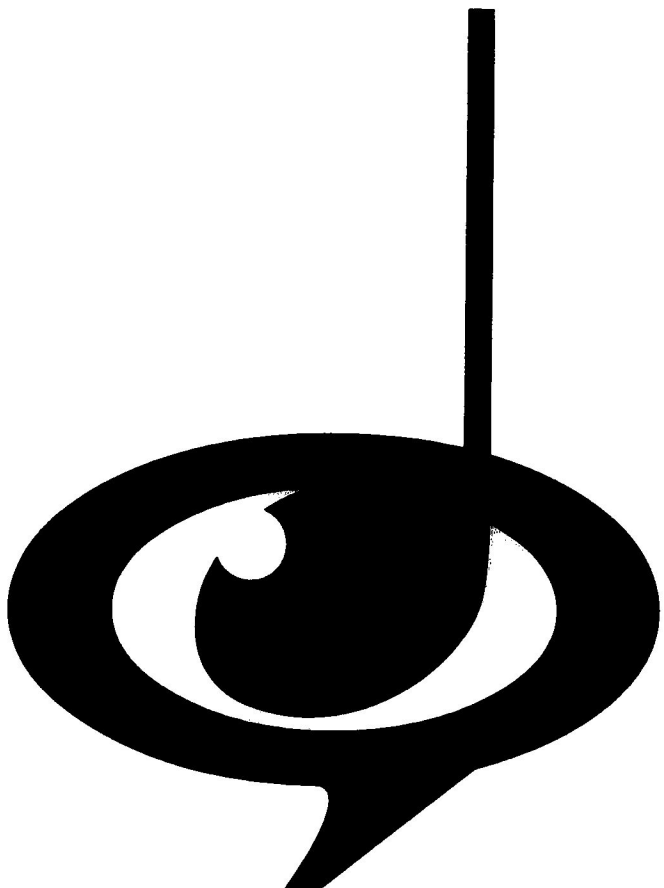
Perception of impacted materials: sound retrieval and synthesis control persepctives

Mitsuko Aramaki, Loic Brancheriau, Richard Kronland-Martinet, Solvi Ystad

PROCEEDINGS



CMMR 2002

Genesis of Meaning in Digital Art

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. 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

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 to determine the sets of sounds that were judged as typical and non typical for each material category. The use of sound continua allowed determining the perceptual borders between these typical and non typical sounds as a function of the position along the continua. A statistical model 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 Determination of sound categories from perceptual tests

We 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. Hence, Glass sounds cannot be transposed to low pitches as they will 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 pitch C4 and C5 and Glass sounds to the pitch 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 by 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.

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.

Finally, 22 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 no known auditory or neurological disorders. They all gave written consent to participate to the test and were paid for their participation.

Participants' responses were collected and were averaged for each sound of all the continua. Based on these responses, 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. From a statistical point of view, this threshold value approximately corresponds to the percentage of values that are within one standard deviation away from the mean value in the case of a normal distribution. We further considered these typical sounds as sounds that were most representative of each material category.

3 Relationship between acoustic descriptors and sound categories

3.1 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:

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

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, 15–17], 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.

3.2 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 described in the previous section: {AT, CGS, SB, SF, R, α}.

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.

Three statistical models of binary logistic regression 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 mix 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:

$$\pi(x) = P(Y = 1/X = x) = \frac{e^{L_{Cat}(x)}}{1 + e^{L_{Cat}(x)}} \quad \text{with} \quad L_{Cat}(x) = \beta_0 + \sum_i \beta_i x_i \quad (2)$$

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 sound on 3 (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 was built by excluding of sound continua that 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 correspond to the set of "typical" sounds that were defined from the results of listening tests (cf. section 2). A stepwise selection method was used and the statistical analysis was conducted with SPSS software (Release 11.0.0, LEAD Technologies).

3.3 Results and discussion

For each category model, the step summary is given in Table 1. 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 on 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 2. 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:

$$\begin{aligned} L_{Wood}(\alpha, CGS, SB) &= -38.5 - 196\alpha - 0.00864CGS + 0.0161SB \\ L_{Metal}(\alpha, SB) &= 14.7 + 322\alpha - 0.00253SB \\ L_{Glass}(SB, CGS, R, \alpha, SF) &= 14.33 - 0.006SB + 0.002CGS - 3.22R \\ &\quad +52.69\alpha - 0.001SF \end{aligned} \tag{3}$$

Table 1. 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

The logistic regression method revealed that the α parameter was the main predictor for Wood (overall percentage correct at 85% at step 1) and Metal (82.7%) categories. This result was in line with several studies showing that the damping was 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, most 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 Glass category for which the perception of the material is intricately associated to glasses (as everyday life objects). The corresponding sounds are generally characterized by high pitches and crystal-clear sounds (few spectral components). Consequently, the discrimination between Glass sounds and the other sound categories can be explained by the spectral properties of Glass sounds (described by SB or CGS) rather than by the damping.

Table 2. 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		
WOOD	Observed \ Predicted	False	True	% correct	False	True	% correct
	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
METAL	Observed \ Predicted	False	True	% correct	False	True	% correct
	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
GLASS	Observed \ Predicted	False	True	% correct	False	True	% correct
	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

4 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 the timbre. For instance, this aspect is of importance for 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 to discuss their actual relevancy from a synthesis point of view.

First, the parameter α (related to the damping) was confirmed as an important predictor in agreement 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 sounds and non typical sounds on each sound continuum allowed us to define characteristic domain (parameter range values) of each material.

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 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. 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. Actually, a separate analysis reflecting the temporal dynamics of the brain processes observed through electrophysiological data (also collected during the listening tests), revealed that R is an adequate descriptor to account for the perception and categorization of these different sound categories [19] and that it offers a more accurate control of the synthesis model. This argument indicates that the descriptors highlighted by statistical analysis as most relevant ones for sound classification may not directly constitute intuitive control parameters for synthesis purposes. To address this issue, we also aim at integrating data from brain imaging that inform of the perceptual/cognitive relevancy of descriptors. Based on these considerations, we are currently investigating the possibilities to define a control space for material categories where the control of R in particular should accurately render the typical dissonant aspect of Metal sounds [20].

5 Acknowledgments

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

References

1. J. M. Martínez. Mpeg-7 overview (version 10). `http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm`, Last checked on March 25th 2008.
2. J. Foote. *Decision-tree probabilty modeling for HMM speech recognition.* PhD thesis, Cornell university, 1994.
3. P. Roy, F. Pachet, and S. Krakowski. 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. M. Slaney. Mixtures of probability experts for audio retrieval and indexing. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, Lausanne, Switzerland, 2002.
5. F. Pachet and A. Zils. Evolving automatically high-level music descriptors from acoustic signals. *1st International symposium on computer music modeling and retrieval*, 2003.
6. S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychological Research*, 58:177–192, 1995.

7. G. Peeters, S. McAdams, and P. Herrera. Instrument sound description in the context of mpeg-7. In *Proceedings of the International Computer Music Conference*, Berlin, Germany, 2000.
8. M. Aramaki and R. Kronland-Martinet. Analysis-synthesis of impact sounds by real-time dynamic filtering. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(2):695–705, 2006.
9. R. Parncutt. *Harmony - A Psychoacoustical Approach*. Springer, Berlin/Heidelberg, 1989.
10. M. Aramaki, H. Baillères, L. Brancheriau, R. Kronland-Martinet, and S. Ystad. Sound quality assessment of wood for xylophone bars. *Journal of the Acoustical Society of America*, 121(4):2407–2420, 2007.
11. R. P. Wildes and W. A. Richards. *Recovering material properties from sound*, chapter 25, pages 356–363. W. A. Richards Ed., MIT Press, Cambridge, 1988.
12. S. McAdams. Perspectives on the contribution of timbre to musical structure. *Computer Music Journal*, 23(3):85–102, 1999.
13. W. A. Sethares. Local consonance and the relationship between timbre and scale. *Journal of the Acoustical Society of America*, 94(3):1218–1228, 1993.
14. P. N. Vassilakis. Sra: A web-based research tool for spectral and roughness analysis of sound signals. In *Proceedings of the 4th Sound and Music Computing (SMC) Conference*, pages 319–325, 2007.
15. S. Tucker and G. J. Brown. 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. B. L. Giordano and S. McAdams. Material identification of real impact sounds: Effects of size variation in steel, wood, and plexiglass plates. *Journal of the Acoustical Society of America*, 119(2):1171–1181, 2006.
17. R. L. Klatzky, D. K. Pai, and E. P. Krotkov. Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environments*, 9(4):399–410, 2000.
18. M. Aramaki, R. Kronland-Martinet, Th. Voinier, and S. Ystad. A percussive sound synthetizer based on physical and perceptual attributes. *Computer Music Journal*, 30(2):32–41, 2006.
19. M. Aramaki, M. Besson, R. Kronland-Martinet, and S. Ystad. Timbre perception of sounds from impacted materials: behavioral, electrophysiological and acoustic approaches. *Journal of the Acoustical Society of America*, submitted, 2008.
20. M. Aramaki, R. Kronland-Martinet, Th. Voinier, and S. Ystad. Timbre control of a real-time percussive synthesizer. In *Proceedings of the 19th International Congress on Acoustics (CD-ROM); ISBN: 84-87985-12-2*, 2007.