.1	77	80.3

The logistic regression method revealed that the α parameter was the main predictor for Wood (overall percentage correct equal to 85% at step 1) and Metal (82.7%)

categories. This result was in line with several studies showing that the damping is an important acoustic feature for the material perception. Following α, the other important descriptors revealed by the analyses were related to the spectral content of the sounds ($\{$CGS, SB$\}$ for Wood and SB for Metal), meaning that spectral information are also important to explain material categorization. For the Glass category, the majority of the descriptors were of equal importance in the classification model (all descriptors were taken into account except AT). Thus, by contrast with the Wood and Metal categories, the membership of the Glass category could not be accurately predicted with few descriptors. Moreover, the most relevant predictors for this category revealed by the analyses were the spectral descriptors, ($\{$SB, CGS, R$\}$), while the temporal α parameter which was the main predictor for Wood and Metal only was relegated at the 4th rank. This may be due to the fact that Glass sounds presented a higher variability in sound decay values than Wood or Metal sounds. More interestingly, another explanation can be found in the specificity of the Glass category for which the material perception is intricately associated to drinking glasses (as everyday life objects). The corresponding sounds are generally characterized by high pitches (associated to small objects) and crystal-clear sounds (few spectral components, due to the structure's characteristics). Consequently, the discrimination between Glass sounds and the other sound categories can be explained by the spectral properties of Glass sounds rather than by the damping.

5 Sound Synthesis Perspectives

In addition to sound classification processes, these results are of importance in the context of synthesis control. In particular, we are currently interested in offering an intuitive control of synthesis models for an easy manipulation of intrinsic sound properties such as timbre. This aspect is for instance of importance within the Virtual Reality domain. Indeed, the use of synthesis models can dramatically be improved in "sonification" processes which generally deal with the choice of optimal synthesis parameters to control sounds directly from a verbal description (in our case, directly from the label of the material category: Wood, Metal or Glass). According to this perspective, we assume that the acoustic descriptors highlighted in the predictive models would constitute a reliable reference. In this section, we propose a discussion on their actual relevancy from a synthesis point of view.

First, the parameter α (related to the damping) was confirmed as an important predictor coherent with previous psychoacoustical studies showing that damping is an essential cue for material perception. The parameter α was kept as an accurate control parameter and was integrated in the control of the percussive synthesizer developed in our group [18]. Moreover, the determination of typical and non typical sounds on each sound continuum allowed us to define characteristic domains of parameter range values for each material category.

In addition to α, the statistical analyses further highlighted CGS and SB as most relevant parameters in the predictive models. These results are in line with post hoc synthesis experiences which revealed that, in addition to the damping, another parameter controlling the spectral content of sounds is necessary for a more complete manipulation of the perceived materials. To address this control, we also aim at integrating data from brain imaging that point out the perceptual/cognitive relevancy of descriptors. Actually, a separate analysis reflecting the temporal dynamics of the brain processes observed through electrophysiological data (also collected during the listening tests), revealed that spectral aspects of sounds may account for the perception and categorization of these different categories (see companion article [19]). This argument supports the fact that the control of the synthesis model can be accurately improved by adding control parameters related to spectral characteristics of sounds. Nevertheless, direct manipulations of the statistically relevant parameters CGS or SB of a given impact sound do not allow intuitive modifications of the nature of the perceived material. Based on these considerations, we are currently investigating the possibilities to define a control space for material categories where the control of spectral characteristics of sounds should particularly render the typical dissonant aspect of Metal sounds [20].

Acknowledgments

This research was supported by a grant from the French National Research Agency (ANR, JC05-41996, "senSons") to Sølvi Ystad.

References

1. Martínez, J.M.: Mpeg-7 overview, version 10, `http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm` (last checked March 25, 2008)
2. Foote, J.: Decision-tree probabilty modeling for HMM speech recognition. Ph.D thesis, Cornell university (1994)
3. Roy, P., Pachet, F., Krakowski, S.: Improving the classification of percussive sounds with analytical features: A case study. In: Proceedings of the 8th International Conference on Music Information Retrieval, Vienna, Austria (2007)
4. Slaney, M.: Mixtures of probability experts for audio retrieval and indexing. In: Proceedings of the IEEE International Conference on Multimedia and Expo, Lausanne, Switzerland (2002)
5. Pachet, F., Zils, A.: Evolving automatically high-level music descriptors from acoustic signals. In: Wiil, U.K. (ed.) CMMR 2003. LNCS, vol. 2771, pp. 42–53. Springer, Heidelberg (2004)
6. 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)
7. Peeters, G., McAdams, S., Herrera, P.: Instrument sound description in the context of mpeg-7. In: Proceedings of the International Computer Music Conference, Berlin, Germany (2000)
8. 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)
9. Parncutt, R.: Harmony - A Psychoacoustical Approach. Springer, Heidelberg (1989)
10. 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)

11. 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)
12. McAdams, S.: Perspectives on the contribution of timbre to musical structure. Computer Music Journal 23(3), 85–102 (1999)
13. Sethares, W.A.: Local consonance and the relationship between timbre and scale. Journal of the Acoustical Society of America 94(3), 1218–1228 (1993)
14. 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)
15. 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)
16. 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)
17. 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)
18. Aramaki, M., Kronland-Martinet, R., Voinier, T., Ystad, S.: A percussive sound synthetizer based on physical and perceptual attributes. Computer Music Journal 30(2), 32–41 (2006)
19. Aramaki, M., Besson, M., Kronland-Martinet, R., Ystad, S.: Computer Music Modeling and Retrieval - Genesis of Meaning of Sound and Music. In: Ystad, S., Kronland-Martinet, R., Jensen, K. (eds.) Timbre perception of sounds from impacted materials: behavioral, electrophysiological and acoustic approaches. LNCS, vol. 5493, pp. 1–17. Springer, Heidelberg (2009)
20. Aramaki, M., Kronland-Martinet, R., Voinier, T., Ystad, S.: Timbre control of a real-time percussive synthesizer. In: Proceedings of the 19th International Congress on Acoustics, CD-ROM (2007) ISBN: 84-87985-12-2