# Perception of Material and Shape of Impacted Everyday Objects

Camille-Eva Rakovec, Mitsuko Aramaki, and Richard Kronland-Martinet

CNRS - Laboratoire de Mécanique et d'Acoustique,
31, chemin Joseph Aiguier 13402 Marseille Cedex 20, France
rakovec@lma.cnrs-mrs.fr

**Abstract.** In this study, we focused on the establishment of perceptual categories of shapes and materials associated with everyday objects, in the perspective of setting up high-level controls in an impact sound synthesizer. For that purpose, impact sounds were recorded from everyday objects covering various physical attributes, i.e., seven types of materials (Metal, Wood, Plastic, Glass, Stone, Ceramic and Cardboard) and five types of shapes (Solid 1D, Hollow 1D, 2D, Solid 3D and Hollow 3D), and resynthesized with signal parameters estimated from a high-resolution method (ESPRIT). Listening tests were conducted on these sounds to define perceptual categories and evaluate their relevance. Perspectives in terms of their acoustic description are finally discussed.

**Keywords:** Sound categorization, material, shape, impact sounds, analysis-synthesis, ESPRIT method, acoustic descriptors.

## 1 Introduction

For sound design and virtual reality, the challenge now lies in making sound synthesis tools accessible to all, offering coherence with complex virtual scenes, and compatible with real time constraints. The design of intuitive control strategies, through the use of high-level descriptors, can meet this demand.

The current study is part of the development of an impact sound synthesizer for which a high-level control of the perceived shape and material of the object is desired [1]. This issue requires knowledge on acoustical properties of sounds that convey relevant information on the source and how they are perceived. Previous studies have investigated the auditory perception of physical attributes such as shape, hollowness or material. In particular, several studies can be found on the dimensions of objects, such as bars [2], [3], rods [4] or spheres [5]. They demonstrated that height-width ratios and lengths could be recovered from sounds with reliable accuracy. In terms of cavity, Lutfi [6] and Rocchesso [7] showed that the perception of hollowness can reasonably be identified from a certain size threshold. The pitch and brightness-related descriptors were found to be relevant for listeners. In terms of shapes, Kunkler-Peck and Turvey [8] investigated the participants' ability to identify circular, triangular or squared plates made of various materials. Performance was almost perfect with only a secondary tendency to associate materials with particular shapes. By contrast, Tucker *et al.* [9] and Giordano [10] found that shape recognition abilities were limited and strongly depended on the material composition. Lastly, the auditory perception of materials was investigated in [11], [12] and [13]. Two gross material categories (wood-plastic, glass-steel) emerged from these studies, where the metal was more often better identified. The damping parameter was found to be one of the main acoustic cues for listeners [14].

Given the perceptual sensitivity of human being to some physical attributes of objects, the challenge now lies in the establishment of representative perceptual categories related to the sound sources as well as in their acoustic description. In our study, we evaluate the perceptual attributes of everyday objects. Based on previous studies, we aim at defining perceptual categories of shapes and also completing and refining the perceptual categories of materials established in [15].

For this purpose, we constituted a sound data bank by recording impacted items recovering a large set of physical categories of material and shape, considered as representative of the everyday life. Hence by recovering a large set of physical attributes, we assumed that a large variety of perceptual categories would be obtained after the listening tests. The recorded sounds were analyzed with a high-resolution method (ESPRIT) and resynthesized on the basis of the estimated parameters. Then, a series of listening tests were conducted to determine the main perceptual categories of material and shape. Finally, in the aim of characterizing the sounds belonging to these perceptual categories from an acoustic point of view, perspectives on acoustic descriptors are discussed.

## 2 Sound Data Bank from Everyday Objects

Lutfi has asked in [6]: *"How might a listener determine from sound that a class of resonant objects is, say, hollow or is made of metal despite differences in the size or shape of individual exemplars or the manner in which they are driven to vibrate ?"* The database we describe in this section is actually intended to enable us to answer this question. Here we first present the different objects that have been impacted and the methodology adopted. In a second part, we discuss the high-resolution ESPRIT method that allowed us to estimate the parameters of the recorded sounds and then make their synthesis.

### 2.1 The Sounding Objects

It is important to remember that the perceptual category that matches a given sound can be totally different from the real category associated with the object that produced the sound. Beyond this consideration, based on previous studies [17], we considered that the fact of covering various physical categories of objects would maximize the variety of perceptual categories obtained after the listening tests. The categorization presented in Figure 1 has therefore emerged as the most exhaustive according to the existing objects encountered in our everyday life.

**What Are the Categories ?** Seven types of materials are investigated: Metal, Wood, Plastic, Glass, Stone, Ceramic and Cardboard. Note that each category is broadly defined. For example, the Metal category includes: iron, steel, aluminum, bronze, alloys, etc. The Wood category includes: beech, fir, pine, chipboard, etc. And the Plastic category includes objects made of different types of polymers (PVC, PET, PP, PS, etc.). Since our previous works on the impact sound synthesizer [15], the categories Plastic, Stone, Ceramic and Cardboard are so new.

Five types of shapes are investigated: Solid 1D, Hollow 1D, 2D, Solid 3D and Hollow 3D. The 1D category denotes objects for which one dimension is much larger than the two others (e.g., rods, beams, bars, tubes, pipes, etc.). The 2D category denotes objects for which two dimensions are of the same order of magnitude compared to the third one (e.g., plates, slabs, sheets, etc.). The 3D category refers to three-dimensional objects for which the three dimensions are of the same order of magnitude (e.g., jars, bottles, bowls, boxes, balls, etc.). The presence or absence of cavity is also considered for 1D and 3D objects and is denoted by the label Hollow (e.g., bell) or Solid (e.g., rod) respectively.

As a summary, 35 categories of objects' attributes have been defined to constitute the complete sound data bank.

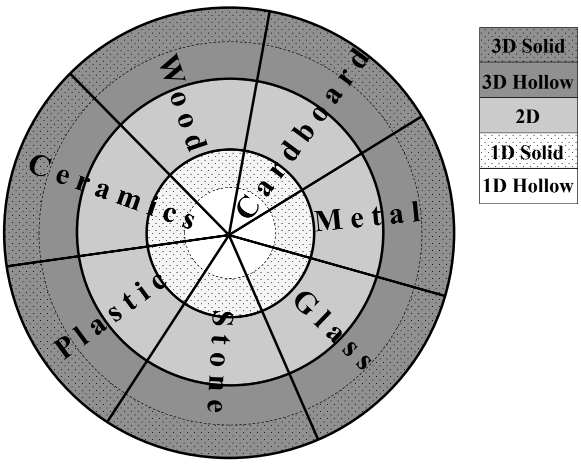


**Fig. 1.** The physical categories of everyday objects : 7 types of materials covering 5 types of shapes, including 2 types of cavities

**Procedure** In practice, 126 objects that match at best the 35 previous categories were collected. Note that for some categories, it was not possible to find corresponding objects because of the unusual physical characteristics (e.g., Hollow 1D objects made of stone). However, this constraint did not fit in conflict with our wish to study representative everyday objects. Since we aimed at collecting 10 sounds by category, the goal of this part of the work was to record about 350 impact sounds.

The recording session was made in an anechoic chamber. The objects were placed on a rigid support and stabilized with hooks and foams (pictures available at [22]). They were then impacted with a Brüel & Kjaer 8202 hammer to record the excitation force. When possible, objects were impacted at three different points (center, edge and section) to obtain various vibratory responses, evoking how we manipulate objects in our everyday life. The sounds were recorded with a Neumann KM183 omnidirectional microphone, placed about 15cm from the object, and a sound card Motu UltraLite mk3 at 44.1kHz sampling frequency.

**The Signal Model** We considered a signal model based on physical considerations to synthesize the recorded impact sounds. From the point of view of physics, an impact sound corresponds to the vibratory response of a structure to an impact excitation. By assuming linear conditions, this structure can be modeled by a mass-spring-damper system, said *harmonic oscillator*. Such a mechanical system can entirely be characterized by its impulse response. Therefore, we assumed that the choice of impact to excite the collected objects would reveal at best the sound sources' attributes investigated in our current study.

In the case of one degree of freedom in a mass-spring-damper system, the equation of motion taking into account the dissipation process is:

$$m\frac{dx^2}{dt^2} + c\frac{dx}{dt} + kx = 0 \ , \tag{1}$$

where $x$ is the displacement, $m$ is the mass, $k$ is the spring stiffness assumed to be constant, and $c$ is the viscous friction coefficient of the object. For $c$ inferior to the critical damping $c_c = 2\sqrt{km}$, the solution of this equation leads to damped oscillations and is written :

$$x = Ae^{-\zeta\omega t} sin(\sqrt{1-\zeta^2}\omega t + \varphi) \ , \tag{2}$$

where $A$ and $\varphi$ are constants, dependent on the initial conditions, $\omega = \sqrt{\frac{k}{m}}$ is the angular frequency and $\zeta = \frac{c}{c_c}$ is the viscous damping coefficient. This solution describes an exponential decay of the oscillation amplitude of the signal. In this study, we therefore considered structures with $K$ degrees of freedom, which corresponding impact sounds were modeled as a sum of $K$ exponentially damped sinusoids, zero for negative time, and could be written as:

$$s(t) = H(t) \sum_{k=0}^{K-1} a_k e^{i\phi_k} e^{-\delta_k t} e^{2i\pi\nu_k t} \tag{3}$$

where $a_k$ is the amplitude at the origin, $\phi_k$ is the phase at the origin, $\delta_k$ is the damping (in $s^{-1}$) and $\nu_k$ is the frequency (in Hz) of the $k^{th}$ component. These signal parameters were estimated using a high resolution method, whose principle is briefly explained below (see [18] for more details).

### 2.2 Signal Parameter Estimation

The recorded sounds were analyzed using the Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [16]. This high-resolution method was chosen because of its high accuracy in the identification of close components, and of its low sensitivity to noise [19].

**The ESPRIT Algorithm** ESPRIT is a parametric method, linked to a signal model where the discrete signal to be analyzed $x(t)$ is discretized and is written:

$$x(l) = s(l) + w(l) \; , \tag{4}$$

where $l$ is the sample index and where the deterministic part $s(l)$, in reference to equation (5), is a sum of $K$ damped sinusoids:

$$s(l) = \sum_{k=0}^{K-1} \alpha_k z_k^l \; , \tag{5}$$

where the complex amplitudes are defined as $\alpha_k = a_k e^{i\phi_k}$, and the poles are defined as $z_k = e^{-\delta_k + 2i\pi\nu_k}$. The stochastic part $w(l)$ is a gaussian white noise of variance $\sigma^2$.

Here is performed a singular value decomposition on an estimate correlation matrix of the signal. The eigenvectors corresponding to the $K$ highest eigenvalues correspond to the so called *signal subspace*, while the remaining vectors correspond to the so called *noise subspace*. The shift invariance property of the signal subspace allows a simple solution for the optimal poles values $z_k$. Then, the amplitudes $a_k$ can be recovered by solving a least square problem.

**ESPRIT in a Gabor Frame** To overcome the high computational complexity of this method and the assumption that the background noise of the analyzed signal is white, a time-frequency representation of the original sounds was computed [20]. This representation consisted in a *Gabor Transform* (GT), which is basically a sub-sampled version of the short-time Discrete Fourier Transform. It was computed within a *Gabor frame* (GF), which forms a discrete paving of the time-frequency plane. The GT allows the expression of $x(l)$ in a given GF $\{g, a, M\}$ which is characterized by a window $g$, a time-step parameter $a$, and a number of frequency channels $M$. Indeed, an Exponentially Damped Sinusoid (EDS) in the original sound still being an EDS inside each band, straightforward time subsampling and sub-band division of the signal through the GT can be achieved [21]. Several configurations of GT parameters $\{g, a, M\}$ were tested. After evaluation of sound quality to computing time ratio, ESPRIT analysis were finally applied with $a = 64$ and $M = 256$ (see [18] for more details).

**The Synthesized Sounds** The sounds were synthesized on the basis of the signal model (5) and the signal parameters estimated from the ESPRIT method.

Synthesis was performed at 44.1kHz sampling frequency, 16-bit resolution, in .WAV format.

Since the estimation of synthesis parameters with ESPRIT method was not always optimal on real sounds [18], each synthesized sound was then compared to the original one from a perceptual point of view. The synthetic sounds that were perceptually closest to the recorded ones were selected and incorporated into the final sound data bank (see Table 1), the others were discarded. Actually, inadequacies between the acoustic morphology of the recorded sound and the chosen signal model according to the assumptions required in the ESPRIT method (i.e., multiple impacts or low signal-to-noise ratio) were the main causes of an unsatisfying synthesis. On the 350 initial recorded sounds, 246 were accurately synthesized and were finally selected. The reader can find examples of original and resynthesized sounds at [22].

**Table 1.** Distribution of the selected synthesized sounds in each real category, forming the sound bank.

| Shapes / Materials | 1D | | 2D | 3D | | Total |
|---|---|---|---|---|---|---|
| | **Solid** | **Hollow** | | **Solid** | **Hollow** | |
| **Metal** | 7 | 10 | 10 | 8 | 16 | 51 |
| **Glass** | - | - | 7 | 4 | 14 | 25 |
| **Stone** | 3 | - | 17 | 5 | 3 | 28 |
| **Ceramic** | - | - | 10 | 3 | 12 | 25 |
| **Plastic** | 8 | 6 | 8 | 7 | 12 | 41 |
| **Wood** | 14 | - | 15 | 8 | 14 | 51 |
| **Cardboard** | 4 | - | 11 | 2 | 8 | 25 |
| **Total** | 36 | 16 | 78 | 37 | 79 | **246** |

## 3 Experiment I: Material and Shape Identification

In this section, we present the listening tests that we conducted in order to determine perceptual categories associated with material and shape attributes of the sound source. In particular, we evaluate the confusion between actual (i.e., physical) and perceptual categories.

### 3.1 Method

**Participants** Seventeen listeners participated in the experiment (7 females, 10 males), aged 19-55 years. All of them reported having normal hearing. None of them was a professional musician.

**Stimuli** The 246 synthesized sounds were evaluated within 5 separated listening tests of about 50 sounds each. The sounds were chosen to cover the different material and shape categories at best in each test. The average distributions are available at [22].

Although we make the assumption that loudness is not a fundamental perception cue for the perception of material or shape, sounds were equalized by gain adjustments to avoid any impact of loudness in the categorization judgments. The gain adjustments were determined by ear by the experimenter with Adobe Audition. To minimize the contribution of the background noise that is naturally present in the recorded sound during the gain amplification process, we spotted the threshold level of appearance of this noise in the concerned sounds and took it as a reference to adjust the other sounds.

**Apparatus and Procedure** The experiments were conducted during almost two months, with a test about every two weeks. Listeners were seated in a quiet room. A graphical interface was developed specifically for these tests with Matlab program (R2011a version) running on a Mac OS 10.7.3 environment. This interface was used to present stimuli to participants and collect responses. Sounds were amplified with an Apogee One sound card and presented through Seinnheiser-HD650 headphones.

To begin, participants were accessing to a training page where they could listen to impact sounds evocative of everyday life objects, as many times as wanted, to familiarize themselves with the categories further tested in the formal session. Then, they were accessing to the test where they were asked to evaluate each sound according to the perceived material and shape. In practice, they had to classify each sound into one of the proposed Material categories (Metal, Glass, Stone, Plastic, Ceramic, Wood, Cardboard or Other) and Shape categories (Hollow 1D, Solid 1D, 2D, Hollow 3D or Solid 3D) by clicking on the corresponding label displayed on the screen (screen-views available at [22]). Note that for the material evaluation, participants were free to propose an additional category by selecting "Other" and by entering the evoked category label with the computer keyboard. The association between response buttons and category labels displayed on the screen was randomized at each trial to avoid any bias linked with the order of presentation of the label categories. Listeners were allowed to replay the sound as many times as wanted during the trial. The order of presentation of the sounds was randomized across participants.

### 3.2 Results

Each session took 20 minutes in average to be completed. In this section, identification rates of the categories are evaluated in order to conclude about their perceptual relevance. Since the material and shape evaluation were performed within a same trial, the interaction between the perceived material and the perceived shape are also investigated. For sake of clarity, the analyses were conducted separately for materials (the seven categories), shapes (1D, 2D, 3D) and cavities (Hollow, Solid).

**Outliers** We first evaluated the participants' responses as a whole to detect possible *outliers*. One of them was found to be *outlier* during the first test. He was asked to effectuate this test again to incorporate the general tendency, that he completed successfully.

**Table 2.** **(a)** Material confusion (%). **(b)** Shape confusion (%). **(c)** Cavity confusion (%). Rates represent the percentages of classification of sounds in the perceived categories with respect to the actual categories of the corresponding physical objects. The reported values are significantly above chance when $^{***}p < 0.001$, except when indicated by $^{**}p < 0.01$, $^{*}p < 0.05$ or $p > 0.05$. In a same row, confused categories ($p > 0.05$) are highlighted in bold.

| (a) | **Perceived** | | | | | | |
|---|---|---|---|---|---|---|---|
| **Real** | **Metal** | **Glass** | **Ceramic** | **Stone** | **Plastic** | **Wood** | **Cardboard** |
| **Metal** | 63.55*** | 15.22*** | 7.27 | 10.03 | 2.42 | 1.38 | 0 |
| **Glass** | **24.47****** | **34.12****** | **15.29****** | **13.88****** | 6.35 | 5.18 | 0.71 |
| **Ceramic** | 14.59 | **19.76****** | 39.29** | **15.76****** | 4.71 | 5.65 | 0.24 |
| **Stone** | 4.20 | 15.76 | 32.77* | 24.16 | 9.66 | 12.61 | 0.21 |
| **Plastic** | 14.35 | 4.60 | 3.16 | **13.34****** | 36.01 | **19.66****** | 8.46 |
| **Wood** | 4.73 | 3.23 | 7.04 | 1.73 | 33.68*** | 47.29*** | 2.08 |
| **Cardboard** | 0.24 | 0 | 0.24 | 3.29 | 31.29*** | 15.76*** | 48.94*** |

| (b) | **Perceived** | | |
|---|---|---|---|
| **Real** | **1D** | **2D** | **3D** |
| **1D** | **37.10****** | 28.73 | **34.16****** |
| **2D** | 17.04 | 46.38*** | 36.58*** |
| **3D** | 20.44 | 25.86 | 53.70** |

| (c) | **Perceived** | |
|---|---|---|
| **Real** | **Solid** | **Hollow** |
| **Solid** | 61.16*** | 38.84 |
| **Hollow** | 39.18 | 60.82*** |

**Table 3.** Results of rANOVA performed with Materials and Shapes as within-participant factors.

| **Factors** | **SS** | **df** | **MS** | **F** | **p** |
|---|---|---|---|---|---|
| **Materials (7 cat.)** | 6628.58 | 7.11 | 946.94 | 31.13 | p¡0.001 |
| **Shapes (5 cat.)** | 5181.69 | 4.64 | 1295.42 | 46.24 | p¡0.001 |
| **Materials x Shapes (35 cat.)** | 9092.90 | 28.45 | 327.75 | 17.19 | p¡0.001 |

**Identification Rates for Material** The rate of the "Other" category being very low (< 1%), it will not be taken into account in the further analyses. Results are shown in the confusion matrix in Table 2 (a) representing the percentages of classification of sounds in the perceived categories with respect to the actual categories of the corresponding physical objects. A t-test with a significance level $\alpha = 0.05$ was conducted in each cell to determine reliable values (i.e., above chance level $= 1\,/\,number\,of\,choices = 12.50\%$). The confusion between categories were then assessed by paired t-tests (T-test 2) between cells in the same row with a significance level $\alpha = 0.05$. In particular, it appears that actual Metal sounds are identified as Metal at 63.55% and as Glass at 15.22%. However, since there is no confusion between the two groups ($p < 0.001$), this material appears to be perceptually clear to identify. Actual Glass sounds were classified in four perceptual categories (Metal, Glass, Ceramic and Stone) at above chance levels ($p < 0.001$). Some perceptual confusion between these categories was revealed since the percentages did not significantly differ from each other ($p > 0.05$). Hence, actual Glass appears to be difficult to identify and classify as Glass. Actual Ceramic sounds were identified as Ceramic at 39.29%, but also as Glass at 19.76% and as Stone at 15.76%, these percentages do not significantly differ from each other ($p > 0.05$). This result confirms the perceptual proximity of these materials. In addition, Stone and Plastic sounds have not been identified as such. Stone was instead perceived, although not significantly, as Ceramic (32.77%) and Plastic, significantly perceived as Stone (13.34%) or Wood (19.66%). Wood and Cardboard were significantly identified as such, but as Plastic also.

**Identification Rates for Shape** Results are shown in the confusion matrix in Table 2 (b). A t-test was used to determine reliable values (chance level = 33.33%). As for material analysis, the confusion were assessed by conducting paired t-tests on rates between cells in the same row of the confusion matrix with a significance level $\alpha = 0.05$. It appears that actual 1D sounds were identified as 1D and as 3D at rates near chance level. Confusion between these two perceived categories has also been highlighted ($p > 0.05$). Actual 2D sounds were significantly identified as 2D at about 46% but also as 3D at 36%. However, these two categories were not found to be confused ($p < 0.001$). Actual 3D sounds were significantly identified as 3D at more than 50% ($p < 0.001$) and were not confused with other categories. The 3D-perceived category is the category that listeners most often chose, followed by the 2D-perceived and the 1D-perceived.

**Identification Rates for Cavities** The confusion matrix for cavity assessment is shown in Table 2 (c). The values reliability was also tested with a T-test (chance level of 50%). It appears that both categories were significantly identified as such with rates over 60% ($p < 0.001$). Moreover, no confusion between categories was found.

**Material and Shape Interaction** At the end of the tests, most participants explained resolving first the material evaluation and then the shape one, because they considered the latter attribute less obvious to evaluate. Thus, the extent to which the perceived material influences the choice of the perceived shape is assessed here.

A chi-squared test with a significance level $\alpha = 0.05$ was performed on the percentages of perceived shapes through the perceived materials in order to evaluate a $H_0$ hypothesis of independence ($\chi^2 = 1.38\,10^3$, $\chi^2_s = 49.8$).

A repeated measures analysis of variance (rANOVA) was also performed on the participants' responses where the Material and Shape were included as within-participant factors (see Table 3). Note that for this analysis, the Shape categories included Solid 1D, Hollow 1D, 2D, Solid 3D and Hollow 3D. Significant effects of main factors were found ($p < 0.001$). Results also revealed a significant Material by Shape interaction ($p < 0.001$). Indeed, it appears that the number of sounds by perceived shape differ according to the perceived materials ($p < 0.001$). Table 4 shows distributions of sounds between shape categories for each perceived material. The numbers of perceived sounds by gross categories (materials, shapes and cavities) are substantially identical to numbers of sounds in each actual gross category of the bank. However, 2D category is more often associated with Metal, Solid 1D with Wood, and Hollow 1D with Plastic, while Solid 3D is more often associated with Stone, and Hollow 3D with Metal.

**Table 4.** Average number of sounds classified by listeners in each category of perceived shape for given categories of perceived material. Maxima per shape are in bold.

| Shapes / Materials | 1D | | 2D | 3D | | Total |
|---|---|---|---|---|---|---|
| | **Solid** | **Hollow** | | **Solid** | **Hollow** | |
| **Metal** | 8 | 4 | **20** | 3 | **17** | 52 |
| **Glass** | 3 | 2 | 7 | 5 | 13 | 30 |
| **Stone** | 2 | 1 | 5 | **17** | 1 | 26 |
| **Ceramic** | 3 | 2 | 15 | 5 | 7 | 32 |
| **Plastic** | 4 | **8** | 14 | 5 | 15 | 46 |
| **Wood** | **13** | 6 | 12 | 3 | 9 | 43 |
| **Cardboard** | 0 | 1 | 7 | 2 | 7 | 17 |
| **Total** | 33 | 24 | 80 | 40 | 69 | **246** |

### 3.3 Discussion

In terms of materials, highest classification rates are obtained in the diagonal of the confusion matrix except for Stone category, meaning that the perception of Material is correlated with the physical category of the impacted object. Metal was the category that was best identified (63%) and the lowest identified category was Stone (24%), which was mostly classified as Ceramic. Results also revealed confusion between some perceptual categories. Based on these confusions, we can highlight two macro-categories: Metal-Glass-Stone-Ceramic on the one hand and Wood-Plastic-Cardboard on the other hand. The obtained classification confirms the findings obtained by Giordano [12] where the author explained that available measures of the mechanical properties of engineering materials report Plastic and Wood (P-W) as strongly different from Metal and Glass (M-G) [25] so that signals originated from P-W objects would always be differentiated from those originated from M-G objects, independently of their geometry.

In terms of shapes, 3D category is best recognized, followed by 2D and 1D. Given the low identification rate for 1D shape and confusions between the three categories of perceived shapes, the 1D category does not seem a relevant attribute at this point. On the other hand, results showed that the 2D and 3D categories can be reasonably well identified and seem intuitive to listeners. However, the fact that the 3D category was most chosen by listeners could be explained by our greater ability to imagine three-dimensional objects than 1D or 2D types objects in all types of materials.

Given the high obtained percentages, the cavity attribute appears to be perceptually relevant to listeners. One can also assume that the Hollow or Solid morphology of an object is acoustically dominant over the nature of its shape (1D, 2D or 3D).

The choice of perceived shape has be shown to be conditioned by the perceived material, which is in line with results of Kunkler-Peck and Turvey [8] and Giordano [12]. This observation may reflect our relative inability to conceptually separate the shape attribute from the material attribute of an object, and to naturally associate specific shape with specific material because of our ecological approach of perception [23]. Listeners thus tend to choose shapes that seem most typical for each material, such as Hollow 3D Metal (e.g., metallic bell) or Solid 1D Wood (e.g., wooden beam), reflecting the physical regularities encountered in our everyday acoustical environment [24]. However, the fact that associations between materials and shapes performed within the tests (see screen-view at [22]) were almost identical to examples presented at the start of the tests lead us to wonder if participants were influenced in their choices or not.

## 4 Experiment II: Shape Perception Regardless of Material Evaluation

Results of Experiment I have shown significant interactions between perceived materials and shapes. We therefore wanted to evaluate to what extent the perception of the shape was influenced by the perceived material. Therefore, we decided to test only one perceptual attribute at a time. In Experiment II, we conducted two listening tests where the material of the impacted objects was fixed and participants were asked to only evaluate the shape of these objects. This Experiment was a way to confirm the relevance of the previous shape categories (1D, 2D and 3D) and to refine their definition by asking participants to describe the perceived objects with labels (names and/or adjectives). The choice of materials was made on the basis of the macro-categories highlighted in Experiment I. We selected a representative category of material from each obtained macro-category and for which a wide variety of everyday objects was represented (see Table 1). According to these criteria, the Metal and Plastic categories were finally chosen.

### 4.1 Method

**Participants** Twenty listeners participated in the experiment (5 females, 15 males), aged 23-57 years. All of them reported having normal hearing. None of them was a professional musician. Eleven participated in Experiment I.

**Stimuli** Two sets of synthesized sounds for Metal and Plastic categories were constituted. Sounds were chosen to cover the shapes categories at best. For more information, the distributions are available at [22]. As in Experiment I (see §3.1), sounds were equalized by gain adjustments to avoid any impact of loudness in the categorization judgments.

**Apparatus and Procedure** Two separate listening tests were conducted. The test dedicated to Metal category was conducted first, for about one month, and the test for Plastic category was conducted then, over a similar period. The apparatus used here was the same as in Experiment I (see §3.1).

To begin each test, participants were accessing to a training page where they could listen to evocative Metal (resp. Plastic) impact sounds of everyday objects, as many times as wanted, to familiarize themselves with the shape categories further tested. Then, they were accessing to the test where they were asked to categorize each sound in one of the propositions (Hollow 1D, Solid 1D, 2D, Hollow 3D or Solid 3D) by clicking on one of the labels displayed on the screen. Listeners were also asked to name (i.e., label) the object evoked by the perceived sound. As in Experiment I, the association between response buttons and category labels was randomized at each trial. Listeners were allowed to replay the sound as many times as wanted during the trial. The order of presentation of the sounds was randomized across participants.

### 4.2 Results

Each session took about 15 minutes to complete in average. Confusion matrices for shapes and cavities are presented Table 5 (a-d). T-test were conducted with a significance level $\alpha = 0.05$ with chance level = 33.33% for shapes and = 50% for cavities. The percentages in brackets are those obtained on the same sounds in Experiment I. Globally, the rates are significantly higher in Experiment II ($p < 0.001$).

For Metal sounds, we observe a significant increase in the 1D identification rate with $+14pts$, while confusion with 2D is not significant ($p < 0.001$ with T-test 2), neither the 3D-perceived rate ($p > 0.05$ with T-test). The identification of 2D is lower ($-7pts$) while confusion with 1D is not significant ($p < 0.001$

with T-test 2), neither the 3D-perceived rate ($p > 0.05$ with T-test). For 3D identification, an increase of $+8pts$ is observed and low confusion with 1D and 2D were found to be significant after T-test 2 ($p > 0.05$). Besides, Solid and Hollow are much better identified with $+9pts$ and $+10pts$ respectively.

For Plastic sounds, the 1D identification rate increased ($+15pts$; $p < 0.005$) but much lesser for 2D ($+2pts$; $p < 0.001$), for which confusion with 1D remains not significant ($p < 0.001$ with T-test 2). Besides, there is no improvement in the 3D identification rate ($-2pts$; $p < 0.001$) but confusion with 2D remains also insignificant ($p < 0.001$ with T-test 2). For Solid and Hollow sounds, identification rate were significantly improved ($p < 0.001$) with $+16pts$ and $+31pts$ respectively.

Here we consider *typical* sounds from Table 6 (i.e., sounds classified in a same category by more than 52% of listeners, this threshold ensuring an acceptable number of sounds within each category). We observe an increase of the number of 1D and Hollow typical sounds that we could explain by the better identification rates observed in those two categories, causing a 2D and 3D typical sounds decrease in both materials. Yet, typical sounds were almost identically distributed within perceived shapes from Experiment I to Experiment II.

**Table 5.** Matrix confusion (%) of: **(a)** Metal shapes, **(b)** Metal cavities, **(c)** Plastic shapes, **(d)** Plastic cavities. The percentages of classification obtained in Experiment I are represented in brackets. The reported values are significantly above chance when $^{***}p < 0.001$, except when indicated by $^{**}p < 0.01$, $^{*}p < 0.05$ or $p > 0.05$. In a same row, confused categories ($p > 0.05$) are highlighted in bold.

| **(a)** | **Perceived** | | |
|---|---|---|---|
| **Real** | **1D** | **2D** | **3D** |
| **1D** | $52.22^{***}$ (38) | $14.72^{***}$ | 33.06 |
| **2D** | $9.09^{***}$ | $50.91^{**}$ (58.20) | 40 |
| **3D** | **$15.24^{***}$** | **$13.81^{***}$** | $70.95^{***}$ (62.70) |

| **(b)** | **Perceived** | |
|---|---|---|
| **R.** | **Solid** | **Hollow** |
| **S.** | $70.71^{***}$ (61.80) | $29.29^{***}$ |
| **H.** | $33.49^{***}$ | $66.51^{**}$(56) |

| **(c)** | **Perceived** | | |
|---|---|---|---|
| **Real** | **1D** | **2D** | **3D** |
| **1D** | $49.64^{**}$ (34.87) | $21.43^{*}$ | 28.93 |
| **2D** | $16.25^{***}$ | $54.37^{***}$ (52.21) | 29.38 |
| **3D** | 33.81 | $17.58^{***}$ | $48.61^{***}$ (51) |

| **(d)** | **Perceived** | |
|---|---|---|
| **R.** | **Solid** | **Hollow** |
| **S.** | $79.25^{***}$ (63.14) | $21.15^{***}$ |
| **H.** | $29.09^{***}$ | $70.91^{***}$ (40.20) |

### 4.3 Discussion

In Experiment II, only the shape of the impacted object was evaluated. Results showed that listeners' choices led to clearer categorizations than in Experiment I. Also, we observed that 1D, Solid and Hollow sounds were much better identified

**Table 6.** Metal and Plastic typical sounds distribution within shapes categories from Experiment I to Experiment II. Sounds were judged *typical* when classified by more than 52% of listeners in a given shape category.

| | **Metal** | | **Plastic** | |
|---|---|---|---|---|
| | **Experiment I** | **Experiment II** | **Experiment I** | **Experiment II** |
| **1D** | 7 | 9 | 5 | 10 |
| **2D** | 12 | 5 | 7 | 4 |
| **3D** | 22 | 21 | 17 | 7 |
| **Solid** | 14 | 15 | 15 | 13 |
| **Hollow** | 15 | 21 | 7 | 10 |

**Table 7.** Labels and adjectives most given by participants in Experiment II to describe the perceived shapes of Metal and Plastic objects.

| | **Labels** | **Adjectives** |
|---|---|---|
| **Solid 1D** | Rod / Bar | Little |
| **Hollow 1D** | Tube | - |
| **2D** | Membrane / Plates | Thin / Flexible / Big |
| **Solid 3D** | Block / Brick | Hard / Rigid / Strong |
| **Hollow 3D** | Bowl / Bell | - |

while confusion between 1D and 3D decreased. However, 2D and 3D identification rates remained identical to those observed in Experiment I. Thus, these results allowed us to actually conclude that the 1D category is relevant from a perceptual point of view. Moreover, Solid and Hollow attributes seemed to be very clearly perceived by listeners, ensuring us about the great interest of these categories for synthesis perspectives. Besides, typical sounds distributions showed us a participants' tendency to perform almost identical associations between materials and shapes to those performed in Experiment I. Consequently, the observed interactions between materials and shapes previously considered as biases, are likely to have a cognitive and not a methodological origin.

In terms of labelling, semantic descriptions most frequently used by participants are listed in Table 7. This brings us to conclude that listeners considered 1D objects as elongated objects such as tubes or bars, 2D objects as flattened objects such as plates or membranes, and 3D objects as volumetric objects such as bricks, bowls or bells. These descriptions are in line with the real categories we defined in our sound bank (see §2.1). A size indication is yet added, showing that 1D objects are most often perceived as little objects, 2D objects most often as big and thin objects, and 3D objects most often perceived as big and strong objects.

## 5 Conclusions and perspectives

The presented study is part of the development of an impact sound synthesizer for which a high-control strategy is desired. More generally, 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 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 convey relevant acoustic information through sounds. According to this perspective, we wanted to define perceptual categories of shapes and materials reflecting everyday objects, with a view to describe them acoustically.

For this purpose, an impact sound bank was created from everyday items classified within seven types of materials (Metal, Wood, Plastic, Glass, Stone, Ceramic and Cardboard) and five types of shapes (Solid 1D, Hollow 1D, 2D, Solid 3D and Hollow 3D). In the aim of defining acoustic descriptors relevant for material and shape identification, recorded sounds were resynthesized on the basis of a signal model expressed as a sum of exponentially damped sinusoids. The synthesis parameters were estimated with a high-resolution (ESPRIT) method. We then set up a first experiment during which participants had to classify the synthesized sounds with respect to the material and shape attributes. In terms of materials, Metal was the best identified and Stone the lesser. From a perceptual point of view, two macro-categories were highlighted: Metal-Glass-Stone-Ceramic on the one hand and Wood-Plastic-Cardboard on the other hand. In terms of shapes, 3D was the most relevant for listeners while 1D was not considered as pertinent at this point. The Hollow and Solid attributes appeared to be quite evocative. Besides, results revealed a certain influence between the perceived material and the perceived shape. In order to evaluate this interaction and the intrinsic relevance of the shapes categories, a second experiment was set up, comprising two listening tests in which only the shape attribute was evaluated. Sets of Metal and Plastic sounds were chosen for their acoustic and clear perceptual distinction. As we also wanted to refine the verbal description of our shape categories, listeners were also asked to label the evoked objects. These mono-tasking tests allowed us to finally conclude on the interest of the 1D category and on the great relevance of the Solid and Hollow categories. In addition, we concluded that participants tended to make their choice from an ecological approach, that is to say, influenced by their daily environment. Though, the semantic description made during the tests were found to be consistent with the shapes categories previously defined.

Perceptual categories of shapes and materials being clarified, the next step is to acoustically describe them. Psychoacoustical studies showed that damping properties are essential cues for material perception (e.g [9], [12], [13]). Recent works ([17] and [15]) proposed a damping model to describe the damping behaviour of the signal components: $\alpha(\omega) = e^{(\alpha_r \omega + \alpha_g)}$, where $\alpha_g$ is the global damping and $\alpha_r$ the relative damping (related to the fact that damping is frequency-dependent). They concluded that these two damping parameters were an important descriptor of materials for synthesis perspective. Roughness, associated to the presence of several frequency components within the limits of a critical band, was also found to be relevant in [15]. On the basis of the estimated signal parameters of the impact sounds analyzed in this study, acoustic description of the materials categories was recently investigated by Sirdey (see his thesis to be defended [18]). In parallel, we began to focus on the acoustic description of shapes categories (see [26]) for which damping and spectral/energy descriptors are foreseen to be relevant. Investigations are currently underway on parameter $\alpha$, roughness, spectral centroid, Mel-Frequency Cepstral Coefficients (MFCC) of modes density and MFCC of energy distribution.

**Acknowledgments.** The authors wish to thank Vincent Germain for the constitution of the sound data bank.

## References

1. Aramaki, M., Gondre, C., Kronland-Martinet, R., Voinier, T., Ystad, S.: Imagine the Sounds: An Intuitive Control of an Impact Sound Synthesizer. In: Auditory Display, pp. 408–421. Springer (2010)
2. Lakatos, S., MacAdams, S., Caussé, R.: The Representation of Auditory Source Characteristics: Simple Geometric Form. Attention, Perception, & Psychophysics. 59(8), pp. 1180–1190 (1997)
3. Lutfi, R. A., Oh, E. L.: Auditory Discrimination of Material Changes in a Struck-Clamped Bar. J. Acoust. Soc. Am. 102(6), pp. 3647–3656 (1997)
4. Carello, C., Anderson, K. L., Kunkler-Peck, A. J.: Perception of Object Length by Sound. Psychological Science. 26(1), pp. 211–214 (1998)
5. Rocchesso, D., Ottaviani, L.: Can One Hear the Volume of a Shape ?. In: 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, pp. 115–118. IEEE (2001)
6. Lutfi, R. A.: Auditory Detection of Hollowness. J. Acoust. Soc. Am. 110(2), pp. 1010–1019 (2001)
7. Rocchesso, D.: Acoustic Cues for 3-D Shape Information. In: Proc. of the 2001 Int. Conf. on Auditory Display, pp. 175–180. Finland (2001)
8. Kunkler-Peck, A. J., Turvey, M. T.: Hearing Shape. J. Exp. Psychol. Hum. Percept. Perform. 26(1), pp. 279–294 (2000)
9. Tucker, S., Brown, G. J.: Investigating the Perception of the Size, Shape and Material of Damped and Free Vibrating Plates. Technical Report, Department of Computer Science, University of Sheffield (2002)
10. Giordano, B. L.: Material Categorization and Hardness Scaling in Real and Synthetic Impact Sounds. In: The Sounding Object, pp. 73–93. Rocchesso, D., Fontana, F. (eds.) Mondo Estremo, Firenze (2003)
11. Klatzky, R. L., Pai, D. K., Krotkov, E. P.: Perception of Material from Contact Sounds. In: Presence: Teleoperators and Virtual Environment 9(4), pp. 399–410 (2000)
12. Giordano, B. L., McAdams, S.: Material Identification of Real Impact Sounds: Effects of Size Variation in Steel, Glass, Wood, and Plexiglass Plates. J. Acoust. Soc. Am. 119(2), pp. 1171–1181 (2006)

13. McAdams, S., Roussarie, V., Chaigne, A., Giordano, B. L.: The Psychomechanics of Simulated Sound Sources: Material Properties of Impacted Thin Plates. J. Acoust. Soc. Am. 128(3), pp. 1401–1413 (2010)
14. Wildes, R., Richards, W.: Recovering Material Properties from Sound. In: Natural Computation, Richards, W. (eds.). pp. 356–363. MIT press, Cambridge (1988)
15. Aramaki, M., Besson, M., Kronland-Martinet, R., Ystad, S.: Controlling the Perceived Material in an Impact Sound Synthesizer. In: IEEE Transactions on Audio, Speech, and Language Processing. 19(2), pp. 301–314 (2011)
16. Roy, R., Kailath, T.: ESPRIT–Estimation of Signal Parameters via Rotational Invariance Techniques. In: IEEE Transactions on Audio, Speech, and Language Processing. 27(7), pp. 984–995 (1989)
17. Aramaki, M., Brancheriau, L., Kronland-Martinet, Ystad, S.: Perception of Impacted Materials: Sound Retrieval and Synthesis Control Perspectives In: Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music, pp.134–146. Springer (2009)
18. Sirdey, A.: Analyse Modale de Sons d'Impact par Méthodes Haute Résolution pour la Catégorisation Perceptive des Matériaux. PhD Thesis. Aix-Marseille Université (2013)
19. Badeau, R.: Méthodes à Haute Résolution pour l'Estimation et le Suivi de Sinusoïdes Modulées. Application aux Signaux de Musique. PhD Thesis. Télécom ParisTech (2005)
20. Sirdey, A., Derrien, O., Kronland-Martinet, R., Aramaki, M.: Modal Analysis of Impact Sounds with Esprit in Gabor Transforms. In: Proc. Int. Conf. on Digital Audio Effects. 88. IRCAM-Paris, France (2011)
21. Sirdey, A., Derrien, O., Kronland-Martinet, R., Aramaki, M.: ESPRIT in Gabor Frames. In: Audio Engin. Soc. 45th Int. Conf.: Applications of Time-Frequency Processing in Audio. Audio Engin. Soc. (2012)
22. Synthesized sounds examples and information on the listening tests available at http://www.lma.cnrs-mrs.fr/∼kronland/rakovec
23. Gaver, W. W.: How Do We Hear in the World ? Explorations in Ecological Acoustics. Ecological Psychology, 5(4), pp. 285–313 (1993)
24. Barlow, H. The Exploitation of Regularities in the Environment by the Brain. Behav. Brain. Sci. 24, pp. 602–607 (2001)
25. Waterman, N. A., Ashby, M. F.: The Materials Selector, 2nd ed. Chapman and Hall, London (1997)
26. Rakovec, C. E.: Étude Perceptive des Objets du Quotidien en Vue du Contrôle Intuitif d'un Synthétiseur de Sons d'Impacts. Master's Thesis. Aix-Marseille Université (2012)