HAL open science

Perception of the object attributes for sound synthesis purposes

Antoine Bourachot, Khoubeib Kanzari, Mitsuko Aramaki, Sølvi Ystad, Richard Kronland-Martinet

To cite this version:

Antoine Bourachot, Khoubeib Kanzari, Mitsuko Aramaki, Sølvi Ystad, Richard Kronland-Martinet. Perception of the object attributes for sound synthesis purposes. Computer Music Multidisciplinary Research 2019, Oct 2019, Marseille, France. hal-02264315

HAL Id: hal-02264315

https://hal.science/hal-02264315v1

Submitted on 7 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Perception of the object attributes for sound synthesis purposes

Antoine Bourachot, Khoubeib Kanzari, Mitsuko Aramaki, Sølvi Ystad and Richard Kronland-Martinet

Aix Marseille Univ, CNRS, PRISM, Marseille, France
bourachot@prism.cnrs.fr

Abstract. This paper presents a work in progress on the perception of the attributes of the shape of a resonant object. As part of the ecological approach to perception – assuming that a sound contains specific morphologies that convey perceptually relevant information responsible for its recognition, called invariants – the PRISM laboratory has developed an environmental sound synthesizer aiming to provide perceptual and intuitive controls for a non-expert user. Following a brief presentation of the different strategies for controlling the perceptual attributes of the object, we present an experiment conducted with calibrated sounds generated by a physically-informed synthesis model. This test focuses on the perception of the shape of the object, more particularly its width and thickness since these attributes, especially the thickness, have not been much studied in the literature from a perceptual point of view. The first results show that the perception of width is difficult for listeners, while the perception of thickness is much easier. This study allows us to validate the proposed control strategy. Further works are planned to better characterize the perceptual invariants relevant for shape perception.

Keywords: Perception of shape, sound synthesis, perceptual invariant, impact sounds, intuitive control

1 Introduction

This paper presents a study on the perception of the shape of an object through listening to the sounds produced by this object. It is a work in progress and takes part of the development of an environmental sound synthesizer (environmental sounds are defined by all naturals sounds other than speech and music [1]).

We know from various studies that it is possible to make links between certain physical parameters of the object and the acoustical parameters of the sounds emitted. For example Lakatos et al. [2] conducted a study on wooden and metal bars of the same length, but of different heights and widths. They showed that subjects were more efficient at judging the height-width ratio when the difference between the bars was large. Carello et al. [3] revealed through the sounds of wooden sticks that there was a strong correlation between perceived and actual size. Tucker et al. [4] studied the perception of the size and shape of different

plates (square, circular, triangular) in different materials. They showed that the shape was difficult to recognize from the impulsive sounds and that the recognition of the material was independent of the shape of the object.

The study presented in this paper is in line with the ecological approach to perception, as proposed by Gibson [5]. First developed as part of the vision, this approach was extended to the hearing by Warren & Verbrugge [6] and then formalized by McAdams & Bigand [7]. The ecological approach to perception stipulates that our perception is not only governed by the capture of "physical stimuli" but also by the environment around the listener. It proposes that our perception is based on invariant structures contained in the stimulus. These invariants are categorized into two groups: structural invariants characterizing the physical properties of a sounding object and transformational invariants describing the action exerted on this object.

Based on the work of W.W. Gaver [8] [9], Aramaki et al. [10] have proposed a paradigm called action-object in which a sound is described as the result of an action performed on an object. This paradigm has been developed to allow intuitive control of sound synthesis. Initially Aramaki et al. [10] focused their studies on the perception of the material in the context of impact sounds and proposed structural invariants describing the perception of materials. Conan et al. [11] identified different transformational invariants, extending actions possibilities to rolling, friction and scratching interactions. Thoret et al. [12] then showed the existence of transformational invariants related to the auditory perception of biological movements.

A synthesizer has been developed at the PRISM laboratory based on these previous studies and serves as an experimental and demonstration platform.

This paper is organized as follows: first, we describe the different control strategies developed in the synthesizer for controlling the perceptual attributes of the object. Then we will move on to a perceptual test conducted by using a physically-inspired model proposed by Pruvost et al. [13]. We considered physical model rather than real sounds to be able to generate parameterized and continuous morphings between different shapes of objects: from a string to a membrane, and to a cube. The goal of this test is to evaluate the perception of width and thickness of an object since these attributes have not been much studied in the literature. The results will be discussed with respect to the information that they provided. Then we will end with propositions for further works to isolate invariants specific to the perceived shape of objects.

2 Physically-Based Control Strategy

The synthesizer developed at PRISM is based on the notion of sound invarians and offers the user an intuitive control interface for sound design. Indeed, the high level control proposed to the user is based on a semantic description of the action and the object involved [14], as shown in figure 1.

The control strategy is based on 3 hierarchical levels allowing to move from the high-level control related to the evocations of sound source attributes to the

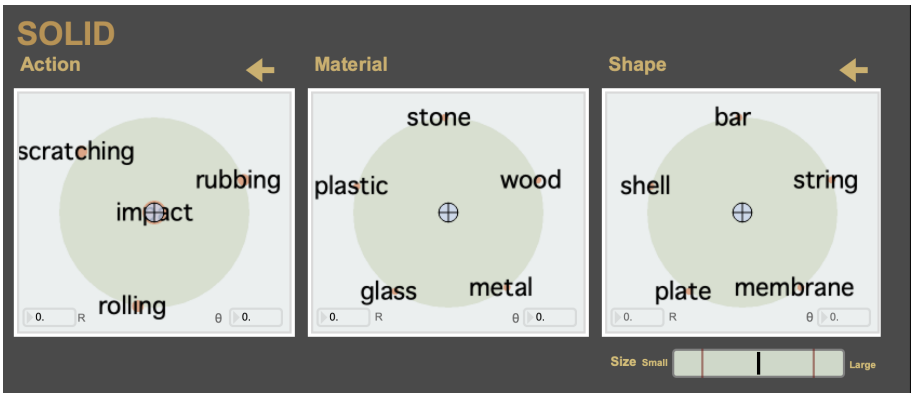

Fig. 1. Presentation of the high level controls of the synthesizer developed at PRISM

low-level of synthesis parameters. Between these two levels is the middle-level linked to the sound descriptors.

The semantic description of the perceived object is done by three control parameters. One for the material, one for the shape and one for the size. Concerning the material and shape controls, the user can navigate in a continuous space between the categories. Material control was mapped by Aramaki et al. [10] The design of the shape control was based on the analytical solutions for predefined shapes (string, bar, plate, membrane, shell) and linear interpolation was done between signal parameters (frequencies, amplitudes) to allow transitions from one label to another. A perceptual study was carried out to test the intuitive aspect of these synthesis' controls proposed. It emerged that the material space was satisfactory while the shape control was not very intuitive when people were asked to reproduce sounds created by the synthesizer. This result was expected since the shape space has not been formally calibrated yet from a perceptual point of view.

Following this study, Pruvost et al. [13] proposed a physics-inspired strategy for the control of the perceptual attributes of a sounding object. This control was for instance used in a framework of interactive and real-time synthesis of solids interaction, driven by a game engine. The controls proposed in this context were no longer semantic, but geometric (see figure 2). The object being defined by its length, width, thickness and so on, which is much more convenient for a control from a game engine.

Control of the Material: Aramaki et al. [10] confirmed previous studies showing that the modal damping plays a leading role in the auditory recognition of the materials. The authors proposed a damping model for the control of the perceived material that follows this equation:

$$\alpha(f) = e^{(\alpha_G + 2\pi f \alpha_R)} \tag{1}$$

The couple α_G and α_R, respectively the global and relative damping, was perceptually calibrated to characterize a given perceived material [10] .

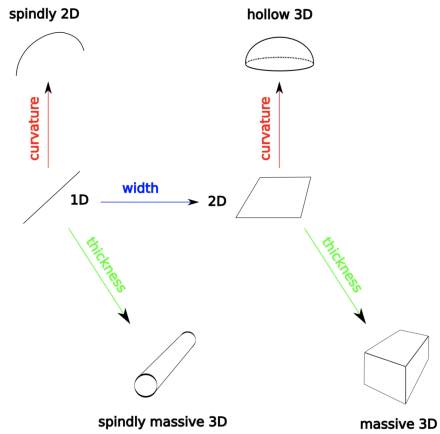

Fig. 2. Control parameters for the object based on its geometrical attributes (scheme taken from Pruvost et al. [13])

Control of the Shape: According to Rakovec et al. [15], the dimension of an object – one dimension of the object is considered when it is much larger than the others – and the massive/hollow aspect of the object are perceptually relevant parameters for the recognition of the shape.

In order to propose a continuous morphing between one-dimensional shape (e.g., a string) and a two-dimensional shape (e.g., a membrane), Pruvost et al. [13] defined that the modal distribution of sound evolves according to the following laws:

$$f_{nm}(\alpha, L) = f_0(L)\sqrt{\frac{m^2}{\alpha^2} + n^2} \tag{2}$$

With L the length (largest dimension) of the membrane. $\alpha \in]0, 1]$ so that the width $W = \alpha L$. This is reflected by an increase of spectral richness of the sound due to the emergence of numerous modal components according to α. We take into account a perceptual point of view when α tends toward 0: since m-related modes tend towards infinity and thus go beyond the limits of human hearing

we do not take into acount these modes and we consider that $f_{nm}(\alpha, L)$ tends toward $f_n(L)$:

$$f_{nm}(\alpha, L) \xrightarrow[\alpha \to 0]{} f_n(L) = f_0(L)n$$

With $f_n(L)$: the harmonic spectrum of the string.

To take into account the thickness of the object, the previous equation is extended to a third parameter, in analogy to the formula for the resonance frequencies of a rectangular cavity.

$$f_{nmp}(\alpha, \beta, L) = f_0(L)\sqrt{\frac{m^2}{\alpha^2} + \frac{p^2}{\beta^2} + n^2} \xrightarrow[\beta \to 0]{} f_{nm}(\alpha, L) \tag{3}$$

With $\beta \in]0, 1]$ the thickness $T = \beta L$ and $T \leq W \leq L$. In the same way, from a perceptual point of view, when $\beta \to 0$ we find the previous equation: (2). To our knowledge, there is no perceptual or physical study concerning the evolution of the damping with respect to the thickness of the object. However, some preliminary experiences led us to suggest that the damping should be modified while the thickness increases [13]. Therefore, we propose that the global damping α_G from equation (1) is increased as follows regarding the thickness linked to β:

$$\tilde{\alpha_G} = 2\beta\alpha_G \tag{4}$$

3 Perceptual Test

We therefore conducted a perceptual test to evaluate the control of shape provided by the previous model. In particular, the purpose of the test is to assess the perceived width and thickness of an object from the produced impacted sound.

3.1 Experimental Setup

Stimuli: In order to test this model, 36 sounds were selected from a "width-thickness" space that has been sampled regularly with a constant width step: $\alpha = 0; 0.2; 0.4; 0.6; 0.8; 1$, The same step was used made for the thickness β. All the possible combinations of α and β, this gives us a total of 36 sounds. The chosen action was an impact and the chosen material was metal to favor the perception of resonances of the object.

Participants: A total of 24 volunteers (14 males, 10 females) participated to the test. All subjects had an audiogram before the test. All the participants presented a normal hearing.

Apparatus: This test was conducted in a quiet room, isolated from ambient noise. Sounds were played with an Apple Mac Pro 6.1 with Seinnheiser HD650 headphones. All the answers of the participants were collected on an interface developed with Max/MSP.

Procedure: After a familiarization session where some typical sounds of the model were presented to the subjects, the formal test began. The sounds were presented only once and in a random order for each subject. For each sound, the subjects were then asked to assess the width and thickness of the object that produced the sound by using sliders. The scale of the sliders for each dimension was between 1 and 100. We chose to present sounds one at a time since we aimed at evaluating whether the width and thickness could be perceived in the absolute (i.e., without reference) or not.

After evaluating all the sounds, the participants were asked to answer a questionnaire at the end of the experiment that contained the following questions:

- "According to you, this test was: very difficult, difficult, average, easy or obvious"
- "What do you think of the dimensions you played on?"
- "Do you have any other dimensional suggestions for your perception of shapes?
- "Do you have any other comments?

3.2 Results

For each set of collected perceptual scores (width and thickness), we performed a repeated measures analysis of variance (ANOVA) using STATISTICA, on all variables with Thickness and Width as factors. A Tukey HSD test was then conducted to specify significant effects. An effect was considered as significant for p-value lower than 0.05.
Participants found the test difficult. Many of them reported that they found the thickness much easier to perceive than the width.

Perceived Width: The ANOVA showed a significant main effect of the Width ($F(5,115)=28.098$, $p < 0.001$). The post-hoc tests showed that the two first levels of perceived width differed from each other ($p = 0.05$) and with the four others levels ($p < 0.001$).

As shown by the figure 3, the analysis also revealed a Width by Thickness interaction. It firstly showed that the width was relatively well perceived for the first thickness level ($\beta = 0$). This can easily be explained by the fact that the model used is based on the frequency distribution of a membrane, i.e. an object generally perceived very thin. In particular, we noticed that the first width, which corresponds to a string in the model, was perfectly recognized. We also see a similar evolution for the second thickness level ($\beta = 0.2$), with a constant overestimation for the first values. However, for the other thickness levels ($\beta > 0.2$), the participants perceived the width in an almost constant way.

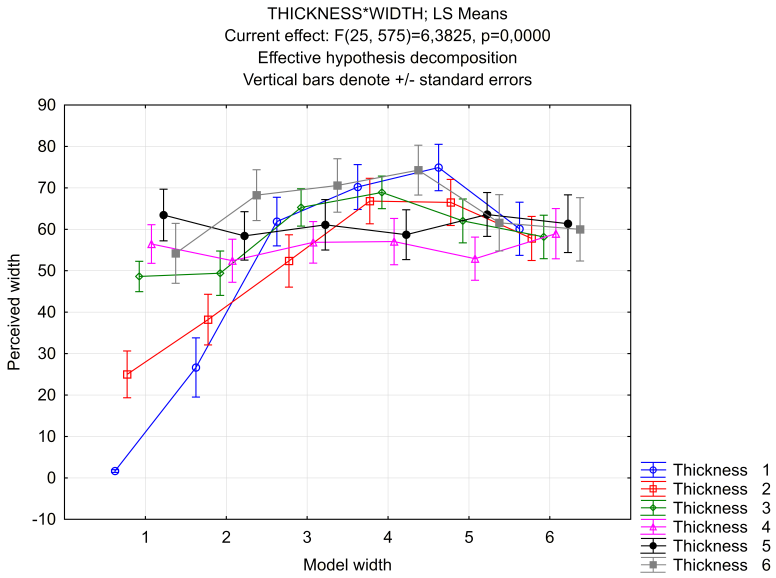

Fig. 3. Evolution of the perceived width regarding model width. Width 1, 2 ... 6 correspond respectively to $\alpha = 0$, 0.2 ... 1. Same for thickness with β

Perceived Thickness: The results are presented in the figure 4.

The perception of the thickness was really good for each level of thickness and width.

Each level of perceived thickness differed from each other significantly ($p < 0.02$): between the two first levels ($\beta = 0$ and 0.2) and between the others ($p < 0.003$).

These results suggest that the listeners were able to perceptually assess the thickness of an object. This consideration allowed us to validate the control strategy of the thickness involving a control of the damping (see equation (4)). This result is supported by self-report made by participants at the end of the experiment. Even if some of them report that it took time to understand the strategy used, none were disturbed by the link between thickness and damping.

4 Discussion and Further Works

The obtained results showed that participants did not perceive the width in an absolute way, except when the thickness is very small. By contrast, the thickness was well perceived for all width conditions. These considerations allow us to calibrate the control for the thickness from a perceptual point of view.

Future work will focus on the perception of width in a relative way, i.e., by comparison between sounds corresponding to different widths. We assume that

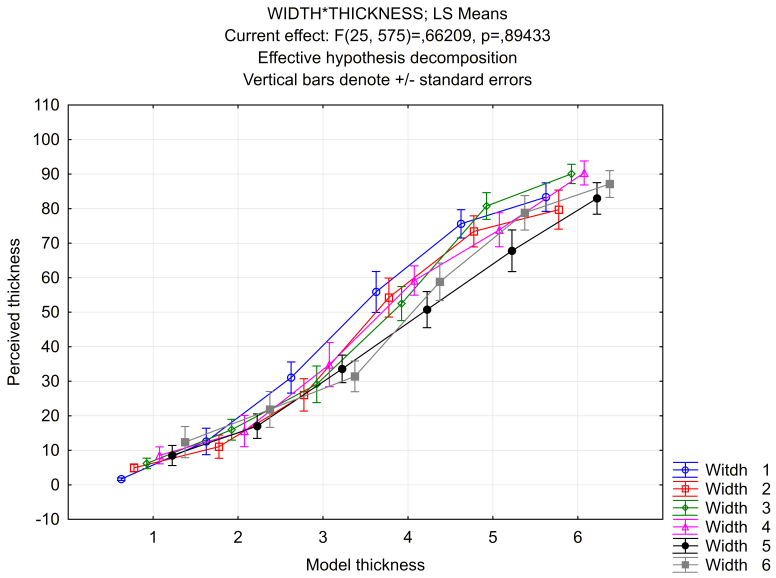

Fig. 4. Evolution of the perceived thickness regarding model thickness. Width 1, 2 ... 6 correspond respectively to $\alpha = 0$, 0.2 ... 1. Same for thickness with β

in this case the evaluation of width could be achieved easier. It will allow us to calibrate the width control in the synthesizer.

In summary, these studies will lead us to better characterize the relevant invariants specific to the perception of the object. To complete these studies, we aim at exploring numerical approach based on the morphology of the eigenmodes of objects. Indeed, to our knowledge, analytical physical models such as the ones used previously have limitations that do not allow us to reach all possible forms, for instance, a morphing between a thin plate and a cubic block. This part of the study is in progress.

The figure 5 shows the evolution of the modal distribution for thin plate - thick plate transition. Computing this evolution for different object sizes and geometries (circular, rectangular, etc.) will allow us to define a generic modal evolution law associated to the evocation of thickness, applicable to a signal synthesis model. Then these models will be assessed perceptually to calibrate the control strategy. Such models will be designed to have a small number of parameters in order to be able to offer intuitive control of the synthesis. The same process will be done for the other attributes of the object (curvature, hollowness, etc.).

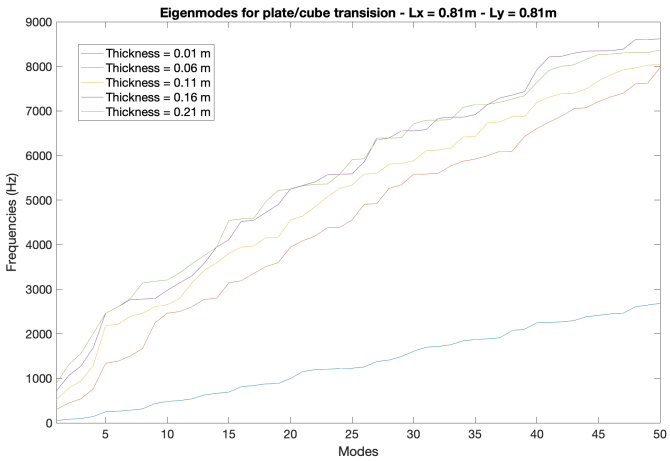

Fig. 5. Modal repartition for a thin plates (length = 0.81m, width = 0.81m, thickness = 0.01m) and different thick plate (constant thickness step: 0.05m)

5 Conclusion

In this paper we have made a brief review of the different synthesis controls proposed in the environmental sound synthesizer developed at the PRISM laboratory. We focused on the control of the shape of the sounding object.

The control of the shape of the object involved two main sound transformations: a 1D-2D string/membrane to bar/plate transition, and a 2D-3D plate/cube transition.

Then we presented a perceptual test related to this control, in particular the control of width and thickness, that had never been evaluated perceptually. In terms of perceptual expectation, the auditors had difficulty evaluating the width of the object, but had no problem evaluating its thickness. These results allowed us to validate the control of thickness in the synthesizer. Further works are underway to better define a control of the perceived width and more generally, to develop the search for the relevant invariants related to the perception of the shape.

References

1. B. Gygi, G. R. Kidd, and C. S. Watson, “Similarity and categorization of environmental sounds,” vol. 69, no. 6, pp. 839–855.
2. S. Lakatos, S. McAdams, and R. Causs, “The representation of auditory source characteristics: Simple geometric form,” vol. 59, no. 8, pp. 1180–1190.
3. C. Carello, K. L. Anderson, and A. J. Kunkler-Peck, “Perception of object length by sound,” vol. 9, no. 3, pp. 211–214.

4. S. Tucker and G. J. Brown, "Investigating the perception of the size, shape and material of damped and free vibrating plates."
5. J. Gibson, *The senses considered as perceptual systems.* Houghton Mifflin.
6. W. H. Warren and R. R. Verbrugge, "Auditory perception of breaking and bouncing events: a case study in ecological acoustics," vol. 10, no. 5, pp. 704–712.
7. S. McAdams and E. Bigand, *Thinking in sound: The cognitive psychology of human audition.*, ser. Oxford science publications. Clarendon Press/Oxford University Press.
8. W. W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," vol. 5, no. 1, pp. 1–29.
9. ——, "How do we hear in the world? explorations in ecological acoustics," vol. 5, no. 1, pp. 285–313.
10. M. Aramaki, M. Besson, R. Kronland-Martinet, and S. Ystad, "Controlling the perceived material in an impact sound synthesizer," vol. 19, no. 2, pp. 301–314.
11. S. Conan, E. Thoret, M. Aramaki, O. Derrien, C. Gondre, S. Ystad, and R. Kronland-Martinet, "An intuitive synthesizer of continuous-interaction sounds: Rubbing, scratching, and rolling," vol. 38, no. 4, pp. 24–37.
12. E. Thoret, M. Aramaki, C. Gondre, R. Kronland-Martinet, and S. Ystad, "Controlling a non linear friction model for evocative sound synthesis applications," in *International Conference on Digital Audio Effects (DAFx)*, p. XX.
13. L. Pruvost, B. Scherrer, M. Aramaki, S. Ystad, and R. Kronland-Martinet, "Perception-based interactive sound synthesis of morphing solids' interactions," in *SIGGRAPH Asia 2015 Technical Briefs.* ACM.
14. M. Aramaki, R. Kronland-Martinet, and S. Ystad, "Perceptual control of environmental sound synthesis," in *Speech, Sound and Music Processing: Embracing Research in India.* Springer, pp. 172–186.
15. C.-E. Rakovec, M. Aramaki, and R. Kronland-Martinet, "Perception of material and shape of impacted everyday objects," pp. 943 – 959.