Perceptual Assessment of Models

Report Version: 2.0
Report Preparation Date: Dec. 29, 2002
Classification: Pub.
Deliverable N. 10

Contract Start Date: 1 Jan. 2001
Duration: 2 years
Project Co-ordinator: University of Verona (UNIVR)
Partners: University of Udine (UNIUD), University of Limerick (UL), Royal Institute of Technology (KTH) - Stockholm

Project funded by the European Community under the ‘Information Society Technologies’ Programme (1998-2002)
# Contents

- **Introduction: using psychophysics to validate sound models**
  - 3

1 Model-based experiments on hardness
  - 5
    - 1.1 Introduction
      - 5
    - 1.2 Free identification
      - 7
        - 1.2.1 Stimuli
          - 7
        - 1.2.2 Procedure - subjects
          - 7
        - 1.2.3 Results
          - 8
        - 1.2.4 Discussion
          - 9
    - 1.3 Forced choice identification
      - 9
      - 1.3.1 Synthesis space investigated
        - 9
      - 1.3.2 Discriminating variations in the percussor from variations in the resonator
        - 10
    - 1.4 Scaling
      - 11
      - 1.4.1 Hardness of the resonator
        - 15
      - 1.4.2 Hardness of the percussor
        - 18
    - 1.5 Conclusions
      - 19

2 Auditory Perception of 3D Object Size
  - 21
    - 2.1 Introduction
      - 21
    - 2.2 Previous experiments
      - 21
        - 2.2.1 Pitch equalization of spheres and cubes
          - 21
        - 2.2.2 Shape recognition
          - 22
        - 2.2.3 Estimation of the sphere's pitch
          - 22
    - 2.3 New experiments about pitch perception related to volume estimation
      - 23
    - 2.4 Conclusions
      - 25

3 Object-based synthesis of distance cues: modelling and validation
  - 27
    - 3.1 Introduction
      - 27
    - 3.2 Acoustics inside a tube
      - 30
    - 3.3 Modeling the listening environment
      - 32
    - 3.4 Model performance
      - 33
    - 3.5 Psychophysical validation
      - 34
        - 3.5.1 Listening by headphones
          - 35
        - 3.5.2 Listening by loudspeakers
          - 37
3.6 Conclusion .................................................. 39

4 Experiments in a 2-D space using the Sonic Browser for validation and cataloguing of Sound Objects 40

4.1 Introduction .................................................. 40

4.2 The Sonic Browser .......................................... 40

4.2.1 Validation of Sound Objects .......................... 41

4.2.2 Cataloguing Sound Objects ............................ 41

4.3 The Sonic Browser - An overview of its functionality .. 42

4.3.1 The Sonic Browser Controls .......................... 43

4.3.2 The Sonic Browser visualisations ...................... 44

4.3.3 The Filtering Mechanisms ............................. 47

4.3.4 The Axis Control Mechanism ......................... 48

4.3.5 The Aura of the Cursor ............................... 49

4.3.6 Accessing the objects .................................. 50

4.3.7 Accessing the related PD patch for the object ......... 51

4.3.8 Dragging the objects ................................... 51

4.3.9 Tagging the objects .................................... 52

4.3.10 Drag And Drop capabilities ........................... 52

4.4 Experiments in the validation scenario using the Sonic Browser .... 52

4.4.1 Brief summary of previous interesting approaches .... 52

4.4.2 The experiments conducted with the Sonic Browser .... 53

4.5 Cataloguing experiment using the Sonic Browser ........ 69

4.5.1 The experiment .......................................... 71

4.6 Conclusions .................................................. 75

5 Multimodal Perception of Model-based Rolling 77

5.1 Introduction .................................................. 77

5.2 The Invisiball: controlling the sound of an invisible ball 77

5.3 Pilot experiment: multimodal perception test ............ 79

5.3.1 Subjects and procedure ................................ 79

5.3.2 Experiment 1: haptic and acoustic feedback .......... 79

5.3.3 Experiment 2: visual and acoustic feedback .......... 79

5.3.4 Experiment 3: acoustic feedback only .............. 80

5.3.5 Results and discussion ................................ 80

5.4 Conclusions .................................................. 82

Bibliography ..................................................... 82
Introduction: using psychophysics to validate sound models

Davide Rocchesso

The role of psychophysics in the Sounding Object (SOOb) project is spread across all phases, from preliminary design to final evaluation. In the early project stages, psychophysics readily followed phenomenology, trying to quantify some relevant dimensions of the perceptual acoustic realm. This had the purpose of obtaining sound models that are easier to control, each control parameter being perceptually salient and properly scaled. Since sounds produced by physically-based models are different from recordings of “actual” sounds, and these microphone/loudspeaker-mediated sounds are different from a genuine everyday experience, the effectiveness of each modelling effort should be measured with similar psychophysical tools.

A thorough assessment and evaluation of all the SOOb sound and control models would require an effort that extends in time much longer than the project life. However, we wanted to investigate a few important phenomena thoroughly and, at least, give indication for a few other lines of future research in psychophysics-based iterative design.

In chapter 1, the most ubiquitous SOOb sound model (i.e., the impact model) has been used to generate a set of sounds that have been tested extensively, to see how the model parameters mediate the impression of object hardness in impacts. Remarkably, the experiments confirm prior studies conducted using actual impacts, and they show a substantial decoupling between the properties of hardness and size. The results open some problems when they are related to prior investigations on categorical perception of material, where internal damping and size concurrently contribute to material perception.

In chapter 2, a conclusive experiment on the perception of the size of 3D objects is presented. It uses only model cavities and it confirms that the volumes of different 3D shapes can be equalized by ear. However, the results of this experiment proved to be highly sensitive on the order of presentation, thus proving that the relationship between object volume, pitch, and shape, is perceptually very complex and difficult to exploit in auditory displays.

Chapter 3 is prototypical of how psychophysics could be used to inspire new designs. The driving idea was that, in order to render the distance of a

\[1\] Dipartimento di Informatica, Università di Verona
virtual sound source, a special reverberating room should be designed. The
intuition that long and thin corridors could be more effective to render distance
in auditory displays was tested by simulation of these environments and by
psychophysical experiments.

Chapter 4 describes a software designed to organize and access large sound
collections in a 2D space. Access is designed to happen by browsing and si-
multaneous spatialized listening. Organization is achieved by sorting and direct
manipulation. A framework for psychophysical testing based on sonic browsing
is proposed.

Chapter 5 presents the experimental evaluation of a demonstrative appli-
cation (the InvisiBall) that was designed around a rolling sound model. The
apparatus allows generation, control, and fruition of rolling sounds in one or
more of the following modalities: haptic, auditory, visual. Results show that
the use of all three modalities does not necessarily increase the perceived re-
alism. In particular, the experiment highlighted a possible mismatch between
the haptic action on the device and the auditory feedback. Indications such as
these are important to drive the iterative design of multimodal interfaces.
Chapter 1

Model-based experiments on hardness

Bruno Giordano and Karin Petrini

1.1 Introduction

Different problems are connected to testing whether a sound synthesis model is effective in modifying the perceived features of the sound source. Previous researches in ecological acoustics have demonstrated that listeners are able to properly scale or recognise various features of the sound source, such as the shape of the percussed objects [38], their size [15], the hardness of the percussors [22], or the type of interaction between objects [72]. Other researches have, however, demonstrated that recognition of other features of the sound source is far from perfect. This happens, for example, when listeners have to recognize the gender of a clapper [58] or when they have to discriminate between certain materials of the percussed objects [28].

Let’s hypothesize that an absence of correspondence between the simulated features of the sound source and the perceived dimensions of the sound source is found. How can one attribute this result to an imperfection in the synthesis model or to a “processing bias” found in the perceptual system? An answer to this question is found only when perceptual data gathered from the synthesis model is compared to perceptual data gathered from sounds generated manipulating real sound sources. Thus an effective sound synthesis model should be able to reproduce the same kind of biases found studying the perception of real sounds. Leaving apart the problem of defining what should a perceptually effective sound synthesis model do, different methodological problems arise. The first one affects both ecological acoustics research conducted on real sounds as well as ecological acoustics research conducted on purely synthetic sounds. This problem concerns the relationship between the range of variation in the synthesis parameters and the experimental results. One could, for example, try to compare Lutfi and Oh [44] experimental results on material perception with

1 Dipartimento di Psicologia Generale, Università di Padova and Dipartimento di Scienze Filosofiche e Storico Sociali, Università di Udine
those by Klatzky [34] and those by Avanzini and Rocchesso [8]. The first one finds that material perception is a function of the fundamental frequency of the signals, despite variations in decay time, while the latter ones find that material perception is dependent both on the fundamental frequency of the signals as well as their decay times. This difference in the experimental results can be explained by analyzing the range of variation of these two variables in the experimental sets investigated in the three researches. As Lutfi and Oh simulate only glass and metallic sounds it’s obvious that they have a lower range of variation in the decay times of their stimuli than that found in Klatzky et al and in Rocchesso and Avanzini sets. From this analysis a general rule concerning the acoustical determinants of the ecological dimensions can be derived: the higher the range of variation in an acoustical feature the higher the probability to find a significant effect on the determination of the investigated ecological dimension. How is this related to the problem of validating a sounds synthesis model? Let’s suppose that a researcher, naïve to these problems, tries to find out if his synthesis model is effective in determining variations in a certain ecological dimensions despite variations in simulated physical dimensions extraneous to the investigated one (or, in other words, if his model manipulates directly the acoustical invariant upon which the perception of the investigated ecological dimension is based). What this researcher will conclude could depend strongly on the chosen range of variation in the extraneous physical dimension (we can call this extraneous dimension perturbation variable). So, he may conclude that his model controls an invariant if he chooses a small range of variation in the perturbation variable, but he will conclude that the model is not effective if he chooses a large range.

In this paper I will try to assess how a synthesis model is effective in determining variations in the perceived hardness of a percussor. This problem have been already addressed by Freed [22] using sounds generated by the interaction between real objects. This author percussed four cooking pans of different diameter using six different mallets made of materials of different hardnesses. Analysis of the performance revealed an appropriate scaling of the hardness of the mallet, independently of the diameter of the percussed pans. Concerning the relationship between the acoustical level and the perceptual estimate of the hardness of the mallets, the author finds that the performance of the subjects could be explained by four acoustical predictors derived from the spectral centroid and from the amplitude. The first two estimated the average spectral centroid and the average spectral level of the signals, while the latter two concerned the variation in time of these two parameters, namely the centroid of the spectral centroid over time (spectral centroid time weighted average), and the slope of the function describing the evolution of the spectral level over time (this latter being an estimate of the decay velocity of the signals). Hardness perception of the mallet will be investigated using a synthesis model that simulates the impact of a non-resonating object (which we will refer to with the terms hammer, percussor and exciter) against a vibrating object, referred with the terms resonator, percussed object and sounding object. Three different experimental techniques will be used in five different experiments.
1.2 Free identification

A first free identification experiment have been run using a previous version of the impact model. The goal of this experiment was to verify if the model manipulated independently the perceived features of the percussed/resonating object and the perceived features of the percussor/hammer.

1.2.1 Stimuli

A set of stimuli was synthesized, by varying both the parameters of the resonator (the frequency of the lowest resonant mode or \( f_0 \), and the internal friction coefficient, later referred as \( T_e \)) and the hammer elasticity parameter. The other synthesis parameters of the model have been fixed. The choice of these hammer mass and hammer strike velocity levels have been guided by the need to avoid rebounds in the synthesized stimuli. For this reason the first one have been set to \(-5\), the second one to \(1e^{-0.05}\). The external friction coefficient has been set to 1.

Two values have been chosen for the frequency of the lowest resonant mode: 50 Hz (\( f_1 \)) and 800 Hz (\( f_2 \)). Two values have been chosen also for the internal friction coefficient: 30 (\( T_1 \)) and 200 (\( T_2 \)). The elasticity parameter have been varied in 4 equally log-spaced steps, from \(1e^{0.02} \) (\( H_1 \)) to \(1e^{0.08} \) (\( H_4 \)), the intermediate levels being \(1e^{0.04} \) (\( H_2 \)) and \(1e^{0.06} \) (\( H_3 \)).

In order to have a higher range of variation of the perceived elasticity the stimuli were equalized by peak rms power. This avoided clipping for high elasticity values and too low intensities for low elasticity values. The combination of these three synthesis parameters produced 16 different sounds.

It was decided to focus the attention of the subjects on the differences between the stimuli, rather than on the isolated sounds. For this reason stimuli were presented in series of 4 sounds: half of the series (4) were built by varying only the resonator parameters, keeping fixed the hammer elasticity, the other half (4) of the series were built by keeping constant the parameters of the resonator while varying the elasticity parameter.

1.2.2 Procedure - subjects

As the focus was on the perceived differences between sounds, the subjects were told to simply describe what was changing between the sounds within each presented series. No suggestions about the modelled physical situation were given. They were allowed to repeat every series as many times as they wanted. Each series was judged once by each subject. The stimuli within each series were presented in randomized order. The order of presentation of the different series was randomized too. Stimuli where presented through Sennheiser HE60 headphones connected to a Sennheiser HEV70 amplifier. The amplifier received the output signal of a Sound Blaster Live! Soundcard. Before the beginning of the experimental session the gain of the amplifier was adjusted so that the rms level of a fixed 1000 Hz reference tone recorded during the recording session was equal to \(-48\) dBu. The rms level of the reference tone was measured by the experimenter that wore both the headphones and a triaxial stereo Sennheiser MK6 2002 microphone. The microphone was connected to an NTI Mynilizer ML1. The experiments were conducted in a silent room. The
computer cabinet was in an adjacent room so that the noise of the fan was not audible. Presentation of the stimuli and data collection was programmed in the Matlab environment. Nine subjects participated to the experiments. All of them reported normal hearing.

1.2.3 Results

Responses were classified in three groups. The first category defined generic responses, such as “4 sounds together” or absent responses. The second category defined descriptions given only in terms of perceptual qualities. The third category defined descriptions given in terms of everyday objects. This latter category was divided into two sub-categories: the first one defined descriptions that involved variations in the percussed/resonating object. The second one defined descriptions that didn’t involve variations in the percussed/resonating object. Table 1.1 gives the frequency of each response category for the two groups of stimuli (resonator variation and hammer variation).

Overall the stimuli have been described in terms of everyday objects 77% of the times, while descriptions based on perceptual qualities alone have been found only 9% of the times. This result agrees with Vanderveer [70] finding that listeners tend to describe sounds in terms of sound producing events, rather than in terms of perceptual properties.

Descriptions that involved different percussed objects were used for 61% of the stimuli generated by varying only the synthesis parameters. Descriptions that involved a single object were instead used for 69% of the stimuli generated by varying only the parameter referred to the elasticity of the hammer. This latter group of responses could be further divided in two groups: the first group contained those responses where the differences between the sounds of the series were not clearly addressed or described in too generic terms: “the sounds were similar” or “the sounds were produced in different ways”. This group contained fourteen of the twenty five descriptions in the “Constant object” group. The second group of responses contained those eleven out of twenty five descriptions where the perceptual variation between sounds were clearly described. Inside this group nine descriptions were based on perceptual qualities, while two were based on variations in the sound source features, namely variations in the external damping exerted over the percussed object. The nine perceptually based descriptions involved variations in pitch, in timbre (“dull”, “bright”) or, in two cases, variations in the attack quality of the sounds. All of these latter responses were highly coherent with the perceptual variations connected to the variations in the hammer elasticity synthesis parameter. In fact as the elasticity parameter is increased, the intensity of the high frequency components gets higher,

<table>
<thead>
<tr>
<th>Description category</th>
<th>Generic</th>
<th>Perceptual</th>
<th>Constant object</th>
<th>Object variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varied Parameter</td>
<td>Resonator</td>
<td>8</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Hammer</td>
<td>1</td>
<td>5</td>
<td>25</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 1.1: Classification of the descriptions given by the subjects (frequencies).
thus producing an increase in timbral brightness and in pitch. An increase in the elasticity parameter also causes an increase in the timbral dimension called “attack hardness”.

1.2.4 Discussion
As pointed out in the introduction, it is impossible to decide whether these results should be caused by a specificity of the way our auditory system processes real sounds or by the features of the sound synthesis model. However, it is plausible that the perceptual salience of a resonator is higher than the perceptual salience of the percussor, as the former is at the origin of a greater amount of information carried in the acoustical signal than the latter one. This can lead subjects to have a relative easiness in describing the resonator used to produce a certain sound, while having more difficulties in describing the hammers/non-resonating exciter used to percuss the resonator.

1.3 Forced choice identification
Two more experiments were designed in order to directly assess the independence of the variations in the resonator versus percussor, and to directly investigate the interpretation of the perceptual variations associated to the variations in the elasticity parameter. These new experiments were conducted using an improved version of the impact module, which allowed greater variations in the elasticity parameter without the need to equalize the amplitude of the synthesized signals.

1.3.1 Synthesis space investigated
The stimuli were different for the two experimental conditions. They were however chosen from the same synthesis space, whose three dimensions where the hammer elasticity parameter, the internal friction coefficient, and the frequency of the lowest resonant mode. The elasticity parameter has been varied using five equally log-spaced levels, from $5e^6$ ($H_1$) to $1e^{10}$ ($H_5$), the intermediate levels being $3.3e^7$ ($H_2$), $2.24e^8$ ($H_3$) and $1.495e^9$ ($H_4$). The internal coefficient has been varied using five equally log-spaced levels, from $10$ ($T_1$) to $160$ ($T_5$), the intermediate levels being $20$ ($T_2$), $40$ ($T_3$) and $80$ ($T_4$). Finally, the frequency of the lowest resonant mode has been varied using five equally log-spaced levels, from $50$ Hz ($F_1$) to $800$ Hz ($F_5$), the intermediate levels being $100$ Hz ($F_2$), $200$ Hz ($F_3$) and $400$ Hz ($F_4$). The other synthesis parameters were fixed as follows:

- hammer strike velocity = $-5$;
- hammer mass = 0.5;
- external friction coefficient = 0.5;
- waveform scale factor = 0.16.
1.3.2 Discriminating variations in the percussor from variations in the resonator

Stimuli

As in the free identification experiment the focus was on the perceptual variation. For this reason, we decided to synthesize 16 different series of sounds where only one dimension of the synthesis space was varied. In half of the series only the elasticity of the hammer was varied in ascending or descending order, along the five levels of the elasticity parameter illustrated above. In half of the series a parameter concerning the resonator was varied: in four of these eight series, the frequency of the lowest resonant mode was varied, in the other four ones, the internal friction coefficient was varied. We will name the investigated series using the following convention: \((f_i, t_j, h_k)\), where \(f\) is the frequency of the lowest resonant mode, \(t\) is the internal friction coefficient and \(h\) is the hammer elasticity parameter that models the hardness of the percussor, while \(i\), \(j\) and \(k\) are the levels of these parameters (see section 1.3.1 for further explanations). A final convention will be used where a level named \(a\) stands for all the levels in ascending or descending order. The following series were investigated: \((f_a, t_2, h_2)\), \((f_a, t_4, h_4)\), \((f_2, t_2, h_2)\), \((f_4, t_a, h_4)\), \((f_2, t_2, h_a)\), \((f_2, t_4, h_a)\), \((f_4, t_2, h_a)\), \((f_4, t_4, h_a)\), where the different parameters were varied both in ascending and descending order.

Procedure - subjects

The subjects were given the following instructions:

“During this experiment you will hear different sounds. Every sound has been generated by the impact between two objects: one vibrating object, called resonator, and a non-vibrating object, called percussor. This setting is found, for example, when a baked clay dish is percussed with a wooden ball. In this case the dish is the resonator, which vibrates after the impact, the ball is the percussor which doesn’t vibrate after the impact. You will hear several series of sounds made of five different sounds. Your task will be to tell if in the series you will hear a change in the resonator or in the percussor”.

Subjects could listen to the sounds as many times as they wanted before giving their response. Each series was judged once by each subject. A second group of subjects was presented with these verbal instructions plus several examples of real objects (3 different resonators made of different materials), percussed using 5 different mallets. The devices used to present the stimuli and to collect the data were the same as in the previous experiment. Twelve subjects participated to the experiment where only verbal instructions were given, 9 participated to the experiment where the sound examples too were given. All of them reported normal hearing.

Results

Table 1.2 shows the observed proportions of choosing “variation in the resonator” for the different series. Data concerning ascending and descending series have been collapsed.

Figure 1.1 plots the proportion of choosing “variation in the resonator” in the four investigated conditions: variation in the resonator synthesis parame-
Table 1.2: Proportion of choosing “variation in the resonator” for the different series and for the different experimental conditions.

<table>
<thead>
<tr>
<th></th>
<th>R - V Series</th>
<th>P Resonator</th>
<th>0.6250</th>
<th>0.6250</th>
<th>0.6667</th>
<th>0.5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>P (Resonator)</td>
<td>f&lt;sub&gt;a&lt;/sub&gt;, f&lt;sub&gt;2&lt;/sub&gt;, h&lt;sub&gt;2&lt;/sub&gt;</td>
<td>0.6250</td>
<td>0.6250</td>
<td>0.6667</td>
<td>0.5000</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P - V Series</th>
<th>0.6111</th>
<th>0.6111</th>
<th>0.7778</th>
<th>0.7778</th>
</tr>
</thead>
<tbody>
<tr>
<td>f&lt;sub&gt;2&lt;/sub&gt;, t&lt;sub&gt;2&lt;/sub&gt;, h&lt;sub&gt;a&lt;/sub&gt;</td>
<td>f&lt;sub&gt;2&lt;/sub&gt;, t&lt;sub&gt;4&lt;/sub&gt;, h&lt;sub&gt;4&lt;/sub&gt;</td>
<td>f&lt;sub&gt;4&lt;/sub&gt;, t&lt;sub&gt;2&lt;/sub&gt;, h&lt;sub&gt;a&lt;/sub&gt;</td>
<td>f&lt;sub&gt;4&lt;/sub&gt;, t&lt;sub&gt;4&lt;/sub&gt;, h&lt;sub&gt;a&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>P (Resonator)</td>
<td>0.2222</td>
<td>0.4444</td>
<td>0.5000</td>
<td>0.4440</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R - V + S Series</th>
<th>0.6111</th>
<th>0.6111</th>
<th>0.7778</th>
<th>0.7778</th>
</tr>
</thead>
<tbody>
<tr>
<td>f&lt;sub&gt;a&lt;/sub&gt;, f&lt;sub&gt;2&lt;/sub&gt;, h&lt;sub&gt;2&lt;/sub&gt;</td>
<td>f&lt;sub&gt;a&lt;/sub&gt;, f&lt;sub&gt;4&lt;/sub&gt;, h&lt;sub&gt;4&lt;/sub&gt;</td>
<td>f&lt;sub&gt;2&lt;/sub&gt;, t&lt;sub&gt;2&lt;/sub&gt;, h&lt;sub&gt;a&lt;/sub&gt;</td>
<td>f&lt;sub&gt;2&lt;/sub&gt;, t&lt;sub&gt;4&lt;/sub&gt;, h&lt;sub&gt;a&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>P (Resonator)</td>
<td>0.2222</td>
<td>0.4444</td>
<td>0.5000</td>
<td>0.4440</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P - V + S Series</th>
<th>0.6111</th>
<th>0.6111</th>
<th>0.7778</th>
<th>0.7778</th>
</tr>
</thead>
<tbody>
<tr>
<td>f&lt;sub&gt;2&lt;/sub&gt;, t&lt;sub&gt;2&lt;/sub&gt;, h&lt;sub&gt;a&lt;/sub&gt;</td>
<td>f&lt;sub&gt;2&lt;/sub&gt;, t&lt;sub&gt;4&lt;/sub&gt;, h&lt;sub&gt;a&lt;/sub&gt;</td>
<td>f&lt;sub&gt;4&lt;/sub&gt;, t&lt;sub&gt;2&lt;/sub&gt;, h&lt;sub&gt;a&lt;/sub&gt;</td>
<td>f&lt;sub&gt;4&lt;/sub&gt;, t&lt;sub&gt;4&lt;/sub&gt;, h&lt;sub&gt;a&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>P (Resonator)</td>
<td>0.2222</td>
<td>0.4444</td>
<td>0.5000</td>
<td>0.4440</td>
</tr>
</tbody>
</table>

Figure 1.1: Proportion of the response “variation on the resonator” in the four investigated conditions.

In all the conditions but the last one, where only the parameters of the hammer were varied and subjects received both oral instructions and the examples, the proportion of choosing “change in the resonator” is higher than chance level (0.5). This may reflect a relative easiness of the subjects to interpret perceptual changes in terms of changes in the percussed/resonating object and, on the other side, the difficulty to interpret those changes as changes in the features of the percussor. This latter tendency is consistent with what emerged from the analysis of the free identification data: when the elasticity of the hammer parameters, verbal instructions (R – V); variation in the hammer/percussor synthesis parameters, verbal instructions (P – V); variation in the resonator synthesis parameters, verbal instructions plus sound examples (R – (V + S)); variation in the hammer/percussor synthesis parameters, verbal instructions plus sound examples (P – (V + S)).
is varied subjects tend to interpret the changes in terms of perceptual dimensions, such as the brightness or the pitch, and not in terms of variations in the percussing objects.

Data were analyzed using a logistic regression model, with a backward elimination procedure, with the significance of the residual chi-square as stopping criterion. Predictors were the type of instructions and the kind of parameters change that specified the series (change in the percussor versus change in the resonator, Res_Perc). The final model contains both the simple effects of the instructions and of the synthesis parameters as well as their interaction (saturated model, residual chi-square $\chi^2 = 5.6745$, $df = 1$, $p$-value = 0.0172). Table 1.3 gives significance tests for the effects in the final model.

### Discussion

As pointed out analyzing the free identification data, subjects have great difficulties in attributing experienced perceptual variations to variations in the percussor. Indeed a strong bias in the direction of attributing perceptual changes to changes in the resonating object is found. This bias, however, decreases when subjects are presented the sounds examples. This could lead us to conclude that variations in the percussor can be recognized but, as in everyday life subject’s attention is directed toward the properties of the vibrating objects, strategies to detect changes in the percussors are not developed.

#### 1.3.3 Identifying variations in the percussor

A second forced choice experiment was run to directly assess the perceptual effectiveness of the hammer elasticity parameter. Subjects were presented different series of sounds where the resonator parameters were fixed, while the hammer elasticity parameter changed. They were asked to tell whether the variation they heard was a variation in the mass of the percussor, in the hardness of the percussor or in the force of the impact of the percussor on the resonator. Even though series with varied impact force and hammer mass were not present in the experimental stimuli, an assumption was made to design this experiment: if the perceptual changes associated to variations in the hammer elasticity parameter were far from being interpretable as a variation in the hammer mass or in the impact force, then these two response categories should be used with a frequency lower or closer to the chance level (33% of the times).

#### Stimuli

The following series of stimuli were presented both with ascending or descending elasticity levels: $(f_1, t_3, h_a)$, $(f_3, t_1, h_a)$, $(f_3, t_3, h_a)$, $(f_3, t_5, h_a)$, $(f_5, t_3, h_a)$.
Table 1.4: Proportion of choose of the different response categories for the different investigated series and for the different instructions conditions. Significance tests for the effects in the final logistic regression model.

The procedure was identical to the one used in the previous experiment, except for the second part of the verbal instructions:

“You will hear several series of sounds made of five different sounds. Every sound has been generated by varying only the features of the percussor. Your task will be to tell which of these three features is changing: the hardness of the percussor, the force of the impact of the percussor on the resonator, the mass of the percussor”.

As before instructions were given only verbally or were followed by real sounds as examples. This experiment was run immediately after the end of the previous one. All the subject that participated to the previous experiment took part to this one, except one of them, in the verbal only instructions condition.

Procedure - subjects

The procedure was identical to the one used in the previous experiment, except for the second part of the verbal instructions:

“You will hear several series of sounds made of five different sounds. Every sound have been generated by varying only the features of the percussor. Your task will be to tell which of these three features is changing: the hardness of the percussor, the force of the impact of the percussor on the resonator, the mass of the percussor”.

As before instructions were given only verbally or were followed by real sounds as examples. This experiment was run immediately after the end of the previous one. All the subject that participated to the previous experiment took part to this one, except one of them, in the verbal only instructions condition.

Results

Table 1.4 gives the proportion of choosing each response category for the different investigated series and for the different instructions conditions. Data concerning ascending and descending series have been collapsed.
Figure 1.2: Proportion of choice of the different response categories for the different investigated series. Verbal instructions conditions.

The same proportions have been plotted in figures 1.2 and 1.3. The chance level is shown with the straight black line.

Significance of the effects have been tested by fitting a multicategory logistic regression model to the observed proportion of response. The final model has been built using backward elimination of predictors with the significance of the residual chi-square as stopping criterion. Stimuli have been encoded using a nominal variable (e.g. a for $(f_1, t_3)$, b for $(f_3, t_1)$ etc.). The final model includes only the effect of the stimuli but not the interaction between the stimuli and the instructions nor the main effect of the instructions (residual $\chi^2 = 4.3940$, $df = 10$, $p$-value = 0.9278). Thus one can reasonably conclude that, under these experimental conditions, showing to subject the “real” examples had no effect on the identification performance.

Discussion

Results show that overall the variation in the hammer elasticity parameter was identified as a variation in the hardness of the percussor. The category “variation in the impact force” was almost chosen at chance level, so that we may conclude that this response category was not representative of the perceived variation in the features of the percussor. The category “hammer mass” was however chosen with probabilities higher than chance level in the series where the fundamental frequency was kept low $(f_1, t_3, h_a)$ and where the fundamental frequency was kept at the intermediate level while the internal friction coefficient was at the lowest used level $(f_3, t_1, h_a)$. This datum shouldn’t however be
Figure 1.3: Proportion of choice of the different response categories for the different investigated series. Verbal instructions plus sound examples condition.

considered as biased, as irrelevant alternatives, as predicted, tended to be chosen at chance level. Indeed it suggests a possible relationship between perceived mass and perceived hardness of the percussor. The absence of a significant effect of the type of instructions for this task is unexplainable at the moment and may be however connected to the use of different statistical models to test the significance of the effects.

1.4 Scaling

Previous experiments showed the existence of a bias that leads subjects to attribute perceptual variations evoked by the investigated sounds to variations in the resonator properties. For this reason a final set of direct scaling experiments was performed to directly measure the same perceptual attribute, namely hardness, on both the resonator and the percussor, using similar sets of stimuli.

1.4.1 Hardness of the resonator

Identification of the material of the resonator has been showed to be based on both the fundamental frequency of the signal and on its decay time. One should then expect that the estimate of a property of the resonator which specifies its material, should be dependent on both these acoustical features. This experiment provide both an initial basis to test this hypothesis as well as a useful reference for the interpretation of the data concerning the scaling the perceived
percussor hardness.

Stimuli
Stimuli have been synthesised by varying all the parameters in the investigated synthesis space, namely the frequency of the lowest resonant mode, the coefficient of internal friction and the hammer elasticity parameter. Three levels for each parameter have been chosen, namely $f_1$, $f_3$, $f_5$, $t_1$, $t_3$, $t_5$, $h_3$, $h_3$ and $h_5$. The combination of the chosen levels thus resulted in a stimuli set made of 27 sounds.

Procedure - subjects
Subjects were asked to rate the perceived hardness of the resonator on a numerical scale ranging from 1 (very soft) to 100 (very hard). Each of the 27 stimuli was judged three times by the same subject. As in the previous experiments instructions were divided in two parts: the first part described the modelled physical situation and was identical to the previous experiments. The second part of the verbal instructions described the task:

“The sounds you will hear have been generated by percussing resonators made of materials of different hardness, with different percussors. Your task will be to estimate the hardness of the material of the resonator. You will use the numbers between 1 and 100, where 1 stands for the very soft materials (e.g. a rubber plate) while 100 stands for the very hard materials (e.g. a steel plate).”

At the end of the verbal instructions subjects where presented the same sound examples used in the forced choice identification tasks. The devices used for stimuli presentation and data collection where the same as in the previous experiments. Ten subjects participated to the experiment. None of them took part to the previous ones. All of them reported normal hearing.

Results
Data have been analyzed by fitting a canonical generalized linear model (link function = identity) using the generalized estimating equations method [76], which assumes correlation of responses emitted by the same subject at the different repetitions of the measure. Predictors have been transformed in natural logarithmical units. A backward elimination procedure have been used to select the final model, starting from the saturated one, which included the three way-interaction between the continuous predictors (frequency of the lowest resonant mode, internal friction coefficient and hammer elasticity). At every step of the backward elimination procedure the least significant and most complex effect was eliminated. The model was then fitted again and significance of the effects was again evaluated. This procedure stopped when all the higher order effects were significant. The final model computed using this procedure contains the 2-way interaction between the elasticity parameter and the internal friction coefficient, as well as the one-way effects of these variables. Table 1.5 gives the significance tests for the final effects in the model.

Figure 1.4 plots observed average resonator hardness estimates as a function of the internal friction coefficient and of the hammer elasticity parameter.
Table 1.5: Significance tests for the effects included in the final regression model.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Wald $\chi^2$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int. frict. Coeff. ($I_f$)</td>
<td>4.65</td>
<td>0.310</td>
</tr>
<tr>
<td>Hammer elasticity ($H_e$)</td>
<td>0.76</td>
<td>0.3822</td>
</tr>
<tr>
<td>$I_f \times H_e$</td>
<td>6.69</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Figure 1.4: Average resonator hardness estimates on a 1-100 scale (1 = very soft 100 = very hard).

Discussion

This data suggest that hardness perception of the resonator is independent of the fundamental frequency of the signals, which better specifies the geometrical properties of the resonator, such as its size. Perceived hardness of the resonator, on the other side, increases for increasing decay times, parameter that probably specifies materials of increasing hardness. The absence of a significant influence of the fundamental frequency poses, however, a problem to the question concerning material perception in impact sounds: material perception seems to be adherent to the physical level when it’s perceptually scaled, but seems to be guided by prototypes when categorized. The effect of the hammer elasticity parameter is a little bit more puzzling: perceived hardness of the resonator grows as a function of the hammer elasticity parameter. A possible solution to this problem is that subjects are not able to completely divide those portions of the percept that give rise to the representation of the resonator from those that give rise to the representation of the percussor. This hypothesis is consistent
with the results of the experiment concerning the discrimination between the variations of the resonator and the variations of the percussor.

### 1.4.2 Hardness of the percussor

The final experiment investigated estimation of the perceived hardness of the percussor upon variations in the material of the resonator (internal friction coefficient) and upon variations in the size of the resonator (frequency of the lowest resonant mode). These perturbation variables have been varied in separate experimental sessions.

#### Stimuli

Two stimuli sets have been synthesized. In both of them the hammer elasticity parameter has been varied along all the five levels found in the synthesis space ($h_1$-$h_5$). In the first set the internal friction coefficient have been fixed to 40 ($t_4$), while the frequency of the lowest resonant mode have been varied using three equally log-spaced levels ($f_1$, $f_3$ and $f_5$). We will refer to this set using the $f_0$-variation term. In the second set the frequency of the lowest resonant mode have been set to 200 Hz ($f_3$), while the internal friction coefficient have been varied using three equally log-spaced levels ($t_1$, $t_3$ and $t_5$). We will refer to this set of stimuli using the $t_e$-variation name. The combination of these synthesis parameters produced two sets of 15 stimuli.

#### Procedure - subjects

Instructions were identical to the ones used in the previous experiments, except for the second part of the verbal instructions:

> "The sounds you will hear have been generated by percussing different resonators with percussors of different hardness. Your task will be to estimate the hardness of the percussor. You will use the numbers between 1 and 100, where 1 stands for the very soft materials (e.g. a felt ball) while 100 stands for the very hard materials (e.g. a steel ball)."

The devices used to present the stimuli and to collect data were the same as those used in the previous experiments. All the stimuli were judged three times by each subject. The two sets of stimuli were presented in separated sessions. The order of the sessions has been counterbalanced across subjects. Ten subjects participated to the experiment. All of them reported normal hearing.

#### Results

Data have been analyzed using the same procedure applied for the data collected in the previous experiment (generalized linear model, generalized estimating equations method, backward selection procedure with significance of the effects as stopping rule). The final model for the $f_0$-variation set includes the one way effects of the hammer elasticity parameter and of the order of the session. Table 1.6 gives the significance tests for the effects in the final model. Figure 1.5 plots the average hardness estimate as a function of the elasticity parameter.
Table 1.6: Significance tests for the effects in the final regression model.  

<table>
<thead>
<tr>
<th>Effect</th>
<th>Wald $\chi^2$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity parameter</td>
<td>6.35</td>
<td>0.0117</td>
</tr>
<tr>
<td>Session order</td>
<td>6.10</td>
<td>0.0135</td>
</tr>
</tbody>
</table>

Figure 1.5: Average percussor hardness estimates on a 1-100 scale (1 = very soft 100 = very hard). $f_0$-variation set.

The final model for the $t_0$-variation set includes all the one-way effects of the elasticity parameter, of the internal friction coefficient and of the session order. Table 1.7 gives significance tests for the final effects in the model.

Figure 1.6 plots the average hammer hardness estimate as a function of the elasticity parameter and of the internal friction coefficient.

1.5 Conclusions

Perceived hardness of the percussor is found to be a linear function of the hammer elasticity parameter. Experimental procedures have been shown to have different problems, as the effect of the order of the sessions was significant for both the investigated stimuli sets. It is plausible that the first session assumed, for each subject, an anchoring role for the estimates given in the second session. Further statistical analyses as well as a deeper inspection of the parameters of the computed statistical models will provide an answer to the anchoring role of the first session. The absence of an effect of the fundamental frequency on the
<table>
<thead>
<tr>
<th>Effect</th>
<th>Wald $\chi^2$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal friction</td>
<td>63.10</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Elasticity</td>
<td>4.93</td>
<td>0.0264</td>
</tr>
<tr>
<td>Session order</td>
<td>9.45</td>
<td>0.0021</td>
</tr>
</tbody>
</table>

Table 1.7: Significance tests for the effects in the final regression model. $t_e$-variation set.

Figure 1.6: Average percussor hardness estimates on a 1-100 scale (1 = very soft 100 = very hard). $t_e$-variation set.

The perceived hardness of the percussor is consistent with Freed’s finding [22] that perceived hardness estimates are independent of the size of the cooking pans he percussed to generate the stimuli. The effect of the coefficient of internal friction on the perceived hardness of the percussor has never been tested in previous researches. As the model manipulation of the material properties is done, in the analyzed model, using previously assessed techniques, there is the possibility to replicate this effect using physical and not computational sound sources. Similarly to the perceived hardness of the resonator, hardness of the percussor has been found to be a linear function of the internal friction coefficient and of the hammer elasticity. This fact again confirms the problems in the discrimination between features of the resonating object and features of the percussor.
Chapter 2

Auditory Perception of 3D Object Size

Laura Ottaviani and Davide Rocchesso

2.1 Introduction

In this chapter, we will introduce our last experiments about size information conveyed by sounds. We don’t want to present the model used for the test, since we have already presented it in conference proceedings [59, 61] and in a previous deliverable [62]. On that occasion, besides the description of the model, we reported the results from some experiments conducted on the auditory information conveyed by the shape of 3D-resonators (spherical and cubic). In particular, we assessed the ability of human beings to distinguish the shapes of the resonators, and we investigated pitch perception and its relationship with the volume of the enclosures. In this deliverable, we examine thoroughly this second topic, introducing further experiments conducted in a more controlled environment, without presenting the sound models and, only briefly, summarizing the previous results.

2.2 Previous experiments

The experiments previously reported [59, 61, 62] concerned 3 main tasks: (i) pitch equalization of sphere and cubes; (ii) shape recognition; (iii) estimation of the sphere’s pitch.

2.2.1 Pitch equalization of spheres and cubes

The first task of the experiments, about the relationship between pitch and volume, reported a good pitch equalization of sounds filtered by different shapes (cubes and spheres) of the same volume.

We conducted two experiments with the constant stimuli approach, using, in each experiment, the impulse response of spherical resonator as standard
stimulus. Two spheres having diameter $d = 36$ cm and $d = 100$ cm were used as references. The comparing sounds were the impulse responses of 13 cubes, one of which had the same volume as the sphere, while six were bigger and six were smaller than the sphere.

The sounds from the 13 cubes were presented, to each participant, in random order, each one preceded by the standard stimulus and repeated 10 times.

We could see that listeners identified the pitch equalization for enclosures of the same volume.

### 2.2.2 Shape recognition

The second task of the experiments concerned the recognition of 3D shape by hearing a snare drum sound filtered by the resonators. We chose that type of sound because it could excite a large part of the frequency response, while keeping its identity. We used 5 spheres with the following diameters: 100 cm, 90 cm, 70 cm, 50 cm, 30 cm, and the corresponding volume-matching cubes, all in random order and repeated 10 times.

The test results reported that the average listener is able to recognize the shape of the enclosure where a sound comes from if the cavity is larger than 50 cm. Otherwise, for smaller cavities, we have seen that the answers converge to random choice.

We specify that, before starting with the test, the subjects attended a training phase, listening to the snare drum pattern filtered by shapes with three different sizes, and different from the ones used during the experiment. The aim of the training was to let the listener only acquire the “method” to distinguish the enclosures shape, and not to learn that particular sound.

### 2.2.3 Estimation of the sphere’s pitch

We performed the third task, about estimation of the sphere’s pitch, using the up-down method or staircase method [42], by comparing the impulse response of the spherical resonator with an exponentially-damped sinusoid. We decided to use the spherical enclosure, because the subjects of the pitch equalization experiment reported a stronger pitch effect in the sphere, as compared to the cubes.

We could observe that one fourth of the participants estimated the pitch close to the lowest partial and that another one fourth estimated the pitch in the second partial neighborhood.

However, the subjects reported the difficulty of this task, due probably to the different identity of the stimuli used in the test. Nevertheless, it seems that where comparing with a damped sinusoid, listeners apply an analytic listening mode. On the other hand, the experiments of sec. 2.2.1 seem to indicate that, when the comparison sound is another complex 3D response, pitch is attributed using a more holistic listening mode.
2.3 New experiments about pitch perception related to volume estimation

We have conducted a new set of experiments on pitch comparison between impulse responses of spherical and cubic resonators. The aim was twofold: (i) validating previous results in a more controlled environment, (ii) investigating the effects of the particular procedure used.

The subjects were 12 volunteers (4 females and 8 males), with age ranging between 21 and 39, all students or faculty members of the University of Verona. Some of them were naive listeners and some trained. All of the subjects reported having normal hearing.

We used the same stimuli as in the previous experiment concerning pitch equalization, that is, the impulse responses of a spherical resonator with diameter equal to 36 cm and the impulse responses of 13 cubes, one of which with the same volume as the sphere.

The test was conducted with the method of constant stimuli. We presented pairwise the sounds from the cubes and the standard stimulus from the sphere. All the couples were presented 4 times in random order. In order to balance the results, we decided to present the sounds with the same position inside the couples, but for half participants with the sphere in first position and for the other half with the sphere in second position.

Besides the balancing issue, this experiment presents another difference from the first one: It is the number of repetitions of the couples in the randomized sequence. We reduced the number of repeated trials, since the number of stimuli available for the test is quite high and we didn’t want the subjects got tired, giving a random answer to the test. In fact, we had seen from the previous test that 4 repetitions would be enough to capture a general trend.

The participants performed the test in quiet noise conditions. The test was conducted by means of a MATLAB environment on a PC Pentium III, with a Creative SoundBlaster Live! soundcard and the stimuli were presented to the listeners over closed headphones (Beyerdynamic DT 770).

In fig. 2.1 we report the probability of the answer “cubic impulse response higher than spherical impulse response” as a function of the cubes size.

From these data, we have calculated the Point of Subjective Equivalence (PSE), that is equal to 0.2871, the Constant Error (CE), equal to -0.0029, and the Differential Lumen (DL), equal to -0.0109.

The PSE indicates the size of the cube judged by the subjects to have the impulse response with the same pitch as the reference sphere (d = 36 cm). We can see that the two enclosures have the same volume.

We calculated the PSE, the CE and the DL values applying the method of interpolation with the least-squares regression line. We estimated also the goodness of fit of the model to the data ($\chi^2 = 18.1033$, $\chi^2_{crit} = 18.3070$, $p < .05$). Therefore, the model fits well the data and the results confirm the previous ones.

It is interesting to report the data and the results obtained by a trial test conducted during the arrangements for this experiment, using another technique for balancing the response results, which consists in providing, to the same subject, and in random order, half trials with one sound in the first position, and the other half trials with that sound in second position. In this condition, we noticed that the subjects got confused.
We can easily see that there isn’t any equalization of pitch, except for random chance, since the estimated PSE, in this case, (PSE = 0.26954) was out of the range of the examined cubes size. Infact, from fig. 2.2, it appears that the probability of responses never crosses the horizontal line corresponding to the 50% of responses.

To understand this fact, it is convenient also to report the individual performances of two subjects. We display individual performances with a representation consisting of two rows, each one for a different response (Lower or Higher) and 13 columns, each one for a different cube. We display the number of responses for each cube with different colours, within a range between black for the minimum number of responses and white for the maximum. Therefore, the maximum level of uncertainty would correspond to the same colour in both rows. This representation is redundant, because one row is the negative of the other, but it gives a clear view of the equalization point, because the maximum uncertainty could be found only by looking for the column with the same hue in both the rows, instead of looking for a particular hue.

In fig. 2.3 and in fig. 2.4 we represent the performances of two subjects, that repeated the test in both conditions, i.e. with the sound positions inverted in half trials, and with the sound positions fixed during the whole experiment. In this second modality, subject n. 1 (fig. 2.3) listened to the sphere always in first position, while for subject n. 2 (fig. 2.4) the sphere was always in second position.

We can observe a different approach of the two subjects to the listening task. Subject n. 1, in the fixed order position case, has a strong pitch equalization perception and he estimates it corresponding to the central cube, which has equal volume to the sphere, while, in the situation of presentation in randomized order, it is clear that he gets confused and cannot give any judgement, showing a lot of uncertainty for all the couples he listened to. On the contrary, subject n. 2, in the randomized order condition, shows a strong perception that the
Figure 2.2: Responses probability as a function of the cubes size, in the randomized sounds order presentation.

sphere’s impulse response is higher in pitch than the one of all the cubes, while in the fixed order position case, he is able to hear, even if not so strongly, a pitch equalization corresponding to the central cube.

Most of the participants to both the experiment conditions referred to the examiner that the task in the fixed order case was easier than the other.

Only one subject, a naive listener, showed difficulties in both conditions, since she was influenced by the variation of another parameter in the two sounds, that she wasn’t able to define but the examiner can identify with the brightness. Indeed, it seems that brightness works as an anchor for listening in the randomized order case.

From the above observations, we can deduce that a “static reference point”, i.e. the order position of the sound, could help a listener, either a trained or a naive one to perform a pitch comparison task. Even if a small number of subjects showed difficulties in the pitch equalization task in both conditions, the performance gap between the two conditions seems to be due only to a change of listening attitude related to the order presentation of the stimuli. However, this has implications in auditory display and sonification, because the ordering of 3D impulse responses of different shapes along pitch scales is not as simple as it may first appear.

2.4 Conclusions

With this last session of experiments, we confirmed what we stated after the previous one: enclosures of the same volume are identified by listeners to have equal pitch.

Moreover, we discover a different listening attitude of the participants with or without a “static reference point” in the presentation of the sounds during the test.
The results of the experiment show that sounds can convey information also on the size and shapes of three-dimensional objects. This feature could be exploited for the multidimensional data representation, and, in particular, for those characteristics that, in a visual representation, would be expressed by using, for example, not only colour properties, but also shape properties.
Chapter 3

Object-based synthesis of distance cues: modelling and validation

Laura Ottaviani, Federico Fontana, and Davide Rocchesso

3.1 Introduction

Previous research in distance perception has mostly investigated the human capability to guess how far a sound source is from a listener. Psychophysical experimentation involving tasks of distance recognition rises up several issues, both methodological and technical. The interpretation and analysis of the subjects’ impressions against sounds, when they are played away from the near-field, is complicate to do as well.

It is probably for the above reasons that, even today, most of the systems aiming to add distance attributes to sounds either rely on reverberation models that were designed for other, more general purposes, or reproduce simplified physical contexts in which sounds are provided with elementary information about distance.

Our approach to distance rendering by synthesis of virtual spaces pursues an objective approach, in this way following the general framework of the project. In fact, the model we come up with certainly reflects a physical rather than perceptual approach. In addition to that, we exploit the inherent versatility of this model to explore physical environments that, independently of their consistency with the everyday experience of the surrounding space, provide to a listener cues that are capable of evoking a well-defined sense of distance.

Psychophysical experiments, conducted through both headphone and loudspeaker listening tests, confirm the effectiveness of this approach.

Humans use several senses simultaneously to explore and experience the environment. On the other hand, technological or human limitations often prevent computer-based systems from providing genuine multimodal displays. Fortunately, the redundancy of our sensory system can be exploited in order

1 Dipartimento di Informatica, Università di Verona
to choose, depending on cost and practical constraints, the display that is the
most convenient for a given application.

Providing access to information by means of audio signals played through
headphones or loudspeakers is very attractive, especially because they can elicit
a high sense of engagement with inexpensive hardware peripherals. Namely, one
may be tempted to transfer spatial information from the visual to the auditory
channel, with the expected benefits of enlarging the perceptual bandwidth and
lowering the load for the visual channel. However, we should bear in mind that
vision and hearing play fundamental but different roles in human perception.
In particular, space is not considered to be an “indispensable attribute” of
perceived sounds [36].

In audio-visual displays, the effectiveness of communication can be maxi-
mized if the visual channel is mainly devoted to spatial (and environmental)
information, while the auditory channel is mainly devoted to temporal (and
event-based) information. However, there are several situations where the spa-
tial attributes of sound become crucial:

1. In auditory warnings, where sounds are used to steer the visual attention;
2. Where it is important to perceive events produced by objects that are
   visually occluded or out of the visual angle;
3. For visually impaired users, where visual information is insufficient or
   absent.

Furthermore, if the “soundscape” of events being conveyed via the auditory
channel is particularly rich, the spatial dislocation of sound sources certainly
helps the tasks of separation and understanding of events, streams, and textures.

Much research has been dedicated to spatial auditory displays, with special
emphasis on directional cues [60], but the literature on the perception and
synthesis of the range of sources is quite limited [43, 74, 49]. Moreover, part
of this literature does not provide clear specifications about the experimental
conditions, particularly concerning the listener’s impressions about the auditory
event and the “blur” experienced during the recognition task [3].

Moreover, severe technical constraints exist on the quality of the equipment
that is usually needed when a psychophysical investigation on distance is per-
formed, in a way that part of the research conducted in this field, in particular
the prior one, can be relied only partially [5].

An interesting approach to distance evaluation [43] compares the auditory
to the visual distance. In some conditions of the experiments the distance eval-
uation were reported by the listeners verbally, whereas under other conditions
the subjects had to walk toward the sound or visual source.

The experiments in [74] have a different goal: to psychophysically investigate
the cues in distance localization, and to understand how they are combined
with different weights for stable estimations. It is shown that the subjects
underestimate far-field distances, while they overestimate near-field distances.
Moreover, a factor called specific distance tendency is introduced to describe
the tendency toward a specific perceived distance value when all distance cues
are removed. The specific distance tendency is estimated by these experiments
to be approximately 1.5 m.

We analyze only static distance cues, since our experiments were conducted
in this condition, i.e., the participants performed their task from a fixed location

28
in space. What most psychoacoustic studies have found is that we significantly underestimate the distance of sources that are located farther than a couple of meters from the subject.

The acoustic cues accounting for distance are mainly monaural:

1. **Intensity** plays a major role, especially with familiar sounds in open space. In the ideal case, intensity in the open space decreases by 6 dB for each doubling of the distance between source and listener [52].

2. **Direct-to-reverberant energy ratio** affects perception in closed spaces or reverberant outdoor environments. The reverberant energy comes from subsequent reflections of the direct sound, each of them having amplitude and time delay that vary with the characteristics of the enclosure, and with the source and listener’s positions.

3. **Spectrum** conveys distance information as well, if the listener has enough familiarity with the original sound. In that case, spectral changes introduced in the direct signal by air loss and/or sound reflection over non-ideal surfaces can be detected by the listener, and hence reconducted to distance information [5].

Also, the existence of binaural cues has been demonstrated, these cues being particularly important in the case of nearby sources [11]. Monaural cues coming from nearby sources have been hypothesized as well [5].

Among the enumerated cues, the third could be exploited to display very large distances, because the spectral cues are relevant only for long paths. The first cue is not very useful in auditory displays and sonification, because it imposes restrictions to the listening level and it may lead to annoying soundscapes.

The second cue can be exploited to synthesize spatial auditory displays of virtual sources in the range of about ten meters. This cue is inherently connected with spatial hearing inside enclosures: in this listening context, a certain amount of reverberant energy is conveyed to the listener. As stated by Guski [29], the effect of reflecting surfaces shows an overall increase in localization accuracy in the case when a sound-reflecting surface is placed on the floor. On the contrary, the percentage of correct estimation decreases with a sound-reflecting surface put on the ceiling, whereas a surface located on one side of the room doesn’t affect the performance. A sound-reflecting surface on the floor increases both the direction and height localization accuracy.

The virtual reproduction of wall reflection is traditionally provided by artificial reverberators, which add reverberant energy to sounds.

The effects of the environment characteristics over the sound are difficult to model and highly context-dependent. Moreover, the psychophysical process that maps the acoustics of a reverberant enclosure to the listener’s impressions on that enclosure is still partially unknown [4]. For this reason, artificial reverberators are typically the result of a perceptual design approach [25], which has had the fundamental advantage of leading to affordable architectures working in real-time, and has resulted in several state-of-the-art realizations, providing high-quality rendering of reverberant environments [33]. Nevertheless, most of these realization do not deal with distance rendering of the sound source.

On the contrary, the structural design philosophy focuses on models whose properties have a direct counterpart in the structural properties that must be
rendered, such as the geometry of an enclosure or the materials the wall surfaces are made of. Thanks to that approach, their driving parameters translate into corresponding model behaviors directly. Unfortunately, structural models resulted to be either too resource-consuming, or, when simplified to accommodate the hardware requirements (i.e., the real-time constraint), excessively poor in the quality of the audio results.

The emerging auditory display field shifts the focus on the usability of the auditory interface rather than on the audio quality per se. For the purpose of enhancing the effectiveness of display, it is often useful to exaggerate some aspects of synthetic sounds. In spatial audio, this has led to the proposal of systems for supernormal auditory localization [16]. In this work, we extend that concept to range localization of virtual sound sources.

As some recent literature pointed out, presenting distance cues to a listener in an auditory display does not necessarily require an individual optimization of that display to the user’s subjective characteristics. In fact, it has been observed that those particular cues are affected by the presence of non-individualized HRTFs in the display only to a small extent [75]. This enables us to focus on the objective mechanisms of distance rendering, neglecting the matter of subjective tuning.

We use the structural approach to reverberation to design a virtual resonator that enhance our perception of distance. As a simple experience, consider a child playing inside one of those tubes that are found in kindergartens. If we listen to the child by staying at one edge of the tube, we have the feeling that she is located somewhere within the tube, but the apparent position turns out to be heavily affected by the acoustics of the tube. Using a virtual acoustic tool, we experimented with several tube sizes and configurations, until we found a virtual tube that seems to be effective for distance rendering. In a personal auditory display [49], where the user wears headphones and hears virtual as well as actual sound sources, these tubes will be oriented in space by means of conventional 3D audio techniques [60], so that the virtual sound sources may be thought to be embedded within virtual acoustic beams departing from the user’s head.

3.2 Acoustics inside a tube

The listening environment we will consider is the interior of a square-section cavity having the aspect of a long tube, sized 9.5×0.45×0.45 meters. The internal surfaces of the tube are modeled to exhibit natural absorption properties against the incident sound pressure waves. The surfaces located at the two edges are modeled to behave as total absorbers (see Fig. 3.1) [39].

A careful investigation on the physical properties of this acoustic system, although not excessively complicated, is beyond the scope of this work. The resonating properties of cavities having similar geometrical and absorbing properties (for example organ pipes) have been previously investigated by researchers in acoustics [21].

Rather, it seems reasonable to think that, although quite artificial, this listening context conveys sounds that acquire noticeable spatial cues during their path from the source to the listener. Given the peculiar geometry of the resonator, these cues should mainly account for distance. The edge surfaces have been set to be totally absorbent in order to avoid echoes originating from subse-
quent reflections of the wavefronts along the main direction of wave propagation. In fact, these echoes would be ineffective for distance recognition in the range specified by the tube size, and would also originate a side-effect annoying for the listener.

The resonating environment is structurally simple although capable of dealing with an interesting range of the physical quantity to be rendered. In particular, it exhibits a convenient trade-off between the provided distance range and the resonator volume.

In this environment, we put a sound source at one end of the tube (labeled with $S$ in Fig. 3.1) along the main axis. Starting from the other end, we move a listening point along 10 positions $x_1, \ldots, x_{10}$ over the main axis, in such a way that, for each step, the source/listener distance is reduced by a factor $\sqrt{2}$. Finally the following set $X$ of distances expressed in meters comes out, as shown also by Fig. 3.1:

$$X = \{x_1, \ldots, x_{10}\} = \{0.42, 0.59, 0.84, 1.19, 1.68, 2.37, 3.36, 4.75, 6.71, 9.5\}$$

An obvious question arise prior to any investigation on distance rendering: why not render distance in an auditory display by simply changing the loudness of sounds as a function of their distance from the listener? The answer is twofold:

- Distance recognition by loudness is as more effective, as more familiar the sound source is. Conversely, a resonator, once become familiar to the listener, adds unique “footprints” to the sound emitted by the source, so that the listener has more chances to perform a recognition of distance also in the case of unfamiliar sounds.

- As introduced above, loudness in open space follows a 6 dB law for each doubling of distance. This means that a wide dynamic range is required for recreating interesting distance ranges in virtual simulations of open spaces. This requirement, apart from inherent technical complications due to hardware constraints, might conflict with the user’s need of hearing other events (possibly loud) in the display. These events would easily mask
Figure 3.2: Particular of a volume section. The lossless scattering junction in the center is connected to other junctions via waveguides 2, 3, 4, and 6. Waveguide 1 leads to a totally absorbing section of wall. Waveguide 5 leads to a partially absorbing section of wall, modeled using a waveguide filter. The filled triangles represent oriented unit delays.

farther virtual sound sources, especially in the case when the auditory display is designed to work with open headphones or ear-phones [37]. In this case, dynamic compression of sounds can be taken into account, even if this would possibly lead to a corresponding reduction of the range perceived by the listener.

Summarizing, the proposed environment should lead to a robust rendering also with respect to unfamiliar sounds, and to a broad perceived range obtained by a compressed loudness-to-distance law.

3.3 Modeling the listening environment

The square tube has been modeled by means of finite-difference schemes. Since these schemes provide a discrete-space and time formulation of the fundamental partial differential equation accounting for three-dimensional wave propagation of pressure waves along an ideal medium [69], they clearly devise a structural approach to the problem of modeling a reverberant environment. Their ease of control is a key feature for this research, that came useful especially during the preliminary informal listening of several tubular environments differing in size, shape and absorption properties.

In particular, a wave-based formulation of the finite-difference scheme has been used, known as the Waveguide Mesh, that makes use of the wave decomposition of a pressure signal \( p \) into its wave components \( p^+ \) and \( p^- \), such that \( p = p^+ + p^- \) [17]. By adopting this formulation the spatial domain is discretized in space into equal cubic volume sections, and each of them is modeled as a lossless junction of ideal waveguides, scattering 6 input wave pressure signals coming from orthogonal directions, \( p_1^+, \ldots, p_6^+ \), into corresponding output waves, \( p_1^-, \ldots, p_6^- \), going to opposite directions (see Fig. 3.2).

It can be shown that, given the length \( dW \) of each waveguide, pressure waves
travel along the Waveguide Mesh at a speed equal to
\[
\frac{1}{\sqrt{3}} d_W F_s \tag{3.2}
\]
where \( F_s \) is the sampling frequency, and the symbol \(<\) means that some spatial frequencies travel slower along the mesh. This effect is called dispersion \[69\], whose main effect is a detuning of high frequencies, which is not considered to be important for the application.

Assuming the velocity of sound in air equal to 343 m/s, and setting \( F_s = 8 \) kHz, we have from (3.2) that each waveguide is about 74.3 mm long. Thus, the mesh needs \( 127 \times 5 \times 5 = 3175 \) scattering nodes to model our tube. Note that the sampling frequency has important effects on the computational requirement of the model. Clearly, our choice is oriented to efficiency rather than sound realism: reliable distance cues should be conveyed also using lower sampling frequencies.

The Waveguide Mesh has already been used in the simulation of reverberant enclosures \[63\]. It has a major advantage in this application case, since it allows to deal with the boundary of the propagation domain quite effectively \[18\]. In fact, it enables to apply waveguide filters at the mesh boundary, that model the reflection properties of the internal surfaces of the tube \[30\].

More in detail, each waveguide branch falling beyond the boundary of the tube is terminated with a spring/damper system, that models a simplified resonating/absorption property of the surface at the waveguide termination. This system is algebraically rearranged into a Waveguide filter, then discretized into a Waveguide Digital filter establishing a transfer function, \( B(z) \), between pressure waves going out from, and incoming to the closest scattering junction:
\[
B(z) = \frac{P^+_i(z)}{P^-_i(z)} . \tag{3.3}
\]

For example, it is \( i = 5 \) in Fig. 3.2. Using the simple physical system seen above, the resulting filter terminations are made of 1st-order pole-zero filters. Despite this simplicity, these filters can be tuned so as to closely approach the reflecting properties of real absorbing walls \[39\].

Considering that the surfaces at the two terminations of the tube have been set to be totally absorbing (this meaning that \( p^+ \equiv 0 \)), the total number of boundary filters is \( 127 \times 5 \times 4 = 2540 \).

### 3.4 Model performance

Measures conducted on tube impulse responses are summarized in Fig. 3.3 and Fig. 3.4. Fig. 3.3 shows spectral differences existing between sounds auditioned nearby and far from the sound source, both for the left and right channel. Fig. 3.4 shows how the signals average magnitudes, defined by the value \( 10 \log \frac{1}{n} \sum \left| s(n) \right|^2 \) where \( s \) is the signal to be studied, and referenced to 0 in correspondence of the the closest position, vary with distance: These variations show that a dynamic range smaller than the 6 dB law is needed for rendering distances using the proposed method.

In particular, Fig. 3.4 shows that the right-channel magnitudes diverge from the left ones, as long as the range becomes greater than about 1 m. This
Figure 3.3: Magnitude spectra of signals 1R, 1L, 10R, 10L, ordered by decreasing magnitude.

divergence does not appear in reverberant sounds taken from real-world environments. This effect can be certainly reordered to the peculiarity of the listening environment proposed here. Nevertheless, a careful analysis of the side-effects coming from using a coarse-grained realization of the Waveguide Mesh as a model of the listening environment should be carried out, to assess the precision of the plots depicted in Fig. 3.4.

3.5 Psychophysical validation

We conducted two experiments in two different listening conditions: one using headphones and one using loudspeakers, in order to evaluate the model response in the two reproduction situations. There is a branch of auditory display that studies the differences existing between headphone and loudspeaker presentation of spatialized sounds [32]. In our model we have not added any specific adaptation to different devices. However, we want to compare the two cases.

The tests apply the magnitude estimation method. This experimental technique uses, sometimes, the modulus [68], that is a comparing stimulus to which the experimenter associates a value for reducing the estimated value range. We decided, during the tests design, to perform them without the modulus. In fact, it has been observed [55, 19] that it could introduce systematic errors both on the values used by the subjects and on the slope of the computed functions.

We investigated how subjects scaled the perceived distance and, hence, whether our model is effective or not.

The setup involved a PC Pentium III, with a Creative SoundBlaster Live! soundcard. During the first experiment sounds were auditioned through Beyerdynamic DT 770 closed headphones; in the second experiment, the participants sat at a distance of 1.5 m from a pair of Genelec 2029B stereo loudspeakers, 1 m far from each other, and a Genelec subwoofer located in between the loudspeakers.
3.5.1 Listening by headphones

The experiment, in the headphone reproduction condition, involved 12 subjects (4 female and 8 males), with age between 22 and 40, who voluntarily participated to the experiment. They study or work at the University of Verona. All of them were naive listeners.

The sound set was synthesized by putting a sound source at one end of the virtual tube, along the main axis, and acquiring ten stereophonic impulse responses along positions $x_{10}, \ldots, x_1$.

The right channel of the stereophonic sound accounts for acquisition points exactly standing on the main axis, whereas the left channel accounts for points displaced two junctions far from that axis, this corresponding to an interaural distance of about 15 cm. The impulse responses obtained in this way have been convolved with a short, anechoic sample of a cowbell. We decided for this type of sound for its simplicity and, additionally, low familiarity with the subject’s normal listening experience.

In a quite, but not isolated listening environment, we presented to the subjects a sequence of 30 sounds, randomly repeating for three times the 10 sounds taken form the set of stimuli. They had to estimate the perceived distance from the sound source, using headphones. Even if Blauert points out that the term “sound source” refers to the physical dimension, while the term “auditory event” concerns the perceptual sphere, and, in our context, it would be more correct to use “auditory event”, we decided to describe the experiment task with a language that is more familiar to the listeners.

The participants had to rate each distance with a value in meters (either integer or decimal), starting from the first one, and associating a value to the

\footnote{It would be interesting to repeat the tests with another type of sound, in order to study how sound familiarity affects the data results.}
Figure 3.5: Headphone listening: Individual distance evaluations together with individual linear regression lines. $a$: intercept. $b$: slope.

other ones, proportionally to the first estimation. The experiment was conducted without training. Moreover, since we did not set a modulus, the collected values define scales that depend on the individual listeners’ judgments. These scales range from 0.2-8 (subject no. 8) to 1-30 (subject no. 5).

The three judgments given for each sound were then geometrically averaged for each subject, and the resulting values were used to calculate a mean average. Subtracting it from the individual averages, we adjusted the listeners’ judgments to obtain a common logarithmic reference scaling [19].

In fig. 3.5, the distance evaluations as functions of the source/listener distance are plotted for each subject, together with the corresponding linear functions obtained by linear regression. The average slope is 0.6093 (standard deviation 0.2062), while the average intercept is 0.4649 (standard deviation 0.2132).

In fig. 3.6, the perceived distance averaged across values is plotted as function of the source/listener distance, together with the relative regression line ($r^2 = 0.7636$, $F(1, 8) = 25.8385$, $F_{crit}(1, 8) = 11.2586$, $\alpha = 0.01$). The $r^2$ coefficient is significant at $\alpha = 0.01$ and, therefore, the regression line fits well with the subjects’ evaluations.

We observed, from this first experiment, that subjects overestimate the distance for sound sources that are close to the listener, and that they reduce this overestimation for greater distances. This result is interesting since it partially conflicts with Zahorik [74], who works with real sounds for the tests, and reports the tendency of listeners to overestimate shorter distances, and underestimate longer distances. In our model, the point of correct estimation is more distant compared with Zahorik. This result can be interpreted as a consequence of the
exaggerated reverberant energy produced by our model.

### 3.5.2 Listening by loudspeakers

The second experiment involved 10 volunteers (4 females and 6 males), 4 of which participated also to the first experiment. They work or study in our department and are aged between 23 and 32.

The set of stimuli was the same as for the previous test, but the reproduction system, in this second experiment, included loudspeakers. The listeners sat 1.5 m far from the loudspeakers and were blindfolded, in order to minimize the influence of factors external to the experiment. Listeners, in a quiet, but not isolated room, had to evaluate the distance of the sound source from the listening point communicating its value to the experimenter, who wrote down the data. The first value, as in the previous test, determined the subjective scale.

It is interesting that, in this phase, four participants were non-naive, because involved also in the first experiment. In this way, we can compare also the responses of different subject categories.

In fig. 3.7 we report, for each subject, the distance evaluations as functions of the sound source/listener distance, together with the corresponding linear functions obtained by linear regression. The average slope is 0.5337 (standard deviation 0.1741), while the average intercept is 0.5034 (standard deviation 0.3573).

In fig. 3.8 the perceived distance averaged across values is plotted as function of the sound source/listener distance, together with the relative regression line ($r^2 = 0.8512$, $F(1, 8) = 45.7603$, $F_{crit}(1, 8) = 11.2586$, $\alpha = 0.01$). The $r^2$ coefficient is significant at $\alpha = 0.01$ and, therefore, the regression line fits well with the subjects’ evaluations.

We can see that the loudspeaker test led to results that are similar compared with the headphone test results. In fact it is evident that, in both cases, there
is a distance overestimation for closer sound sources, that reduces as long as the distance increases.

There is only one subject (no. 10) whose individual scale ranges between 0.1-2, and who perceived all the sound sources to be closer compared with the other listeners. However, during the talk/questionnaire taken after the test, this participant did not reported any difficulty in performing the required task.

Furthermore, there is no evident difference between judgments of naive participants, and subjects “trained” by the previous experiment.

A comparison between the two experiments gives interesting hints. First of all, the subjects’ responses are similar in both the reproduction conditions. Although our model does not deal with the issue of sound reproduction, it fits both reproduction systems [32, 75].

Moreover, there is an exaggeration especially in rendering close sound sources, probably due to the amount of reverberant energy existing in that case. The point of correct estimation, in both the reproduction scenarios, is far away from results obtained by other researchers [75]. For this reason, our virtual resonating environment could be adopted in the setup of auditory displays where sounds in the far-field must be presented, without any particular requirement on the reproduction device.
Figure 3.8: Loudspeaker listening: Average distance evaluation together with linear regression line. $a$: intercept. $b$: slope.

3.6 Conclusion

A virtual listening environment capable of sonifying sources located in the far-field has been presented, along with a versatile way to model it. Listening tests show that it actually conveys exaggerated range cues, nevertheless the model can be easily re-parameterized to account for different psychophysical scales. In this perspective our model is prone to further optimization, as new reverberant spaces can be explored and straightforwardly validated through psychophysical experiments similar to the ones performed during this research.
Chapter 4

Experiments in a 2-D space using the Sonic Browser for validation and cataloguing of Sound Objects

Eoin Brazil†, Laura Ottaviani‡, and Mikael Fernström††

4.1 Introduction

In this chapter we describe a software tool developed in the Interaction Design Centre at the University of Limerick, which can be used for multiple purposes relevant to the Sounding Object Project. We also discuss two scenarios for psychoacoustic validation and for cataloguing of Sound Objects using the aforementioned software tool and we present the results of some experiments on these scenarios recently conducted at the IDC.

4.2 The Sonic Browser

From 1996 onwards, a software tool has been under development [9] [10] at the IDC at UL, for auditory and multimodal browsing. The initial aim was to explore if auditory display could be augmented by providing multiple and concurrent auditory streams in a highly interactive environment, supporting users to develop mental models of sonic information spaces. For musical content, the approach proved to be successful, and the concepts have been further developed and enhanced into the current version of the Sonic Browser. Originally, the main feature of the Sonic Browser was the ability interactively to choose and to listen to multiple audio streams of music concurrently. This concept has now been improved with a number of interactive visualisations and filtering mecha-

† Interaction Design Centre, University of Limerick
‡ Dipartimento di Informatica, Università di Verona.
nisms. The current version is also very promising as a client-server application that might be usable over local area networks and, perhaps, the Internet.

Currently, we see two main scenarios of use for the Sonic Browser in the context of the Sounding Object Project. Our first scenario is to use the tool for psychoacoustic validation of Sound Models. The second scenario is for sound designers to use the tool for managing catalogues of Sound Objects. In the following subsections we introduce briefly the two scenarios. In section 4.3, we describe the tool, while in section 4.4 and section 4.5 we present the results of the experiments conducted in the two contexts respectively. Finally, in section 2.4 we report our conclusions and future investigations.

4.2.1 Validation of Sound Objects

The development of Sound Objects in this project has been a progression of studies ranging from recordings of real sounds, listening tests, phenomenology, psychophysical scaling, to physical modelling of sound sources and actions.

In this chapter we outline an approach for validation of sound models, i.e. how Sound Objects compare to recordings of real sounds in terms of scaling, quality and realism, and we present the results of the recent experiments.

We have informally noted the difficulty of representing sound, both as recordings and as models, as the experience is always mediated either by microphones, recording/playback mechanisms and loudspeakers, or, by the models themselves. As the arguments about the relation between electronically mediated sounds and real sounds are outside the scope of the project, we henceforth disregard this somewhat philosophical discussion and look at the possibilities of empirical validation of Sound Models.

In the real world, we can quite easily determine the properties of auditory events, e.g. the size and distance to a bouncing object, its material properties, if it breaks. Real sounds and people are situated in real acoustics, which helps us to pick up acoustic events. Sound Objects are, so far, not situated in an intrinsic environment and therefore has to be seen as hyper-real representations of acoustic events, although in their final application they will be somehow encapsulated into objects and environments that will make them situated.

In section 4.4 we propose to empirically validate the properties and quality of Sound Objects using the Sonic Browser and by having the subjects sort recordings of real acoustic events and Sound Object events. In that section, we present the results of the experiments based on this approach and conducted recently by the authors.

4.2.2 Cataloguing Sound Objects

Cataloguing sounds is traditionally problematic. Many, or all, Foley artists know their collections of sounds and most often they appear to be indexed and sorted based on events and sound sources in text format. While this is the currently accepted practice for cataloguing sounds, it can be induced that it is quite inefficient. If you are working as a sound designer, reading about a sound, without hearing it or having a recollection of having heard the particular sound, is as productive as swimming a bicycle.

The same problem applies to other sonic workers such as electroacoustic composers who work with collections of thousands of sounds and quite often
think of sounds in an acousmatic sense, which makes verbal descriptions of sounds somewhat meaningless.

With the current version of the Sonic Browser, users can explore corpora of sounds by hearing. Visualisations are simultaneously available only to support interactive spatial organisation and indexing, based on the perceived auditory properties and qualities of the sounds.

In section 4.5 we introduce the results of the recent experiment in the cataloguing scenario.

4.3 The Sonic Browser - An overview of its functionality

The Sonic Browser current supports four visualisation mechanisms for the representation of audio data. These mechanisms are TreeMap, HyperTree, TouchGraph and SOBGrid (Starfield display). These mechanisms are supplemented by the use of multistream audio and in certain cases an “aura” as well as three filter mechanisms.

A metaphor for a user controllable function that makes it visible to the user is the aura [20]. An aura, in this context, is a function that defines the user’s range of perception in a domain. The aura is the receiver of information in the domain, shown as a grey circle surrounding the cursor in fig. 4.1, the cursor is surrounded by the aura. All sonic objects within the aura play simultaneously, panned out in a stereo-space around the cursor.

The filtering mechanisms are based on both the intrinsic and extrinsic object properties and include text filtering, dynamic sliders and colour filtering. The filters offer simple ‘And’/’Or’ type filtering of the objects as well the ability to load and save of the current filter settings. It is important to state that all these mechanisms are dynamic and function in real-time. This work is based on Shneiderman’s mantra for design of direct manipulation and interactive visualisation interfaces “overview first, zoom and filter, then detail on demand” in mind [65].

Mappings that are used include acoustic/perceptual, onomatopoeia, source, sound type and event. Acoustic/perceptual mappings describe a sound’s physical attributes such as brightness, dullness or pitch. Onomatopoeia is where sounds are described by the way they sound, e.g. hammering could be “thunk-thunk” [71]. Source is the actual type of object associated with producing that sound or sounds. Sound type is used to break sound into various subcategories such as Speech, Music and Environmental. Audio events/actions are associated with what action or event has occurred, for example, a sound of a car braking would be “braking”.

Each object can be recognised by its visual and aural properties. The objects, which are under the aura of the cursors, are played. Sounds are panned out in a stereo field controlled by the visual location of the tunes nearest the aura. The volume of the tunes playing concurrently is proportional to the visual distance between the objects and the cursor. The maximum volume will be heard when the particular object is centered under the aura. The auditory aspects are controlled by the aura or the cursor (if the aura is off or unavailable). The rest of this introduction will focus on particular aspects of the Sonic Browser’s
4.3.1 The Sonic Browser Controls

The left side or control side of the Sonic Browser offers various buttons and mechanisms for dynamic control of the application as shown in fig. 4.2. The slider is used to dynamically control the size of the aura. The buttons below this slider are the available visualisation buttons. These buttons activated the desired visualisation when pressed and display the particular visualisation in the right side/pane of the application. The next set of controls are the filter control mechanisms. They offer ‘OR’, ‘AND’, removal of the current filter settings, clearing the current filter settings as well as saving and loading of these settings.
The current filter settings are displayed in the gray area to the right of these buttons. These control mechanisms are governed by the selections from the particular filtering mechanisms. The filtering mechanisms are shown directly below the controls and are dependent on the currently selected tabbed pane. These mechanisms are used to add particular filters and/or combinations of filters to the list of currently selected filters. The final area in fig. 4.3 represents the status area where information on the current user identification number, the connection state to the sound server and the connection state to the database are displayed.

Figure 4.3: The Sonic Browser Controls

4.3.2 The Sonic Browser visualisations

This section provides a brief overview of the available visualisations used in the Sonic Browser to illustrate these visual aspects and their purposes. Information visualisation techniques offer many components, which can be applied to a limited display surface used in accessing large volumes of data. There are two distinct categories: distortion-oriented and nondistortion-oriented [41]. The Sonic Browser uses mechanisms from both categories. The techniques used
are mainly from the branch of distortion-oriented techniques known as focus +
context. Analyses of these techniques have been written by Noik [54] and by
Leung [41], which we will not duplicate here. The various visualisations offer
several viewpoints of the dataset these allow users to gain a greater insight into
the dataset by exploring through the various visualisations. Different domains
can be better illustrated by specific visualisations by offering various visualisa-
tions; a more domain generic application is created.

The TreeMap visualisation

The Treemap [31, 66] is a space-filling layout that is generated automatically,
used primarily for overviews of document collections and their meta-data and is
shown in fig. 4.4. Creating the Treemap involves dividing and subdividing the
screen into rectangular areas of alternate horizontal and vertical divisions (“slice
and dice”) to represent subordinate nodes. The rectangles are colour coded
to some object attribute to achieve the rectangular method used to display
information objects. These rectangles can have associated text or labelling
describing the information object it represents. A disadvantage of the Treemap
is while being good in representing the attribute of the information structure
portrayed through rectangle area (usually size), it is not so good at conveying
the structure of the hierarchy.

![Figure 4.4: The Sonic Browser with the TreeMap visualisation](image)

The HyperTree visualisation

The HyperTree technique was pioneered at the Xerox Corporation by Lamping
et al [40] and is based on Hyperbolic geometry which was one of the non-
Euclidean geometries developed at the turn of the century [14]. It is a focus
and context view that lays out the hierarchy dataset in a uniform way on a
hyperbolic plane and then maps this plane onto a circular display region. It
allocates the same amount of room for each of the nodes in a tree while still
avoiding collisions because there is an exponential amount of room available
in hyperbolic space. With a single still image, a projection from hyperbolic space looks similar to a Euclidean scene projected through a fisheye lens. The projection to hyperbolic space serves the purpose of a Degree of Interest function as required by Furnas [23]. The advantages of this technique are that any node of interest can be moved into the centre so its detail can be examined using simple dragging of the node. As this node becomes the focus of attention the entire tree is appropriately repositioned with this node as the root node. A node’s context can be easily seen, as it is view from all directions of the tree with its parent, siblings and children show in close proximity. The Sonic Browser HyperTree visualisation was already shown in fig. 4.2

The TouchGraph Visualisation

The TouchGraph visualisation is a simple Zoomable User Interface (ZUI). ZUIs are based on alternative user interface paradigm using zooming and a single large information surface. The earlier foundations for ZUIs are found in work by Furnas and Bederson on multiscale interfaces called Space Scale Diagrams [24]. Detail can be shown without losing context, since the user can always rediscover context by zooming out. The TouchGraph visualisation uses geometric zooming where all the objects change only in their size so it simply provides a blown up vision of the dataset in combination with physics based forces. It also uses simple semantic zooming by selecting the number of related nodes and children of those nodes, which are displayed. The Sonic Browser TouchGraph visualisation is shown in fig. 4.5

Figure 4.5: The Sonic Browser with the TouchGraph visualisation

The SObGrid visualisation

The SObGrid visualisation is based upon the work by Williamson and Shneiderman [73] on dynamic queries. Dynamic queries allow for rapid, incremental and reversible changes to query parameters, often simply by dragging a slider, users can explore and gain feedback from a display in a few tenths of a second. The
Starfield display is essentially a two-dimensional scatter plot using structured result sets and zooming to reduce clutter. Ahlberg and Shneiderman [2] discuss this in the context of visual information seeking (VIS) which is a methodology specially created to support visual browsing tasks. The Starfield display provides an overview of the dataset with the information objects being represented by coloured dots or shapes. The properties of the information objects are used to map to the screen location, colour and shape for each of the objects and its related graphical representation. The number of visible screen objects is controlled through the use of dynamic queries; in the case of the Sonic Browser these are the various filtering mechanisms such as text filtering, dynamic sliders or colour filtering. The Sonic Browser SObGrid visualisation was already shown in fig. 4.1.

4.3.3 The Filtering Mechanisms

The Sonic Browser offers three different distinctive mechanisms for dynamic filtering of objects. The three mechanisms are the text filtering, dynamic sliders and colour filtering. These mechanisms operate on information about the objects that has been stored in the underlying database. In the case of the text filter and the dynamic sliders this information is a set of text descriptions based on arbitrary classification of the object’s properties. These descriptions of arbitrary classifications can be used in both classification and identification tasks.

Text Filtering

The text filtering mechanisms uses user-specified properties of the objects to filter them. The user-specified properties of the objects, they may not all apply to any particular object as they are based on a general approach to include the largest number of possible properties. The possible shape properties of an object are shown below in fig. 4.6. It offers several categories of filtering on object properties from filename to music genre, it does not currently offer a automatic classification of the object to these categories but this idea has already been considered for a future interface mechanism as described by Brazil [10].

Dynamic Sliders

The dynamic sliders mechanism is again based on the user-specified properties of the objects. The dynamic sliders are based on the Alphaslider [1]. Alphasliders are used for navigating a large number of ordered objects where the scale range corresponds to the range in which the items exist. The dynamic sliders are shown in fig. 4.7. The dynamic sliders can be described as a modification of the existing slider interface component to include both changing of the position within the scale and also of the scale itself.

Colour Filtering

The colour filtering mechanism is based on the current visual properties of the objects. This mechanism is more experimental, its goal is to gauge the usefulness of filtering objects by their colour properties. It is a modified GUI component as shown in fig. 4.8 and is based on the normal colour chooser component. The
Figure 4.6: The Sonic Browser’s Text Filtering Mechanism

Figure 4.7: The Sonic Browser’s Dynamic Sliders Mechanism

preview pane is retained to ensure that the current colour is obvious. Further experimentation will validate this mechanisms advantages and disadvantages in the context of dynamic filtering of object.

4.3.4 The Axis Control Mechanism

The SObGrid component offers the ability to map a property of the object dataset to the X or the Y dimensions and as such requires a mechanism to perform this task. The Axis Editor is the component that performs this task.
and is shown in fig. 4.9. The current axis configuration allows only for a 2D display but the functionality for a 3D display has been included for a possible future release. The axis control mechanism is based on both arbitrary and non-arbitrary object properties to allow for a better exploration of the intrinsic properties of the objects.

Figure 4.8: The Sonic Browser’s Colour Filtering Mechanism

Figure 4.9: The Sonic Browser’s Axis Editor

4.3.5 The Aura of the Cursor

The aura within the Sonic Browser can be controlled by one of three mechanisms. These mechanisms are via the keyboard, via the Aura slider or via the Aura Size Dialog. The keyboard mechanism is included as a set of four shortcut keys, which are the same as in previous releases of the Sonic Browser and are illustrated in Table 4.1.

The second mechanism for control of the aura is via its slider. This is a slider with the available size range of the aura mapped to the slider. It is shown
Table 4.1: Aura Keyboard Shortcuts

<table>
<thead>
<tr>
<th>Key</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>Increase the Aura size by 1</td>
</tr>
<tr>
<td>-</td>
<td>Decrease the Aura size by 1</td>
</tr>
<tr>
<td>1</td>
<td>Activate the Aura</td>
</tr>
<tr>
<td>0</td>
<td>Deactivate the Aura</td>
</tr>
</tbody>
</table>

Figure 4.10: The Aura Slider

The third mechanism for controlling the aura is via the Aura Dialog. This can set the current size as well as the maximum and minimum sizes for the aura. This dialog is dynamically linked with the existing on screen aura and the aura slider to ensure that any changes made within the Dialog are immediately passed to these elements. The Aura Size Dialog is shown in fig. 4.11.

Figure 4.11: The Aura Size Dialog

### 4.3.6 Accessing the objects

The Sonic Node Viewer Window allows for an object’s properties to be accessed and modified dynamically. It can be opened by a control - right click on an object. The Sonic Node Viewer Window allows the user to browse all the properties of the objects stored in the current collection. The user can change any of the object’s properties from the Node Viewer, shown in fig. 4.12. It also includes a list of the other objects within the collection, which can be accessed by simply double clicking on their name within the list to update the Node Viewer to show the details of that particular object. The Node Viewer shows
various properties for each object and it also describes the categories (field of values for the objects) and the graphical attributes of the object. The properties of an object can be easily changed and any changes will be reflected within the collection.

![Properties of an Object under the Aura](image)

Figure 4.12: Properties of an Object under the Aura

### 4.3.7 Accessing the related PD patch for the object

Accessing the related PD patches for any object within the Sonic Browser is easy. It can be opened by an alt - right click on an object with a related PD patch. The sound file within the Sonic Browser is an auditory representation of a defined set of values for that particular PD patch and will open the particular PD patch with those values. At this point of the implementation there is no return feedback available from PD to the Sonic Browser, so that any changes to the PD patch will not be reflected in the related object within the Sonic Browser.

### 4.3.8 Dragging the objects

The Sonic Browser offers two methods of dragging objects around the screen. The first method is a simple single object select; the second is a group object select which is currently only available in the SOBGrid. Single objects can be moved around the screen by holding down the left mouse button while an object is directly under the centre of the aura. This can be done repeatedly; a log of this movement is kept in database under the table ‘profiles’. It can be used to determine where objects were positioned on the screen after being dragged from their original position/s as well as if there were changes to any of the objects properties. Group selecting object is where the mouse is clicked on the screen but not on a object and dragged to the appropriate point and released. A square / rectangular region is highlighted any nodes within this region are now currently selected. To drag the selected nodes, you click within the region.
highlighted and drag the mouse to the point where you wish to drop the nodes then release the mouse to drop the nodes at that point.

4.3.9 Tagging the objects

The Sonic Browser offers a simple method of tagging files of interest. An object can be tagged by simple choosing the ‘Tagged’ option in the Sonic Node Viewer Window as shown in fig. 4.12. The visual representation of the object is then updated with a halo around it corresponding to the object current shape. Saving the current list of tagged files is achieved by selecting the File Menu and then the option ‘Save a list of tagged sounds to file’.

4.3.10 Drag And Drop capabilities

The Sonic Browser offers the ability to ‘Drag And Drop’ one or more sound files directly onto the application. These files will then be copied and placed within the current collection and will be fully accessible as with any other object within the collection.

4.4 Experiments in the validation scenario using the Sonic Browser

4.4.1 Brief summary of previous interesting approaches

A fundamental aspect of conducting psychoacoustic experiments is the choice of the right type of test, in order to collect all the necessary data with the least amount of effort by the user.

There are different approaches proposed for different goals. We know the classical methods in psychoacoustics, such as the methods of limits, of adjustment and of constant stimuli. These methods, sometimes, lack certain features. Therefore, there is a wide research field trying to study and develop new approaches for psychoacoustic experiments.

This is why, for instance, the Signal Detection Theory was born, which focuses not on the stimulus thresholds estimation, but on the reasons underlying one particular subject decision. The STD estimate judgement process elements not considered by the classical psychophysical methods.

An interesting method is the Stimulus Sample Discrimination (SSD) method, proposed by Mellody and Wakefield [50], based on previous studies on the stimulus sampling procedure by Sorkin et al [67] and Lutfi [45, 46, 47]. It focuses on psychophysical discrimination experiments by using samples from a context distribution, which, according to the particular task, could be considered as additional information provider or as distraction component. The researchers report two main scenarios where this method could be applied: investigating on informational masking - that is the influence of the context on the listener’s decision - and studying the subject discrimination between two distributions. In their paper, they present an application of the SSD for evaluating the preservation of the singer identity in low-order syntesis.

Another important branch of experimental methods is classified as unidimensional and multidimensional scaling methods. They aim at estimating psycho-
logical scales which allow to compare the physical to the perceptual parameters, referring respectively to one and more dimensions.

An improvement of the classical multidimensional scaling technique is the one proposed by Scavone et al. [64], i.e. to use an interactive program, the Sonic Mapper, in order to allow the listeners to arrange, by drag-and-drop, the stimuli in a bi-dimensional space according to the similarities they find among them. The advantages of this approach are a decrease of the users fatigue, since it is possible to apply less comparisons than in the classical methods, and, as a consequence this can lead to an increase in attention and consistency of the listeners decisions. Moreover, as the user is able to compare all the stimuli interactively, they are able to appreciate the full set of stimuli more so than can in a pairwise comparison task.

4.4.2 The experiments conducted with the Sonic Browser

In the validation scenario we conducted a psychophysical experiment to compare real sounds to Sound Objects and to investigate the perceptual scaling of the physical parameters that control the sound models. The aim of the experiment is to begin to understand how the synthesized sounds produced by our models are scaled in comparison with the physical dimensions. We focused on two dimensions: perceived height of the object drop and perceived size of dropped objects. Our exploration was not limited only to the scaling task, but also encompassed the perceived realism of the event. Therefore, we divided the experiment in two phases, one concerned with the scaling task per se and the other one focused on the realism judgement. Moreover, as we wanted to compare the sound rendering of two different approaches in the sound modelling, the stimuli set included, besides recorded events, Sound Objects from both of these modelling approaches.

The experiment was preceded by a pilot experiment, which used only one modelling approach. The pilot probe allowed for a first observation of the type of results and it highlighted which sounds were most suitable to focus on in the main experiment.

The experiment used the Sonic Browser, that allows the users to listen to the stimuli and to drag-and-drop them according to their judgements within a specified bi-dimensional scale. The specified scale was a 2-D plot with the perceived size of the dropped object on the X axis and the perceived height of the object drop on the Y axis.

The experimental data collection involved two techniques. First, data logging was collected by the application for the object positioning in the 2-D space. Second, the user was asked to comment aloud on the thinking process, as it is established by the Thinking Aloud Protocol.

The Thinking-Aloud Protocol is one of the most widely used methods in usability testing and it represents a way for the experimenter to have a “look” in the participants’ thought processes [6]. In this approach, the users are asked to talk during the test, expressing all their thoughts, movement and decisions, trying to think-aloud, without paying much attention to the coherency of the sentences, “as if alone in the room”.

We applied this protocol in our formative usability probe. Employing this protocol, we were able to collect not only the data concerning the stimuli positions in the 2-D space of the Sonic Browser, but also the comments of the users
during the experiment, which expressed the reasons, for instance, of a particular judgement or their appreciation of the Sound Objects realism. The tests were all recorded by a video-camera.

In the next subsections we will present both the pilot and the main experiment, introducing procedures and results of the test.

The pilot probe

Participants. The pilot probe involved 4 volunteers, all students at the Computer Science Department of the University of Limerick. All of them referred to have neither hearing nor sight problems and all of them have a musical training (5, 4, 2 and 10 years respectively).

Stimuli. The stimuli set included 9 recorded sounds and 9 Sound Objects, derived from the same modelling approach.

The recorded sounds were produced by 3 steel balls, weighting 6, 12 and 24 g, and falling on a wood board of 1500 x 500 x 20 mm from a height of 10, 20 and 40 cm, respectively, by positioning the microphone at 3 different distances: 20 - 40 - 80 cm, respectively. Recordings used a MKH20 Sennheiser microphone, and a sound card sampling at 44.1 kHz rate.

These stimuli were used in previous experiments conducted by the SOb project on the perception of impact sounds [12]. In this study, Burro found the relationship between the physical quantities of weight, distance and height and the relative perceptual quantities. He argues that manipulating one of the physical parameters affects more than one of the perceptual quantities.

In the pilot probe, we decided to keep the height of the dropped balls constant (h=20 cm).

The synthesized sounds were all designed with the PD-modules modelling impact interactions of two modal resonators [57], simplified returning only one mode, and they used either glass or wood as the material property.

In the tab. 4.2, we report the list of the stimuli used in the pilot probe, with the relative short name and the sound type.

In tab. 4.3, we report the values of the parameters used in the PD-modules for synthesizing the stimuli. For a reference on the main features of the models and the meaning of the parameters, we suggest to consult the previous deliverables, e.g. [57].

Procedure. The probe was conducted in the isolation room of the Computer Science Department at UL. The stimuli were presented by stereo headphones to the user through the Sonic Browser, which allowed the users to listen to the sounds as many times as they wanted and to drag-and-drop them according to the dimensions on the axis. The experiment was conducted applying the Thinking-Aloud Protocol and the participants sessions were all recorded on video-tapes.

After the perception estimation task, the participants, during the second phase of the test, were asked to tag the sounds that they thought were unrealistic.

At the end of each session, a questionnaire was presented to the participants in order to gain an insight into their perceptions of the performed task.

The users estimated the data positions in the bi-dimensional scale without a comparison stimulus or a reference scale. Despite being pre-defined, i.e. being limited to the screen, the ranges of perceptual evaluations were relative to each
<table>
<thead>
<tr>
<th>Short Name</th>
<th>Sound File</th>
<th>Sound Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>sound1</td>
<td>d20-w12-h20.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound2</td>
<td>d20-w24-h20.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound3</td>
<td>d20-w6-h20.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound4</td>
<td>d40-w12-h20.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound5</td>
<td>d40-w24-h20.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound6</td>
<td>d40-w6-h20.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound7</td>
<td>d80-w12-h20.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound8</td>
<td>d80-w24-h20.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound9</td>
<td>d80-w6-h20.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound10</td>
<td>small-bouncing-glass-ball-1-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound11</td>
<td>small-bouncing-glass-ball-2-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound12</td>
<td>small-bouncing-glass-ball-3-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound13</td>
<td>small-bouncing-glass-ball-4-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound14</td>
<td>small-bouncing-glass-ball-5-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound15</td>
<td>small-bouncing-wooden-ball-1-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound16</td>
<td>small-bouncing-wooden-ball-2-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound17</td>
<td>small-bouncing-wooden-ball-3-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound18</td>
<td>small-bouncing-wooden-ball-4-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
</tbody>
</table>

Table 4.2: List of the stimuli used in the pilot probe.

<table>
<thead>
<tr>
<th>Short Name</th>
<th>elasticity</th>
<th>damping</th>
<th>gravity force</th>
<th>strike velocity</th>
<th>freq (Hz)</th>
<th>decay time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sound10</td>
<td>15000</td>
<td>46.4159</td>
<td>990</td>
<td>630.957</td>
<td>1758.52</td>
<td>0.233307</td>
</tr>
<tr>
<td>sound11</td>
<td>5540.1</td>
<td>8.57696</td>
<td>990</td>
<td>1318.26</td>
<td>1782.52</td>
<td>0.090315</td>
</tr>
<tr>
<td>sound12</td>
<td>15000</td>
<td>21.5443</td>
<td>950</td>
<td>1584.89</td>
<td>1388.82</td>
<td>0.090315</td>
</tr>
<tr>
<td>sound13</td>
<td>3161.6</td>
<td>21.5443</td>
<td>580</td>
<td>2290.87</td>
<td>1388.82</td>
<td>0.090315</td>
</tr>
<tr>
<td>sound14</td>
<td>15000</td>
<td>46.4159</td>
<td>450</td>
<td>630.957</td>
<td>1782.52</td>
<td>0.043070</td>
</tr>
<tr>
<td>sound15</td>
<td>15000</td>
<td>46.4159</td>
<td>450</td>
<td>630.957</td>
<td>1758.52</td>
<td>0.043070</td>
</tr>
<tr>
<td>sound16</td>
<td>15000</td>
<td>63.0957</td>
<td>940</td>
<td>912.011</td>
<td>1113.23</td>
<td>0.603386</td>
</tr>
<tr>
<td>sound17</td>
<td>11395</td>
<td>2.92864</td>
<td>860</td>
<td>301.995</td>
<td>1294.33</td>
<td>0.752992</td>
</tr>
<tr>
<td>sound18</td>
<td>1309.5</td>
<td>4.64159</td>
<td>970</td>
<td>436.516</td>
<td>1294.33</td>
<td>0.784488</td>
</tr>
</tbody>
</table>

Table 4.3: Values of the parameters for the synthesized sounds used in the pilot probe.
user. The perceptual space boundaries were considered by all the users, as they reported at the end of the task, relative to their maximum value. In fact, we noticed an initial difficulty by the participants of referring to the screen space. On the contrary, they showed a preference of defining their own boundaries. In order to be able to compare the results of each participant, we decided to normalize the data coordinates, which identify the locations in the 2-D space, between 0 and 1.

**Results and Observations.** In fig. 4.13, we report the representation of the perceptual scaling and tagging information of all the users and all the stimuli in one graph. It is evident from this representation, that the participants estimate correctly the height from the real sounds, \(h=20\) cm for all of them, since most of the real sounds, barring five outliers, are positioned in the central area of the evaluation space.

![Figure 4.13: Pilot probe: representation of the perceptual scaling and tagging information of all the users and all the stimuli.](image)

An interesting observation arises from fig. 4.14, which represents the individual perceptual scaling and tagging information sorted by users. Two participants in particular (users n. 2 and n. 3) made an obvious distinction between real and synthetic sounds.

In fig. 4.15, we represent the individual perceptual scaling and tagging information sorted by stimuli. We can see that, while, as we have already said, the height of the real sounds is perceived correctly, the size estimation varies to a degree between users. This could be influenced by either the distance and/or the conditions in which the real sounds were recorded.

Looking at the synthesized sounds, we noticed that for most of them, the participants agreed in the scaling task, at least for one of the two dimensions. Only for two stimuli (small-bouncing-glass-ball-1, sound10, and small-wooden-ball-2, sound16) the perceptual scaling was spread across the evaluation space.

It is interesting to observe the single stimuli, looking at fig. 4.16 which rep-
Figure 4.14: Pilot probe: representation of the individual perceptual scaling and tagging information sorted by users.

As it arises from these plots, two stimuli (small-bouncing-glass-ball-1-pd, sound10 and small-wooden-ball-2-pd, sound16) were hardly judged in a uniform way by all the users and the perceptual scaling is spread across the evaluation space, while the perceptual scaling of small-wooden-ball-3-pd, sound17, is slightly spread, but the latter was tagged as unrealistic by all the participants to the probe. Therefore, the data spread could be due to the lack of realism provided by the sound.

On the other hand, there is one sound small-wooden-ball-4-pd, sound18, which is judged uniformly by all the users, and especially for the size dimension. We can see it more clearly in fig. 4.17.

Finally, the other five stimuli of the synthesized set were all judged uniformly in one dimensions. In particular, the sounds small-bouncing-glass-ball-2-pd, small-bouncing-glass-ball-3-pd and small-bouncing-glass-ball-4-pd, i.e. respectively sound11, sound12 and sound13, are perceived with the same height, while small-bouncing-glass-ball-5-pd and small-bouncing-wooden-ball-1-pd, i.e. sound14 and sound15, are perceived with the same size. In fig. 4.18 and fig. 4.19 we report, respectively, the box plots of small-bouncing-glass-ball-3-pd, sound12, and small-bouncing-wooden-ball-1-pd, sound15, examples of the uniform perceptual scaling in only one dimension.

We noticed that the participants didn’t agree in the results of the tagging
task. Only one stimulus, small-wooden-ball-3-pd, sound17, was defined by all the users as unrealistic. On the other hand, only two, i.e. small-bouncing-glass-ball-2-pd and small-bouncing-glass-ball-4-pd, that are sound11 and sound13, were judged to be unrealistic by 3 participants. All the other stimuli received only two mentions. Moreover, it was observed that no participant identified any of the real sounds to be unrealistic. This is due to the presence in the real sounds of the room acoustic that the synthesized stimuli lack.

Another interesting observation regards the type of approach taken by each participant to the task. Some of them preferred not to use the aura, but it was found to be useful for comparing the stimuli and checking the estimation of the whole group by those participants who used it.

At the end of each session as part of the participant debriefing, a questionnaire was presented to the participants in order to gain an insight into their perceptions of the performed tasks and about the Sonic Browser.

In the debriefing phase, a seven point Likert scale questionnaire with six sets of semantic differentials was filled out by the participants who were asked to express their responses to the interfaces and to the tasks. (from 0 to 6, where
0 is “poor” and 6 is “excellent”). In fig. 4.20 the results of the questionnaire with cumulative participant responses displayed per question can be seen with 0 representing a negative result to the question and 6 represent a positive result to the question.

Question one deals with the perceived difficulty in performing the task by the participant and the results show that it was found to be a non trivial task. Question two deals with the ease of use of the Sonic Browser for these tasks which was found to be above average ease of use. Questions three and four dealt with the quality and realism of the sounds. The results from these questions show that sounds were found to be both of high quality and realistic by the
Figure 4.17: Pilot probe: representation, by a box plot, of the perceptual scaling of the height and size for stimulus n.18. Uniform perceptual scaling in both the dimensions.

participants. Questions five concerned a technical issue which arose from the piloting phase of this experiment. There is a slight delay of up to .3 of a second when playing an audio file with the Sonic Browser. The result of this question was an acceptable but noticeable delay when playing sounds within the Sonic Browser.

The rich verbal protocol returned several interesting results during the experiment. The lack of room acoustics or background recording noise was commented by one participant who stated that some of the sounds did not have any “room effect”. Several participants found that the “speed of bouncing was directly related to the realism of the sounds”. The participants were found to use one of three strategies for scaling the sounds. These strategies were to “rough order the objects to the height scale first” or to “sort according to size initially” or to “sort them into real or synthesized sounds”.

The aura was only found useful by half of the participants. Participants also commented that “the longer I spent working with the sounds, the more difficult it was to sort them”. This relates to a greater working knowledge of the sound collection and the difficulty in maintaining a consistent scale across multiple sounds.
The main experiment

Participants. The participants to the main experiment are 5 volunteers, all students or workers at UL. All the participants referred to have musical training in average of 8 years, with a minimum of 6 and a maximum of 10 years. Two participants require glasses for reading, but no participant reported to have hearing problems.

Stimuli The stimuli set included 6 real sounds and 12 synthetic, 6 of which are designed with the PD-modules modelling impact interactions of two modal resonators [57], simplified returning only one mode, while the other 6 with the PD-modules modelling impact interactions of two modal resonators as well as the dropping event.

The real sounds are the same used in the pilot probe, but, in this case, we kept constant the distance (d = 80 cm), while changing the height.

For the synthetic sounds, that belong to two groups, we preferred to keep constant the material, since we noticed during the pilot probe some difficulties by the users to evaluate and compare the dimensions of events involving different material. We decided on wood as the material, even if it is not clear if the wood is the material of the impactor or of the surface. In fact, even if the
real sounds come from steel balls, they were referred to by the participants as wooden balls. This perception arose from the bigger influence of the surface material in certain cases.

In the tab. 4.4, we report the list of the stimuli used in the experiment, with the relative short name and the sound type.

In tab. 4.5 and tab. 4.6 we report the values of the parameters used in the PD-patches of the two modelling approaches for synthesizing the stimuli sets. In tab. 4.6 we report only the values that we changed for each sound. The other ones were kept constant at the following values: elasticity = 1e+007, alpha = 1.02882, lambda = 1e-006, strike velocity = -1.44544, minimum & = 3.16228, maximum (regular) interval = 1000, multiplication factor = 0.88, interval deviation = 1, value deviation = 1. For a reference on the main features of the models and the meaning of the parameters, we suggest to consult the previous deliverables, e.g. [57].

Procedure. The experiment was conducted in the same isolation room as the pilot probe. The stimuli were presented by stereo headphones to the users through the Sonic Browser. As in pilot experiment, the Thinking-Aloud Protocol was applied and all the users’ performances were video-taped.

After the perception estimation task, the participants were asked to tag the
<table>
<thead>
<tr>
<th>Short Name</th>
<th>Sound File</th>
<th>Sound Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>sound1</td>
<td>d80-w12-h10.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound2</td>
<td>d80-w12-h20.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound3</td>
<td>d80-w12-h40.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound4</td>
<td>d80-w24-h10.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound5</td>
<td>d80-w24-h20.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound6</td>
<td>d80-w24-h40.wav</td>
<td>Real</td>
</tr>
<tr>
<td>sound7</td>
<td>small-bouncing-wooden-ball-1-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound8</td>
<td>small-bouncing-wooden-ball-2-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound9</td>
<td>small-bouncing-wooden-ball-3-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound10</td>
<td>small-bouncing-wooden-ball-4-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound11</td>
<td>small-bouncing-wooden-ball-5-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound12</td>
<td>small-bouncing-wooden-ball-6-pd.wav</td>
<td>Sound Object - 1 Mode</td>
</tr>
<tr>
<td>sound13</td>
<td>w12-h10-pd.wav</td>
<td>Sound Object - 2 Modes</td>
</tr>
<tr>
<td>sound14</td>
<td>w12-h20-pd.wav</td>
<td>Sound Object - 2 Modes</td>
</tr>
<tr>
<td>sound15</td>
<td>w12-h40-pd.wav</td>
<td>Sound Object - 2 Modes</td>
</tr>
<tr>
<td>sound16</td>
<td>w24-h10-pd.wav</td>
<td>Sound Object - 2 Modes</td>
</tr>
<tr>
<td>sound17</td>
<td>w24-h20-pd.wav</td>
<td>Sound Object - 2 Modes</td>
</tr>
<tr>
<td>sound18</td>
<td>w24-h40-pd.wav</td>
<td>Sound Object - 2 Modes</td>
</tr>
</tbody>
</table>

Table 4.4: List of the stimuli used in the experiment.

<table>
<thead>
<tr>
<th>Short Name</th>
<th>elasticity</th>
<th>damping</th>
<th>gravity force</th>
<th>strike velocity</th>
<th>freq (Hz)</th>
<th>decay time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sound7</td>
<td>15000</td>
<td>46.4159</td>
<td>450</td>
<td>630.957</td>
<td>1758.52</td>
<td>0.233307</td>
</tr>
<tr>
<td>sound8</td>
<td>15000</td>
<td>63.0957</td>
<td>940</td>
<td>912.011</td>
<td>1113.23</td>
<td>0.603386</td>
</tr>
<tr>
<td>sound9</td>
<td>11395</td>
<td>2.92864</td>
<td>860</td>
<td>301.995</td>
<td>1294.33</td>
<td>0.752992</td>
</tr>
<tr>
<td>sound10</td>
<td>1309.5</td>
<td>4.64159</td>
<td>970</td>
<td>436.516</td>
<td>1294.33</td>
<td>0.784488</td>
</tr>
<tr>
<td>sound11</td>
<td>1309.5</td>
<td>8.57696</td>
<td>990</td>
<td>1318.26</td>
<td>1254.95</td>
<td>0.784488</td>
</tr>
<tr>
<td>sound12</td>
<td>3162.28</td>
<td>25.1189</td>
<td>900</td>
<td>524.807</td>
<td>1322.83</td>
<td>0.233307</td>
</tr>
</tbody>
</table>

Table 4.5: Values of the parameters for the synthesized sounds with the first approach.

<table>
<thead>
<tr>
<th>Short Name</th>
<th>hammer mass</th>
<th>initial interval (ms)</th>
<th>ac/deceleration</th>
<th>initial value</th>
</tr>
</thead>
<tbody>
<tr>
<td>sound13</td>
<td>0.0215443</td>
<td>228.530</td>
<td>0.76</td>
<td>0.56</td>
</tr>
<tr>
<td>sound14</td>
<td>0.0215443</td>
<td>306.516</td>
<td>0.74</td>
<td>0.75</td>
</tr>
<tr>
<td>sound15</td>
<td>0.0398107</td>
<td>207.223</td>
<td>0.72</td>
<td>0.57</td>
</tr>
<tr>
<td>sound16</td>
<td>0.0398107</td>
<td>207.223</td>
<td>0.72</td>
<td>0.57</td>
</tr>
<tr>
<td>sound17</td>
<td>0.0398107</td>
<td>277.939</td>
<td>0.70</td>
<td>0.75</td>
</tr>
<tr>
<td>sound18</td>
<td>0.0398107</td>
<td>372.786</td>
<td>0.70</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 4.6: Values of the parameters for the synthesized sounds with the second approach.
sounds that they thought were unrealistic, and at the end, the participants were asked to fill out a questionnaire.

As in the pilot probe, the ranges of perceptual evaluations were relative to each user. We decided to normalize the data coordinates between 0 and 1, for comparing the results of each participant.

**Results and Observations.** In fig. 4.21 we report the representation of the individual perceptual scaling and tagging information sorted by users. As for the pilot experiment and also in this case, we can see the classification by sound groups. Moreover, we notice that two of the participants (user n.1 and user n.2) only performed minor judgements on size of the real sounds. They referred, in fact, that they perceived other parameters changing, such as distance and material. This complex influence of the three parameters has been already discussed by Burro [12].

In fig. 4.22 we represent the perceptual scaling and tagging information sorted by stimuli. We can see a different approach of the users to the two groups. In fact, with the sounds synthesized by the first model, it seems that the users are much able to estimate the two parameters, since four of the six sounds (sound7, sound10, sound11 and sound14) are estimated coherently by the most of the participants. On the contrary, the sounds designed with the second model are not clearly estimated and provide a spread of answers across the participants. A possible explanation of this spread of estimations may be due to the presence of a “buzz” tail, that conveys an unnatural perception of the event.

It is interesting to observe the single stimuli, as we did for the pilot probe, looking at fig. 4.23 which represent, through a box plot, the individual perceptual scaling of height and size respectively for each stimuli. As with the previous experiment we again have focused on the synthesized sounds.

We can observe that there is more uniformity in perceptual scaling in two dimensions, than in the pilot experiment. For instance, the stimuli small-
bouncing-wooden-ball-1-pd, sound7, and small-wooden-ball-5-pd, sound11, have a strong uniformity in both the dimensions, despite an outlier for small-bouncing-wooden-ball-1-pd, sound7, concerning the perception of its height. In fig. 4.24 and fig. 4.25, we report their individual box plots.

Considering another two stimuli, i.e. small-bouncing-wooden-ball-4-pd, sound10, and w12-h20-pd, sound14, (fig. 4.26 and fig. 4.27), we can see that there is a slightly uniformity in either dimension.

Four of the stimuli were perceived uniformly in one dimension, despite the presence of one outliers in most of the cases. Specifically, small-bouncing-wooden-ball-3-pd, sound9, and w12-h40-pd, sound15, where judged uniform in the size dimension, and small-bouncing-wooden-ball-6-pd, sound12, and w12-h10-pd, sound13, where judged uniform in the height dimension.

There were three stimuli with dispersed perceptions of their estimations (small-bouncing-wooden-ball-2-pd, w24-h10-pd and w24-h40-pd, i.e. sound8, sound16 and sound18) and a highly dispersed perception of one sound in particular (w24-h20-pd, sound17), whose individual box plot is represented in fig. 4.28.

Contrary to the results of the tagging task in the pilot probe, there is no stimuli, in this experiment, that was tagged by all the participants. The maximum users consensus, regarding unrealistic stimuli, was achieved by 3 users. The real sounds were perceived again as realistic, duplicating the results of our pilot study.

At the end of each session as part of the participant debriefing, a questionnaire was presented to the participants in order to gain an insight into their perceptions of the performed tasks and into the Sonic Browser.

In the debriefing phase, a seven point Likert scale questionnaire with six sets of semantic differentials was filled out by the participants who were asked to express their responses to the interfaces and to the tasks. (from 0 to 6, where 0 is “poor” and 6 is “excellent”). In fig. 4.29 and in fig. 4.30 the results of the questionnaire with cumulative participant responses displayed per question can
be seen with 0 representing a negative result to the question and 6 represent a positive result to the question. In fig. 4.30 three additional questions were added to the questionnaire after the pilot probe and these asked about the learnability, interpretation of the application as it applied to the task and the difficulty in replaying the last sound.

Question one deals with the perceived difficulty in performing the task by the participant and the results show that it was found to be a non trivial task. Question two deals with the ease of use of the Sonic Browser for these tasks which was found have only an average ease of use. Questions three and four dealt with the quality and realism of the sounds. The results from these questions show that sounds were found to be of a high quality but that did not seem particularly realistic by the participants. This can be attributed to the inclusion of two different types of sound objects containing either one or two modes as well as the lack of room acoustics within the sound object sounds and the presence of a “buzz tail” at the end of the two mode sound object sounds. Questions five concerned a technical issue which aroused from the piloting phase of this experiment. There is a slight delay of up to .3 of a second when playing an audio
Figure 4.23: Representation, by a box plot, of the perceptual scaling of the height and size sorted by stimuli.
etition as characteristic of height” was found to be helpful in classifying the height of a sound. Another common problem was that the “metallic zips distracts/confuses the classification of sounds”, this refers to ending of each of the two mode sounds. Another issue illustrated by participants was that a “detailed comparison without reference points is very difficult and would be much easier with only a single scale” and this illustrates the cognitive load of scaling the sounds within a bi-dimensional space. The aura was found to be particular useful as “it allows me to see which is higher or which is lower by using pitch. The aura now gives me a comparison for similar sounds”. Another important issue highlighted by participants was that the sound collection consisted of “three divisions (small, medium, large) and that it was very hard to compare between divisions but it was easy to compare within a division”. The divisions refer to the three types of sounds within the space, real sounds, one mode sound objects and two mode sound objects. The participants also spoke about the different materials and surfaces as they found that the “different surfaces are very noticeable”.

One participant (user n.3) performed the task in a very short period compared to the other participants. This compared with other participants who found that “the longer I spent working with the sounds, the more difficult it
was to sort them”. This relates to a greater working knowledge of the sound collection and the difficulty in maintaining a consistent scale across multiple sounds. By concentrating on an initial reaction with a continuous exploration and classification of the sound collection it is possible to complete the scaling very quickly but the results showed that quality of the results were only of average quality compared to the other participants as shown in fig. 4.21.

4.5 Cataloguing experiment using the Sonic Browser

In the cataloguing scenario we conducted an experiment with a set of tasks to explore common issues arising from the browsing and management of sound object collections. The aim of the experiment is to investigate the usability of the Sonic Browser as a suitable tool for sound object collections and to gain user feedback with regarding the interface which can suggest future improvements or illuminate issues arising from this type of scenario. Specifically this experiment was an exploratory probe designed to further our understanding of managing a large sound collection of sound objects and sound files as well to help elaborate upon suitable visual and spatial organisations for this type of scenario. In
Figure 4.26: Representation, by a box plot, of the perceptual scaling of the height and size for stimulus n.10. Slightly uniform perceptual scaling in both the dimensions.

In particular, we collected formative data relevant to the understanding of auditory browsing and participant filtering of sounds.

The experiment was preceded by a pilot experiment of the tasks with two participants. The pilot probe allowed for a preliminary observation of the issues arising from the different type of tasks and it highlighted which types and sound categories were most suitable to focus on in the experiment.

The experiment used the Sonic Browser with three visualisation mechanisms for the visual representations of the sound collection. The three types of visualisation mechanisms used were those of a HyperTree, a TouchGraph and a SObGrid (Starfield display).

A similar experimental setup to the psychophysical experiment was used with both video capture of participants actions and application data logging. The video capture was supplemented by active participant feedback gathered by the Thinking Aloud Protocol as previously described in the validation scenario. This type of feedback and commentary from the participants allowed for a greater insight into their perspective of both the tasks and issues encountered within the exploratory probe. The qualitative analysis technique used on the video was a critical incident analysis [48]. The video analysis when used in conjunction with the data logging provides a clear image as to the user actions and intentions in
Figure 4.27: Representation, by a box plot, of the perceptual scaling of the height and size for stimulus n.14. Slightly uniform perceptual scaling in both the dimensions.

the context of the scenario and each particular task.

In the next section, we will present the cataloguing experiment, introducing both the experiment and the results.

4.5.1 The experiment

Participants. The pilot experiment involved 6 volunteers, all students at the Computer Science Department of the University of Limerick. Five of the participants referred to have musical training in average of 6 years, with a minimum of 4 and a maximum of 12 years. Two of the participants require glasses for reading and one participant reported to have hearing problems with very low tones.

Stimuli. The stimuli set included 57 recorded sounds and 10 Sound Objects, designed with PD-modules modelling impact interactions of two modal resonators [57], simplified so as to return only one mode. The recorded sounds used in this experiment were drawn from eight sources, seven commercial sound effects CD’s and a local collection of ecological sounds. The length of the sounds varied from 0.1 to 50 seconds. The synthesized sounds used consisted of the same synthesized sounds used in the pilot probe of the validation scenario.
Figure 4.28: Representation, by a box plot, of the perceptual scaling of the height and size for stimulus n.17. Highly Spread perceptual scaling.

Figure 4.29: Results of the questionnaire for the main experiment.
Figure 4.30: Results of the questionnaire for the main experiment. These are the data concerning additional questions after the pilot probe.

**Procedure.** The experiment was conducted in the isolation room of the Computer Science Department at UL. The sound collection was browsed using stereo headphones and via the Sonic Browser. The experiment was conducted applying the Thinking-Aloud Protocol and the participants sessions were videotaped.

The users were asked to browse the sound collection for a selection of sounds matching a particular property or object. The tasks included searching for a specific sound such as the ‘cry of a seagull’ and to broader categories such as find all the sounds of ‘cats meowing’. In each specific task, the participants were allowed to move the cursor around freely in the GUI trying to find target sounds as well as change the current visualisation at will to compare relationships between sounds within different visualisations. Overall, for the eight auditory tasks, several interesting browsing behaviours were observed. As part of each task, the participants were asked to tag the sounds that they thought fulfilled the criteria of the task.

At the end of each session as part of the participant debriefing, a questionnaire was presented to the participants in order to gain an insight into their perceptions of the performed tasks and into the Sonic Browser.

**Results and Observations.** In the debriefing phase, a seven point Likert scale questionnaire with six sets of semantic differentials was filled out by the participants who were asked to express their responses to the interfaces and to the tasks. (from 0 to 6, where 0 is “poor” and 6 is “excellent”). In fig. 4.31 the results of the questionnaire with cumulative participant responses displayed per question can be seen with 0 representing a negative result to the question and 6 represent a positive result to the question.

Questions one, two and seven deal with aesthetics, interpretation and learnability of the Sonic Browser. The results of these questions show that the users’ find the Sonic Browser easy to learn and use. Questions three to six deal with
the filtering mechanisms of the Sonic Browser. The results of these questions, confirmed by video analysis, illuminate several items such as it is always easier to find a sound when you know its filename and that the filtering mechanisms were found to be easy to use. Question eight concerns a technical issue which arose from the piloting phase of this experiment. There is a slight delay of up to .3 of a second when playing an audio file with the Sonic Browser. The results of this question show that in a cataloguing scenario this delay was not appreciable and did not affect the task. This juxtaposes with the results of this question in the validation scenario which finds that participants had a very noticeable appreciation of the delay. This allows us to say that for tasks involving many sound-to-sound comparisons, play delay of the sound should be kept to an absolute minimum but in a cataloguing scenario that while this is still an important factor a greater play delay is acceptable. Questions nine and ten deal with realism and quality of the sounds which were found to be excellent by participants.

The rich verbal protocol returned several interesting results during the experiment. The play delay was only highlighted by one participant who “expected sound to be instantaneous, not take ages” but this was actually related to the issues of silence at the start of the sound or a sound beginning with a very low volume. The HyperTree visualisation was found to be preferred by half of the participants. Other issues were discovered in the debriefing and through user comments during testing, mostly relate to future improvements of the Sonic Browser such as the “Text lookahead should look for ‘Flushing toilet’ as well as ‘Toilet flushing’ ” and the addition of a “Right click context menu with options for setting object properties and for group of objects properties”. 

Figure 4.31: Results of the questionnaire for the cataloguing experiment.
4.6 Conclusions

Examining the results of our validation scenario we can state that sound convey information about dimensions even if they have only one mode. Apart from one case in the pilot probe, the unrealistic perception of sounds did not affect the participant’s perception of the sound’s dimensions. This illustrates that the "realism" of a sound does not affect the amount of information extracted by a participant. Our studies have shown the difficulty in conveying information in more than one dimension, which is similar to the difficulty encountered with the parameterising of auditory icons [26, 27]. A similar result was shown by Braida’s [9] work on multidimensional vibrotactile stimulus which illustrated that when two dimensions must both be identified, that the sensitivity for a particular stimulus is often reduced. Another possible area of difficulty may be due to the orthogonality [35] wherein a change in one parameter of the auditory stream may cause a perception change in another variable. This can arise from a change in pitch or loudness of one of the sound which then affects the perception of the other sounds. Perception distortions can easily be affected by changes to low level acoustic dimensions such as frequency or intensity as discussed by Neuhoff et al [53]. The differing results between the two modelling approaches of sound objects highlight the need for a further investigation of these sound objects exploring any possible perceptual distortions. The results of our study show that unrealistic synthetic sounds can be recognized as unrealistic events but that their high-level parameters can still be extracted and evaluated. This again highlights the technique of sound cartoonification which caricatures certain aspects of a sound event while discarding other aspects. Rath et al [56] have already discussed this technique in greater detail with a focus on sound objects in combination with cartoonification.

Another interesting finding was that no sounds were deemed unrealistic by all participants and also that none of the real sounds were selected as being unrealistic. There are several possible explanations for these findings. Our experimental analysis and our user debriefing suggest three aspects which should be further investigated. The first aspect is the inclusion of “room acoustics” and the necessary elements of reverberation within a sound object to allow for a more natural sounding event. The second facet is the material and surface perception by participants which should be further examined as the participants stated that the “different surfaces are very noticeable”. The third area is the distractors found within the sounds used in the experiments. The distractors are split into two issues. Firstly, the how the participants related the speed and temporal pattern of the bouncing to the realism of the sound. The second is the “metallic zips” occurring at the end of each of the two mode sound objects. These distractors illustrate the need for a further refinement of the perceptual parameters within the sound models to prevent user confusion when judging a sound’s properties. Further experiments into the perception of object elasticity and other physical parameters should also be investigated for a greater understanding of the perceptual scaling of Sound Objects.

The cataloguing scenario results allow us to assert that the management of a sound collection is a difficult task but it can be made easier through the use of dynamic filtering combined with both direct manipulation and direct sonification of the sounds within the collection. The Sonic Browser has been successful in providing a new interface for sound collections which allows for
the exploration of a sound collection without the users feeling lost or confused. We plan to continue to enhance the Sonic Browser system building upon our experiences to date as well as those of similar projects. The system and its source will made available as open-source software and its development will continue. The next stage of this development will be a exploration of the Sonic Browser as a tool for accessing network collections of sound objects, in parallel with the development of greater linkages between PD and the creation of linkage with other audio authoring tools such as open-source tools like Audacity or professional tools like SoundForge, Cubase VST or Protools. Even without these planned enhancements, we believe our approach marks a major step forward in audio browsing interfaces and can serve as a model to others implementing similar systems.
Chapter 5

Multimodal Perception of Model-based Rolling

Roberto Bresin and Kjetil Falkenberg Hansen

5.1 Introduction

“Rolling” sounds form a category that seems to be characteristic from the auditory viewpoint. Everyday experience tells that the sound produced by a rolling object is often recognizable as such, and in general clearly distinct from sounds of slipping, sliding or scratching interactions, even of the same objects.

Prior experiments showed that listeners can discriminate differences in size and speed of wooden rolling balls on the basis of recorded sounds [51]. Results from perceptual experiments, performed in the framework of the Sounding Object Project [13], demonstrated that listeners are able to perceive the size and weight of a ball under different conditions. These conditions were: (1) steel ball falling on a wooden plate, (2) wooden ball falling on a ceramic plate.

5.2 The Invisiball: controlling the sound of an invisible ball

The impact model implemented in the framework of the Sounding Object Project [57] was controlled with the mechanics equation of a rolling ball. The sound model was controlled by simply feeding it with the X, Y, and Z coordinates of a target position in a 3-dimensional space.

In the demonstration set-up in the experiment presented in this paper, the space is given by the dimensions of an haptic interface that was constructed for controlling the Max Mathews’ Radio Baton [7]. A radio signal is sent from a fingertip. For this purpose a “finger-sender” device was constructed (see Figure 5.1). The interface is made of stretching material and it is placed over the receiving antenna. Finger position in the 3-dimensional space is detected

1Department of Speech, Music, and Hearing, Royal Institute of Technology, Stockholm
in real-time and it is feeded into the algorithm controlling the rolling movement of a ball. By pushing the membrane with the “finger-sender”, users can make the ball rolling towards the finger by moving it from rest position in the 3-dimensional space (see Figure 5.2). The position of the rolling ball as a projection on the XY plane, i.e. as seen from above, is visualized on the computer screen. The position is represented as a colored disk assuming colors in the red-range at high speed (hot ball) and blue-range at low speed (cold ball).

This new interface allows users an interaction by using three different types of feedback:

- Acoustic: the sound model of the rolling ball
- Haptic: control of the position of the ball by pressing the elastic membrane with a finger
- Visual: graphical projection of the position of the ball

Figure 5.1: The “finger-sender”: a finger-based controller for sending a radio signal to the receiving antenna.

Figure 5.2: Haptic interface for the control of the Max Mathews’ Radio Baton. The interface is placed over the receiving antenna. Finger position in the 3-dimensional space is detected in real-time.
5.3 Pilot experiment: multimodal perception test

The interface described in the previous section was used to test the quality of the sound model from a perceptual point of view. In particular we were interested in how realistic is the sound of the rolling ball controlled with this interface. Since the interface allows three different kinds of feedback – acoustic, haptic and visual – three different experiments were conducted. The three experiments were run in parallel at the same time so that 3 subjects at a time could listen to exactly the same sound, as we will explain in the following.

5.3.1 Subjects and procedure

The subjects were twelve, 6 females and 6 males. Their average age was 30. The subjects all worked at the Speech Music Hearing Department of KTH, Stockholm.

Subjects listened to the examples individually over infrared headphones adjusted to a comfortable level. In this way three subjects at a time could take part each to a different experiment without seeing each other. Each subject was instructed to estimate, for each example, its degree of realism. The responses were recorded on a scale on paper, from 0 to 10, with “unrealistic” and “realistic” as extremes. Stimuli were presented twice in a random order.

5.3.2 Experiment 1: haptic and acoustic feedback

Subjects had an acoustic feedback through headphones and an haptic feedback from the finger-controlled interface presented in the previous section.

Stimuli

Nine sound model set-ups were used. They were obtained by combining 3 sizes of the ball with 3 damping values, thus producing 9 different set-ups. In this way the influence of both size and damping could be tested on the classification of the corresponding sounds stimuli as being realistic. These 9 set-ups were presented twice to the subjects. This correspond to a factorial design (3 sizes) x (3 damping) x (2 repetitions).

By controlling the haptic interface with the “finger controller” subjects produced 9 acoustic stimuli with 2 repetitions, for a total of 18 stimuli, of the duration of about 20 seconds each.

The sound of each stimulus was that of a rolling ball.

5.3.3 Experiment 2: visual and acoustic feedback

In this experiment subjects had an acoustic feedback through headphones and a visual feedback from a computer screen.

Stimuli

The acoustic stimuli were those produced at the same time by the subject controlling the haptic controller in Experiment 1. In addition, a visual stimulus was synchronized with the acoustic stimuli. The visual feedback was presented
on a computer screen and it represented the real-time moving position of the rolling ball in the 2-dimensional space.

5.3.4 Experiment 3: acoustic feedback only

In this third experiment subjects had only an acoustic feedback through headphones.

Stimuli

The acoustic stimuli were the same as those in Experiment 1 and Experiment 2. They were produced at the same time by the subject controlling the haptic controller in Experiment 1.

5.3.5 Results and discussion

In the following a preliminary analysis of the results from the three experiments is presented.

A repeated measurements ANOVA was conducted on answers collected in Experiment 1. There was no significant effect of neither size nor damping parameters. However it can be observed that subjects classified as more realistic sound stimuli with low and medium damping produced by the sound model of a rolling ball of medium size (see Figure 5.3). A high within-subjects variability emerged for stimuli obtained with a medium value of the damping factor. Values ranging from 0 to 8 were recorded.

![Figure 5.3: Effect of damping and size as resulted from the analysis of the responses in experiment 1.](image)

A repeated measurements ANOVA was conducted on answers collected in Experiment 2. There was no significant effect of neither size nor damping parameters. Nevertheless a closer observation of the results suggests that subjects tended classifying as more realistic sound stimuli associated to low and medium
damping and to large size of the rolling ball (see Figure 5.4). Also, subjects gave a higher preference rate to the most realistic stimuli as compared with best ratings given by subjects in Experiment 1. A high within-subjects variability emerged for stimuli obtained with a large value of the damping factor. Ratings ranging from 0 to 7 were recorded.

A repeated measurements ANOVA was conducted on answers collected in Experiment 3. There was no significant effect of neither size nor damping parameters. A preliminary observation of the results suggests that subjects with only the acoustic feedback tended classifying as more realistic sound stimuli associated to low and medium damping and large size of the rolling ball (see Figure 5.5). Subjects gave a higher preference rate to the most realistic stimuli as compared with best ratings given by subjects in Experiment 1. These results are comparable to those obtained in Experiment 2. A high within-subjects variability emerged for stimuli obtained with a small value of the damping factor and large size of the ball. Values ranging from 0 to 9 were recorded.

A repeated measurements ANOVA was conducted on all answers collected in the three experiments. Main results were a significant effect of the ball size parameter, $F(2, 22)=6.6175$, $p=.00562$, and a significant effect of the damping factor of the sound model, $F(2, 22)=5.5417$, $p=.01124$ (see Figures 5.6 and 5.7). There was no significant interaction between the size of the ball and the damping parameter.

The average rating for each feedback modality across all stimuli is presented in Figure 5.8. The “acoustic & haptic” modality, Experiment 1, resulted as worst and the “acoustic & visual”, Experiment 2, was classified as best, according to ratings given by the subjects in all three experiments. There is no significant difference between average ratings of the “acoustic” modality and of the “acoustic & visual” modality.

Responses to stimuli can be averaged through all three experiments, as shown in Figure 5.9. It can be observed as, in all damping conditions, stimuli with
small size are classified as less realistic. In general, stimuli with both medium size and medium damping were classified as more realistic.

5.4 Conclusions

In this work we proposed a new device for controlling physics-based sound models through direct manipulation of an haptic interface, the finger control radio baton. The “realistic” property of sounds produced by acting on the interfaced was analyzed in three experiments.

Main result was that sound stimuli corresponding to balls with large and medium size and low and medium damping were classified as more realistic.

At our knowledge, this is the first time that an haptic controller producing only an audio feedback is tested. This interface allows a gestural control of a sounding object in a 3-dimensional space by direct manipulation.

As an overall result, sound stimuli were classified as more realistic by subjects using only “acoustic” feedback or “acoustic and visual” feedback. It seems that this is due to the difficulty in controlling the haptic interface and the sound metaphor associated to it. Some of the subjects reported that it was difficult to imagine the ball rolling towards their finger. Nevertheless some of the subjects in Experiment 1 gave high rates to stimuli corresponding to balls of medium size. These results suggest that the haptic controller and/or the testing application can be better designed.
Figure 5.6: Effect of size obtained by averaging across all three experiments.

Figure 5.7: Effect of damping obtained by averaging across all three experiments.
Figure 5.8: Effect of modality on the classification of the sound stimuli in the “realistic” scale.

Figure 5.9: Average classification values for each sound stimulus used in the multimodal experiments.
Bibliography


86


89


