Università degli Studi di Verona
Dipartimento di Informatica

DOTTORATO DI RICERCA IN INFORMATICA
CICLO XVI

Representation of multi-dimensional data through auditory display

Laura Ottaviani

Coordinatore: Prof. Andrea Masini
Supervisore: Prof. Davide Rocchesso

Ai miei genitori e a Claudio

Abstract

We aim at investigating the representation of multi-dimensional data through Auditory Display. As pointed out by Kubovy and Van Valkenburg [KV01], while the visual system deals with surfaces that reflect light, the auditory system deals with sources of sounds and auditory events. If we want to investigate perceptual dimensions that are new to auditory display, we should focus on perceptual features of sound sources and of auditory events, rather than on physical properties of the sound signals per se. Sound sources and auditory events have been recently considered effective in designing auditory display [HR99, HHR01].

In this research work we study, with an ecological perspective, some perceptual properties, that were little or even never considered in the auditory display field. Our purpose is to base the investigations on the ecological approach, trying to find features useful for creating a "natural auditory environment" for the users of the auditory representation. In fact, we think that it is better and more involving for the users to work with auditory representations which use everyday sounds rather then unnatural sounds, because they could be less distracted and less irritated. Moreover, by using everyday sounds we think that the auditory display could be more general and therefore applicable to a variety of situations. On the contrary, if an application uses abstract sounds, the mapping design should be more related to that precise data set in order to have a metaphorical relationship with the data it represents.

We decide to use synthetic models of the sound sources and of the sound events/processes we want to investigate. An experimenter, by using physical models for generating the stimuli used in the experimental framework, is able to control directly each physical attribute [LCS00]. Since all these investigations are conducted for being, in the future, applied to auditory displays or multimodal applications which could enjoy their benefits, it could be more flexible to include sound models inside applications rather than pre-recorded sounds and, in particular, by using the sound models developed by the the EU-funded project "the Sounding Object" (SOb)[1], it could be computationally convenient. The aim of our dissertation is to investigate the perceptual features of sound sources and of auditory events/processes synthesized with these models able to convey information, and in particular quantifiable information, in order to auditorilly represent multi-dimensional data. Sound sources and auditory events have been recently considered effective in designing auditory display [HR99, HHR01].

[1] `http://www.soundobject.org`

In the dissertation, we present our studies and the experiments conducted, including test procedures and results, concerning perceptual dimensions of both sound sources and sound events, by highlighting reasons, aims and results obtained. We see if these features could be used for providing information and if they are ecologically relevant. We check if they can be noticed by human listeners, if they make objects and events recognizable and if they can provide quantifiable information.

In particular, we present some perceptual experiments we conducted by means of simulated 3-D resonators with cubic and spherical shape in order to study whether some object dimensions, that are commonly "seen", can auditorilly provide information. The investigated features, i.e. the size of the resonators and their shape, resulted to be usable only with strict requirements. From the results of these experiments, an interesting perceptual effect, that is the order effect, has been found. It affects the listening attitude of the listeners and it can cause performance gaps between users.

Then, some investigations about sound sources and auditory events are presented. They focus on some perceptual dimensions of two sound event categories that are common in the everyday auditory environment, i.e. the impact/bounce event and the rolling process. Even if there is a complex relationship between the perceptual space and the parameters space, the investigations highlight the perceptual dimensions that "pop-out" more or, at least, they find how to get more uniform perceptual scaling by means of a proper parameterization. Some experimental methods for analyzing the collected data of perceptual scaling experiments are presented. They focus on finding useful representations and summary plots of the complex data gathered with the experiments for comparing physical parameters values of sound models and the scaling of perceptual parameters of the sound events they synthesize.

Sommario

Il nostro obiettivo è lo studio della reppresentazione di dati multi-dimensionali attraverso l'Auditory Display. Come messo in luce da Kubovy and Van Valkenburg [KV01], mentre il sistema visivo si occupa di superfici che riflettono la luce, il sistema uditivo si occupa di sorgenti di suoni e di eventi uditivi. Se noi vogliamo occuparci di dimensioni percettive che siano nuove per l'auditory display, dovremmo focalizzare la nostra attenzione su caratteristiche percettive di sorgenti di suono e di eventi uditivi, piuttosto che su proprietà fisiche dei segnali sonori in sè. Sorgenti di suono ed eventi uditivi sono stati recentemente considerati efficaci per lo sviluppo di auditory display [HR99, HHR01].

In questo lavoro di ricerca studiamo, dal punto di vista ecologico, alcune proprietà percettive, che furono poco o persino mai considerate nel campo dell' auditory display. Il nostro scopo consiste nel basare la nostra ricerca sull'approccio ecologico, cercando di trovare caratteristiche utili per creare un "ambiente uditivo naturale" per l'utente della rappresentazione uditiva. Infatti, noi pensiamo che sia meglio e più coinvolgente per gli utenti lavorare con rappresentazioni uditive che usino "everyday sounds" piuttosto che suoni non naturali, poiché potrebbero risultare meno distratti e meno irritati da tale tipo di suoni. Inoltre, usando everyday sounds pensiamo che l'auditory display possa essere più generale e perciò applicabile ad un maggior numero di situazioni. Al contrario, se una applicazione usa suoni astratti, la progettazione dell'auditory display, per avere una relazione metaforica con i dati che esso rappresenta, dovrebbe essere più legata a quel preciso insieme di dati.

Decidiamo di usare modelli sintetici delle sorgenti di suono e degli eventi/processi uditivi che vogliamo investigare. Uno sperimentatore, usando i modelli fisici per generare gli stimoli usati nell'ambito dei suoi esperimenti è in grado di controllare direttamente ogni attributo fisico [LCS00]. Dal momento che tutti questi studi sono condotti per essere in futuro applicati ad auditory display e applicazioni multimodali che possano godere dei loro vantaggi, potrebbe essere più flessibile inserire modelli di suoni all'interno delle applicazioni piuttosto che suoni pre-registrati e, in particolare, usare i modelli di suono sviluppati del progetto Europeo "the Sounding Object" (SOb)[2] potrebbe essere conveniente dal punto di vista computazionale. Lo scopo della nostra tesi è studiare le caratteristiche percettive di sorgenti di suono e di eventi/processi uditivi sintetizzati con questi modelli in grado di fornire informazione, e in particolare informazione di tipo quantitativo, in modo da rappresentare

[2]http://www.soundobject.org

dati multi-dimensionali tramite il canale uditivo. Sorgenti di suono ed eventi uditivi sono stati recentemente considerati efficaci nello sviluppo di auditori diplay [HR99, HHR01].

Nella tesi presentiamo i nostri studi e gli esperimenti condotti, inclusi procedure dei test e risultati, riguardanti dimensioni percettive sia di sorgenti di suoni che di eventi sonori, evidenziando ragioni, obiettivi e risultati ottenuti. Vediamo se tali caratteristiche potrebbero essere utili per fornire informazioni e se sono ecologicamente rilevanti. Proviamo a dimostrare se possono essere notate da un ascoltatore, se rendono oggetti ed eventi riconoscibili e se possono fornire informazioni di carattere quantitativo.

In particolare, presentiamo alcuni esperimenti percettivi che abbiamo condotto utilizzando risonatori 3-D sintetici con forma di cubica e sferica, per studiare se alcune dimensioni di oggetti, che di solito sono percepite visivamente, possono fornire informazione anche tramite il canale uditivo. Le caratteristiche investigate, cioè la dimensione dei risonatori e la loro forma, sono risultate essere utilizzabili solo a condizioni limitate. Dai risultati di tali esperimenti, è emerso un interessante effetto percettivo, che è l'effetto d'ordine ("order effect"). Esso influenza la modalità di ascolto degli ascoltatori e può causare delle differenze tra le prestazioni dei vari utenti.

Poi, vengono presentate alcuni studi riguardo sorgenti di suono ed eventi uditivi. Essi si focalizzano su alcune dimensioni percettive di due categorie di eventi sonori che sono comuni nell'ambiente uditivo di tutti i giorni, ossia l'evento impatto/rimbalzo e il processo di rotolamento. Anche se esiste una relazione complessa tra lo spazio percettivo e lo spazio dei parametri, gli esperimenti evidenziano le dimensioni percettive che "risaltano" ("pop-out") maggiormente, o, almeno, trovano come sia possibile ottenere uno scaling percettivo più uniforme tramite una parametrizzazione appropriata. Vengono presentati alcuni metodi sperimentali per analizzare i dati raccolti dagli esperimenti di scaling percettivo. Essi puntano sul trovare delle rappresentazioni e dei grafici riassuntivi utili dei dati complessi raccolti tramite gli esperimenti per confrontare i valori dei parametri fisici dei modelli di suono e lo scaling dei parametri percettivi degli eventi di suono che essi sintetizzano.

Acknowledgments

I would like to take the opportunity of this dissertation for remembering and thanking all the people who helped me in the shaping of my personality, and who, with their friendship, stayed at my side and from whom I had the chance to learn so much!

First of all, I would like to say “Thank you!” to my parents. With them I share all the beautiful and pleasant moments, they give me always precious advice and they support me in every difficulties. Above all, without them I would not be as I am: I would not be the sunny person that I am and I would not have so many interests as I always have. I cannot forget the role that the interest for music played in my life: my parents patiently helped me, when I was a child, in learning to love music and developing my attitude.

I would like also to thank my uncle who was the first to introduce me to the Computer and who made me love mathematics. I thank my aunt and my uncle, as well, who are always happy to help me and to give me hospitality...and not only in Milano: I spent very pleasant holidays with them! And I cannot forget to mention my relatives who always show me all their love and affection.

There are many people that I cannot forget because from them I learnt a lot.

It was a great opportunity for me to have the chance to be a member of the VIPS lab, because during our weekly meetings I could listen interesting debates not only about sound processing but also about image processing, that is a field which always fascinated me. For this reason I have to express my gratefulness to Vittorio Murino, Andrea Fusiello, Manuele Bicego, Umberto Castellani, Marco Cristani, Andrea Colombari.

I thank the student Germana Olivieri for the interesting results achieved by her studies. It is interesting to supervise her thesis.

To my colleagues Debora, Fausto, Giuditta, Graziano, Arnaud, Isabella, Rosalba, Barbara, Alessandro, Nicola, Francesca, I wish their luck for their future.

I am grateful to Stefano De Marchi and Paolo Fiorini for the chance they gave me in studying with them interesting topics.

The five months I spent at the Interaction Design Centre of the University of Limerick were great! It was a very interesting experience for me both from the research point of view, because I had the opportunity to work in a different research group and to learn a lot, and from the personal point of view. I have met many friends there. Eoin Brazil, he was always there for me and with him I worked very well. Luigina Ciolfi and Michael Cook...thank you for the enjoyable dinners! Breege Moynihan, Annette Aboulafia, Marilyn Lennon, and

Tony Hall, with them I had pleasant conversations. Mark Marshall and Bruce Richardson, very good friends, with whom I spent nice breakfasts! Enrique Franco and Krispin Leydon, with them I had enjoyable time. Kieran Ferris, Paul Gallagher and Alan Keenan, always ready to have amusing conversations with me. Many thanks to Colm McGettrick who helped me frequently for many technical problems and to Anne Murphy for her kindness and her attentive presence.

A special thank you to Liam Bannon for his hospitality and for everything I had the opportunity to learn from him. I cannot forget Mikael Fernström: the long talks in his office helped me to clarify my ideas and taught me the importance of deepening everything I read and study. I am very pleased to know you!

A great experience and a very good opportunity for me was to collaborate to the EU-funded project "the Sounding Object" (SOb). First of all, I would like to say "Thank you!" to Nicola Bernardini for the privilege of working with you. I would like to thank Giovanni Bruno Vicario: I am honoured to have met him. I cannot forget the cheerful attitude of Roberto Bresin, even during hard work periods and the precious expertise of Gianpaolo Borin.

I would like to thank Roberto Burro, Bruno Giordano and Massimo Grassi for their kindness in giving me hints and advice. A special thank you to Federico Avanzini and Carlo Drioli for the friendship they showed me. I cannot forget Sofia Dahl and Kjetil Falkenberg Hansen for the amusing moments.

Thank you very much to Federico Fontana, who I consider a very good friend. I enjoyed to talk with him both about work and about personal topics. Thanks Federico!

I would like to thank also Matthias Rath. I spent three nice years sharing with him the office and I am happy to have known him! And I cannot forget Amalia De Götzen for her friendship and her kindness.

I am deeply indebted to Davide Rocchesso. I would have never reached this goal without him. He always supported me with his precious advice and he taught me how a good research activity should be conducted. He has been a model for me and I am very grateful for the friendship he showed me.

Finally, I would like to thank my husband Claudio, because with his love and his presence is able to make all my days bright.

Ringraziamenti

Vorrei cogliere l'occasione di questa tesi per ricordare e ringraziare tutte le persone che mi hanno aiutato nel formare la mia personalità e che, con la loro amicizia, sono state al mio fianco e dalle quali ho avuto l'opportunità di imparare così tanto!

Innanzitutto, vorrei dire "Grazie!" ai miei genitori. Con loro condivido tutti i momenti belli e piacevoli, mi danno sempre preziosi consigli e mi sostengono sempre in ogni difficoltà. Sopratutto, senza di loro non sarei così come sono: non sarei la persona solare che sono e non avrei così tanti interessi come sempre ho. Non posso scordare il ruolo che l'interesse per la musica ha occupato durante la mia vita: i miei genitori mi hanno pazientemente aiutato, quando ero bambina, ad imparare ad amare la musica e a sviluppare le mie attitudini.

Vorrei anche ringraziare mio zio che è stato il primo ad introdurmi al Computer e a farmi amare la matematica. Ringrazio anche mia zia e mio zio che sono sempre lieti di aiutarmi e darmi ospitalità...e non solo a Milano: ho trascorso con loro vacanze molto piacevoli! E non posso dimenticare di menzionare i miei parenti che mi dimostrano sempre il loro amore ed affetto.

Ci sono molte persone che non posso dimenticare perché da loro ho imparato molto.

E' stata un'opportunità grandiosa per me avere la possibilità di essere un membro del laboratorio VIPS, perchè durante i nostri incontri settimanali ho potuto ascoltare interessanti dibattiti non solo riguardanti l'elaborazione del suono ma anche riguardanti l'elaborazione delle immagini, campo che mi ha sempre affascinato. Per questa ragione devo esprimere le mia gratitudine a Vittorio Murino, Andrea Fusiello, Manuele Bicego, Umberto Castellani, Marco Cristani, Andrea Colombari.

Ringrazio la studentessa Germana Olivieri per gli interessanti risultati raggiunti dai suoi studi. E' interessante supervisionare la sua tesi.

Ai miei colleghi Debora, Fausto, Giuditta, Graziano, Arnaud, Isabella, Rosalba, Barbara, Alessandro, Nicola, Francesca, auguro buona fortuna per il loro futuro.

Sono grata a Stefano De Marchi e Paolo Fiorini per l'opportunità che mi hanno data nello studiare con loro argomenti interessanti.

I cinque mesi che ho trascorso all'Interaction Design Centre dell'Università di Limerick sono stati grandiosi! E' stata un'esperienza molto interessante per me sia dal punto di vista della ricerca, perché ho avuto l'opportunità di lavorare in un gruppo di ricerca diverso e di imparare molto, che dal punto di vista personale. Ho incontrato molti amici lì. Eoin Brazil, per me era sempre disponibile e con lui ho lavorato molto bene. Luigina Ciolfi e Michael

Cook...grazie per le piacevoli cene! Breege Moynihan, Annette Aboulafia, Marilyn Lennon, e Tony Hall, con loro ho conversato piacevolmente. Mark Marshall e Bruce Richardson, buoni amici, con i quali ho fatto delle buone colazioni! Enrique Franco e Krispin Leydon, con loro ho trascorso divertenti serate. Kieran Ferris, Paul Gallagher e Alan Keenan, sempre pronti ad avere divertenti conversazioni con me. Molte grazie a Colm McGettrick che mi ha aiutato frequentemente per molti problemi tecnici e ad Anne Murphy per la sua gentilezza e la sua presenza attenta ai bisogni e necessità altrui.

Un ringraziamento speciale a Liam Bannon per la sua ospitalità e per tutto ciò che ho avuto l'opportunità di imparare da lui. Non posso dimenticare Mikael Fernström: i lunghi incontri e discussioni nel suo ufficio hanno aiutato a chiarificare le mie idee e mi hanno insegnato l'importanza di approfondire tutto ciò che leggo e studio. Sono molto lieta di conoscerti!

Poter collaborare al progetto Europeo "the Sounding Object" (SOb) è stata per me una grandiosa esperienza e un'opportunità molto buona. Innanzitutto, vorrei dire "Grazie!" a Nicola Bernardini per il privilegio che è stato lavorare con te. Vorrei ringraziare Giovanni Bruno Vicario: sono onorata di averlo conosciuto. Non posso dimenticare l'atteggiamento positivo e allegro di Roberto Bresin, anche durante i periodi di duro lavoro e la preziosa esperienza di Gianpaolo Borin.

Vorrei ringraziare Roberto Burro, Bruno Giordano e Massimo Grassi per la loro gentilezza nel darmi suggerimenti e consigli. Un grazie speciale a Federico Avanzini e Carlo Drioli per l'amicizia che mi hanno manifestato. Non posso dimenticare Sofia Dahl e Kjetil Falkenberg Hansen per i momenti divertenti.

Mille grazie a Federico Fontana, che considero veramente un buon amico. Mi è piaciuto parlare con lui sia di lavoro che di argomenti personali. Grazie Federico!

Vorrei ringraziare anche Matthias Rath. Ho trascorso tre anni piacevoli dividendo con lui l'ufficio e sono contenta di averlo conosciuto! E non posso dimenticare Amalia De Götzen per la sua amicizia e la sua gentilezza.

Sono profondamente grata a Davide Rocchesso. Non avrei mai raggiunto questo traguardo senza di lui. Mi ha sempre sostenuto con i suoi preziosi consigli e mi ha insegnato come deve essere condotta una buona attività di ricerca. E' stato un modello per me e gli sono molto grata per l'amicizia che mi ha manifestato.

Infine, vorrei ringraziare mio marito Claudio, perchè con il suo amore e la sua presenza è in grado di rendere luminose tutte le mie giornate.

Contents

Chapter 1

How can we represent information?

In our daily life, we are surrounded by every kind of information and data and we usually need to represent them in a clearer way, in order to summarize them, to enhance the most relevant facts and to allow people to remember them more easily. There are many examples of data with a completely different nature each one from the others, and, according to the main characteristics of the data, we have to choose the appropriate way to represent them. For instance, if we think of some items that we want to buy, the obvious choice in order to remember them is a shopping list; but if we are investigating some statistical data, it could be better to use graphical plots, like bar charts, pie charts or others, choosing the most suitable plot for that case. In the literature this concept goes under the wider definition of cognitive artifact, which deals not only with data representations but it includes all those "aids" for the human beings to increase their cognitive activities.

In a formal way, *cognition* is the intellectual process through which information is obtained, transformed, stored, retrieved and used. The mind can be considered as an Information Processor. Human beings need external aids or external representations in order to enhance their cognitive abilities, i.e. to increase memory, thought, and reasoning, as pointed out by Norman [Nor93]. He calls these "things that make us smart" with the name *cognitive artifacts*.

> ...we used sounds, gestures, and symbols to refer to objects, things, and concepts. The sound, gesture, or symbol is not the thing itself; rather, it stands for or refers to the thing: It represents it.
>
> The powers of cognition come from abstraction and representation: [...] ...we can make marks or symbols that represent something else and then do our reasoning by using those marks [Nor93].

We could think of cognitive artifacts such as an address book, which provides us the telephone numbers we need, and therefore without having to memorize them, or a sketch map, for instance, to easily explain to someone where is our house.

The way to augment human intellect was studied thirty years before Norman by Engelbart [Eng63] with the focus on investigating a new approach for organizing documents in a hierarchic structure, enhancing the aspects of group creation and problem solving [BGBG95].

> By "augmenting man's intellect" we mean increasing the capability of a man to approach a complex problem situation, gain comprehension to suit his particular needs, and to derive solutions to problems. [...]. Our culture has evolved means for us to organize and utilize our basic capabilities so that we can comprehend truly complex situations and accomplish the process of devising and implementing problem solutions [Eng63].

Engelbart calls *augmentation means* the ways in which human capabilities are extended. But his studies aim at improving a system, an ideal physical system, that could augment people abilities, a sort of ideal-ancestor of the modern digital computer. Therefore, the means he is speaking about are physical means. On the contrary, Norman [Nor93], by cognitive artifacts, refers both to physical and abstract concepts. He defines the *experiential artifacts* as those artifacts which allow the users to experience events as if they were there, for example a telescope, a movie or the petrol gauge in the car, while the *reflective artifacts* as those artifacts allowing the users to ignore the real world and concentrate on artificial worlds, for instance a graph or a scientific essay.

As we already noted, the concept of cognitive artifact is wide. We are interested in the topic of data representation. We are talking about data representation in order to provide information to users. Data is not Information. It is possible to have *data*, but not *information*.

> There are many situations in which data is available, sometimes in very large quantities, and where some human insight into the data is required. [...] They wish to be *informed* by the data. [...] ...it is the derivation of information (or understanding, or insight) from the data that is difficult [Spe01].

As outlined by Alexander in [Ale02], there are three perspectives for defining the terms "data" and "information", i.e. an objective, a subjective and an intersubjective point of view.

For the objective point of view, *data* are factual, resulting from recording of measurable events or objects, and *information* is the result of a computer program, which associates data with other data in structures or which summarizes some data to produce information. The output of the computer program, whose usefulness has been established by the system analyst, will be useful and meaningful regardless who the recipient is [Ale02].

For the subjective point of view, *data* can record also subjective opinion, not only facts. They acquire meaning and they become *information* only when appropriated by a human recipient. As pointed out in [Ale02], the output from any computer program is still data. It is the human recipient who, by interpreting them, translate them into information. Therefore, while for the subjective perspective the meaning of information is subjective and related to the human recipient, for the objective perspective the meaning of information is fixed and not open to interpretation.

The third point of view mentioned in [Ale02] is the intersubjective point of view, for which the *data* are recorded in a formalized structure, depending on the data type. For instance, if the data are in the form of text, the structure is defined by language syntax and semantics, while if the data are numeric or symbolic, the structure is defined by a predefined database or form. Within this perspective, *information* is related to a context where to "read" the data available.

In the context we work, we prefer the subjective perspective, as usually considered in the visualization research field [Spe01]. We define *data* to become *information* when they can be analyzed by the human recipient. We can have many data, but if they cannot be processed, we are not informed by them: We have just raw data. For each type of data there should be an appropriate cognitive artifact for managing it. In our dissertation we focus on the problem of representing a particular type of information, i.e. multi-dimensional data, through ad-hoc cognitive artifacts. The most obvious approach would be to use *visualization*, that is the graphical representation of data or concepts [War00] and that will be briefly introduced in section 1.1. On the contrary, we aim at investigating the representation of multi-dimensional data through Auditory Display, i.e. to represent data by means of sounds, that will be introduced in section 1.2. In fact, as the nature of the human beings is multisensorial, why don't we exploit it for achieving better results?

There have been many research works trying to explore the world of auditory display, but few of them concerned multi-dimensional data. For starting our investigations, we wonder how human perception deals with auditory signals and how a listener is able to distinguish one sound among others and identify it. As we will describe in section 1.2, two of the approaches to perception are the *ecological approach* and the *cognitive approach*, where the former considers the experience to come directly from the surrounding environment, while the latter states that perception can be understood only with the aid of cognitive processes, that is using individuals' knowledge and memory. These two approaches seem incompatible, even if there are some cases of interaction between the two theories [Gav88].

We think that distinguishing firmly the two approaches could not reflect all the reality faces, because human perception tends to adapt to the particular situation where it is and to change approach according to it. We prefer to think that the two approaches we have presented do inform each other.

> Ecological theories will be most useful in describing what people perceive in normal circumstances, and in describing the basis of their perceptions in terms of the proximal stimuli. Cognitive theories will be useful in describing details of the mechanisms of perception, and the effects of pathological conditions for perception. In addition, ecological theories may be expected to provide the raw material and basic phenomena for description by cognitive theories. Such a division may allow the peaceful coexistence of these theories despite their fundamental incompatibilities [Gav88].

Considering whatever perceptual sense, people want to understand what is happening around them and, therefore, in every situation human nature tries to understand the surrounding events, even if it fails, and it wants to achieve this goal with the less effort needed.

Starting from this point of view, when subjects face an event, the first perceptual approach that they apply is the ecological, by trying to directly perceive and understand the information provided by that event. If they don't succeed, the alternative is to use their knowledge or their past experience in order to achieve the goal. All these trials should happen unconsciously. Only if they cannot unconsciously understand the event in few instants, everything is moved at conscious level.

Looking at this theory from the auditory point of view, if a sound is presented to the auditory channel, the perceptual system tries to identify the sound source directly, but if the perceptual dimensions are not "ecologically" known to the listener, memory and knowledge are used for recognizing the source.

What happens if the auditory event is complex and it comprises some sounds simultaneously generated by different sound sources? We think that the perceptual system applies the ecological approach, but, recognizing the presence of many sounds together, it applies the Auditory Scene Analysis principles in order to "segment" and "group" the different streams, and it applies the ecological approach to the stream the listener is interested in. Cognition will be used only if it is needed.

For designing an auditory display able to provide information about multi-dimensional data, we start to investigate ecologically some features of sound events that were little or even never studied in the auditory display field and to see if they could be used for providing information, because if they are ecologically relevant, they could be caught immediately and directly from the user perceptual system. Therefore, we will check if these features can be noticed by human listeners, if they make objects and events recognizable and if they can provide quantifiable information. We would like to study them from an ecological perspective, trying to find features useful for creating a "natural auditory environment" for the users of the auditory representation. In fact, we think that it is better and more involving for the users to work with auditory representations which use everyday sounds rather then unnatural sounds, because they could be less distracted and less irritated. Moreover, by using everyday sounds we think that the auditory display could be more general and therefore applicable to a variety of situations. On the contrary, if an application uses abstract sounds, the mapping design should be more related to that precise data set in order to have a metaphorical relationship with the data it represents,

> [...] a knowledge of the data structure is crucial for the design of an effective synthesis display [Bar97].

although we know that in some cases it is a compelled choice the use of abstract sounds, due to the data nature and the absence of a realistic mapping for them.

In our dissertation, we aim at studying the features from a perceptive point of view, with the only goal of enhancing their pros and cons. The presentation of rules for a good auditory display design is beyond the scope of the dissertation. Nevertheless, even if we focus on auditory display and on sound sources and events perception with the purpose of representing multi-dimensional data, we think that an efficient representation, in particular a complex one, such as for multi-dimensional data, should include not only a visual display or an auditory display, but it should be multimodal. Multimodality is not covered by our research activity, but, for completeness of introduction to our work, we like to mention it.

A *multimodal display* can take advantage of the single senses according to the current needs. Even if there are examples of bad interaction of the visual and auditory display [TSHK87], a multimodal representation can provide information to the users brain through different communication channels, allowing the users brain to decide which channel is faster for that particular information. Moreover, by providing redundant information, multimodal display can exploit all human capacities for acquiring the meaning. In addition, it can add power to the representation. In fact, if we think of multi-dimensional data, on a paper we can represent only bi-dimensional plots. For being able to represent multi-dimensional data, we need to "add dimensions". The research on this topic is wide, but the results consist always in complex graphical representations. This complexity would be reduced if auditory display would be added.

Each sense has its advantage. They can support each other and they can affect each other, sometimes positively and sometimes negatively, as well. There are various investigations on their relationships. It is interesting the work of Kubovy and Valkenburg on auditory and visual objects where they "propose a new, cross-modal conception of objecthood" [KV01]. They divide the auditory system in auditory "where" and "what" subsystems, as Milner and Goodale [MG95] did for the visual system. While the auditory "what" subsystem, similarly to the visual counterpart, deals with auditory objects and their segregation, grouping and edge identification, the auditory "where" subsystem is demonstrated to be at the service of the visual localization. In fact, auditory localization can be affected by the visual one, which means that the visual objects we see control what we are hearing, but the opposite cannot happen. The visual "where" subsytem controls the auditory "where" subsystem. It happens, for instance, in the ventriloquist effect, where the location of an auditory stimulus is perceived to come from the same location as a simultaneously presented visual stimulus.

> Typically, the modality with the best spatial resolution (e.g. vision's superiority over audition) has the greatest influence on the location of the fused percept [CBI98].

Talking about crossmodal influences, we cannot avoid mentioning the McGurk effect[1], where the two modes affect each other and produce an output that is the result of the combination of the two inputs. There are also cases where auditory signals can influence visual stimuli and create a visual illusion, for instance, when a single visual flash is presented to subjects together with multiple auditory beeps, they report to see multiple flashes [SKS00].

[1]The McGurk effect consists in presenting to the subjects an audible syllable, e.g. 'ba', and the lip movements of another syllable, e.g. 'ga'. The subjects typically report to hear another syllable, usually 'da'.

This is an example where the auditory "what" subsystem controls the visual "what" subsystem. In this case, the auditory temporal resolution is higher than the visual one. For a summary about cross-modal interactions, we suggest to refer to [SS01].

Despite these examples of negative influences, there are many examples of positive interactions between the two modes and there is a wide research on this topic. Therefore, by providing multimodal display we think to be able to combine the advantages of the modes and to create an efficient system. Before moving to the kernel of the dissertation, we would like to underline that our major interest regards some specific perceptual dimensions of sounds and sound events and these observations about multimodal displays are just general ideas, without entering into details, for possible future applications of our studies.

In section 1.1 we will present the basic concepts of visualization, while in section 1.2 we will focus on auditory display and how it is possible to represent data by means of sounds. Finally, in section 1.3 we will present the dissertation, pointing out the topic of each chapter and their purposes.

1.1 Representing data through the visual channel

In science, *visualization* is the graphical representation of data or concepts [War00]. It is the process of exploring, transforming, and viewing data as images to gain understanding and insight into the data [SML98]. Visualization constitutes a cognitive artifact that supports human beings in exploring and analyzing data by using visual processing and in decision making.

> In other words, visualizations assist humans with data analysis by representing information visually. This assistance may be called *cognitive support*. Visualizations can provide cognitive support through a number of mechanisms [...]. These mechanisms can exploit advantages of human perception, such as parallel visual processing, and compensate for cognitive deficiencies, such as limited working memory [TM04].

Recently, this field has been divided into two different research subfields: Information Visualization (IV) and Scientific Visualization, even if this distinction is not accepted by all. As they are distinguished by [Mun00], *Information Visualization* (IV) concerns finding a spatial mapping of data that is not inherently spatial, while in *Scientific Visualization* (SV) the mapping is based on a spatial layout that is implicit in the data. Scientific Visualization usually represents physical objects and phenomena with a precise and realistic representation, while Information Visualization deals with abstract data, where the representation is more intuitive and schematic rather than realistic and precise.

The advantages of visualization are many. As highlighted by Ware [War00] they include: (i) allowing the user to comprehend huge amounts of data; (ii) enhancing properties that were not anticipated; (iii) highlighting errors and artifacts in the data itself. Chittaro [Chi01] adds the following ones to the list: (i) allowing users to explore available data at various level of abstraction; (ii) giving users a greater sense of engagement with data; (iii) supporting the recognition of relevant patterns by exploiting the visual recognition capabilities of users.

Visualization is beyond the scope of our investigations. For a detailed introduction of the research field, we suggest to refer to [War00]. In this section, we will present briefly the ecological approach to visual perception (subsection 1.1.1), which is based on Gestalt psychology and from which their auditory counterparts had origin. We will, moreover, present the research in the visualization field aiming at representing multi-dimensional data (subsection 1.1.2), that is the counterpart of the problem we want to deal with from the auditory point of view.

1.1.1 Gestalt psychology and ecological approach in vision

The ecological approach, that we try to apply to auditory perception, was born with Gibson who was a student of the Gestalt psychologist Koffka.

The term *Gestalt* is derived from the German term ‘gestalten’ or ‘organized wholes’ and the *Gestalt-Theorie* was officially initiated in 1912 in an article by Wertheimer [Wer12] on the phi-phenomenon, that is a perceptual illusion in which two stationary but alternately flashing lights appear to be a single light moving from one location to another [Gre00].

Gestalt psychologists state that human brain is innately able to organize perceptions. Their work was born in the field of vision perception, but, as we will see later (subsection 1.2.1), it can be related also to auditory perception. Gestalt psychology is based on two guiding principles: (i) the whole is greater than the sum of the parts; (ii) the parts are defined by the whole as much as vice versa. These were demonstrated by Navon [Nav77], who, in his experiments, presented briefly to the subjects large outlines of letters composed of smaller letters and found that the large letters were identified despite the identity of the small ones, while the perception of the identity of these was affected by the large letters.

> Whereas the identity of the small characters had no effect on recognition of the large ones, global cues which conflicted with the local ones did inhibit the responses to the local level.[...] The results of this experiment indicate that the global pattern is responded to faster than the elements. Moreover, whereas people can voluntarily attend to the global pattern without being affected by the local features, they are not able to process the local features without being aware of the whole [Nav77].

Gestalt psychologists developed a set of Principles of Organization, known as the "laws of Pragnanz", which constitute the rules by means of which human beings manage the perceptual field and form mental patterns in their brain. These principles are:

Proximity: elements that are closer together will be perceived as a coherent object;

Similarity: elements that look similar will be perceived as part of the same form;

Common Fate: a stimulus will be organized into a figure as more symmetrical, regular and simpler as possible;

Good Continuation: human perception tends to continue contours whenever the elements of the pattern establish an implied direction. For instance, if two lines intersect then the two lines may realistically be any combination of four divergences from the meeting point, but they are perceived as being the combination of the two most similarly orientated;

Figure/Ground: in any visual array there is always one object which will be perceived as the object and everything else as a background to it. The object is usually identified because it is in some way different, whereas the background is undistinctive;

Closure: human perception tends to enclose a space by completing a contour and ignoring gaps in the figure.

The Gestalt rules, that define how perceptual scenes are grouped and segmented, are based on the physical properties of the visual objects. In 1950 Gibson started the so-called *ecological approach* to visual perception, which, on the contrary, considers the perceptual scenes to be based on the cues of the objects related to the environment where they are.

> We are told that vision depends on the eye, which is connected to the brain. I shall suggest that natural vision depends on the eyes in the head on a body supported by the ground, the brain being only the central organ of a complete visual system [Gib86].

He considers perception to be direct, i.e. to be "not mediated by retinal pictures, neural pictures, or mental pictures" [Gib86]. Moreover, he defines the *invariants*,

> The shape of a growing child is relatively permanent for some features and changing for others. An observer can recognize the same room on different occasions while perceiving the change of arrangement, or the same child at different ages while noticing her growth. The permanence underlies the change [Gib86].

Gibson, presenting the concept of invariants in ecological optics in [Gib86], defines four kinds of invariants:

- **Invariants of optical structure under changing illumination.** If conditions of illumination change, i.e. amount, direction or spectral composition of light change, the surfaces are always perceived to be the same;
- **Invariants of optical structure under change of the point of observation.** If the point of observation changes, and, therefore, some transformations in the scene happen, such as occlusions or new objects appear, etc., the scene is always perceived to be the same;
- **Invariants across the sampling of the ambient optic array.** By looking around, human beings have some samples of a scene but they still can recognize the identity of the whole scene;
- **Local invariants of the ambient array under local disturbances of its structure.** The local disturbances of optical structures are those related to the local events such as deformations of surfaces, changing of expression in the face, etc., i.e. everything that changes, and their identity is still perceived and it cannot be classified in the previous categories (for example, a rolling ball or rippling water).

1.1.2 Multi-dimensional data visualization

As defined by Healey [Hea96], *multi-dimensional data visualization* involves representation of multi-dimensional data elements in a low dimensional environment, such as a paper or a computer screen.

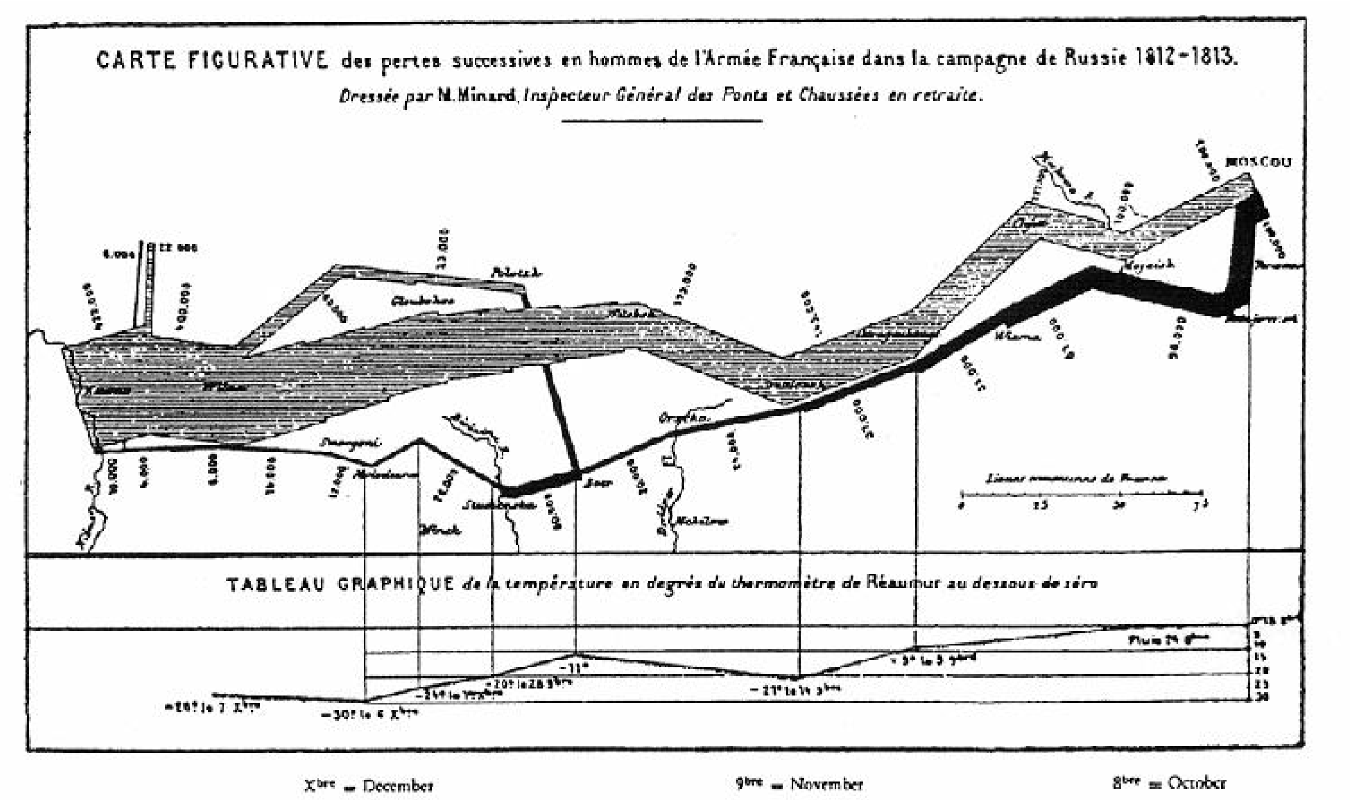


Figure 1.1: *Map drawn by Charles Joseph Minard about Napoleon's 1812 Russian Campaign* [Map]. In this map there are six variables represented: the size of the army, the latitude and longitude of its movements, the direction that the army was traveling, the location of the army with respect to certain dates and the temperature along the path of retreat [Tuf83].

A famous example of multi-dimensional representation is the map (Figure 1.1) drawn by the French engineer Charles Joseph Minard in 1861, which shows the losses of Napoleon's army during his Russian Campaign of 1812. In this map there are six variables represented: the size of the army, the latitude and longitude of its movements, the direction that the army was traveling, the location of the army with respect to certain dates and the temperature along the path of retreat [Tuf83].

> Minard's graphic tells a rich, coherent story with its multivariate data, far more enlightening than just a single number bouncing along over time [Tuf83].

It is a very good multi-dimensional representation because it is able to represent six variables within the same graph, but the user has to learn the conventions used in the graph, before being able of understanding its meaning. Moreover, even if, from the historical point of view, it is a very good example, it is a complicated representation, that seems impossible to be automatically generated by a machine, since, as pointed out by Tufte during a talk [Gol02], the map is a combination of human drama, multivariate information, and minimal extraneous data,

> "It may be the best statistical graphic ever drawn." Why? Because Minard was able to depict in a simple image the brutal fate of an army of 442,000 men, just 10,000 of whom survived the retreat from Moscow through sub-zero winter temperatures. Napoleon's death march looks, in Minard's graphic, like a trunk of a felled redwood, doubled back on itself, stout and powerful at the roots, wispy and brittle at the top. Miraculously, the map shows you everything you need to know: how many men perished, where and when, and under exactly what conditions.
>
> "Minard made the map because he hated war," [...] "Something subtle is going on here: The word Napoleon does not appear on this chart. Minard's point is not to celebrate surviving celebrities, but to memorialize the dead soldiers." [...] *Guernica* and Minard's map of Napoleon's march — both of them belong in the same museum of forever anti-war art [Gol02].

In visualization literature there is a long list of representations of multi-dimensional data, from the parallel coordinate plots, to starplots to Chernoff's faces, to many others. We don't want to give a list of all these examples. But, we think that all these approaches are complicated and, even if they are able to summarize many dimensions in one representation, it is difficult to focus on only one dimension, and, sometimes, it is difficult to have an overview, as well. Healey [Hea96], in order to solve this problem, based his data visualization technique on *pre-attentive processing*, a field of cognitive psychology which aims at studying the visual features that are detected rapidly and with little effort by the human visual system, i.e. the features that "pop-out" from their surroundings [War00]. In fact, pre-attentive processing refers to an initial organization of the visual system based on operations believed to be rapid, automatic, and spatially parallel [HBE95].

> [...] visual properties that are "preattentively" processed [...] are detected immediately by the visual system. Viewers do not have to focus their attention on an image to determine whether elements with a given property are present or absent [Hea96].

Features that are preattentively processed, as explained in [War00], can be organized in the following categories: form, color, motion, and spatial position. For instance, in the form category we can find line orientation, line length, curvature, spatial grouping, etc.; in the color category there are hue and intensity, while in the motion category there are flicker and direction of motion; in the spatial position category we can find 2D position, convex/concave shape from shading etc.

From the visualization point of view, pre-attentive processing is important to be studied, because it could be used in designing applications, in order to enhance some aspects of the graphical representation.

> Properties that are processed preattentively can be used to highlight important image characteristics. [...] Research in visualization has shown that preattentive features allow users to better perform visual tasks such as grouping of similar data elements, detection of elements with a unique characteristic, and estimation of the number of elements with a given value or range of values [...] [Hea96].

Within our scope of multi-dimensional data representation through Auditory Display, we will discuss if we can find in auditory processing similar properties that could help us in designing auditory representations.

1.2 Representing data through the auditory channel

Auditory Display concerns the representation of data by means of sounds. It can be used both to develop multimodal interfaces, which exploit the multi-sensorial nature of the human beings for providing them with information, and to represent efficiently huge amounts of data, in which case we still suggest to use multimodal catalogs in order to achieve better results. For an overview to auditory display we suggest to consult the introduction by Kramer [Kra94a].

There is a wide research area in auditory display aiming at finding the rules and principles for representing data efficiently

> ...to generate reliable and meaningful auditory displays. We must discover the set of truly useful auditory parameters and understand their perceptual transfer functions so that displays can be designed to take advantage of them [Bar97].

As Barrass [Bar97] summarizes, there are seven approaches to the design of sounds in order to provide information:

- the **syntactic method** where information is provided by the organization of the auditory elements (e.g. Morse code, earcons);
- the **semantic method** where information is carried by the auditory element itself (e.g. auditory icons);
- the **pragmatic method** where the focus is on the form of the signifier, i.e. it is based on psychoacoustic investigations for allowing the discrimination of the sounds (e.g. Auditory Scene Analysis and Gestalt theory);

- the **perceptual method** where information is provided by the relations between the sounds. It is usually classified in *audification*[2], and *sonification*[3];
- the **task-oriented method** who assumes that the meaning of a sound depends on the context in which it is used;
- the **connotative method** which focuses on the cultural and aesthetic implications of the sounds;
- the **device-oriented method** that aims at designing sounds that are transportable on different devices, i.e. that are device-independent.

This view could be defined as "design-purposed", because its goal is to define a set of rules and guidelines that allow auditory display designers to develop efficient interfaces and representations.

The focus of our work, on the contrary, is not on the design issues of auditory interfaces or displays, but on the general issue of the perception of sound events, in order to represent multi-dimensional data. We aim at investigating some general acoustic features of auditory events that could be used for representing multi-dimensional data through auditory display. Therefore, the emphasis of this work is on the perceptual aspects of sound events rather than on practical rules for auditory design. Nevertheless, our investigations on the perception of the auditory events could define the basis for a new design of auditory displays for multi-dimensional data.

Within an overview of the organizing principles for representing data with sound, Kramer [Kra94b] emphasizes the problems that designers have to face when they develop a multivariate auditory display. He states that the parameters overlapping could cause difficulties to users to distinguish the different variables and, moreover, that the users could be distracted and influenced by unbalanced displays. In fact, as he says

> Even if sound parameters that do not directly interfere with each other are used, the fact that our attention is drawn more to certain variables than others makes the design of a balanced, or unbiased, auditory display virtually impossible [Kra94b].

In his article, Kramer refers to abstract sounds or to synthesized sounds, even realistic, but combined to form an abstract auditory scene. On the contrary, we think that representing data through synthesized sounds that reproduce everyday sound events could provide information even on more than two dimensions. In particular, we aim at displaying data by

[2] *Audification*: direct translation of the data into sounds, i.e. the data are listened (for instance, the Auditory Seismology [Dom01], where the seismograms are shifted into an audible frequency range and are played).

[3] *Sonification*: translation of the relations between data into relations between auditory signals (for instance, a pitch variation over time could represent abstract data relations).

means of the “natural” perceptual dimensions encompassed by few auditory events, rather than using many events simultaneously, each one representing a variable.

The basic assumption for our investigation is that the human ear is able to recognize many aspects of the surrounding environment

> It is apparent from everyday experience that listeners can detect significant aspects of the environment by ear, from a knock at the door to the condition of an automobile engine and the gait of an approaching friend [WV84].

and, moreover, that the human auditory system is able to perceive many attributes and characteristics both of the sounds themselves and of the sources by which they are produced. For instance, listeners can easily order sounds, produced by a musical instrument, into a musical scale and they can also recognize the instrument by which the sounds were produced. From the same sound, people can derive different information, depending on their listening attitude. Gaver [Gav93a, Gav93b] distinguishes between musical listening, that concerns the perceptual dimensions and attributes of the sound itself, and everyday listening, that deals with the perception of sound-producing events.

> The distinction between everyday and musical listening is between experiences, not sounds (nor even psychological approaches). [...] ...while walking down a city street we are likely to listen to the sources of sounds — the size of an approaching car, how close it is and how quickly it is approaching — but occasionally we might listen to the world as we do music — to the humming pitch of a ventilator punctuated by a syncopated birdcall, to the interplay and harmony of the sounds around us [Gav93b].

We can find two contrasting approaches to perception: the *cognitive approach* and the *ecological approach*, assuming *perception* to be the experience coming directly from the surrounding environment, while *cognition* to be the experience that is “filtered” by memory and knowledge.

> The dispute between cognitive and ecological approaches is apparently quite simple. At its heart is a disagreement about where perception ends and cognition begins. In other words, in explaining the organism as a system that perceives and interacts with its environment, how much of that explanation can come from outside the system, and how much from within? In general, the cognitive answer to this is that cognition plays a major role even in what appears to be simple perception. The ecological approach argues that perception alone can explain far more, including many aspects of experiences that seem to require cognition [Gav88].

Their main difference consists on their views about the adequacy of information for perception. In the cognitive approach the information must come from higher processing, since the stimuli per se don’t provide enough information for the human percept, which, on the contrary, in the ecological approach, can receive enough information by the sounds per se, without the requirement of high-level processing.

For our purposes, we want only to cite the two approaches. We refer who is interested in examining in detail this topic to Gaver's PhD thesis — second chapter [Gav88].

1.2.1 Auditory Scene Analysis, Gestalt psychology and ecological approach in auditory perception

The hearing system, even in front of complex auditory scenes and in unfavourable conditions, is able to separate and recognize auditory events accurately. A great deal of effort has gone into the understanding of how, after having captured the acoustical data, the human auditory system processes them.

> ...our ears are rarely presented with the sound of a single speaker in isolation, but more often with a combination of several speech and nonspeech sounds which may also have been further altered by the acoustic environment. Faced with such a mixture, the listener evidently needs to consider each source separately, and this process of information segregation is known as *auditory organization* or *auditory scene analysis* [Ell98].

The word *Auditory Scene Analysis (ASA)* was introduced by Bregman to denote the psychoacoustic field founded by him with the objective of understanding how the auditory system and the human brain process the *complex auditory scenes*. In order to know what it means, we can refer to Scheirer [Sch00], who, with reference to McAdams and Yost, defines *auditory image* as what a subject, listening to many sources simultaneously, perceives as a single source. In fact, a sound set is divided in auditory images that the listener can imagine to hear independently. A sound group perceived as made of auditory images is called *auditory scene* and the perceptive process applied to an auditory scene is called *auditory scene analysis*.

Auditory Scene Analysis considers the auditory processing divided into two steps: *segmentation*, which separates the sounds reaching the listeners ears into sensorial elements defined as *segments*; and *grouping*, which collects the elements, that are likely generated by the same sound source, in a structure called *stream*.

Bregman [Bre90] identifies two types of auditory grouping: primitive grouping and schema-driven grouping. While *primitive grouping* is driven by the incoming acoustic data, and it is probably innate, *schema-driven grouping* employs the knowledge of familiar patterns and concepts that have been acquired through experience of acoustic environments [Bro00].

Auditory Scene Analysis is thought to follow Gestalt Principles of Organization, that we have already presented from the visual point of view in subsection 1.1.1, and which from an auditory perspective are:

Proximity: sounds that are close together over time will be perceived in the same stream;

Similarity: sounds that are similar in timbre, pitch, loudness or close in apparent location or time will be perceived as part of the same stream;

Common Fate: a complex auditory scene will be organized into a stream where the components that are fused are those who are coordinated;

Good Continuation: natural sounds usually change gradually their frequency, intensity, location or timbre. A new stream is perceived when an abrupt change occurs;

Figure/Ground: in a complex auditory scene it is possible to follow only a stream per time and everything else is considered as a background to it;

Closure: human perception tends to perceive a sound continuing even if it is obscured or absent (e.g. auditory tunnel effect [Vic60] or phonemic restoration effect [War70]).

A different approach to auditory perception is studied in *ecological psychology*, that we have already briefly presented in visual perception (subsection 1.1.1), and which studies acoustic phenomena by observing the physical characteristics of a sound event, the high-order configuration of variables, and the listener's ability to detect the information provided by the event [Kel99]. In the ecological perspective, each event is characterized by its *structural invariants*, that specify the kind of object and its properties under change, and its *transformational invariants*, which specify the style of change of the object's properties [WV84]. For instance, what allows us to understand that we are listening to the same melody played by different musical instruments, or on different registers, or by different players?

There is a relatively wide literature about experiments on perception with the ecological approach. For example, Repp [Rep87], investigating on clapping by means of acoustical analysis and perceptual tests, states that people can hear spectral properties of clapping and they can recognize the hands' configuration with good performances. Therefore, as he points out, a sound can convey information about the sound source configuration, as it is for speech, which reflects the changing state of the vocal tract. Other sound sources analyzed in ecological perception studies are the footsteps [LLP91] and the acoustic properties that differentiate male and female walking sounds. Also, in this case, it is demonstrated that a listener can understand some sound source characteristics by the features of the sound signal produced by it. Another ecological approach we cite is Freed's work [Fre90], whose goal is to find a function able to predict the perceived mallet hardness. Also, in this third case, we can see how it is possible to gain information about the sources just from the cues of the sounds they produce.

These are examples of studies about the perception of certain sound sources and their capacity to provide information to users, that could be useful in auditory display, and, in particular, for multi-dimensional data representation, that is our main interest. It is worth extending the investigations to other sound sources and to sound events, in order to define a sound object set for designing new multi-dimensional data display systems. Within this frame we aim at studying some perceptual dimensions of everyday sounds, that were never considered for providing information in auditory display, and we aim at investigating if they can be used for representing auditorilly multi-dimensional data. In section 1.3, we will clearly define the framework of our research activity.

1.2.2 Multi-dimensional data auditory representation

Various approaches tried to solve the problem of representing multi-dimensional data. For exemplifying, we refer to a couple of typical applications of sonification. The first case is an auditory display for physiological data, by Fitch and Kramer [FK94], that could be used in a operating room setting and that is able to represent eight variables simultaneously. In order to represent the data efficiently the authors decided to use two different approaches, according to the data type. For four parameters, they used a synthesis technique for generating sounds recalling the real ones, which the authors defined as "self-labeling sounds" (e.g. heart rate, respiratory rate, etc.) On the contrary, four other variables weren't related to any sound; For instance, which sound has the blood pressure? Therefore, they used abstract sounds that had to be learnt by the users of the system before starting to work with it. The learning process, however, took less than one hour. By using the two different mappings the authors state that the system is efficient, because

> this approach combines the ease of use of everyday sound with the power and flexibility of nonrepresentational sound. Our results suggest that this is a fruitful marriage [FK94].

The second example is a sonification of weather data [FWGK01]. In this application the authors use abstract sounds, even if they try to recall the sounds of the weather events, and they decided the mappings according to metaphorical reasons, by relating the heaviness and the amount of fallen rain or snow to the frequency of the "plinks" and the number of them played in succession.

These two examples are mentioned just for showing what we mean when we talk about auditory design of applications for a precise representational problem, that is within a precise and pre-defined context.

Besides these applications that are practical solutions to a precise display problem, there are studies concerning the general principles of auditory design and the rules and guidelines for an effective auditory design.

In our dissertation we do not want to focus on specific representational problems, but on more general issues, from the perceptual point of view. There is an extensive work by Barrass [Bar97] about auditory information design. It develops the Information-Sound Space (ISS), a 3D-spatial organization of auditory relations, that helps in facing the design process in a very direct manner, without a strong experience in sound design theory or principles. It is the auditory equivalent to the Hue, Saturation, Lightness (HSL) color model for the visual perception field.

As in other studies, the approach of Barrass concerns abstract sounds and it refers to three-dimensional data. Although we are interested in multi-dimensional spaces — not only 3-D — by using everyday sounds properties to convey information, we can begin our investigations looking at the visual perception field, as Barrass did.

In subsection 1.1.2, we have defined the concept of pre-attentive processing for the visual field. It seems interesting to study a similar concept within the auditory field, in order to exploit characteristics that "pop-out" for representing multi-dimensional data more efficiently. In fact, it could provide a solution to the problem of unbalanced displays that Kramer [Kra94b] mentions while he is presenting multivariate auditory displays and that we presented in section 1.2. We recall that by unbalanced displays Kramer means those displays where the acoustic parameters that represent data have different levels of forcefulness (attention-compelling characteristics for the listener) [Kra94b], and therefore they are influenced by the perceptual impact of some auditory variables rather than others. Pop-out cues could be used for enhancing a sound or a particular perceptual feature of a sound within an auditory representation, as Ware states for vision,

> In displaying information, it is often useful to be able to show things "at a glance". If you want people to be able to instantaneously identify some mark on a map as being of type A, it should be differentiated from all other marks in a pre-attentive way [War00].

Within visual processing, there is a wide research field studying and conducting experiments in order to find pop-out features and cues that can be processed pre-attentively. These works start from the basic visual properties, such as form (e.g. line orientation, length, and width, curvature of elements, spatial grouping), color (e.g. hue and intensity), spatial position, motion (e.g. flicker, direction), etc. These properties are the primitive features that are extracted in early visual processing. They are well known and the list is long. In order to investigate pre-attentive processing and pop-out cues, the only thing to be done in this case is examining each characteristics and identifying their properties.

Within the auditory processing research, on the contrary, we think that the perceptual auditory features available to the human being for extracting information from an auditory representation are not completely known, especially from the ecological point of view. Therefore, in our research work, we do not focus on the features classification yet, but on finding interesting features, which could be used for effective auditory display.

1.3 Introduction to the dissertation

Our research activity deals, from the auditory perspective, with the issue of how it is possible to represent multi-dimensional data in order to provide information, by considering the data to become information, as we have already defined in the introduction to this chapter, when they can be processed and understood by the users.

Organizing the perceptual relations and properties in a multi-dimensional space makes visual/auditory designers understand and control them easier. For instance, in a visual representation, we can exploit different colors to group data and, among these, we can distinguish various categories by means of geometric shapes. Moving to the auditory field, we think that it could be useful to base multi-dimensional data representation on the human auditory system capabilities to perceive and distinguish, not only the auditory cues of a sound itself, but also the characteristics and configurations of the sources producing them, as we stated above, by applying the ecological approach. We know the important role that the structural invariants and the transformational invariants play in the auditory perception of sound events within the ecological perspective. If we think, for instance, to a rolling sound, the structural invariants are shape, size, material, while the transformational invariants are velocity and accelleration. Human beings can recognize, by just listening to a rolling sound, that an object is rolling and, even if rolling object changes its qualities, they are still able to recognize that the process (i.e. rolling) remains the same. Or we can think of the temporal patterns that allow listeners to distinguish between breaking and bouncing categories and the transformational invariants characterizing these sound events, as studied by Warren and Verbrugge [WV84].

> There has been a steady growth in the consideration of this view as characterizing aspects of the auditory system with several experiments having been conducted into the possible existence of invariants as well as speculations as to their signal properties (Gaver 1993, 1994; VanDerveer 1979; Warren and Verbrugge 1984; Wildes and Richards 1988). The general results of this body of work suggest that certain structures of sound events are lawfully and invariantly related to fundamental properties of physical systems and force interactions, and that human auditory perception may be directly sensitive to such structures [Cas98].

The main issue faced by the dissertation is the possibility of using other perceptual dimensions, for auditorilly representing multi-dimensional data, rather than those usually considered such as pitch, loudness, timbre, rhythm, etc. In fact, we observed that human beings are able to recognize, through their auditory systems, other perceptual dimensions besides those proper of the auditory domain, such as physical and geometric properties of objects. In particular, we would like to investigate on perceptual dimensions that could be ecologically relevant and that, therefore, could provide users with information through a "direct input channel", without the need of using knowledge or memory processes for gaining information from the auditory display. We point out that, although our goal is to use a "direct input channel", our main issue is primarily concerned with providing information *to the users* of the auditory display. Therefore, even if, by applying the ecological approach, we try to exclude knowledge and memory from the processes users might apply for gaining information from the auditory display, we can still talk about data and information within the perspective we have defined above, since we focus on the users of the auditory display in relationship with the data we would like to represent, rather than on the data by themselves.

We conducted different psychoacoustic experiments trying to cover some features of sound sources and common sound events, investigating how they convey information to the human auditory system.

> By listening to sounds, we are able to extract information about the sound source, the location, and the environment in which the sound is produced. Although nonspeech sounds are a familiar and natural medium to get information, they are barely used in information and communication technology. In order to create suitable auditory interfaces based on everyday sounds, we have to better understand how people perceive these sounds [HS02].

As pointed out by Kubovy and Van Valkenburg [KV01], while the visual system deals with surfaces that reflect light, the auditory system deals with sources of sounds and auditory events. If we want to investigate perceptual dimensions that are new to auditory display, we should focus on perceptual features of sound sources and of auditory events, rather than on properties of the sound signals per se. Sound sources and auditory events have been recently considered effective in designing auditory display. The model-based sonification [HR99] approach is built on the idea of establishing a virtual scenario from the data to be sonified and of defining the model for the interaction of the single elements of the system. For an efficient implementation of the model, physical model sound-synthesis techniques are recommended. The user, then, can explore the data by interactively exciting the system and by listening to the sounds generated.

> The user then might explore the data set by shaking, plucking or hitting the virtual data material [HHR01].

For our perceptual experiments, we decided to use synthetic models of the sound sources and of the sound events/processes we wanted to investigate. An experimenter, by using physical models for generating the stimuli used in the experimental framework, is able to control directly each physical attribute [LCS00].

> If physical properties of sound sources can be presented auditorially, can we attend selectively to them? This is a difficult question to address with natural sources, since it is often cumbersome to isolate individual physical attributes. Recently developed tools that can permit such manipulations, however, are physical models of sound sources implemented digitally [LCS00].

Since all these investigations were conducted for being, in the future, applied to auditory displays or multimodal applications which could enjoy their benefits, it could be more flexible to include sound models inside applications rather than pre-recorded sounds and, in particular, by using the sound models developed by the the EU-funded project "the Sounding Object" (SOb)[4], it could be computationally convenient.

We started our investigations from some spatial attributes. We investigated the auditory perception of the shape of 3D-resonators (spherical and cubic). In particular, we studied the ability of human beings to distinguish the shapes of the resonators and we focused on pitch

[4] `http://www.soundobject.org`

perception and its relationship with the volume of the enclosures. We have seen that in this case listeners need very precise conditions in order to be able to distinguish this perceptual dimension.

Then, we moved to another spatial attribute, an important physical dimension which affects the auditory perception of the sound event, that is the distance of the sound source from the listener. It is useful to be able to convey distance cues in auditory display applications. Our research group developed a virtual resonating environment, which reproduces the acoustics inside a tube, with the aim of enhancing the perception of the distance. Within this framework, we conducted psychoacoustical experiments, in order to validate the information about distance conveyed by the aforementioned virtual environment.

Afterwards, we studied the perceptual dimensions of some specific sound events and processes, whose sound models were investigated and developed by the EU-funded SOb project. These sound objects are characterized by their *cartoonification*, i.e. they are simplifications which exaggerate certain features and which provide information, despite their caricature aspect. We focused on two specific sound object categories — the impact/bouncing event and the rolling process[5]. We aimed at studying the auditory perception of the single dimensions, as well as the complex relationships existing between each perceptual dimension. We conducted our experiments on the impact/bounce event by means of the Sonic Browser, a software for navigating among sounds, in a bi-dimensional space, using the hearing system, that was born at the Interaction Design Centre in the University of Limerick — Ireland — in 1996 and in continuous improvement since then [BF01, BFTC02]. On the contrary, the rolling process was studied by using an interface designed in the Matlab environment for this purpose, since from informal experiments we noticed that the Sonic Browser was suitable for events rather than for processes.

Besides the aim of investigating some interesting perceptual dimensions of the sound events from an ecological point of view, our work is useful for studying other auditory features that were never studied before and which could be processed pre-attentively, constituting pop-out cues. These features could represent a solution to the problem of unbalanced displays arisen by Kramer [Kra94b].

In the following chapters we will describe each research activity, by presenting the purposes of each work, the experiments conducted, including test procedures and results, and by concluding the dissertation with some final remarks and observations. In **chapter 2**, we will introduce our studies about the perception of shape of 3-D (spherical and cubic) resonators and about the relationship between the pitch and the volume of the enclosures, conducted with synthetic resonators. In **chapter 3**, we will present the work conducted in order to validate the rendering of the distance cue by means of a virtual environment. Then, in **chapter 4**, we will describe the experiments conducted using the Sonic Browser for investigating some important perceptual dimensions of the impact/bouncing events and we will show their results. In particular we will present our studies on the perceptual di-

[5]We refer to a sound object as an 'event' when the time scale is small, while a 'process' covers a longer time scale. Therefore, an impact will be an event, while a bouncing and a rolling will be a process.

mensions of perceived height of the object drop versus perceived size of the dropped object and of perceived elasticity of the impact/bounce event versus perceived force "throwing" the object. In **chapter 5**, we will introduce the experiments conducted on the perception of the rolling process. In particular, we will briefly report the results of the tests on the perceptual dimensions of perceived velocity versus perceived size of the rolling object. Finally, in **chapter 6**, we will summarize the results achieved by the dissertation focusing on their advantages and their practical utility, suggesting ideas for further investigations.

Chapter 2

Auditory perception of shape and 3-D size

Probably, the most relevant visual features of an object are size and shape. We wonder if, besides the visual system, they are perceptually relevant also for the auditory system. We started our research activity with the following questions: Are we able to distinguish the shape of the sound sources? Which relationship is there between pitch perception of a sound source in an enclosure and the size of the enclosure itself?

We observed that the shape of 3-D cavities affects the timbral quality of sound sources located inside them. There is a sort of pitch due to the enclosures resonances, and the pitch height is somehow related to the size of the cavity. In this context, we are interested in investigating how differently-shaped enclosures give rise to different perceived pitches [RO01].

It would be interesting from the auditory display design point of view to find the relationship between pitch perception of a sound source inside a resonator and its size and to prove the ability of human listeners in classifying resonators shape. It would be possible to create virtual environments by synthesizing resonators with different shapes and sizes. But, in order to use these perceptual dimensions for conveying information in an auditory display application, we need to be sure that the users can classify and scale these features.

Recent psychoacoustic studies concerned similar topics, investigating the perception of physical features, such as shape, size, or material of the resonators and how human beings are sensitive to these characteristics. For example, Carello et al [CAKP98] showed that listeners are able to reliably evaluate the length of rods without any particular training. Lakatos et al. [LMC97] showed that the hearing system can estimate the rectangular cross section of struck bars, and Kunkler-Peck and Turvey [KPT00] did similar experiments using suspended rectangular plates, but, while Lakatos et al. conducted the experiments with a cross-modal visual matching procedure, Kunkler-Peck and Turvey used only the auditory channel. Wildes and Richards [WR88] noticed that material properties are related to the damping characteristics of the vibrating object, and Klatzky et al. [KPK00] confirmed this fact, providing listening tests involving stimuli that were obtained using additive synthesis of damped sinusoids. Djoharian [Djo00], working on viscoelastic models in the context of modal synthesis, showed that finite difference models of resonators can be "dressed" with specific material qualities.

Related to these topics, we studied the perception of 3-D resonators. In particular, we worked with spherical and cubic enclosures, and we focused on the relationship between pitch and size perception of them.

We have to stress the difference with previous works and investigations about auditory perception of shape and volume. As Cabrera [Cab99] points out, the term "volume" must not be confused. In fact, many prior studies dealt with the so-called tonal volume, meant as an attribute of auditory sensation produced by the sound itself, and not by the sound source size. According to a definition due to Perrott et al. [PMS80], the term "volume" refers to the apparent size of extensity of the sound image.

> For example, if one presents a high-frequency pure tone via earphones at a moderate intensity level the listener will typically report hearing a relative small image somewhere inside the head. Either an increase in stimulus amplitude or a decrease in stimulus frequency generally results in an increase in image size [PMS80].

The studies of Perrott at al. refer to auditory spatial perception. The "volume" they mention is not the three-dimensional feature of a certain object, but the impression the sound gives to the listener. For example, they point out the "expanding-image" effect, where the apparent size of an auditory image grows or expands as a function of stimulus duration [PMS80]. On the contrary, we are interested in perception of sound pitch and its relationship with the *physical* source size, in order to be able to represent multi-dimensional data in a more ecological way, as it happens in everyday listening.

A clarification about the word "pitch" is mandatory at this point, as the impulse responses of 3-D resonators are typically inharmonic and they don't produce a strong pitch sensation. However, as it emerges from experiments, listeners are capable of ordering them into scales, with an accuracy comparable to JND's of pitch perception for sine waves. Therefore, a standard definition of pitch can be applied to these sounds [ANS].

In order to be able to control directly the physical dimensions of the enclosures we were investigating, we used synthetic models of the 3-D resonators. As we already said, the sound models could be also flexible and computationally convenient for future applications within auditory displays or multimodal systems.

After an overview of the sound models we used for synthesizing the 3-D resonators (section 2.1), we will report the experiments [Roc01, RO01] concerning 3 main tasks: (i) pitch equalization of spheres and cubes (section 2.2); (ii) shape classification (section 2.3); (iii) estimation of the sphere's pitch (section 2.4). Within the first task, moreover, we will investigate the effects of the particular procedure used in conducting the experiments (subsection 2.2.5).

2.1 The sound models we used

We conducted our experiments by using a simplified version of the 3-D resonator model that was developed by Rocchesso [RD01] and that was based on a feedback delay network where each delay line corresponds to a set of normal modes of the enclosure.

The resonance frequencies distribution for rectangular and spherical enclosures is defined as follows. A rectangular resonator has a frequency response that is the superposition of harmonic combs, each having a fundamental frequency

$$f_{0,lmn} = \frac{c}{2}\sqrt{(l/X)^2 + (m/Y)^2 + (n/Z)^2}, \tag{2.1}$$

where c is the speed of sound, l, m, n is a triple of positive integers with no common divisor, and X, Y, Z are the edge lengths of the box [MI68].

A spherical resonator has a frequency response that is the superposition of inharmonic combs, each having peaks at the extremal points of spherical Bessel functions. Namely, said z_{ns} the s^{th} root of the derivative of the n^{th} Bessel function, the resonance frequencies are found at

$$f_{ns} = \frac{c}{2\pi a} z_{ns}, \tag{2.2}$$

where a is the radius of the sphere [MMG86].

For the experimental framework, the impulse responses of a sphere or a rectangular box have been modeled by summing the contributions of exponentially damped sinusoids, each tuned at the position of a theoretical resonance frequency. In the models, it is possible to choose the size of the resonator and the material of its enclosure among marble, wood and drape, each material being specified by frequency-dependent absorption curves.

The absorption coefficient, together with volume and enclosure surface area, are used to compute the decay rate of each partial, according to the Sabine reverberation formula

$$T = 0.163\frac{V}{\alpha A}, \tag{2.3}$$

where V is volume, A is surface area, and α is the absorption coefficient.

2.2 Pitch equalization of spheres and cubes

For examining the relationship between pitch and volume, we conducted two experiments, where the latter was performed in a more controlled environment. In this section we will introduce both, by describing the procedure applied and by commenting the results obtained.

The first experiment was the starting point of our research. It comprised two settings, that applied the same procedure by using a different stimuli set and that will be both presented.

The purpose of the second experiment consisted in repeating the first one in a more controlled environment, in order to verify the results of the previous and to check whether a different procedure could affect or not the experimental results.

2.2.1 Participants

The participants to the first experiment were 19 students at the University of Verona (2 females and 17 males, with age ranging between 23 and 36) who volunteerly participated to the test, 7 of which participated to the first setting and 14 to the second one, i.e. two of the subjects participated to both the tests.

All the participants referred to have no hearing problems. No one of the first participants group classified himself as a musician, while two of the second group are trained musicians.

On the other hand, the subjects of the second experiment were 15 volunteers (8 females and 7 males), with age ranging between 21 and 39, all students or faculty members of the University of Verona. They were naïve listeners but two that were trained, as they participated to the first experiment as well. All of the participants reported having normal hearing.

2.2.2 Stimuli

The stimuli set used for each test comprised 14 sounds: an impulse response of a spherical resonator of a certain size and 13 cubes, one of which had the same volume as the sphere, while six were bigger and six were smaller than the sphere. The cubes differed in size by a minimal edge length difference obtained by converting frequency JNDs into length JNDs by means of the formula

$$\Delta l = \frac{c}{2}\left(\frac{1}{f_0 - \Delta f}\right) - l_0, \tag{2.4}$$

where c is the speed of sound in the enclosure, and $l_0 = \frac{c}{2f_0}$ is the reference size length.

The diameter of the spherical resonator in the first experiment was d = 36 cm and d = 100 cm for each setting respectively, while in the second experiment was d = 36 cm.

2.2.3 Procedure

The experiments were conducted in a quiet, but not isolated, room. The stimuli were presented to the subjects through closed headphones (Beyerdynamic DT-770) and, by applying the constant stimuli approach (see appendix A), the sounds were presented pairwise: the impulse response of the spherical resonator as standard stimulus and the impulse response of a cubic enclosure. The participants were asked to judge whether the second sound was higher or lower in pitch than the first sound.

In the first experiment the sphere was always presented to the subjects in first position, while the cubes always in second position. The pairs were repeated 10 times, in random order for each subject, giving a whole set of 130 comparisons.

On the other hand, in the second experiment, each pair was repeated just 4 times in random order, instead of 10 times, since 4 repetitions are enough for capturing a general trend. Moreover, for balancing the results, we decided to present the sphere and the cubes sounds with the same order within the pairs related to the shape of the enclosure, but for half participants with the spherical enclosure as first stimulus, and the other half with the reference sound in second position.

At the end of each session, the subjects were asked to fill a questionnaire concerning their musical training and their opinions about the listened sounds and about the difficulties they found in completing the task.

The experiments were conducted by means of a MATLAB environment on a PC Pentium III with a Creative SoundBlaster Live! soundcard.

2.2.4 Results and Observations

In Figure 2.1 and Figure 2.2 we report the results of the first experiment: the probability of the answer "spherical impulse response higher than cubic impulse response" as a function of the cubes volume for the two test settings respectively (spheres with diameter d = 36 cm and d = 100 cm). In the plots, the standard deviation bars are reported.

The resulting data of the two settings are the following (see appendix A for a brief introduction to the measures calculated):

d = 36 **cm (volume** = 0.0244 $\mathbf{m^3}$**)**: Point of Subjective Equivalence $PSE_{36} = 0.0240$ m^3, Constant Error $CE_{36} = -0.0003$, and Differential Limen $DL_{36} = -0.0016$ ($\chi^2 = 19.0505$, df = 10, $p > .02$);

d = 100 **cm (volume** = 0.5236 $\mathbf{m^3}$**)**: Point of Subjective Equivalence $PSE_{100} = 0.5307$ m^3, Constant Error $CE_{100} = 0.0071$, and Differential Limen $DL_{100} = -0.0952$ ($\chi^2 = 18.9718$, df = 10, $p > .02$).

We observe that, while for the first setting there is a slight underestimation (CE_{36} negative), in the second setting there is an overestimation (CE_{100} positive).

We noticed a participant to the second test performing differently from the general trend, who ranked all cubes as lower in pitch than the sphere. This fact could be due to an analytic rather than holistic listening approach, since the subject is an expert listener and referred to be confused by hearing two pitches in the cube. On the contrary, the other subjects were able to distinguish the pitches of the two shapes and, on average, to equalize the pitch of the sphere with one of the central cubes. Anyway, we decided to keep all the subjects for the data analysis.

Despite the slight irregularities in the general trend caused by that outlier and maybe some others, the Point of Subjective Equivalence is estimated close to the cube with volume

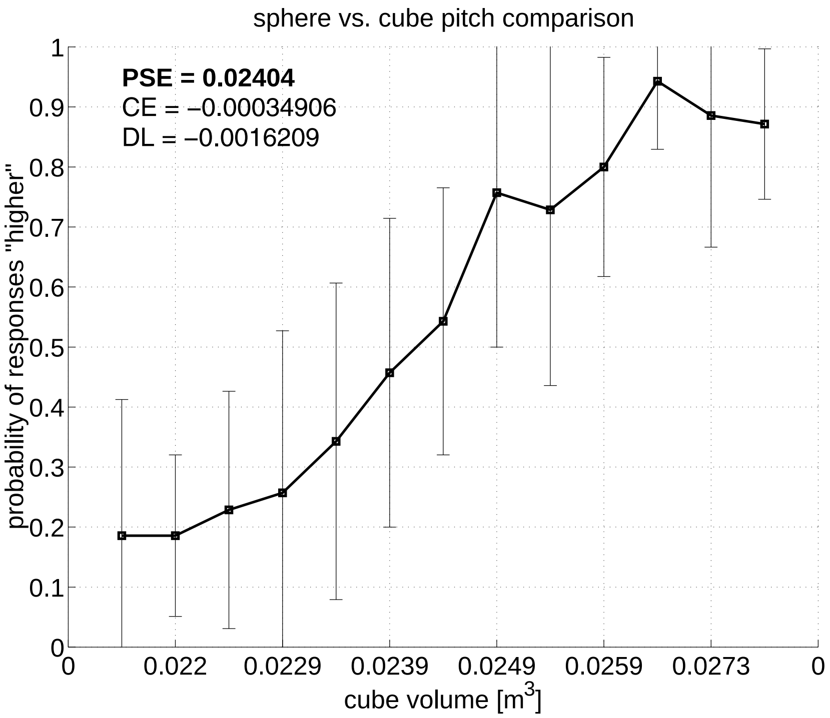


Figure 2.1: *Pitch equalization results. First experiment. Responses probability as a function of the cubes volume and standard deviation bars. Standard stimulus: sphere with diameter* $d = 36$ *cm, volume* $= 0.0244$ m^3.

equal to the reference sphere of each setting, with Constant Error $\mathrm{CE}_{36} = -0.0003$ m^3 and $\mathrm{CE}_{100} = 0.0071$ m^3 respectively. 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. Therefore, as resulting from the first experiment, pitch equalization seems to occur for equal volumes.

Moving to the second experiment, as we did for the previous one, in Figure 2.3 we report the response probability as a function of the cubes volume with the standard deviation bars.

In this case the Point of Subjective Equivalence is PSE = 0.0237 m^3, the Constant Error is CE = −0.0006, and the Differential Limen is DL = −0.0027 ($\chi^2 = 10.6439$, df = 10, $p > .3$). We can notice that the data in the plot are less aligned than in the first experiment and the model hardly fits the data.

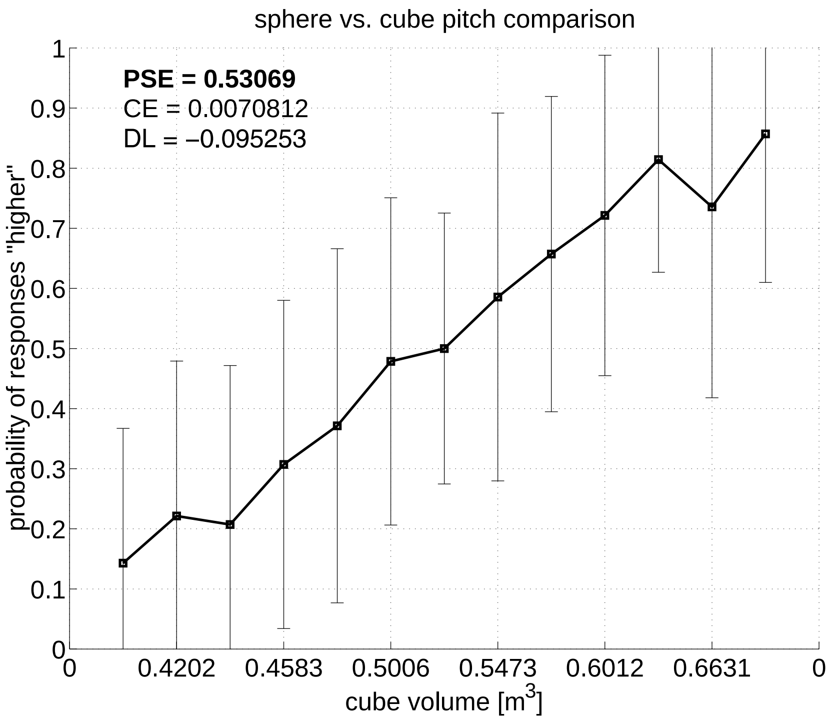


Figure 2.2: *Pitch equalization results. First experiment. Responses probability as a function of the cubes volume and standard deviation bars. Standard stimulus: sphere with diameter* $d = 100\ cm$*, volume* $= 0.5236\ m^3$.

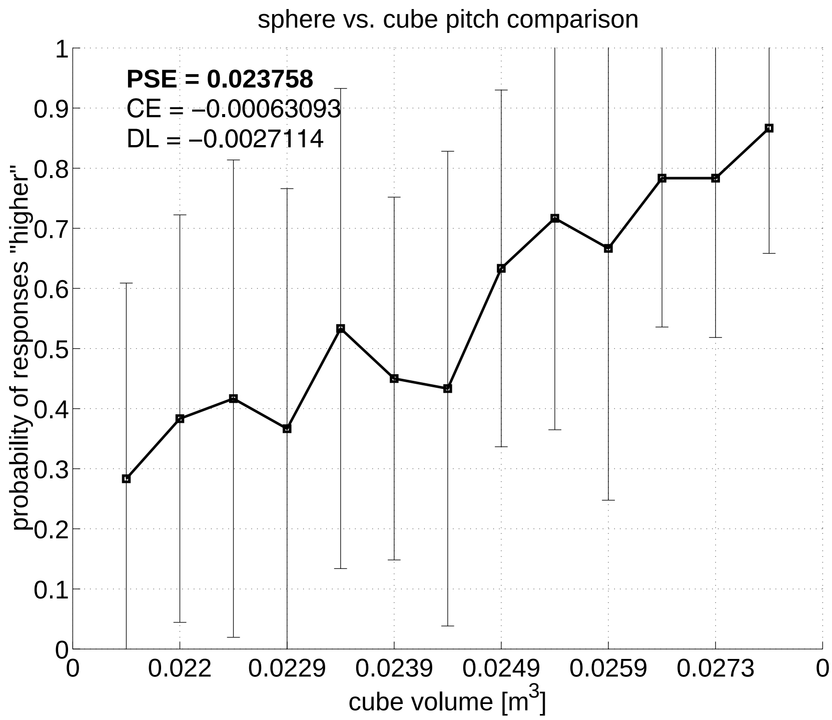


Figure 2.3: *Pitch equalization results. Second experiment. Responses probability as a function of the cubes volume and standard deviation bars. Standard stimulus: sphere with diameter* $d = 36\ cm$, *volume* $= 0.0244\ m^3$.

It is interesting to note that a participant to this experiment who defined himself as a naive listener since he has never practised music, during the debriefing phase at the end of his performance, expressed his difficulties in completing the task. "I've sometimes heard several pitches in the same sound and then it was hard to decide what was the main 'high' of the sound". Anyway, his performance results show that he clearly heard a pitch equalization corresponding to the central cube. In Figure 2.4 we report his individual performance.

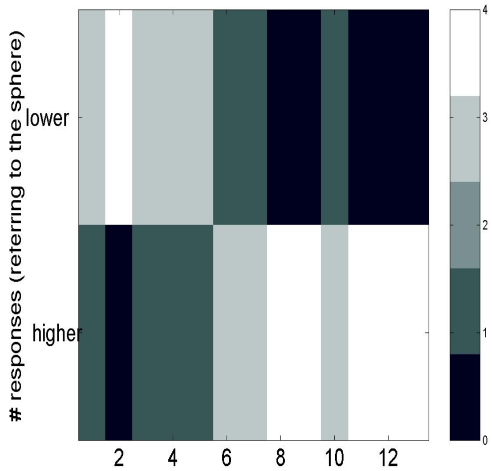


Figure 2.4: *Individual performance of a subject. Display of the number of responses "the sphere is higher/lower in pitch than the cube". The number of responses is represented for each cube with different colors, within a range between black — the minimum, and white — the maximum. The maximum level of uncertainty corresponds to the same color in both rows. This performance results show that the subject, even if he is a naive listener who never practised music, clearly heard a pitch equalization corresponding to the central cube, despite judging the task difficult.*

We display individual performances with a representation consisting of two rows, each one for a different response (the sphere is Lower or Higher in pitch than the cube) and 13 columns, one for each cube. We display the number of responses for each cube with different colors, within a range between black — the minimum, and white — the maximum. Therefore, the maximum level of uncertainty would correspond to the same color 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 can be found only by looking for the column with the same hue in both the rows, instead of looking for a particular hue.

In Figure 2.5, we report the same data of Figure 2.3 without two outliers: a participant who admitted to have had a big difficulty in judging some pairs and to have answered by chance to some of them and a naive listener who was influenced by the variation, besides pitch, of another parameter in the two sounds, that she wasn't able to define but the examiner can identify with the brightness, which is a feature clearly changing, in these sounds, according to the shape of the resonators and which we will analyze afterwards.

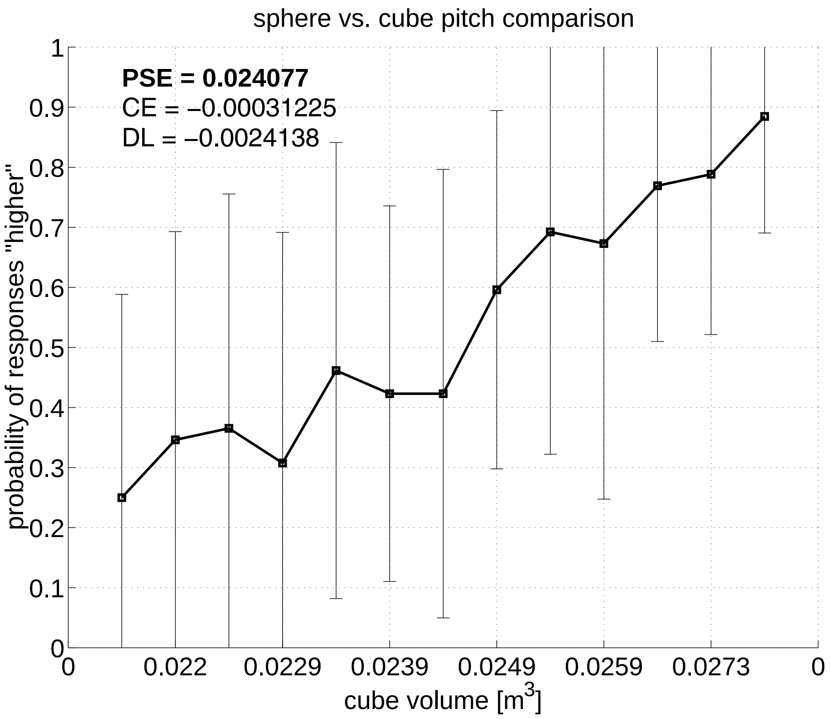


Figure 2.5: *Pitch equalization results. Second experiment. Responses probability as a function of the cubes volume and standard deviation bars. Standard stimulus: sphere with diameter* $d = 36$ *cm, volume* $= 0.0244$ m^3. *The data don't include two outliers.*

From the data without outliers, the Point of Subjective Equivalence is PSE = 0.0241 m^3, the Constant Error is CE = −0.0003, and the Differential Limen is DL = −0.0024 ($\chi^2 =$ 7.6322, df = 10, $p > .6$). 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 — volume = 0.0244 m^3). We can see that, if we don't consider the outliers who were biased by other factors, the two enclosures, that were judged to have the same pitch, have the same volume as well.

This second experiment confirmed the results of the previous one, i.e. cubic and spherical enclosures of the same volume are identified by listeners to have equal pitch.

In Figure 2.6 we report the frequency responses, limited to the low-frequency range, of the spherical resonator with diameter d = 36 cm, volume = 0.0244 m^3 and the cubic resonator with the same volume. In the plot, the first resonances of each shape are highlighted. The

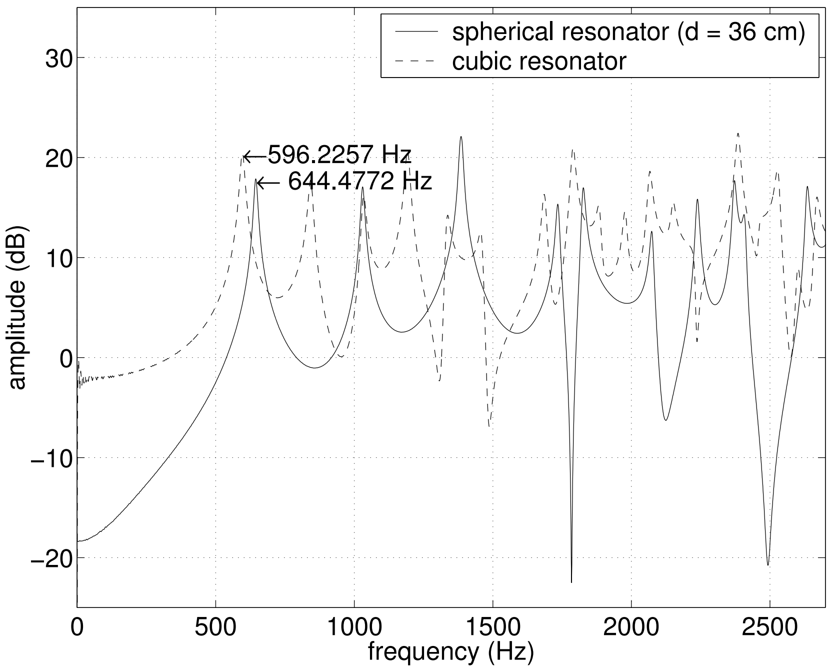


Figure 2.6: *Frequency responses, limited to the low-frequency range, of the sphere with diameter d = 36 cm, volume = 0.0244 m*^{3} *and the cube with the same volume. The first resonances of each shape are highlighted.*

first partials f_0_sphere = 644.5 Hz and f_0_cube = 596.2 Hz are more than 48 Hz far away. The JND at these frequencies is lower than 4 Hz. Therefore, the pitch of the two resonators cannot be associated just to their fundamental frequencies, since their distance is more than the JND, while their pitch is estimated to be equal.

In Figure 2.7 we report the similar plot for the case of the sphere with diameter d = 100 cm, volume = 0.5236 m^3 and the volume-matching cube.

In this case the first partials f_0_sphere = 232.0 Hz and f_0_cube = 214.6 Hz are more than 17 Hz far away. The JND at these frequencies is a bit more than 3 Hz and this confirm what resulted above.

It is interesting to measure the brightness of the stimuli, in order to check if this feature influenced the subjects' performance; in fact, the two shape sounds are clearly different in brightness. For measuring this feature we computed the spectral centroid by adapting to our requirements a MATLAB routine available on [AKZ02], which calculates the centroid for each frame obtained by windowing the signal with a Hamming window [OS89].

In Figure 2.8 and Figure 2.9 we report the centroids of the impulse responses of the spherical and cubic resonator with equal volume, respectively for the cases d = 36 cm, volume = 0.0244 m^3 and d = 100 cm, volume = 0.5236 m^3, represented as a function of

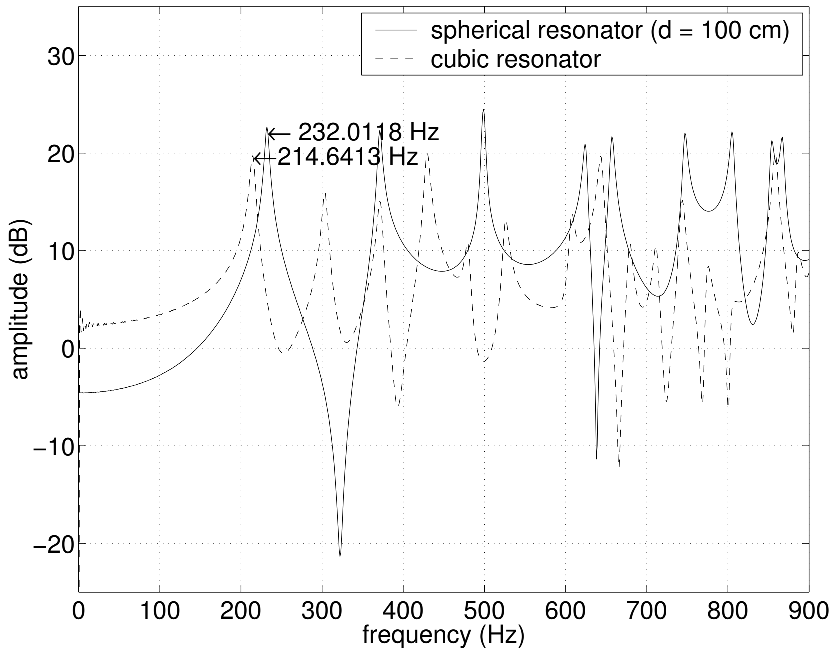


Figure 2.7: *Frequency responses, limited to the low-frequency range, of the sphere with diameter d = 100 cm, volume = 0.5236 m^3 and the cube with the same volume. The first resonances of each shape are highlighted.*

the frame number. We used an Hanning window of 1024 samples with an hop size of 245 samples and a sample rate of 22050 [OS89].

The average values of the centroids are 6152 Hz and 5633 Hz respectively for the sphere with diameter d = 36 cm and its volume-matching cube, while 5811 Hz and 5277 Hz for the sphere with diameter d = 100 cm and its volume-matching cube. We notice that the spheres considered, because of their smaller surface area in comparison with the relative cubes and the consequent lower effect of the absorption coefficient for the spheres rather than for the cubes, resulted to be systematically brighter than the cubes themselves. In fact, if we consider the sphere with diameter d = 0.36 m and volume = 0.0244 m^3, its surface area will be S = 0.4072 m^2 and it is smaller than the surface area of all the cubes with edge $l > 0.2605$ m and volume > 0.0177 m^3. The same reasoning can be applied to the sphere with diameter d = 1 m. This results about the relationship between brightness and shape is enhanced in Figure 2.10 where we report the average centroid value of each cube and we indicate the average centroid value of the sphere by means of an horizontal line. We can see a clear separation between the centroids of the shapes.

Although the subjects were able to hear the different brightness of the sounds, as reported by some of them, they were able, except one outlier, to compare their pitch. In Figure 2.11 we report the individual performance of a participant who is a trained singer and who referred

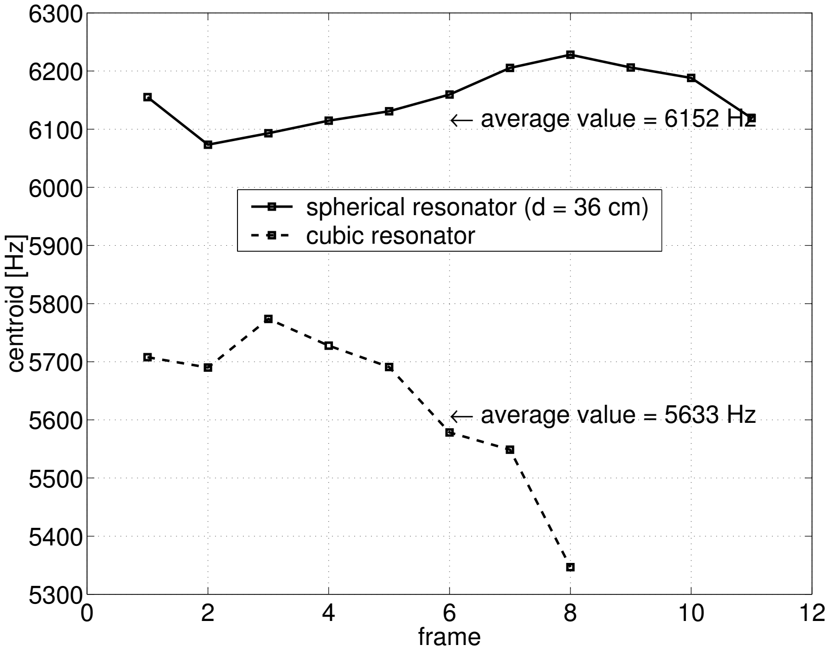


Figure 2.8: *Centroids of the impulse responses of the sphere with diameter* $d = 36$ *cm, volume* $= 0.0244$ m^3 *and the cube with the same volume, represented as a function of the frame number. We can see that the sphere is brighter than the cube.*

to have heard the sounds with a different brightness. We can see that the brightness didn't affect her judgment and she was just a bit confused for the small cubes.

We could claim that brightness does not play a role in pitch equalization. However, as we will explain in the next subsection (subsection 2.2.5), under particular listening conditions, subjects use brightness as a "reference point" for comparing the sounds.

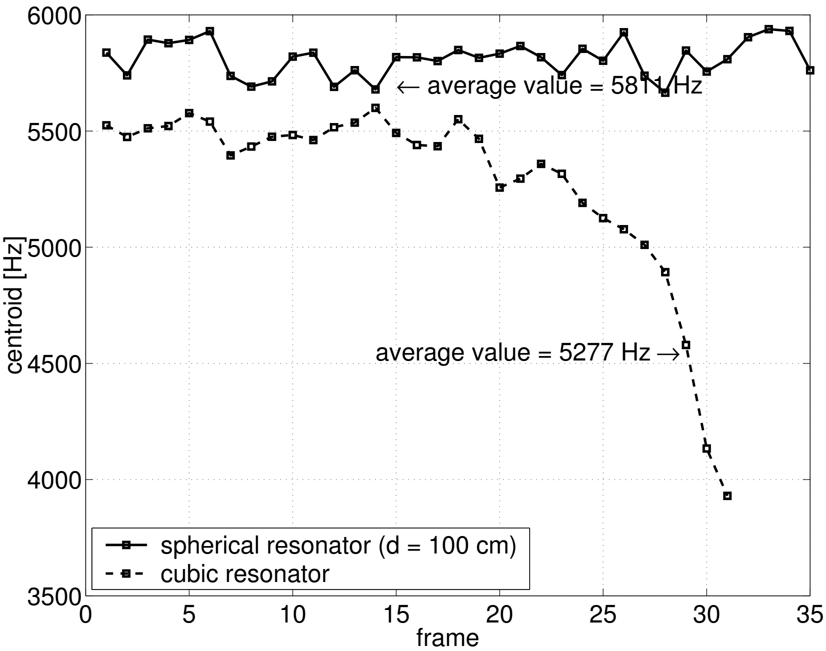


Figure 2.9: *Centroids of the impulse responses of the sphere with diameter* $d = 100$ *cm, volume* $= 0.5236$ m^3 *and the cube with the same volume, represented as a function of the frame number. We can see that the sphere is brighter than the cube.*

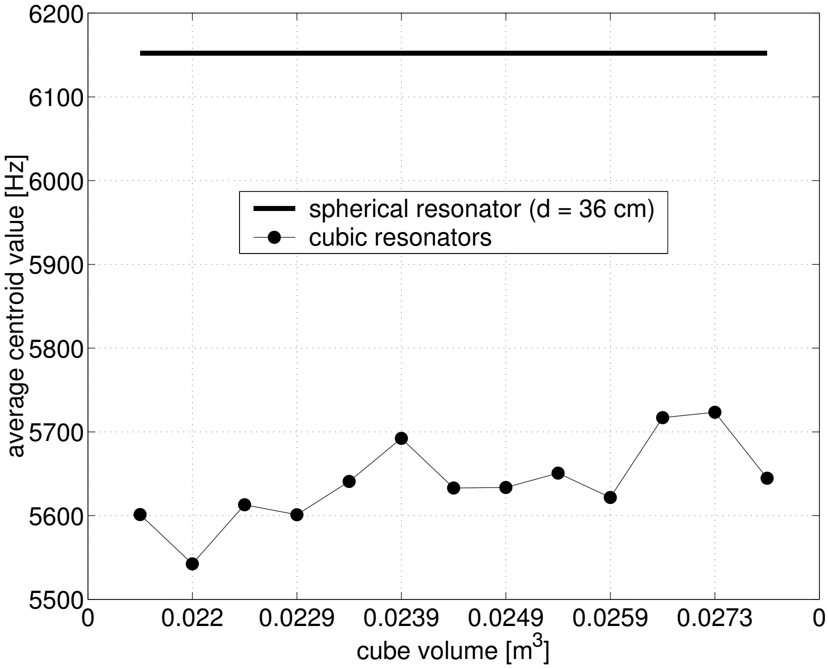


Figure 2.10: *Average centroid values of the cubes in comparison with that of the sphere with diameter* $d = 36$ *cm, volume* $= 0.0244$ m^3*. We can see that the sphere is brighter than all the cubes.*

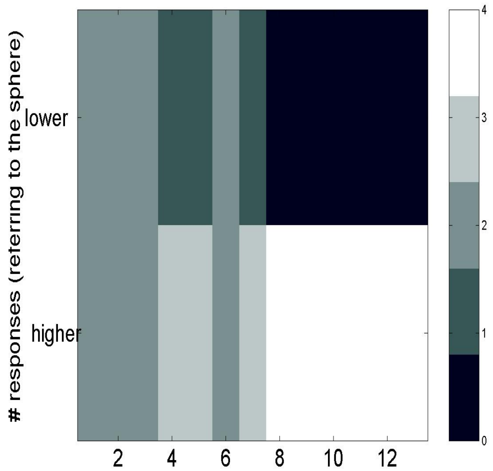


Figure 2.11: *Individual performance of a subject. Display of the number of responses "the sphere is higher/lower in pitch than the cube". The number of responses is represented for each cube with different colors, within a range between black — the minimum, and white — the maximum. The maximum level of uncertainty corresponds to the same color in both rows. This performance results show that the judgments of the subject, who is a trained singer, weren't affected by the brightness, even if she referred to have heard the sounds with a different brightness. She was just a bit confused for the small cubes.*

2.2.5 Order effect

It is interesting to describe a trial test conducted, during the arrangements for the second experiment, in order to check if the order of presentation of the stimuli within each pair would have had influences on the results.

Participants

The participants to this trial test were 17 (5 females and 12 males), all studying or working at the University of Verona, with age between 20 and 55. Two of them are expert listeners. Nobody referred to have hearing problems.

Stimuli

The stimuli set was the same as those in the aforementioned tests, i.e. a sphere of d = 36 cm and 13 cubes.

Procedure

The experiment was conducted in a quiet, but not isolated room. The stimuli were presented to the subjects through closed headphones (Beyerdynamic DT-770), by means of a MATLAB environment, on a PC Pentium III with a Creative SoundBlaster Live! soundcard.

We applied the constant stimuli approach, as in the other experiments, but in this case we used another technique for balancing the response results, which consists in providing, to the same participant, and in random order, half trials with one sound in the first position, and the other half trials with that shape sound in second position.

Results and Observations

By applying the balancing technique of "randomized order position", we noticed that the subjects got confused.

From Figure 2.12, it appears that the probability of responses "sphere is higher" never crosses the horizontal line corresponding to 50% of the responses: It remains above it. Since the estimated PSE is out of the range of the examined cube sizes (PSE = 0.0196 m^3) ($\chi^2 = 16.0675$, df = 10, $p > .05$), there is no pitch equalization in such range.

To understand this fact, in Figure 2.13 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 (Figure 2.13 (a)) listened to the sphere always in first position, while for subject n. 2 (Figure 2.13 (b)) the sphere was always in second position.

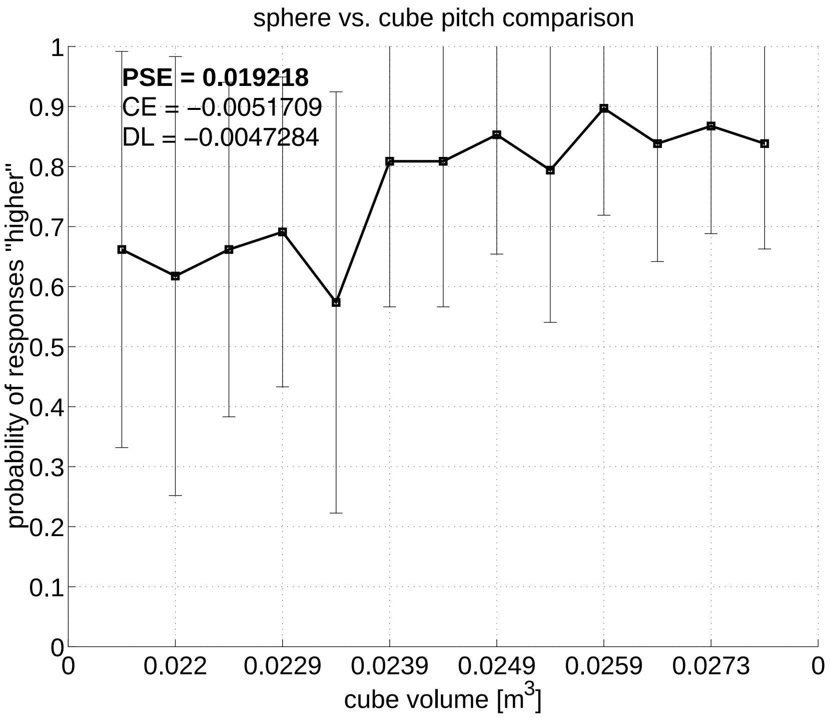
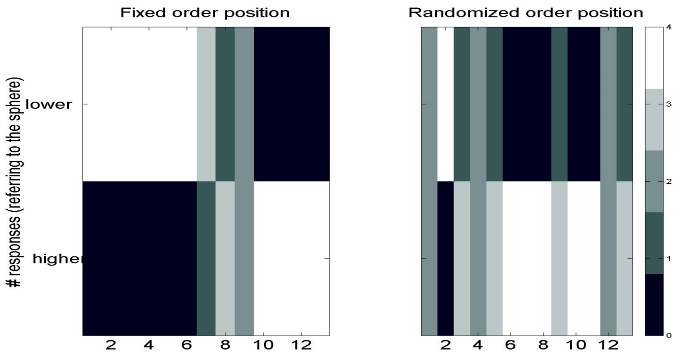


Figure 2.12: *Order effect. Responses probability as a function of the cubes volume and standard deviation bars. Standard stimulus: sphere with diameter* $d = 36$ *cm, volume* $= 0.0244\ m^3$.

(a)



(b)

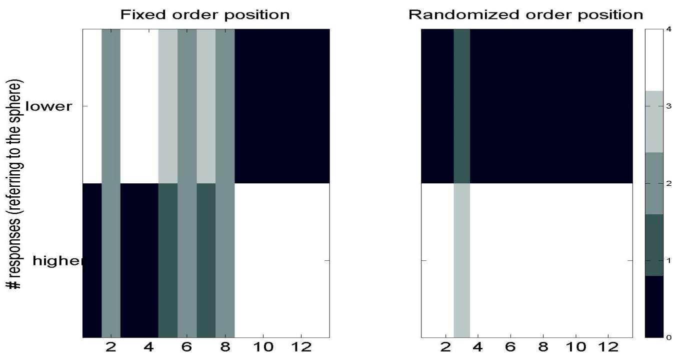


Figure 2.13: *Order effect. Display of the number of responses "the sphere is higher/lower in pitch than the cube". The number of responses is represented for each cube with different colors, within a range between black — the minimum, and white — the maximum. The maximum level of uncertainty corresponds to the same color in both rows.* Individual performances of subjects n. 1 (a) and n. 2 (b) in the two different test conditions, i.e. with the sound positions fixed during the whole experiment and with the sound positions inverted in half trials. *Subject n. 1, in the fixed order position case, has a strong perception of pitch equalization and he estimates it corresponding to the central cube, while, in the situation of randomized presentation, it is clear that he gets confused and cannot give any judgment, showing a lot of uncertainty for all the pairs he listened to. Subject n. 2, in the randomized presentation condition, shows a strong perception that the sphere's impulse response is higher in pitch than any cube response. Conversely, in the fixed order position case, he is able to hear, even if not as strongly as subject n. 1, a pitch equalization corresponding to the central cube.*

We can observe a different approach of the two subjects to the listening task, although both of them found the test with the fixed order position easier to be performed as "the pitch was more distinct" rather than in the randomized order position. The same observation has been referred by another subject who "found the task easier than last time — i.e. in the randomized order position — even if I had some difficulties with certain sounds". Subject n. 1, in the fixed order position case, has a strong perception of pitch equalization and he estimates it corresponding to the central cube, which has equal volume to the sphere, while, in the situation of randomized presentation, it is clear that he gets confused and cannot give any judgment, showing a lot of uncertainty for all the pairs he listened to. On the contrary, subject n. 2, in the randomized presentation condition, shows a strong perception that the sphere's impulse response is higher in pitch than any cube response. Conversely, in the fixed order position case, he is able to hear, even if not as strongly as subject n. 1, a pitch equalization corresponding to the central cube.

Most 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 participant, a naive listener, that we consider one of the outliers, showed difficulties in both conditions, since she was influenced, as we already said, by the variation of the brightness. Indeed, it seems that brightness works as an anchor for listening in the randomized order case.

Therefore, randomizing the presentation order seems to entail a change in the listening attitude, which could be due to the lack of a "reference point", and the consequent changing of focus of attention. We could compare this perceptual ambiguity to that created by the famous Boring's figure — old/young woman (Figure 2.14 on the left) or by the Rubin's vase (Figure 2.14 on the right). In these examples, an image offers two or more interpretations, which could be perceived alternatively by the observers, according to the cues they focus on. Similarly, while keeping the presentation order fixed doesn't affect the listener's capacity of judgment, the randomization of the presentation order can produce an ambiguity and therefore the subjects could get confused in performing the required task.

Although a small number of subjects showed difficulties in the pitch equalization task in both conditions, the order effect could have implications in auditory display, since the change of listening attitudes could cause performance gaps between users, representing a problem for auditory designers. Moreover, a setting designed for avoiding such an effect could represent a strict requirement that could be too much for an auditory display designer.

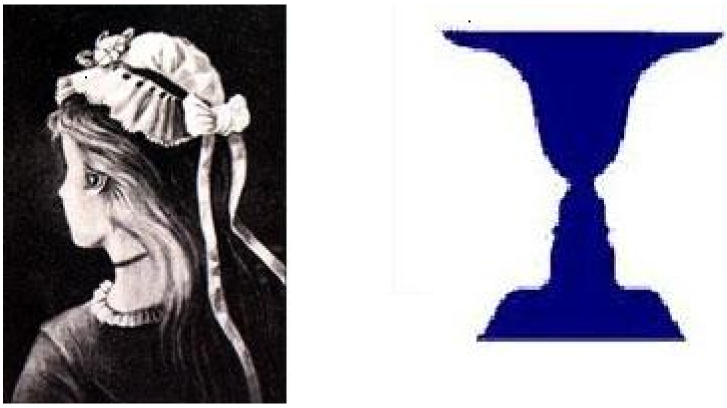


Figure 2.14: Boring's figure (on the left). *Perceptual ambiguity: it is possible to alternatively perceive an old and a young woman.* Rubin's vase (on the right). *Perceptual ambiguity: it is possible to alternatively perceive a vase or two silhouettes. Example of figure/background.*

2.3 Shape classification

Another task we wanted to investigate, besides the relationship between size and pitch, was shape classification, i.e. whether a listener would have been able to auditorilly classify the shape of an enclosure by hearing a sound filtered by it.

2.3.1 Participants

Nineteen subjects participated volunteerly to this experiment — 16 males and 3 females, in the range between 22 and 34 years old. They were students in computer science or they worked at the University of Verona. They didn't report any hearing problem. Two of the participants were expert listeners.

2.3.2 Stimuli

In the stimuli set we decided to convolve a suitable sound with the impulse responses of the enclosures, instead of using the impulse responses per se, since listeners are not used to hear impulse responses of cavities. We chose a snare drum sound to be filtered by the resonators, 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.

2.3.3 Procedure

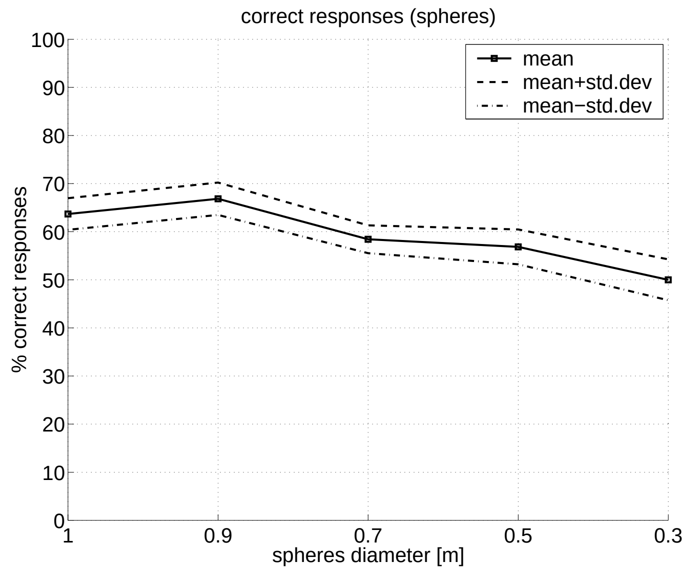
The experiment was conducted on a PC Pentium III with a Creative SoundBlaster Live! soundcard in a MATLAB environment, by presenting to the subjects all the stimuli in random order and repeated 10 times. They performed the test in a quiet, but not isolated room and they were asked to classify the shape of the resonator, i.e. whether it was a sphere or a cube, in a 2-Alternative Forced-Choice (2AFC) task setting.

The test was preceded by a training phase, where the subjects could listen to pairs of stimuli, consisting of the snare drum pattern filtered respectively by a spherical and a cubic resonator of the same volume. We chose spheres with diameters 106 cm, 60 cm, 36 cm, that were different from those used during the experiment. In fact, the aim of the training was to let the listener only acquire the "method" to distinguish the enclosing shapes, and not to learn *that* particular sound.

2.3.4 Results and observations

In Figure 2.15, we report the percentage of correct classification of the resonators shape as a function of the resonators size. In Figure 2.15 (a), we plot the data regarding the spherical resonators, while in Figure 2.15 (b), those of the cubic enclosures.

(a)



(b)

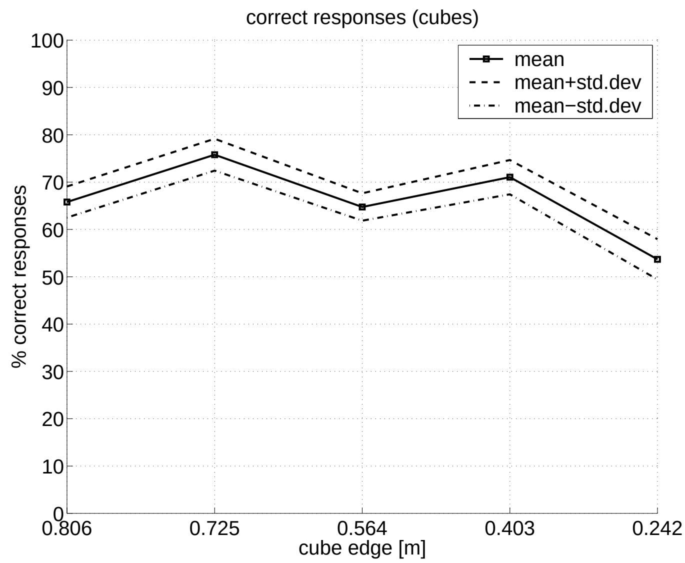


Figure 2.15: *Shape classification results. Percentage of correct classification of the resonators shape as a function of the resonators size. (a) spheres; (b) cubes. The performance in the shape classification task is not independent of the resonator size, but it is related instead to it. It seems to be more difficult to classify small rather then large resonators, probably related to the common experience of sounds filtered by larger enclosures.*

We can see that the performance in the shape classification task is not independent of the resonator size, but it is related instead to it (for the spheres: $\chi^2 = 13.3140$, df = 4, $p < .001$; for the cubes: $\chi^2 = 23.3090$, df = 4, $p < .000109$). In particular, it seems to be more difficult to classify small rather then large resonators, probably related to the common experience of sounds filtered by larger enclosures. Moreover, by removing the four poorest performances, even if the data are more variable and size-dependent, we can see a slight improvement of 5% - 10% for some enclosures.

Anyway, we kept all the data for the analysis, even if we noticed some outliers. In particular, some subjects identified resonators of certain sizes consistently with the same label. It happened especially for the smaller enclosures and it could be due to a non-accurate training or to a mental connection of the two perceptual features pitch and shape.

However, there were some very good performances, as that reported in Figure 2.16. As in the general plot (Figure 2.15), even in this example of good performance, we can see a difficulty for the subject in the classification of the smaller resonators.

We conclude that, after training, the average listener turned out to be able to classify 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. Therefore, we can see that the identification of the shape feature is strictly related to the listening setting.

2.4 Estimation of the sphere's pitch

In the third task we estimated the enclosure pitch, in order to be able to quantify what the subjects perceived performing the first task reported in section 2.2.

2.4.1 Participants

The participants to this task were 28 subjects (24 males and 4 females with age ranging between 24 and 36), all computer science students, but 3 that were members of the Department — one of the subject was the author. All of them referred to have no hearing problems.

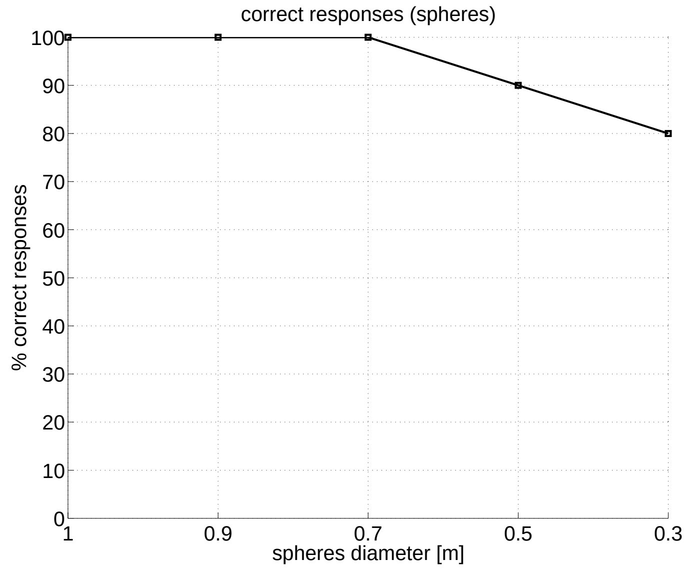
2.4.2 Stimuli

We decided to evaluate the pitch of the spherical enclosure, since during the previous experiments the sphere's pitch arouse to be perceived relatively with ease.

We compared the impulse response of a spherical resonator of diameter d = 50 cm with exponentially-damped sinusoids of varying frequencies.

The decay time of the sine wave was tuned to the same value as that of the spherical resonator at that frequency.

(a)



(b)

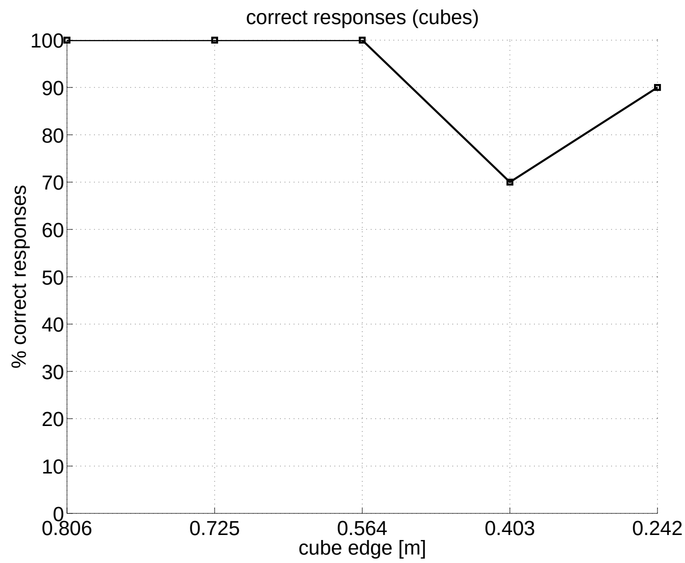


Figure 2.16: *A good performance of a subject in the shape classification task (a) spheres; (b) cubes. Even in this example of good performance, we can see a difficulty for the subject in the classification of the smaller resonators.*

2.4.3 Procedure

We conducted the test in a quiet, but not isolated room on a PC Pentium III with a Creative SoundBlaster Live! soundcard. By means of a MATLAB environment, presenting the stimuli to the subjects through closed headphones (Beyerdynamic DT-770) we applied the simple up-down or staircase method [Lev70].

The stimuli were presented pairwise with the standard stimulus as first, i.e. the impulse response of the spherical resonator, followed by a sine wave, and the participants were asked whether the first sound was higher in pitch than the second sound.

We decided to make 8 runs before terminating one subject's test and, for each test, to choose randomly the initial value of the frequency of the sine wave among the following possibilities:

$$initial\_f = [300, 350, 400, 450, 500, 550, 600] \text{ Hz}. \tag{2.5}$$

Then, during the test, the step size on trial n was set equal to c/n, where c is a constant value set as $c = 200$ Hz.

2.4.4 Results and observations

In Figure 2.17, we report, by means of an histogram, the distribution of the frequency values resulted from the pitch estimation task performances. On the x-axis there are the pitch estimate ranges, while on the y-axis there are the values corresponding to the number of responses. Each bar of the histogram represents the number of responses relative to a certain pitch estimate range and, since for each subject we collected one response, the values of the histogram correspond to the number of subjects who estimated the pitch of the impulse response of the spherical resonator to be within a certain range.

One fourth of the participants (7 subjects) estimated the pitch in the range $[450, 480]$ Hz. The lowest partial of the frequency response of the sphere is found at 464 Hz. Therefore, one fourth of the subjects estimated the pitch close to the lowest partial. The frequency responses, limited to the low-frequency range, of the sphere and the volume-matching cube are reported in Figure 2.18. In the plot, the first resonances of each shape are highlighted.

Moreover, we observe another one fourth (7 subjects) which estimated the pitch in the range $[650, 750]$ Hz, that is in the second partial neighborhood, found at 742 Hz.

However, the subjects found this task difficult, due probably to the different identity of the stimuli used in the test.

Nevertheless, it seems that when comparing a frequency response of a resonator with a damped sinusoid, listeners apply a analytic listening mode which consists in the ability to hear a component in a mixture, i.e. in extracting several elementary components from a given sound, leading to the perception of several simultaneous sound objects [Met96].

On the other hand, the experiments of section 2.2 seem to indicate that, when the standard sound is another complex 3-D response, pitch is attributed using a more holistic

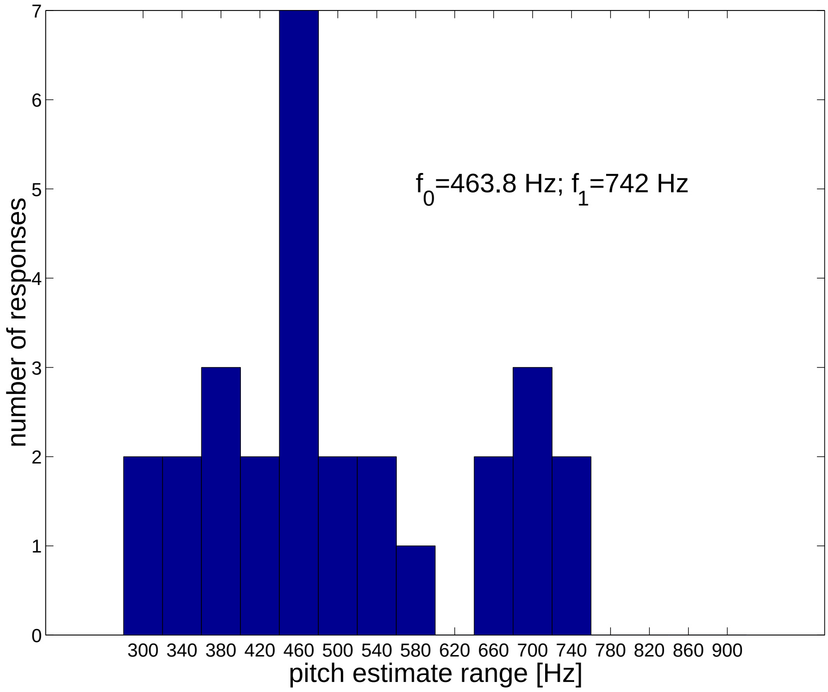


Figure 2.17: *Pitch estimation results. Distribution of the pitch estimation values of the impulse response of a spherical resonator (d* $= 50$ *cm). Each bar of the histogram represents the number of responses relative to a certain pitch estimate range and, since for each subject we collected one response, the values of the histogram correspond to the number of subjects who estimated the pitch of the impulse response of the spherical resonator to be within a certain range. One fourth of the participants (7 subjects) estimated the pitch in the range* $[450, 480]$ *Hz and, therefore, close to the lowest partial.*

listening mode, or synthetic listening mode, which consists in hearing properties of a mixture, leading to the perception of a single complex sound object [Met96]. In fact, as we can see, for instance, in Figure 2.18, although there is a significant difference between the values of the first and second partials of the spherical enclosure — that are respectively 464 Hz and 742 Hz — and those of the cubic enclosure with the same volume — that are respectively 429.3 Hz and 607.1 Hz, the two resonators are judged to have same pitch, as we showed in the previous experiment (section 2.2).

2.5 Conclusions

In this chapter, we have presented our investigations concerning the auditory perception of shape and size of three-dimensional resonators. In particular, we studied, with synthetic

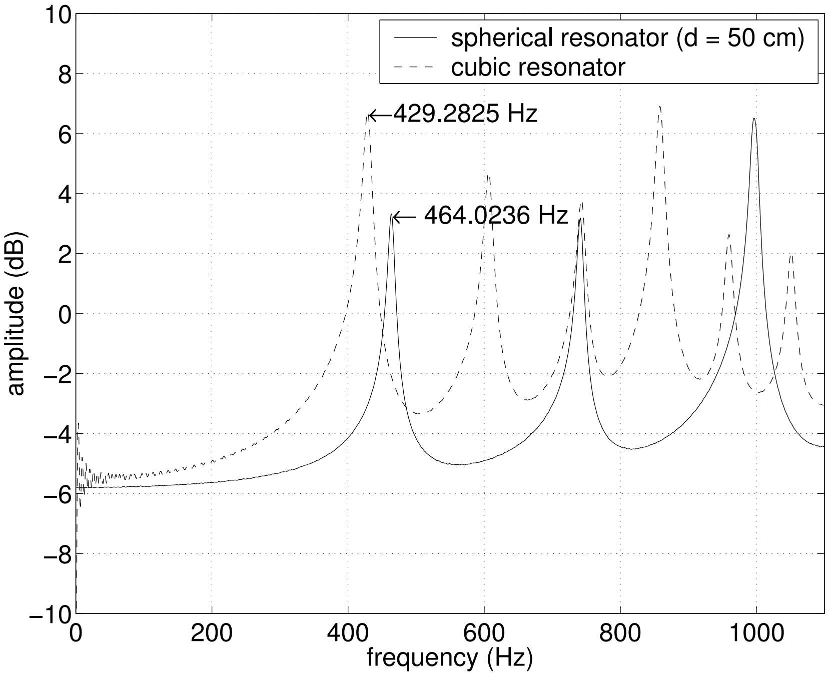


Figure 2.18: *Frequency responses, limited to the low-frequency range, of the sphere with diameter* $d = 50$ *cm, volume* = 0.0654 m^3 *and the cube with the same volume. The first resonances of each shape are highlighted.*

resonators, the relationship between pitch and size of enclosures, the classification of the resonators shape by listening to the sound it filters and the estimation of the pitch of an enclosure.

Even if we achieved some interesting results, they are strictly dependent on the listening conditions or the stimuli setting. For instance, we have observed that subjects seem to estimate the resonators pitch with an holistic rather than analytic listening attitude, as they equalize the pitch of volume-matching enclosures, instead of identifying the pitch according to their partials distribution. In this context, our investigations are quite interesting. In fact, even if the partial distribution is different, subjects perceive equal pitch, when they listen to volume-matching resonators and, as a consequence, it is better in this case to analyze the fact in an ecological perspective, rather than in an analytic perspective, applying the Fourier decomposition. Anyway, the subjects performance seems to be anchored to the perception of a clear feature, the brightness, which plays the role of "reference point" only if it keeps a fixed position in the presentation order. Therefore, since the relationship between pitch and volume of the cavities requires the anchor of the brightness in order to be identified, we cannot consider it for a multi-dimensional auditory display.

Another feature we investigated has been the shape of the resonator and whether it is possible to classify it. We discovered that, with a training, the shape could be identifiable, but with the requirement of the cavity to be large enough. We say that it has to be larger than 50 cm.

Therefore, although we found some interesting properties, they require too much strict conditions for being used inside an efficient multi-dimensional auditory display. The shape feature seems to be the less constrained; but, it requires at least a training and, therefore, a cognitive process. For avoiding this requirement, since our initial aim was to find perceptual features that are ecologically relevant, we prefer to investigate other properties that guarantee an ecological approach rather than cognitive. In fact, our aim is to find some properties which are able to convey information and which can be easily recognized and quantified by the "typical" user[1].

To go back to what Kubovy and Van Valkenburg [KV01] define the auditory system dealing with, i.e. sources of sounds and auditory events, as we said in the introduction to this chapter (chapter 2), we have investigated two perceptual dimensions of sound sources by now. Unfortunately, these perceptual features introduce strict requirements in the auditory display design. In order to reach our goal of efficient, meaningful and ecological auditory display, we investigate the other element considered by Kubovy and Van Valkenburg, that are auditory events. We will deal with auditory events in chapter 4 and chapter 5. In particular, in chapter 4, we will focus on bouncing/impact events, by investigating some perceptual dimensions of them, while chapter 5 will study two perceptual dimensions of the rolling process. On the other hand, in the next chapter (chapter 3), we will wonder whether it is possible to represent information in auditory display through distance rendering, by validating a virtual environment realized within our research group.

[1]For "typical" user we mean a subject without a particular training and without particular abilities.

Chapter 3

Distance perception

After the investigation in the auditory perceptual domain of two physical dimensions of sound sources, such as shape and size, we move to another important feature, the distance. Contrary to the aforementioned dimensions, the distance cue is related to the sound source *within* the surrounding environment rather than to the sound source *per se*. As we already pointed out in chapter 1, we assume that the human hearing system is capable of getting information both of the sound sources and their environment.

Distance is one of the most evident perceptual dimension that allows the listener to locate a sound source. The location of a sound source with respect to the listener is defined by three coordinates: azimuth, i.e. the angular distance measured along the horizon; elevation, i.e. the angular distance measured above the horizon; and distance.

Distance is studied as perceptual spatial feature both in visual and in auditory perception. In fact, space plays a different role for the two perceptual channels, but for each one it is important with different connotations.

Kubovy and Van Valkenburg define an attribute to be *indispensable* if and only if it is a prerequisite of perceptual numerosity [KV01].

> The criterion that an attribute must satisfy to be an IA (i.e. indispensable attribute) is that in its absence (and the absence of any other IA for that modality) perceptual numerosity is impossible [VK03].

They state that there are different indispensable attributes for the visual and the hearing channel. For the visual channel, space and time are indispensable attributes. For the auditory channel, this role is played by frequency and time, that become indispensable attributes for hearing numerosity [VK03]. As they exemplify in [KV01], if we consider two lights pointing to a surface, even if they have the same colour or different colours, if they are directed to the same point, they will be perceived as one spot light, with the colour depending on the two colours of the sources' light. Only if the spots of light have different spatial locations, they will be perceived as distinct. On the other hand, from the auditory point of view, the spatial location of two sound sources doesn't affect the perception of their numerosity. In fact, even if a sound is played on two loudspeakers, it will be perceived as one sound. Only if the sounds have different frequencies, even if the same spatial location, they will be perceived as two distinct sounds.

However, even if spatial attributes are not as important for hearing as they are for sight, range cues become crucial in several situations. For instance, through hearing, human beings, even without using the visual channel, are able to perceive events visually occluded or out of sight. Moreover, hearing is omni-directional, while in order to see something it is necessary to look in a precise direction and, therefore, the auditory channel can be used for orientating the visual sense,

> ...it allows us to monitor and identify sources of information from all possible locations, not just the direction of gaze. [...] recall how we naturally use audition to gain information and explore the environment; that is “the function of the ears is to point the eyes” [Wen92].

These advantages are widely exploited in auditory display design. For example, auditory localization properties have been used in auditory warning systems, especially when visual cues are limited and workload is high, as in Air Traffic Control (ATC) systems and Traffic Collision Avoidance Systems (TCASs) [BWSM96]. They have been applied also to interfaces for visually impaired users [Myn95, MK96, MKS99, Sem01], to systems for teleconferencing [Beg99], to multi-modal workstation systems [BSB02] and to applications for teleoperation [Nem96].

So far the research in auditory perception of spatial attributes has focused especially on 3-D audio localization in order to synthesize a realistic auditory environment [SMK98], on the conditions that determine the perceived distance [Gar68, Zah01, Col62] and on the influence that different experimental conditions have on the collected results [LKPG98].

In this research activity we want to validate a synthetic resonating environment developed within our research group [FRO02], in order to be able to enhance, by means of reverberation, the distance perception that the users have from the sound source location [OFR02]. This validation is interesting from our point of view, as we can estimate the usefulness of distance rendering with that particular model of virtual resonator, in order to check if it can be used efficiently for other auditory display applications.

There are two approaches that model the effects of the environment characteristics. The *perceptual approach* [Gar98] aims at reproducing the reverberation effects, at the listener's point. This approach provides high-quality rendering, regardless of the physical parameters, given that the psychophysical process that maps the acoustics of a reverberant enclosure is still partially unknown [Ber92]. Moreover, it leads to affordable architectures working in real-time. Nevertheless, most of these realizations do not deal with distance rendering of sound sources.

On the contrary, the *structural approach* aims at modeling environments, focusing on the structural properties that must be rendered, such as the geometry of an enclosure or the materials the wall surfaces are made of. The reverberation effects result as a consequence. Unfortunately, structural models result to be either too resource-consuming or, if the system is simplified to accommodate the hardware requirements, excessively poor in the quality of the audio results.

In this chapter, we will work with a virtual resonating environment that we modeled with the aim of enhancing distance perception by means of reverberation. In section 3.1, we will briefly introduce the acoustics inside a tube that we modeled using the structural approach, by presenting the key aspects of the resonator design, whose advantage consists in a simple and a relatively light computational structure. Then, in the next two sections, section 3.2 and section 3.3, we will focus on its perceptual validation describing two psychophysical experiments we conducted using this model in two different experimental conditions. Finally, in section 3.4, we will compare the results achieved with the two experiments.

3.1 The model of the resonating environment

For modeling our resonating environment, consisting in a tubular cavity, with square section and size equal to 9.5×0.45×0.45 m, we applied the structural approach.

The tube size was set according to prior investigations, in order to convey an interesting range of distance cues by using a resonating environment that is structurally simple and computationally relatively light [FRO02].

The resonator was modeled by means of the Waveguide Mesh [DS93], a particular formulation of the finite difference scheme where wave components traveling along the mesh sum to produce physical (pressure) quantities.

For our purposes, we focus on the psychoacoustical experiments we conducted for this evaluation. For the technical details on the model of the virtual resonating environment, we suggest to read Fontana's PhD thesis [Fon03].

3.2 Listening by headphones

In this section we present one of the two experiments we conducted aiming at investigating how subjects scaled the perceived distance of the stimuli filtered by the virtual resonator and, hence, whether the model is effective or not. We will introduce the experimental setting, whose main characteristic that differentiate it from the second experiment consists in presenting the stimuli to the subjects by means of closed headphones, instead of using loudspeakers; then, we will show the collected data and the results.

3.2.1 Participants

The experiment involved 12 volunteers (4 female and 8 males), with age between 22 and 40. They study or work at the University of Verona. All of them were naive listeners.

3.2.2 Stimuli

The sound set was synthesized using the following technique. By putting a sound source at one end of the virtual tube, along the main axis, we acquired ten stereophonic impulse responses along positions $x_{10}, \ldots, x_1$, where each one got closer to the sound source by a factor of $\sqrt{2}$, and the first one was distant 9.5 m. Therefore, the final set X of distances expressed in meters was:

$$X = \{x_i,\, i = 1, \ldots, 10\} = \\ \{0.42, 0.59, 0.84, 1.19, 1.68, 2.37, 3.36, 4.75, 6.71, 9.5\} \tag{3.1}$$

The right channel of the stereophonic sound accounted for acquisition points exactly standing on the main axis, whereas the left channel accounted 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 this sound in particular in order to work with a type of sound which is quite uncommon in our everyday life and, therefore, that could be judged unfamiliar. In fact, we wanted to avoid an influence factor due to a familiar sound or to a sound usually related to a particularly defined distance.

> Listener familiarity with the particular source signals being localized may also be a significant factor in auditory distance perception [Zah02b]

It would be interesting to repeat the tests with another type of sound, in order to study how sound familiarity affects the data results, even if we know that experience has a strong effect on auditory distance assessment [NW01].

Each stimulus in the set was repeated 3 times in random order, leading to a group of 30 sounds for the experiment.

3.2.3 Procedure

The experiment was conducted in a quiet, but not isolated room, by applying the magnitude estimation method without modulus, that is a comparing stimulus to which the experimenter associates a value for reducing the estimated value range.

Each stimulus was presented to the subjects within a MATLAB environment on a PC Pentium III with a Creative SoundBlaster Live! soundcard. In this experiment, the participants listened to the sounds through Beyerdynamic DT-770 closed headphones.

The subjects were asked to estimate the perceived distance of the listening point from the sound source by rating each distance with a value in meters (either integer or decimal), starting from the first stimulus of the random sequence, and associating a value to the other ones, proportionally to the first estimation.

The experiment was conducted without training. Moreover, it was conducted, as we have already said, without a modulus, i.e. we did not set a comparing stimulus with a pre-defined value. The subjects could use for comparison just the first estimation they gave or the previous estimations, without being allowed to listen again to previous stimuli. Therefore, the collected values defined scales that depended on the individual listeners' judgments. These scales ranged from 0.2-8 (subject no. 8) to 1-30 (subject no. 5).

In order to define a common scale and to be able to compare the different evaluation scales, we applied a technique found in [Pur97], who refers to [Eng71]. It consists in computing the geometric average of the three judgments given for each sound by each subject, and in calculating, with these values, a mean average of all the stimuli for each subject and then a mean average of all the stimuli and all the subjects. To clarify this concept, if we think of a table $m \times n$, where m is the number of participants to the experiment and n is the number of stimuli, the cell (i, j) will contain the geometric average of the three judgments given by subject i for the stimulus j. The technique applied consists in calculating the mean average for each row of the table (let us put the results in a new $(n + 1)$th column), and then the mean average of this last column of the table. For obtaining a common logarithmic reference scaling, it is required to subtract the mean average of the $(n + 1)$th column from each value of that column (let us put the results in a new $(n + 2)$th column), and finally to sum the values of each cell in the table with the values in the relative rows of the $(n + 2)$th column. By applying this method, we obtained a common scale and we were able to compare the different evaluation scales of each participant.

3.2.4 Results and Observations

In Figure 3.1 the distance evaluations for each listener are shown as functions of the source distance, together with the corresponding linear functions obtained by linear regression and the standard deviation bars. The average slope is 0.61 (standard deviation 0.21), while the average intercept is 0.46 (standard deviation 0.21).

In Figure 3.2 the perceived distance averaged across subjects is plotted as function of the source distance, together with the relative regression line ($r^2 = 0.76$, $F(1, 8) = 25.84$, $p < 0.01$).

From the results of this first experiment we observed that users overestimate the distance of close sound sources, and that they reduce this overestimation for greater distances, leading to a perceptual estimation curve which tends to equalize greater distances. This is confirmed by the box plot reported in Figure 3.3, where this general trend is enhanced, although few stimuli received quite spread assessments (sound sources at 0.42 m and 0.84 m). This result is interesting since it partially contradicts Zahorik [Zah02a], who worked with real sounds for the tests, and reported the tendency of listeners to overestimate short distances, and underestimate long distances. In our model, the point of correct estimation is more distant compared with Zahorik. More precisely, our regression line has an offset upward, in a way that the point of correct estimation moves toward longer distances. This result can be interpreted as a consequence of the exaggerated reverberant energy produced by our model. Therefore, this synthesizing approach to the resonating environment seems particularly suitable for auditorilly rendering sounds at long distances.

In the box plot of Figure 3.3, we notice three outliers in three different distance cases, two of which refer to the closest distances, i.e. 0.42 m and 0.59 m, while the third concerns the sound sources at 4.75 m.

In the next experiment (section 3.3) we will check if the listening conditions affect the results of the validation of the model, since we would like to propose a model "irrespective on the chosen reproduction format over loudspeakers or headphones" [Jot99].

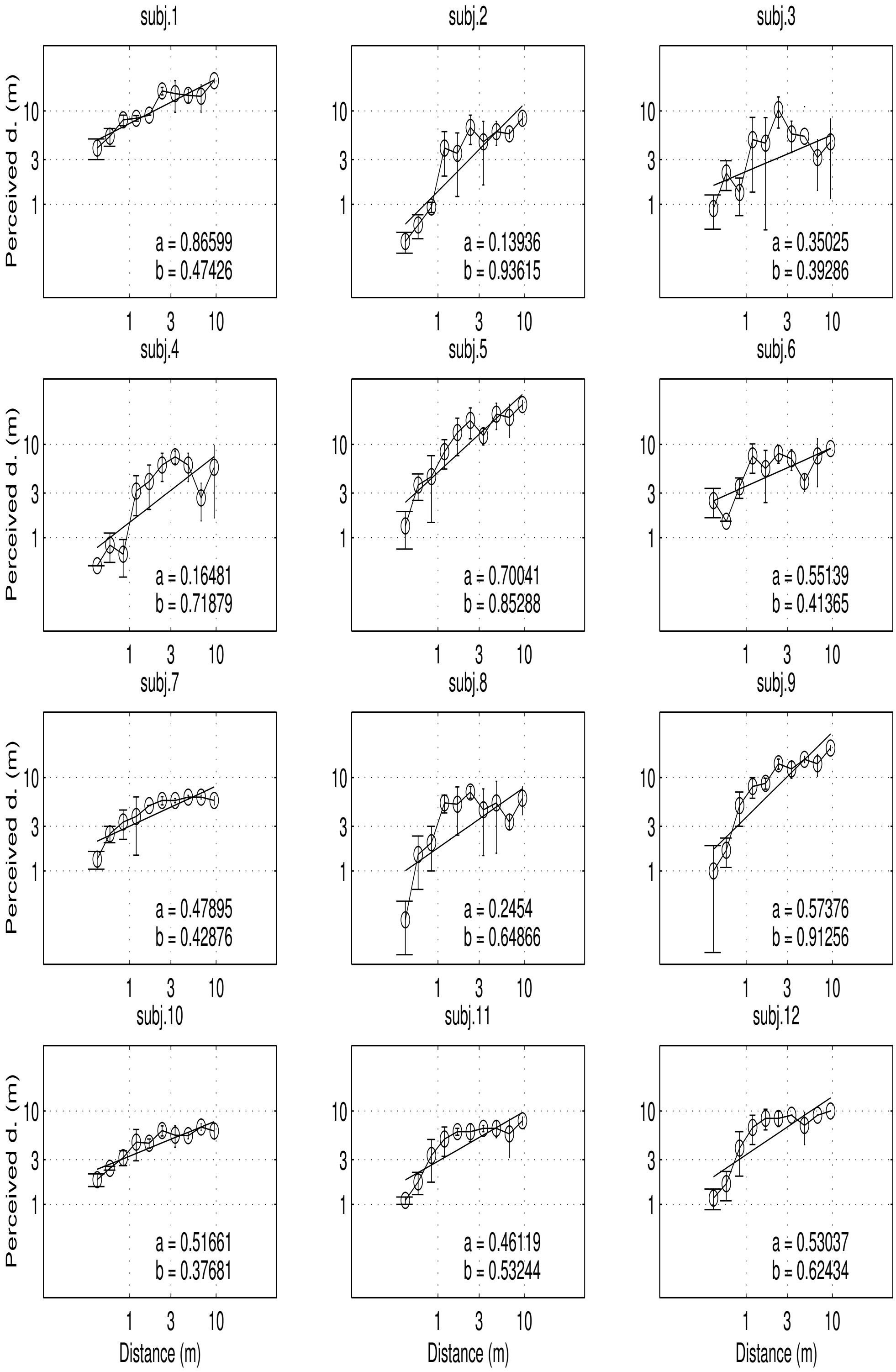


Figure 3.1: *Headphones listening: Individual distance evaluations together with individual linear regression lines and standard deviation bars. a: intercept. b: slope.*

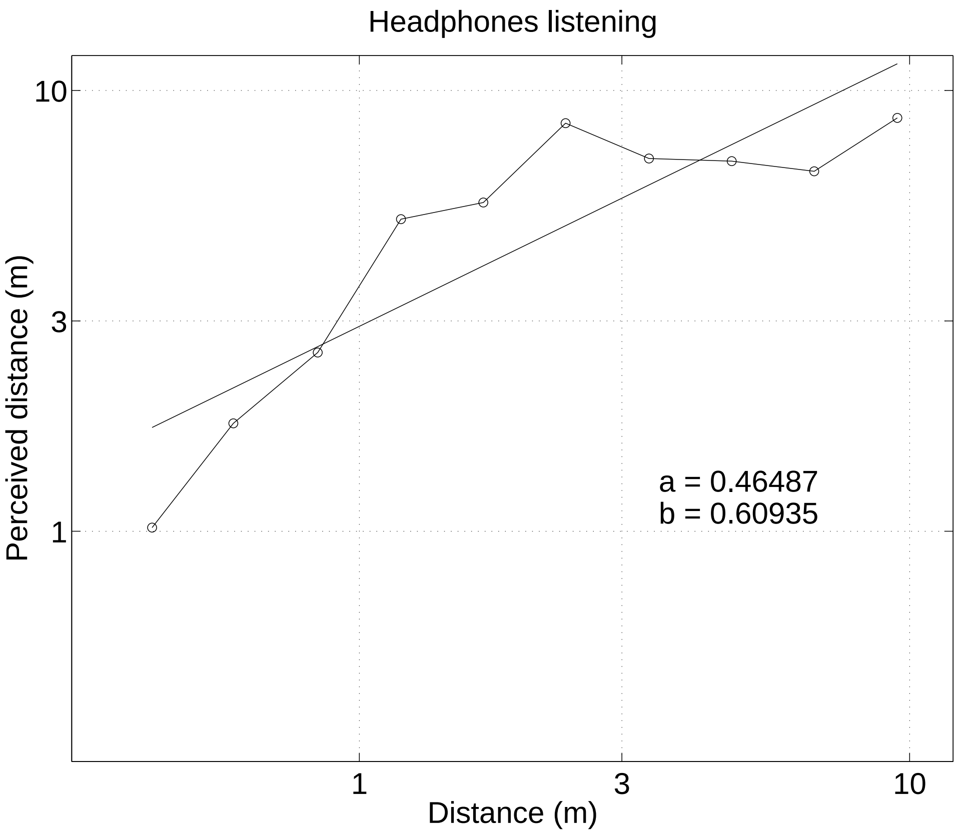


Figure 3.2: *Headphones listening: Average distance evaluation together with linear regression line. a: intercept. b: slope.*

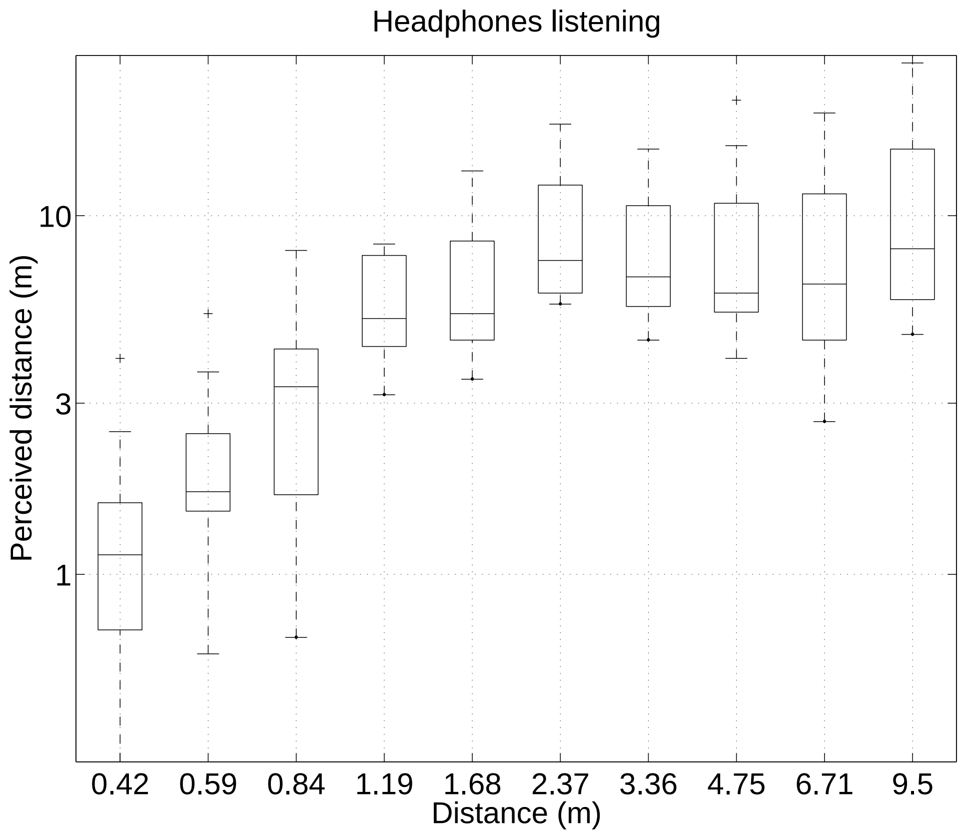


Figure 3.3: *Headphones listening: Distance evaluations, represented by a box plot.*

3.3 Listening by loudspeakers

The second experiment aimed at repeating the previous one, in a different listening setting, in order to check if our model of resonating environment is independent of the listening conditions or if it has particular requirements for being effective. The setting, in this second experiment, consisted in using loudspeakers.

3.3.1 Participants

The subjects of the second experiment were 10 (6 males and 4 females) who volunteerly participated. They were in the range between 25 and 34 years old. They all referred to have no hearing problems.

3.3.2 Stimuli

The stimuli set was the same as for the first experiment. Therefore, it consisted of 10 sounds, repeated 3 times in random order, coming from sound sources at different distances from the listening point. The distances we used are reported in equation 3.1.

The sound filtered by the virtual resonator was, as for the previous test, an anechoic sample of a cowbell, for avoiding experience effects and for repeating the same test in another condition setting.

3.3.3 Procedure

As in the previous test, for conducting this experiment we used a PC Pentium III, with a Creative SoundBlaster Live! soundcard and the room we used was quiet, but not isolated. In this case, 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.

The users were blindfolded, in order to minimize the influence of factors external to the experiment. They were asked to evaluate metrically the distance of the sound source from the listening point communicating its value to the experimenter. The first value, as in the previous test, determined the subjective scale.

In order to calculate a common logarithmic reference scaling, we applied the method explained in subsection 3.2.3.

3.3.4 Results and Observations

In Figure 3.4 we report, for each participant, the distance evaluations as functions of the source-listener distance, together with the corresponding linear functions obtained by linear regression and the standard deviation bars. The average slope is 0.53 (standard deviation 0.17), while the average intercept is 0.50 (standard deviation 0.36).

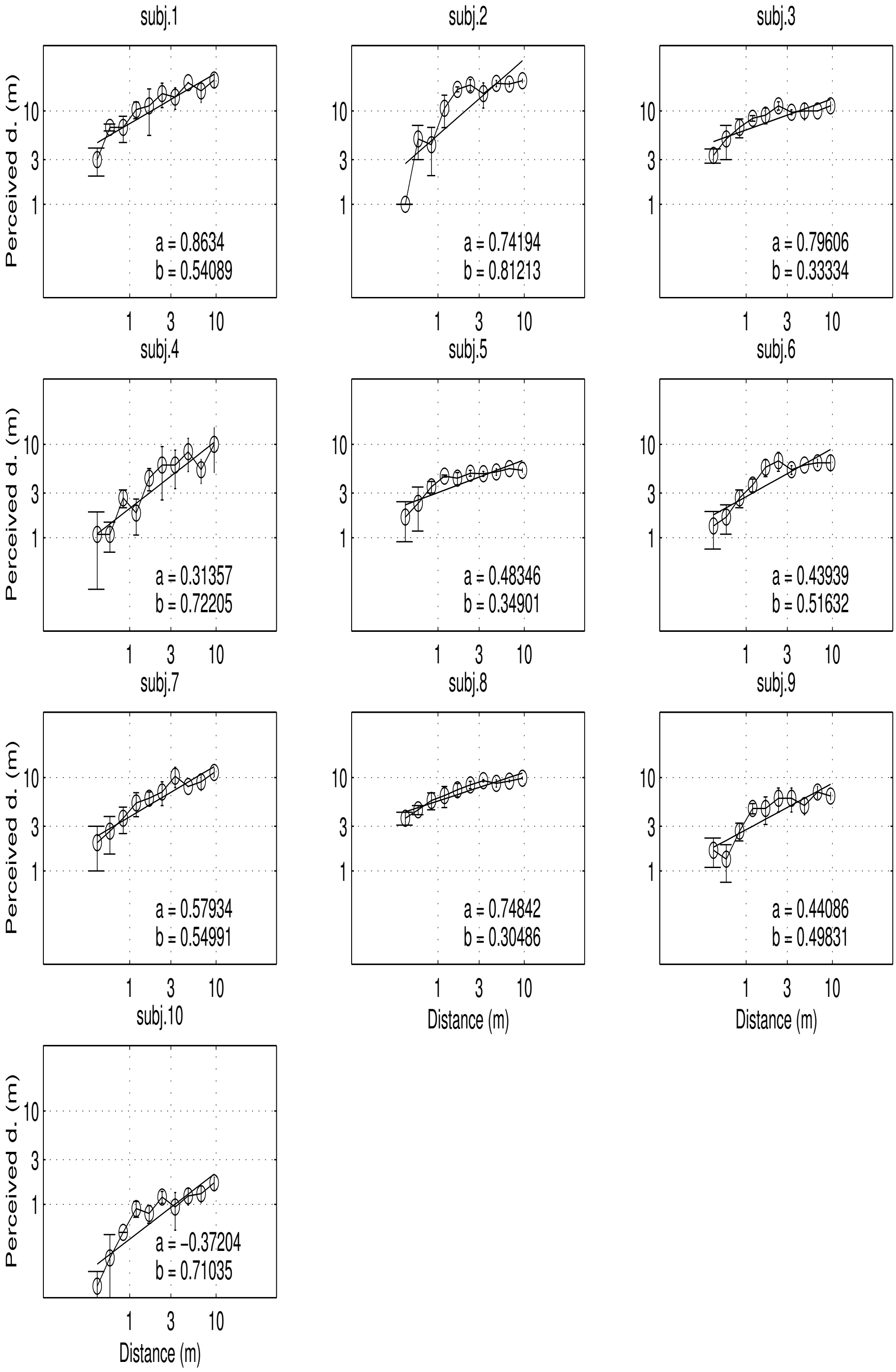


Figure 3.4: *Loudspeakers listening: Individual distance evaluations together with individual linear regression lines and standard deviation bars. a: intercept. b: slope.*

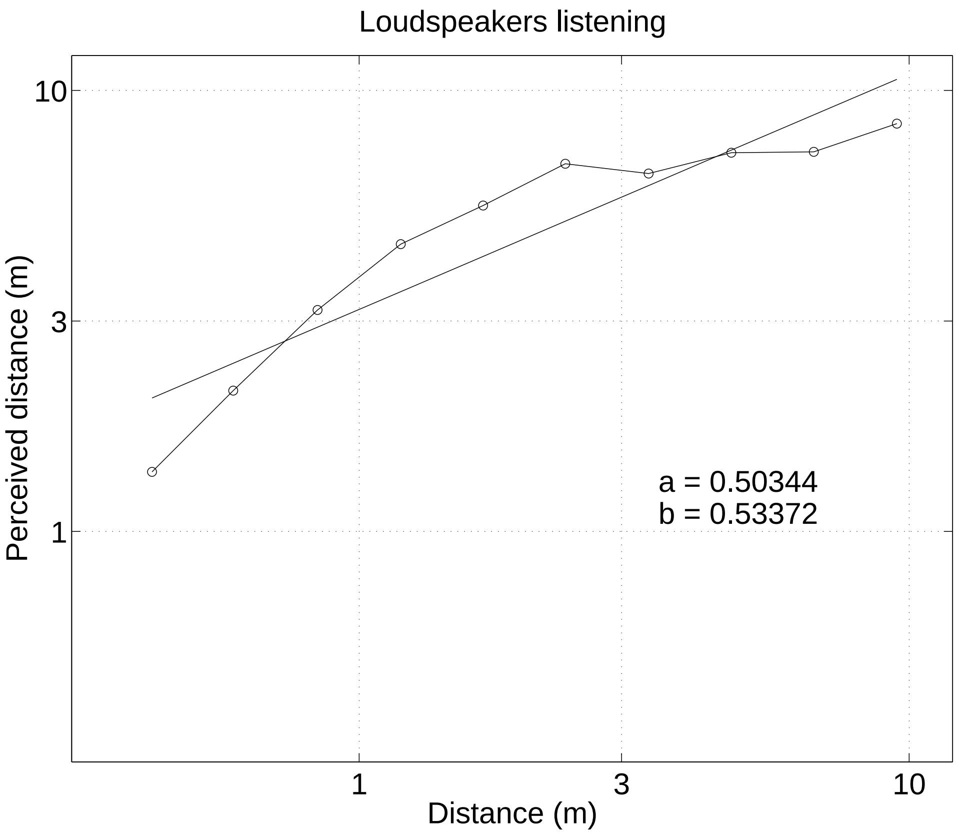


Figure 3.5: *Loudspeakers listening: Average distance evaluation together with linear regression line. a: intercept. b: slope.*

In Figure 3.5 the perceived distance averaged across subjects is plotted as function of the source-listener distance, together with the relative regression line ($r^2 = 0.85$, $F(1, 8) = 45.76$, $p < 0.01$).

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, leading, also in this case, to a perceptual estimation curve which tends to equalize greater distances.

There is only one participant (subject no. 10) whose individual scale ranged between 0.1-2 meters, and who perceived all the sound sources to be closer compared to the other listeners. Despite this outlier, we can see that the trend resulted in the first experiment, is enhanced also in this phase, as it can be seen in the box plot of Figure 3.6 as well. Therefore, the virtual environment synthesized for these experiments can be applied to applications which have to render sounds at long distances.

Looking at the box plot of Figure 3.6, we can see some outliers, but, in comparison to the headphone test results, they are relative to greater distances. In particular, there is an

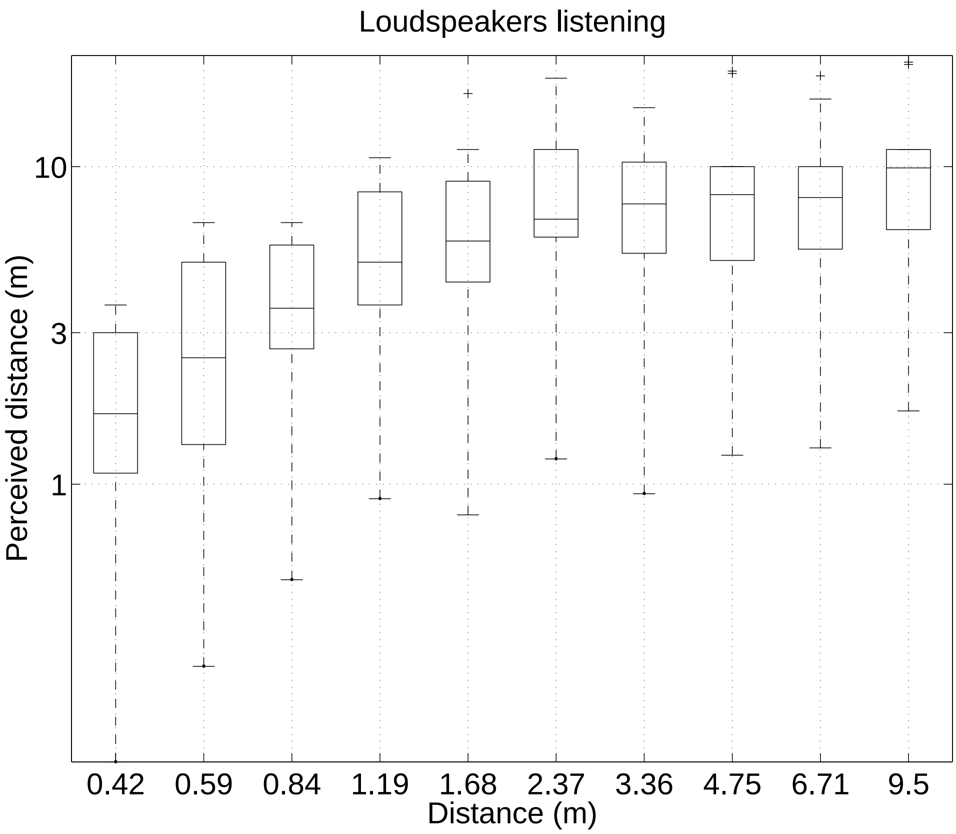


Figure 3.6: *Loudspeakers listening: Distance evaluations, represented by a box plot.*

outlier for the sound sources at 1.68 m, 6.71 m, 9.5 m and 4.75 m, which had an outlier also in the headphone listening condition. The difference in the outliers location relative to the sound sources distance could be related to the type of reproduction condition. In fact, while in the headphones listening test, the outliers are relative to close distances, in the loudpeakers listening test they concern greater distances. In any case, these observations regard few outliers, which maybe could be more influenced than others by the reproduction system.

We can compare the two listening conditions and evaluate if they are significantly different. With the Wilcoxon rank sum test, applied by means of the Statistical Toolbox of Matlab, we can see that the listening setting doesn't affect the subjects' performance, since the average distance evaluations in the two cases resulted to be not significantly different (Wilcoxon rank sum test, $p = 0.88$). In order to compare the two listening settings, we report (Figure 3.7) also the box plot regarding the average data in the two cases. The data distributions are not significantly different, as resulted from the Wilcoxon rank sum test as well. Nevertheless, we can notice that the data distribution in the loudspeakers listening case is slightly asymmetric, while the distribution of the distance evaluations in the headphones

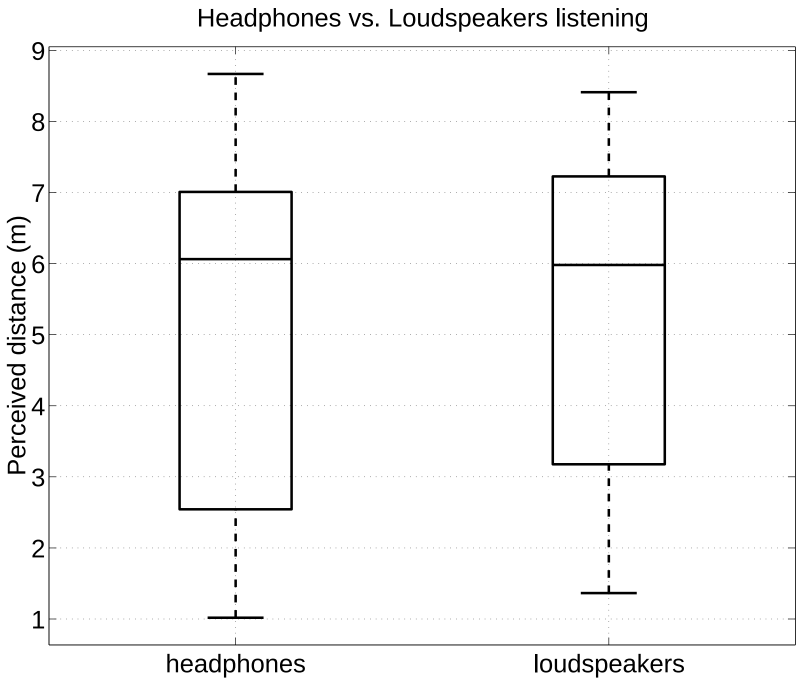


Figure 3.7: *Comparison of the distribution of the average distance evaluations in the two listening settings, represented by a box plot.*

case is symmetrical. Moreover, the distance evaluations in the headphone listening setting are slightly more spread than in the headphone case, but the distribution seems to be skewed to the left.

3.4 Conclusions

In this chapter, we have evaluated a virtual resonating environment that could be efficiently used in auditory display, which lacks of efficient distance rendering applications.

> Many commercially available spatial auditory displays do not do a very good job of creating realistic distance percepts [Zah02b].

We have psychophysically validated it, by means of two experiments, with two different listening settings. The former test was conducted by means of closed headphones, while the latter by means of loudspeakers.

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 [Jot99, Zah02b], allowing similar performances by users both in headphones listening conditions and in loudspeakers listening conditions.

There is an exaggeration especially in rendering close sound sources, probably due to the amount of reverberant energy existing in that case. 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.

Therefore, these experiments were useful in order to find a possible synthesis approach of a model for a new virtual environment where to locate the different "sound objects" of an alternative auditory display.

Chapter 4

Perceptual dimensions of the impact/bounce event

After our studies about size and shape as auditory perceptual dimensions of sound sources (chapter 2), and distance as perceptual feature of the sources *within* the surrounding environment (chapter 3), in this chapter we will focus on the auditory cues of particular sound events, in order to identify the auditory dimensions that could provide information and, therefore, that could be used in multi-dimensional auditory display.

We conducted some experiments with sound events modeled by the EU-funded project "the Sounding Object" (SOb)[1]. Our aims were twofold: to empirically validate the properties and quality of Sound Objects developed and to investigate the perceptual scaling of the physical parameters that control the sound models, in order to explore which auditory dimension could be used to convey information. The use of sound models allowed us to control directly the physical parameters of the event, in order to synthesize, in a flexible and computationally convenient way, a sound space able to auditorilly convey information.

We started this research activity from studying the impact/bounce event.

> Many of the sounds we hear in the everyday world involve one solid impacting against another. Tapping on an object, placing it against another, letting it fall — all involve impact sounds [Gav93a].

We analyzed pairwise the four auditory dimensions that are mostly associated and easily identified by subjects when they listen to an impact/bounce event. We studied perceived height of the object drop versus perceived size of dropped objects, and, then, perceived elasticity of the impact/bounce versus perceived force "throwing" the object. With perceptual force "throwing" the object we mean both the perception of the object just dropped without any force and the perception of the object thrown by applying some force.

For conducting our psychoacoustical experiments on the impact/bounce event we used the Sonic Browser, a software tool which was developed in the Interaction Design Centre at

[1] `http://www.soundobject.org`

the University of Limerick in 1996 and which has been under improvement since then [BF01, BFTC02]. It allows the user, represented by the cursor, to navigate a bi-dimensional multimedia space primarily through listening and to listen simultaneously to all the sounds included in the aura, that is the circular area surrounding the cursor. The Sonic Browser is introduced and briefly described in appendix B.

Before showing our experiments and reporting their results and conclusions in section 4.2 and section 4.3, we briefly summarize the previous approaches for conducting psychoacoustical experiments related to our goals (section 4.1).

4.1 Brief summary of previous interesting approaches

A fundamental issue in 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.

Different approaches are suitable for different goals. We know the classical methods in psychoacoustics, such as the methods of limits, adjustment and constant stimuli. These methods lack sometimes 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 (SDT) was born, which focuses not on the stimulus thresholds estimation, but on the reasons underlying one particular subject decision. The SDT estimates judgment processes elements not considered by the classical psychophysical methods.

An interesting method is the Stimulus Sample Discrimination (SSD) method, proposed by Mellody and Wakefield [MW01], based on previous studies on the stimulus sampling procedure by Sorkin et al [SRB87] and Lutfi [Lut89, Lut90, Lut92]. 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 providers or as distraction components. 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 synthesis.

Another important branch of experimental methods is classified as unidimensional and multidimensional scaling methods. They aim at estimating psychological 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. [SLH02], 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 a consequent increase in attention and consistency of the listeners decisions. Moreover, as the users are able to compare all the stimuli interactively, they can appreciate the full set of stimuli more than in a pairwise comparison task.

4.2 Experiment: perceived height of the object drop versus perceived size of the dropped object

The experiment aims to begin to understand how the synthesized sounds produced by our models are scaled in comparison with physical dimensions, by means of the Sonic Browser. In this experiment we focused on two dimensions: perceived height of the object drop and perceived size of dropped objects, that specified the axes of the 2-D plot in the application. 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 judgment. Moreover, as we wanted to compare the sound rendering of two different approaches in the sound modeling, the stimuli set included, besides recorded events, Sound Objects from both of these modeling approaches.

The experiment was preceded by a pilot probe, which used only one modeling 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.

In this section, we will present both the pilot probe and the main experiment, showing the procedure applied and the results obtained.

4.2.1 Participants

The pilot probe and the main experiment involved respectively 4 and 5 volunteers, all students or workers at the Computer Science Department of the University of Limerick (UL).

All the participants referred to have a musical training, for the pilot probe in average of 5 years, with a minimum of 2 and a maximum of 10 years, while for the main experiment in average of 8 years, with a minimum of 6 and a maximum of 10 years.

Nobody of the subjects of the pilot probe referred to have hearing or sight problems; on the other hand, two participants to the main experiment require glasses for reading but no participant reported to have hearing problems.

4.2.2 Stimuli

The stimuli used in both the pilot probe and the main experiment were not of a single collision, but of more than one bounce. The stimuli sets of both the experiments included recorded sounds and Sound Objects and consisted of 18 sounds, but in a different proportion. The sounds of the pilot probe from *sound1* to *sound9* were recorded while from *sound10* to *sound18* were synthesized. On the other hand, in the main experiment we decided to include just the first 6 sounds as real and 12, from *sound7* to *sound18*, as synthesized sounds, since we wanted to focus on the Sound Objects rather than recorded sounds and to keep them only for comparison.

The recorded sounds were produced by 3 steel balls, weighing 6, 12 and 24 g, with diameters of 11.3, 14.3 and 18 mm respectively, and falling on a wooden 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, from the board at the surface level. 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 [GB03]. In this study, Burro found the relationship between the physical quantities of weight, distance and height and the relative perceptual quantities. He argued 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), while in the main experiment we kept constant distance (d = 80 cm), while changing the height.

All the synthesized sounds in the pilot probe and 6 — from *sound7* to *sound12* - in the main experiment were designed with the PD-modules modeling impact interactions of two modal resonators [RF03], simplified returning only one mode, and they used either glass or wood as the material property. On the contrary, the remaining 6 synthesized stimuli of the main experiment were designed with the complete model of the impact interactions and the dropping event, as well. In this latter case, we preferred to keep the material constant, since we noticed some difficulties during the pilot probe for 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 wood. This perception arouse from the bigger influence of the surface material in certain cases.

In Table 4.1 we report the values of the parameters used in the PD-modules for synthesizing the stimuli of the pilot probe. In Table 4.2 and Table 4.3 we report the values of the parameters used in the PD-patches of the two different parameterizations for synthesizing the stimuli sets of the main experiment. In Table 4.3, 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. Just as an overview of the most important parameters that we mention, the elasticity parameter is the hardness of the impact and the lambda parameter is the force damping weight, accounting for the dissipation of energy during the contact. The alpha parameter, that we kept constant, is related to the local geometry around the contact area. The gravity force parameter represents the gravity force applied to the "striker", and

Short Name	**Elasticity** k	**Damping** λ	**Gravity force**	**Strike velocity**	**Frequency** (Hz)	**Decay time** (s)
sound10	15000	46.4159	990	630.957	1758.52	0.043070
sound11	5540.1	8.57696	990	1318.26	1782.52	0.043070
sound12	15000	21.5443	950	1584.89	1388.82	0.090315
sound13	3161.6	21.5443	580	2290.87	1388.82	0.090315
sound14	15000	46.4159	450	630.957	1782.52	0.043070
sound15	15000	46.4159	450	630.957	1758.52	0.233307
sound16	15000	63.0957	940	912.011	1113.23	0.603386
sound17	11395	2.92864	860	301.995	1294.33	0.752992
sound18	1309.5	4.64159	970	436.516	1294.33	0.784488

Table 4.1: *Pilot probe. Values of the parameters for the synthesized sounds.*

Short Name	**Elasticity** k	**Damping** λ	**Gravity force**	**Strike velocity**	**Frequency** (Hz)	**Decay time** (s)
sound7	15000	46.4159	450	630.957	1758.52	0.233307
sound8	15000	63.0957	940	912.011	1113.23	0.603386
sound9	11395	2.92864	860	301.995	1294.33	0.752992
sound10	1309.5	4.64159	970	436.516	1294.33	0.784488
sound11	1309.5	8.57696	990	1318.26	1254.95	0.784488
sound12	3162.28	25.1189	900	524.807	1322.83	0.233307

Table 4.2: *Main experiment. First parameterization. Values of the parameters for the synthesized sounds.*

the strike velocity is the velocity of the object during the contact, i.e. when the distance between the two colliding objects is zero. The decay time parameter controls one of the most important features for material perception [AR01], i.e. the decay time, and the frequency parameter controls the resonant frequencies of the resonator. For a reference on the main features of the models and a detailed meaning of the parameters, we refer to [RF03].

The mentioned parameters represent the set of controllable parameters in the PD-modules we used, and the values of the parameterizations were decided according to informal trials in order to include in the stimuli set the most different sound types as possible. During the synthesizing phase, we focused on the perceptual quality of the resulting sounds rather than on the parameter values. On the contrary, during the synthesizing phase of the next experiment (see subsection 4.3.2), we designed the stimuli set paying attention to the parameter values as well.

Short Name	Hammer mass	Initial interval (ms)	Acceleration/ Deceleration	Initial value
sound13	0.0215443	228.530	0.76	0.56
sound14	0.0215443	306.516	0.74	0.75
sound15	0.0398107	207.223	0.72	0.57
sound16	0.0398107	207.223	0.76	0.57
sound17	0.0398107	277.939	0.70	0.75
sound18	0.0398107	372.786	0.70	1.00

Table 4.3: *Main experiment. Second parameterization. Values of the parameters for the synthesized sounds.*

4.2.3 Procedure

Both the pilot probe and the main experiment were conducted in the isolation room of the Computer Science Department at UL. The stimuli were presented by stereo headphones to the participant through the Sonic Browser.

The stimuli were represented in the Sonic Browser with coloured shapes and they were put randomly within the 2-D plot identified by the two axes labeled according to the dimensions under investigation, i.e. perceived height of the object drop on the y-axis and perceived size of the dropped object on the x-axis.

The cursor was surrounded by the aura, which could be resized or turned off and on directly by the subjects and how many times they wanted.

The subjects could listen to the sounds by moving the cursor on them and, if the aura was large enough, they could listen simultaneously to all the sounds it was surrounding. The participants could also move the stimuli in the perceptual space by dragging-and-dropping the shapes which were representing the stimuli.

The experiment sessions were divided in two phases, each one with a different task — scaling task and realism judgment — and, finally, a debriefing phase. In the first phase, the participants were asked to listen to the sounds, by moving the cursor on each stimulus, and to estimate the perceptual values of the dimensions under investigation, by moving the sounds in the perceptual space identified by the two axes — in this case perceived height of the object drop on the y-axis and perceived size of the dropped object on the x-axis. In this way, they could create their own perceptual space. During the second phase of the sessions, the subjects were asked to listen again to the sounds they had scaled during the first phase and to tag the sounds that they thought to be unrealistic.

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

The Thinking-Aloud Protocol is widely used in usability testing and it represents a way for the experimenter to have a "look" at the participants' thought processes [BR00]. In this approach, the users are asked to talk during the test, expressing all their thoughts, movements and decisions, trying to think-aloud, without paying much attention to the consistency of the sentences, "as if alone in the room".

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 judgment or their appreciation of the Sound Objects realism. The tests were all recorded by a video-camera.

In the scaling task the subjects 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 participant. The perceptual space boundaries were considered by all the subjects, 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.

At the end of each session, each participant filled out a 7-point Likert scale questionnaire, from a "poor" evaluation (0) to a "excellent" evaluation (6). In the main experiment, three questions, asking about the learnability, interpretation of the application and the difficulty in replaying the last sound, were added to the questionnaire after the pilot probe.

4.2.4 Results and Observations

Results of the pilot probe

In Figure 4.1 we represent the perceptual scaling and the tagging information of all the users and all the stimuli. The two axes represent the perceptual dimensions estimated, that are perceptual size of the object and perceptual height of the drop. The 2-D plot identified by the two axes represents the perceptual space, where we report the locations of all the stimuli for all the participants to the experiment. As we already said in subsection 4.2.3, we decided to normalize the data coordinates between 0 and 1. In the plot, we highlight the tagging information by means of different shapes. With black circles, we represent the real sounds, that were all judged to be realistic by all the subjects. With empty squares and crosses we represent the synthetic sounds, that have been judged respectively to be unrealistic and realistic.

From a global observation of the collected data represented in Figure 4.1, we can notice that the participants estimated correctly the height from the real sounds, h=20 cm for all of them, since most of the real sounds, barring five outliers, were positioned by the users in the central area of the evaluation space. On the other hand, the size estimation varies to a

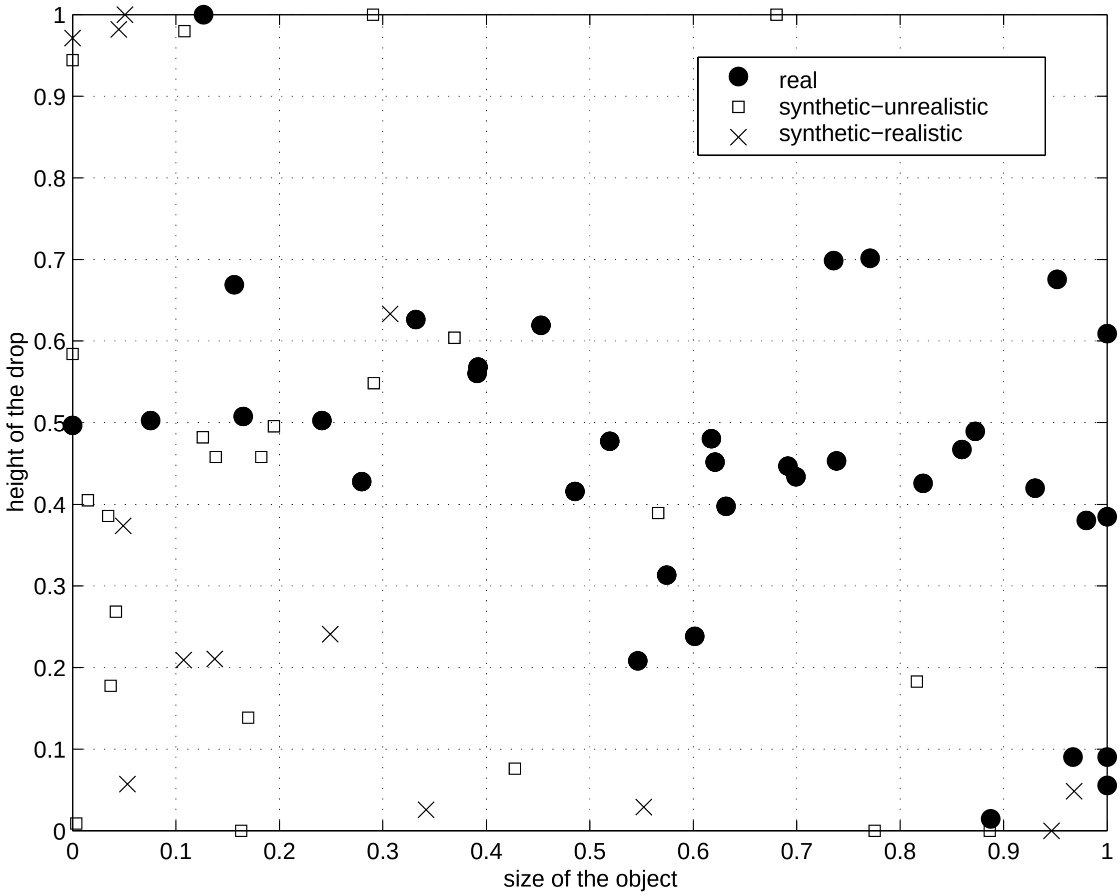


Figure 4.1: *Pilot probe. Representation of the perceptual scaling and tagging information of all the users and all the stimuli. The 2-D plot represents the perceptual space, where the locations of all the stimuli for all the participants to the experiment are reported. The perceptual scaling values are normalized between 0 and 1. The tagging information: Black circles represent the real sounds, all judged to be realistic by all the subjects; Empty squares and crosses represent the synthetic sounds, that have been judged respectively to be unrealistic and realistic.*

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 individual perceptual scaling and tagging information sorted by users, reported in Figure 4.2 where we use the same conventions as those of Figure 4.1, we notice that two participants in particular (users 2 and 3) made an obvious distinction between real and synthetic sounds.

It is interesting to observe the single stimuli. In Figure 4.3, we represent, with the same conventions as those used in the previous figures, the individual perceptual scaling and

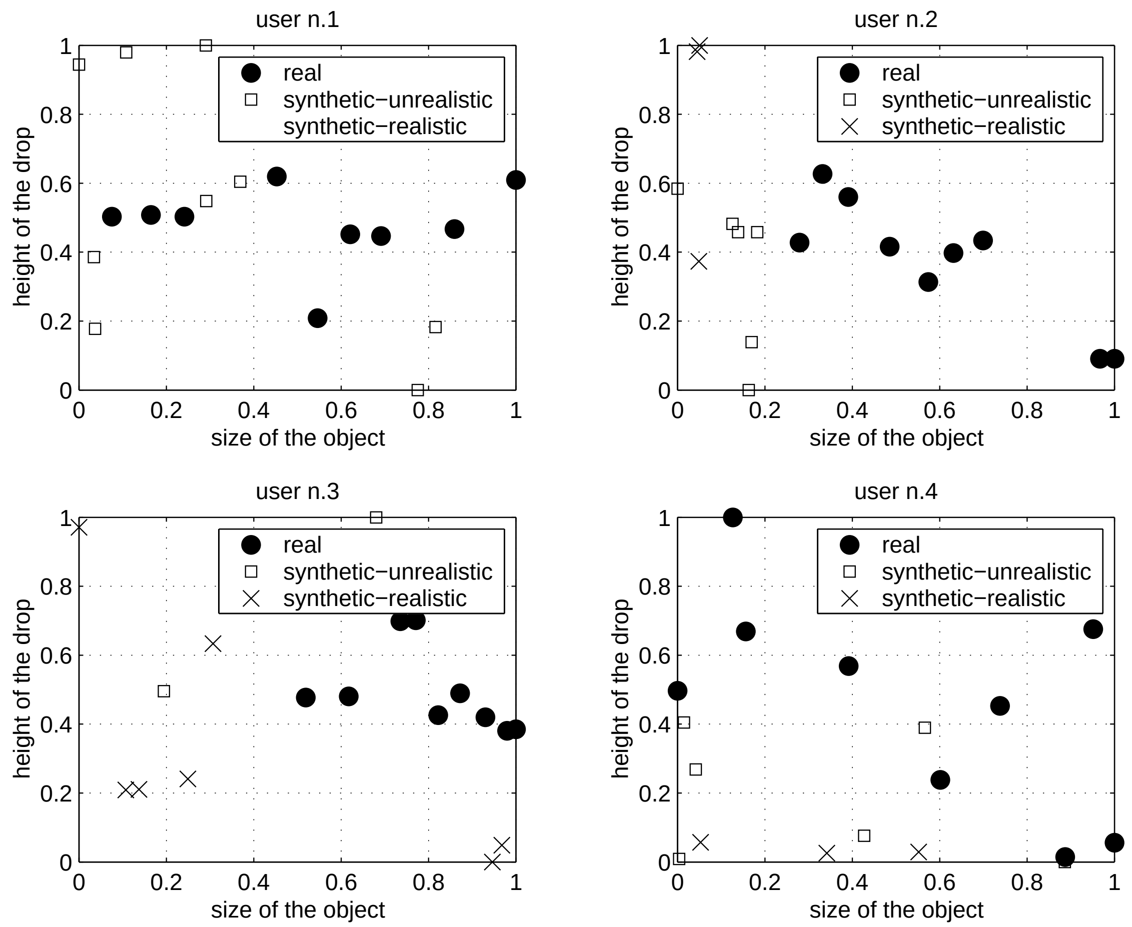


Figure 4.2: *Pilot probe. Representation of the individual perceptual scaling and tagging information sorted by users. For each user, it is reported the individual perceptual space, where the locations of all the stimuli for that user are reported. The perceptual scaling values are normalized between 0 and 1. The tagging information: Black circles represent the real sounds, all judged to be realistic by all the subjects; Empty squares and crosses represent the synthetic sounds, that have been judged respectively to be unrealistic and realistic.*

tagging information sorted by stimuli. For our purposes, we will focus on the synthesized sounds. In Figure 4.4, we report the box plot of the individual perceptual scaling of height and size respectively sorted by stimuli. Here, we consider just the synthesized stimuli.

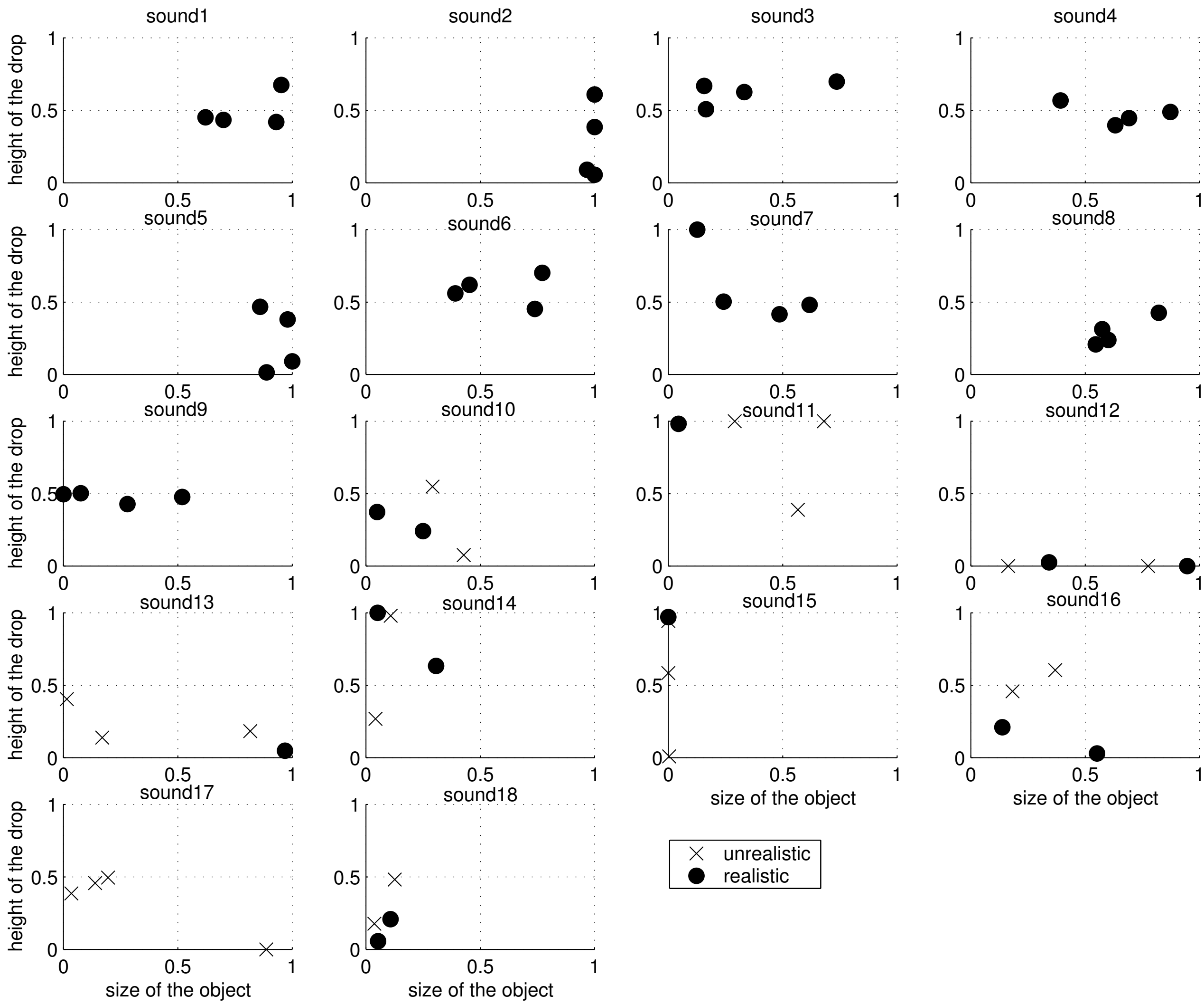


Figure 4.3: *Pilot probe. Representation of the perceptual scaling and tagging information sorted by stimuli. For each stimulus, it is reported the perceptual space, where the locations of that stimulus for all the users are reported. The perceptual scaling values are normalized between 0 and 1. The tagging information: Black circles represent 'realistic' judgments, while crosses represent 'unrealistic' judgements.*

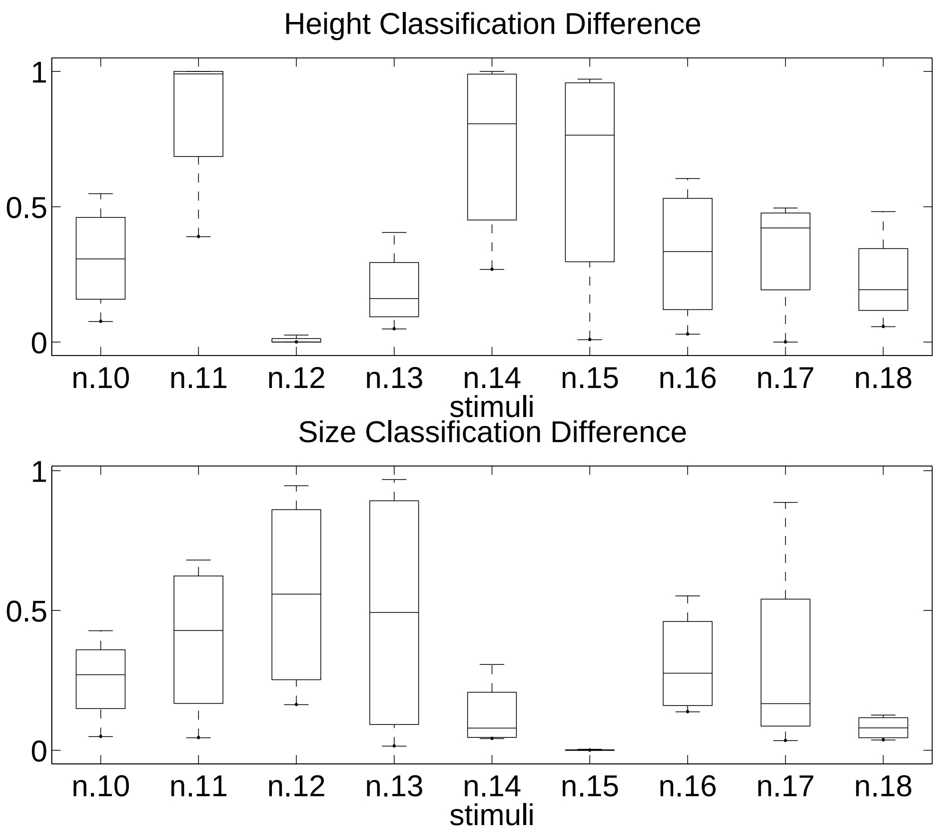


Figure 4.4: *Pilot probe. Representation, by a box plot, of the perceptual scaling of the height and size sorted by stimuli. All the sounds are synthesized.*

From these plots, we can see different perceptual scaling by the users along the two dimensions. In fact, we can find some sounds that were judged uniformly by most of the users, at least along one dimension, while others (*sound10*, *sound16* and *sound17*) whose scaling is spread across the perceptual estimation space, showing a difficulty by the participant to estimate them. It is interesting to notice that *sound17* 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, *sound18* was judged uniformly by all the users, and especially for the size dimension. Finally, the other five stimuli of the synthesized set were all judged uniformly in one dimension. In particular, *sound11*, *sound12*, and *sound13*, are perceived with the same height, while *sound14* and *sound15* are perceived with the same size.

We noticed that the participants did not agree in the results of the tagging task. Only one stimulus, *sound17*, was defined by all the subjects as unrealistic. On the other hand, only two, i.e. *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 probably due to the presence of some reverberation (room acoustics) in the real sounds, that the synthesized stimuli lack.

Another interesting observation regards the type of approach taken by each participant to the task. One of them preferred not to use the aura, but the other three decided to turn it on after an initial period of "self-training"and they found it useful for comparing the stimuli and checking the estimation of the whole group.

It is worth to look at the relationship between the perceptual scaling of the stimuli and their parameters setting. In Figure 4.5 and Figure 4.6 we plot, sorted by parameters, the centroids of the scaling of the perceived size and of the perceived height, respectively, versus the parameters values. Therefore, in the plots, we can compare the stimuli mean positions within the perceptual space and the stimuli positions within the parameters space. Stimuli parameters with the same value are highlighted with horizontal dashed lines.

From the plots, we can see that the relationship between the parameters space and the perceptual space is complex and, even if we cannot generalize our observations since they are based on few sounds with a certain parameterization, some trends seem to appear.

We can notice, barring one or two outliers, an increase in perceived size of the object with the increase of gravity force and strike velocity. There seems to be a similar relationship also related to the damping parameter, even if the sounds appear to be divided into two groups: those with damping value less than 40, and those in the range 40-70. As the elasticity parameter increases, the perceived size seems to decrease, except for *sound12* and *sound18*. In the case of decay time, we cannot see any simple relationship with the perceived size, while only for frequency values lower than 400 Hz, the perceived size increases with the frequency values.

As far as the perceived height of the drop is concerned, it seems to decrease as the gravity force parameter and the decay time parameter increase, but for values of the latter parameter lower than 0.1 s, the sounds don't have a particular configuration in the perceptual space.

The perceived height of the drop shows a sort of decreasing tendency as the strike velocity parameter increases, but it is not clear and strong. The last three parameters, i.e. elasticity, damping and frequency, seem not to be involved, at least in a simple way, in the perceptual scaling of the height of the drop.

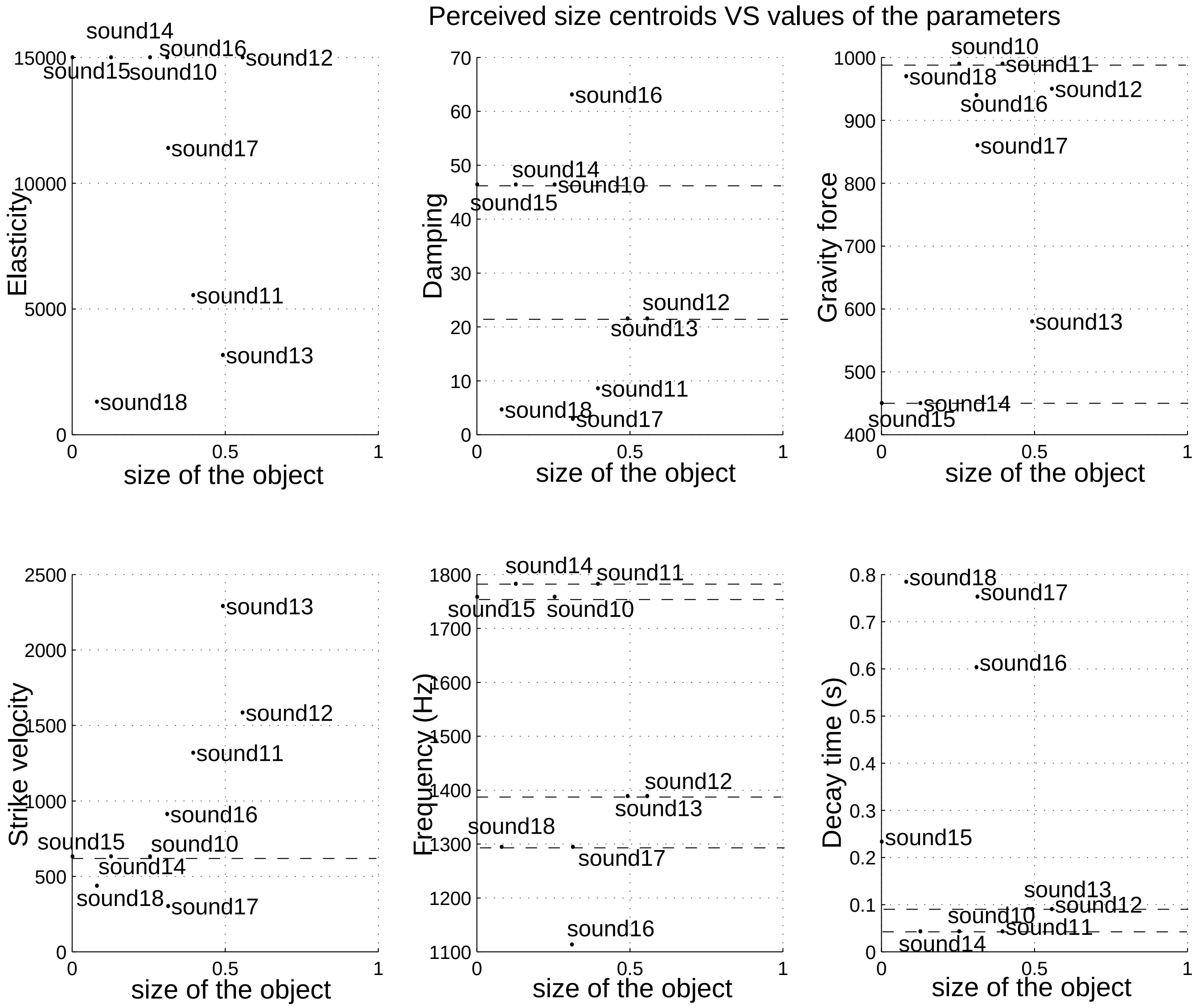


Figure 4.5: *Pilot probe. Centroids of the scaling of the perceived size versus the parameters values, sorted by parameters. The perceptual scaling values are normalized between 0 and 1. Stimuli parameters with the same value are highlighted with horizontal dashed lines.*

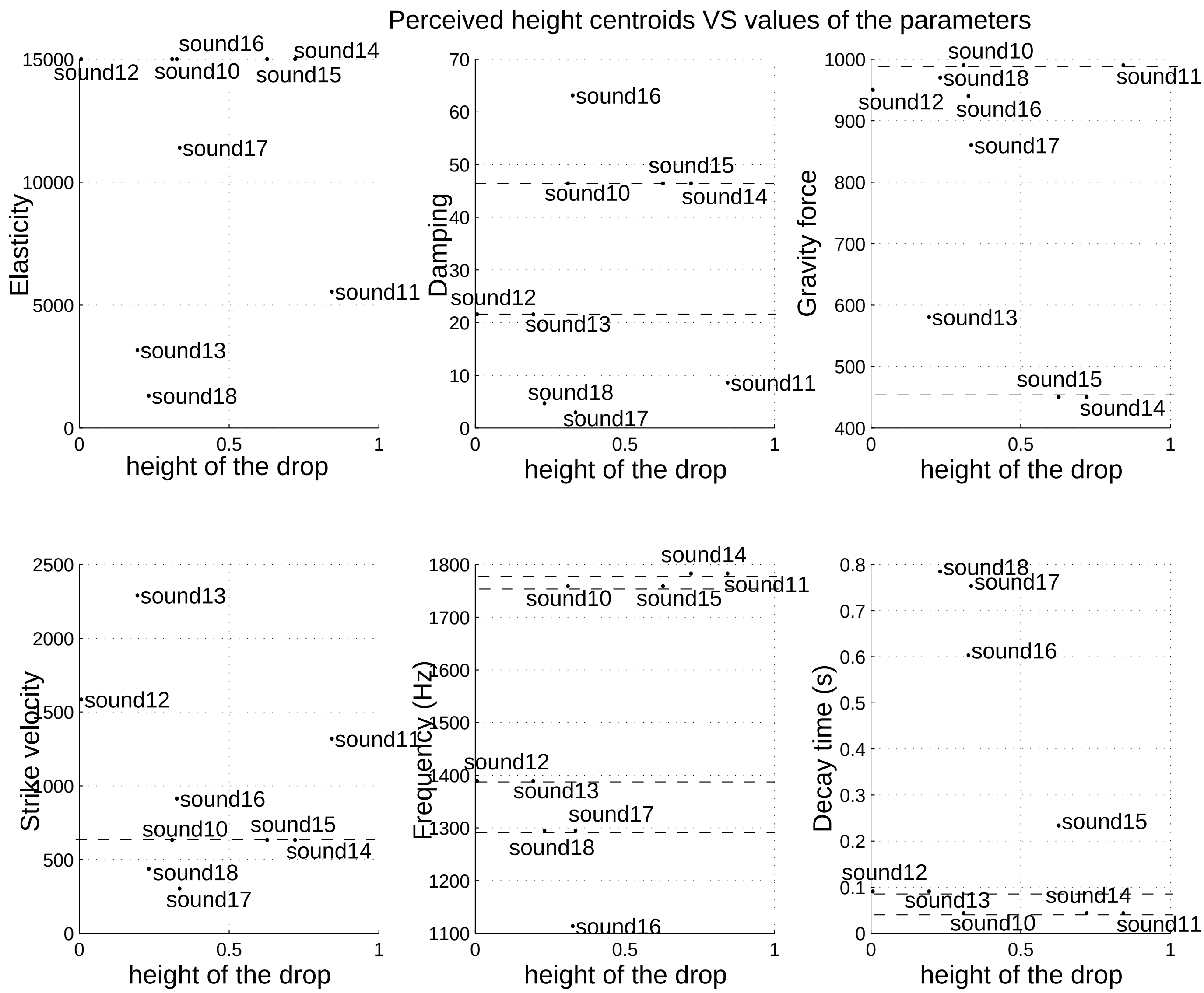


Figure 4.6: *Pilot probe. Centroids of the scaling of the perceived height versus the parameters values, sorted by parameters. The perceptual scaling values are normalized between 0 and 1. Stimuli parameters with the same value are highlighted with horizontal dashed lines.*

In Figure 4.7, the results of the questionnaire, with cumulative participant responses represented through a bar chart, can be seen, with -3 representing a negative result to the question and 3 a positive one, while 0 represents the average result.

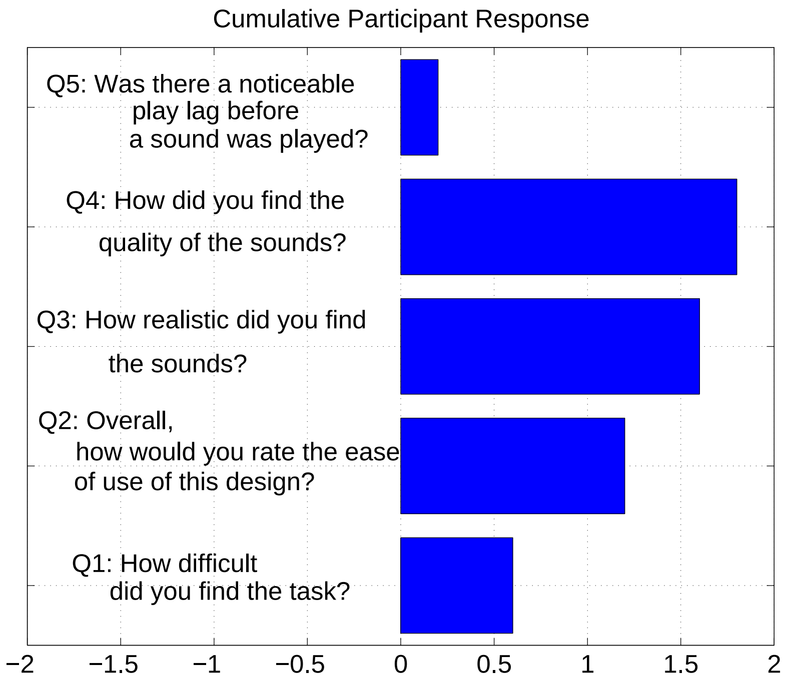


Figure 4.7: *Pilot probe. Results of the questionnaire filled out during the debriefing phase: Cumulative Participant Response.*

We can see that while the task was found to be non trivial (question 1), the users rated the ease of use of the application above average (question 2). The participants judged the sounds to be realistic (question 3) and of high quality (question 4). In the application, there is a slight delay of up to 0.3 of a second when playing an audio file. The users found it to be acceptable but noticeable (question 5).

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”. Most 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”.

Results of the main experiment

In Figure 4.8, we report the data from the main experiment sorted by users. For each user, it is reported the individual perceptual space, where the location of all the stimuli is reported. We highlight with different shapes the tagging information. Black circles represent the real sounds, all judged to be realistic by all the subjects, while empty squares and crosses represent the synthetic sounds, that have been judged respectively to be unrealistic and realistic.

As for the pilot probe, we can see the classification by sound groups. Moreover, we notice that two of the participants (user 1 and user 2) only performed minor judgments 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 already been discussed by Burro [GB03].

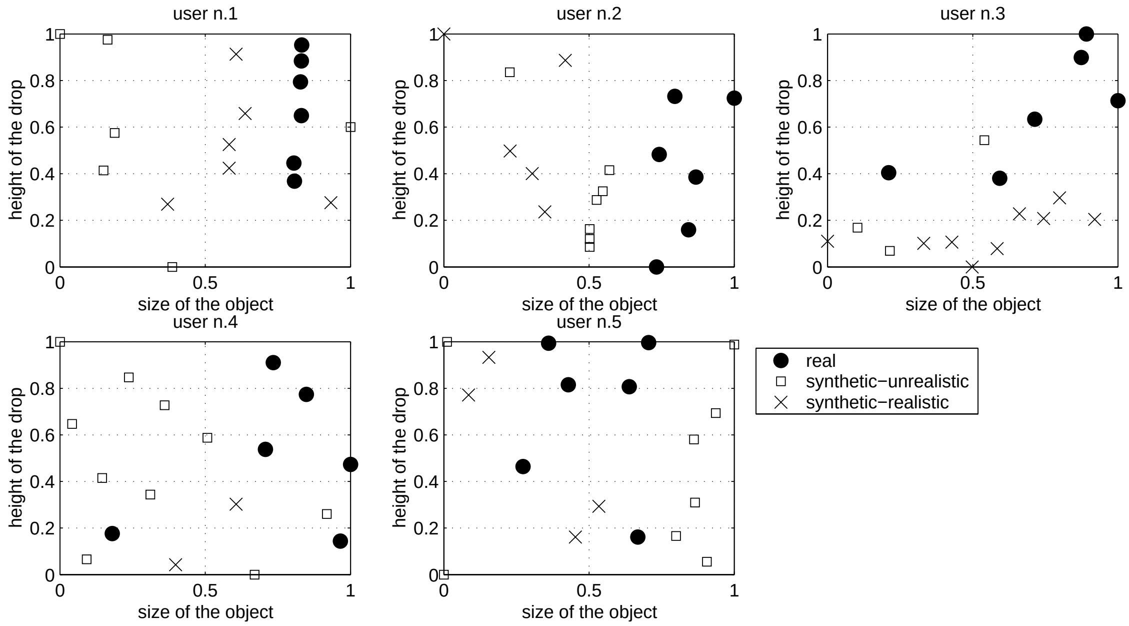


Figure 4.8: *Main experiment. Representation of the individual perceptual scaling and tagging information sorted by users. For each user, it is reported the individual perceptual space, where the locations of all the stimuli for that user are reported. The perceptual scaling values are normalized between 0 and 1. The tagging information: Black circles represent the real sounds, all judged to be realistic by all the subjects; Empty squares and crosses represent the synthetic sounds, that have been judged respectively to be unrealistic and realistic.*

It is interesting to observe the single stimuli, as we did for the pilot probe, looking at Figure 4.9, where we represent the individual perceptual scaling and tagging information sorted by stimuli. For each stimulus, it is reported the perceptual space, where the locations of that stimulus for all the users are reported. We highlight with different shapes the tagging information. Black circles represent 'realistic' judgments, while crosses represent 'unrealistic' judgements. We again focus on the synthesized sounds and, in Figure 4.10, we report the box plot of just the synthesized stimuli.

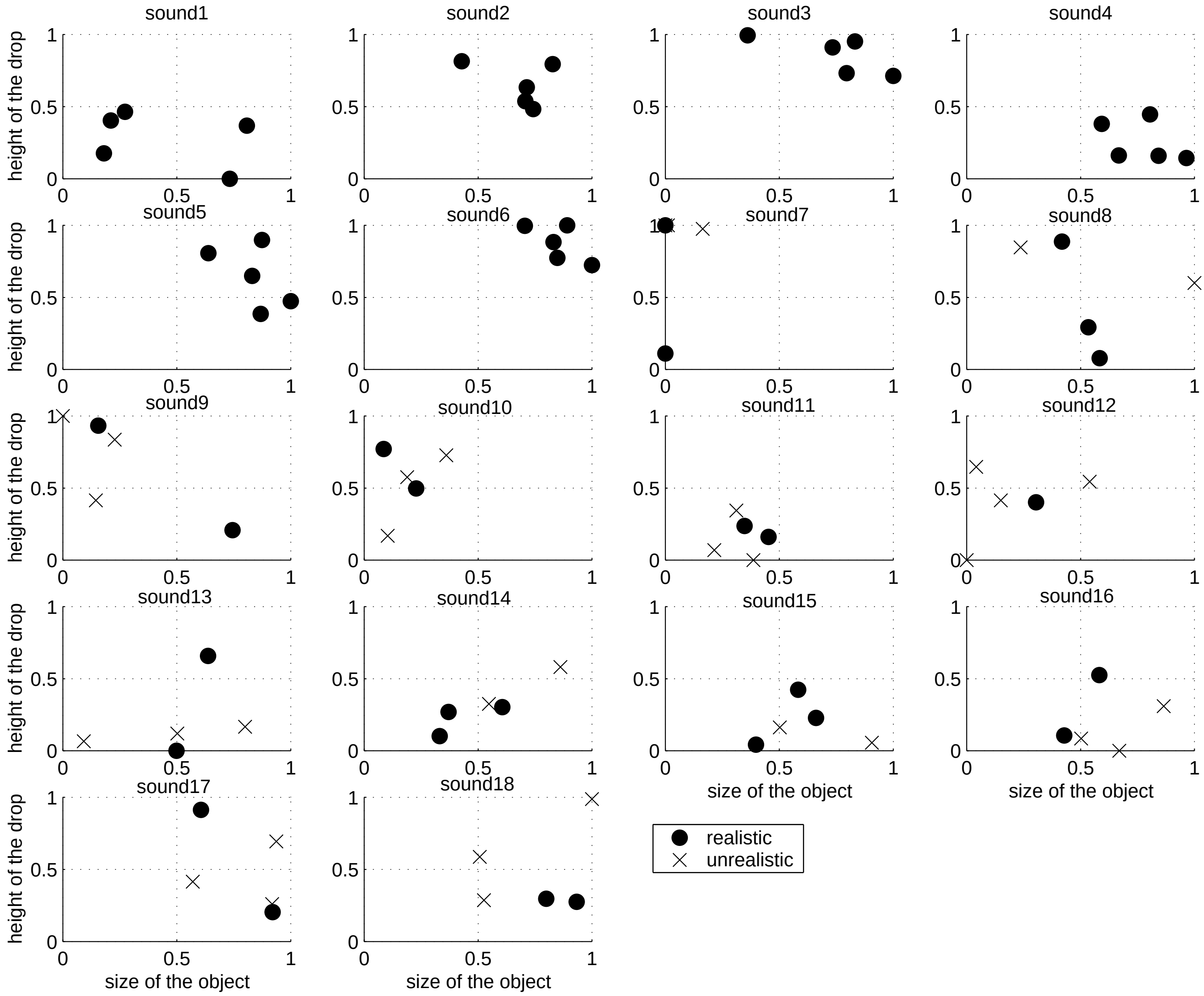


Figure 4.9: *Main experiment. Representation of the perceptual scaling and tagging information sorted by stimuli. For each stimulus, it is reported the perceptual space, where the locations of that stimulus for all the users are reported. The perceptual scaling values are normalized between 0 and 1. The tagging information: Black circles represent 'realistic' judgments, while crosses represent 'unrealistic' judgements.*

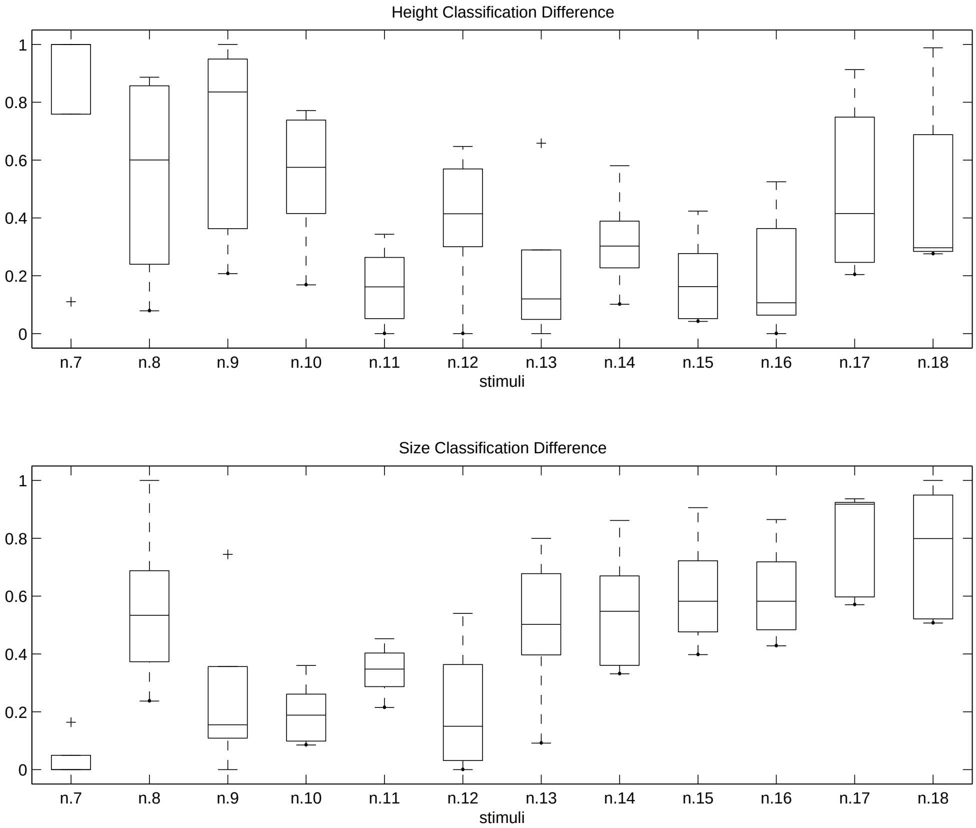


Figure 4.10: *Main experiment. Representation, by a box plot, of the perceptual scaling of the height and size sorted by stimuli. All the sounds are synthesized.*

As compared to the pilot probe, we can observe that more users agreed in the perceptual scaling of the two dimensions. For instance, four stimuli were judged uniformly by the participants, where *sound7* and *sound11* could be considered to show a strong uniformity in both dimensions, while *sound10* and *sound14* a slight uniformity in either dimension. Moreover, four of the stimuli were perceived uniformly in one dimension (*sound9*, *sound12*, *sound13* and *sound15*), and, finally, three stimuli (*sound8*, *sound16* and *sound18*) were dispersed in the perceptual estimation space and one, *sound17*, highly dispersed.

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 user consensus regarding unrealistic stimuli was achieved by 3 users. The real sounds were perceived again as realistic, duplicating the results of our pilot probe.

It is interesting to look at the complex relationship between the perceptual scaling of the stimuli and their parameters setting. In Figure 4.11 and Figure 4.12 we plot, sorted by parameters, the centroids of the scaling of the perceived size and of the perceived height, respectively, versus the parameters values, for the first parameterization. In Figure 4.13 and Figure 4.14, we report the same plots calculated for the second parameterization. In the plots, we can compare the stimuli mean positions within the perceptual space and the stimuli positions within the parameters space. Stimuli parameters with the same value are highlighted with horizontal dashed lines.

By looking at the two parameterizations separately, we point out, as we did for the pilot probe, the simpler relationships between the parameters space and the perceptual space, since it is evident that the two spaces are related in a complex way. Even if we cannot generalize our observations since they are based on few sounds with a certain parameterization, some trends seem to appear.

In the first parameterization, we can see an increase in perceived size of the object with the increase of gravity force and strike velocity, as it happened in the pilot probe results (see Figure 4.5), even if it is not as clear as in the pilot probe results, due to the values setting. We can notice, also, a trend for the damping parameter similar to that of the pilot probe, that is an increase in perceived size with the increase of the damping values, which are divided into two groups: those with damping value less than 40, and those in the range 40-70. The elasticity parameter shows results similar to the pilot probe, i.e. as it increases, the perceived size seems to decrease, except for *sound8* and, also in this case, the last two parameters, that are decay time and frequency, don't show any simple relationship with the perceived size.

As far as the perceived height of the drop is concerned, we can notice similar trends to those of the pilot probe, as it happened for the perceived size. As the gravity force parameter and the decay time parameter increase, the perceived height of the drop seems to decrease, but for the decay time parameter there are few outliers.

The perceived height of the drop shows a sort of decreasing tendency as the strike velocity parameter increases, and it is clearer and stronger than in the pilot probe, if we divided the sounds between those with values lower than 600 and those with values in the range 600-

1400. Contrary to the pilot probe, there is an increasing tendency of the perceived height of the drop as the elasticity parameter increases, barring on or two outliers, while the last two parameters, i.e. damping and frequency, seem not to be involved, at least in a simple way, in the perceptual scaling of the height of the drop, as it happened for the pilot probe.

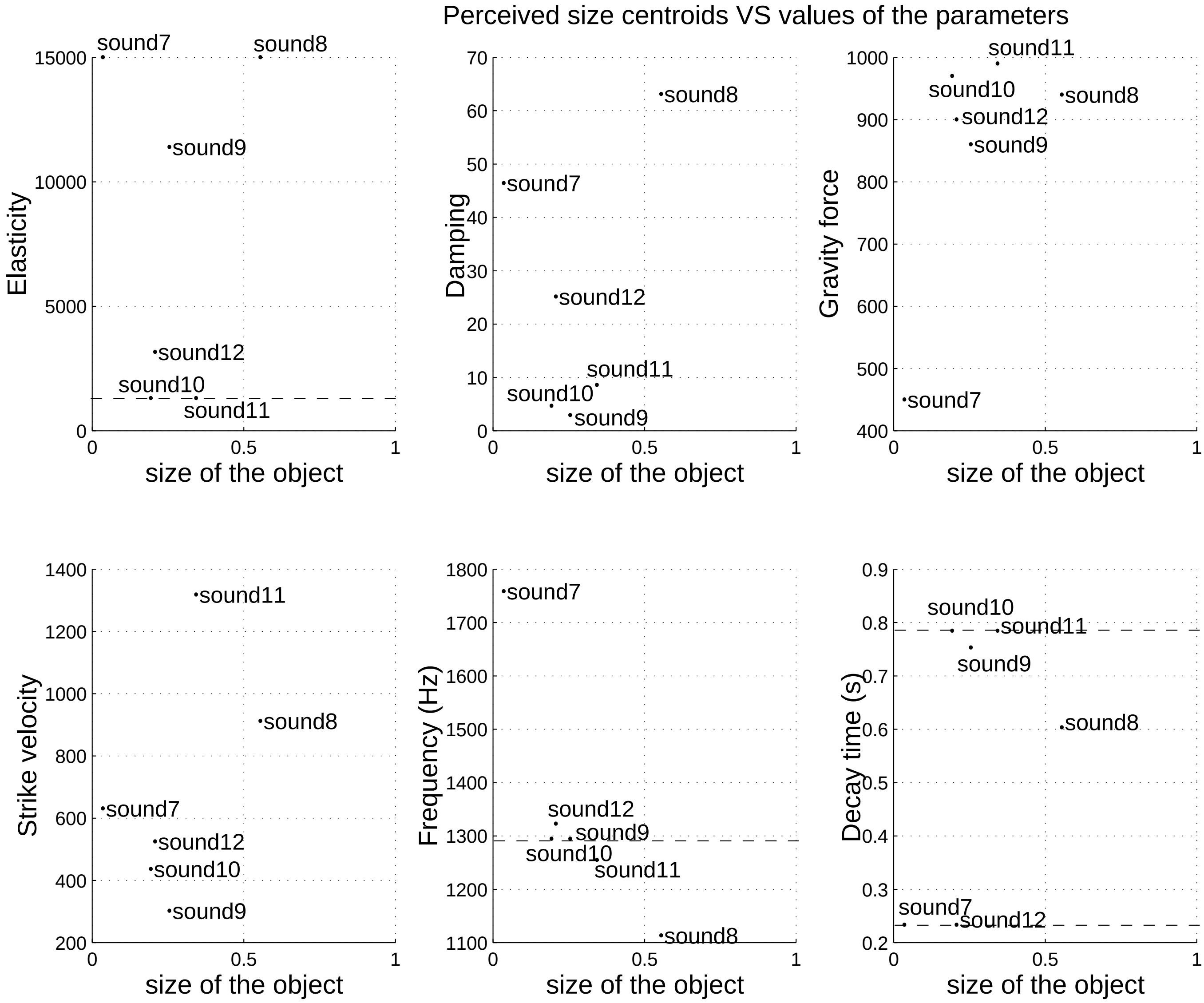


Figure 4.11: *Main experiment. Centroids of the scaling of the perceived size versus the parameters values, sorted by parameters. The perceptual scaling values are normalized between 0 and 1. The parameters values are only those of the first parameterization. Stimuli parameters with the same value are highlighted with horizontal dashed lines.*

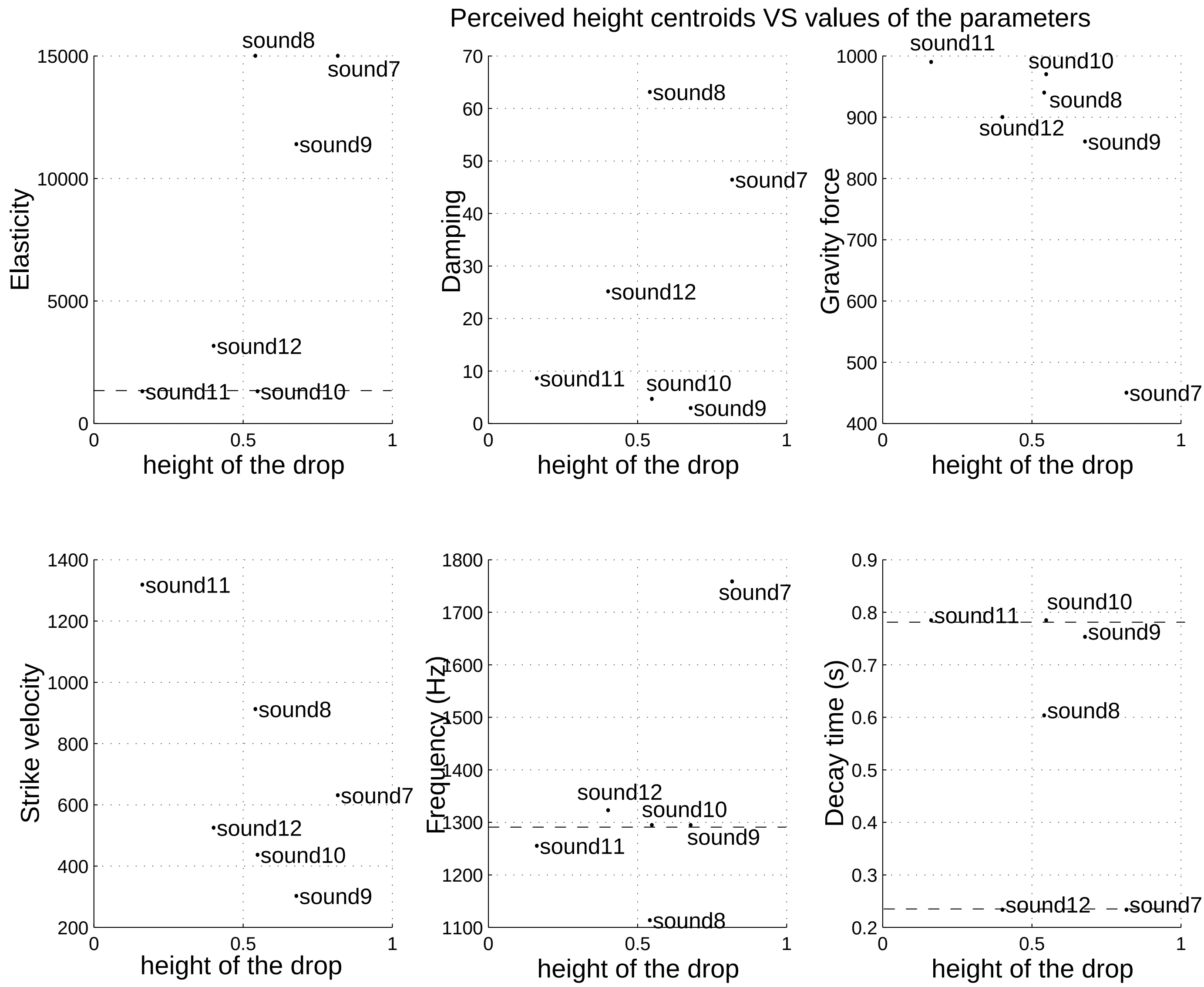


Figure 4.12: *Main experiment. Centroids of the scaling of the perceived height versus the parameters values, sorted by parameters. The perceptual scaling values are normalized between 0 and 1. The parameters values are only those of the first parameterization. Stimuli parameters with the same value are highlighted with horizontal dashed lines.*

Looking at the plots concerning the second parameterization, we can see that the hammer mass seems not to create a particular configuration in the perceptual space of both the size of the object and the height of the drop. On the contrary, the last three parameters show similar trends for both the perceptual space, i.e. a decreasing trend for the acceleration/deceleration parameter and an increasing trend for the other two parameters, except an outlier in both of them.

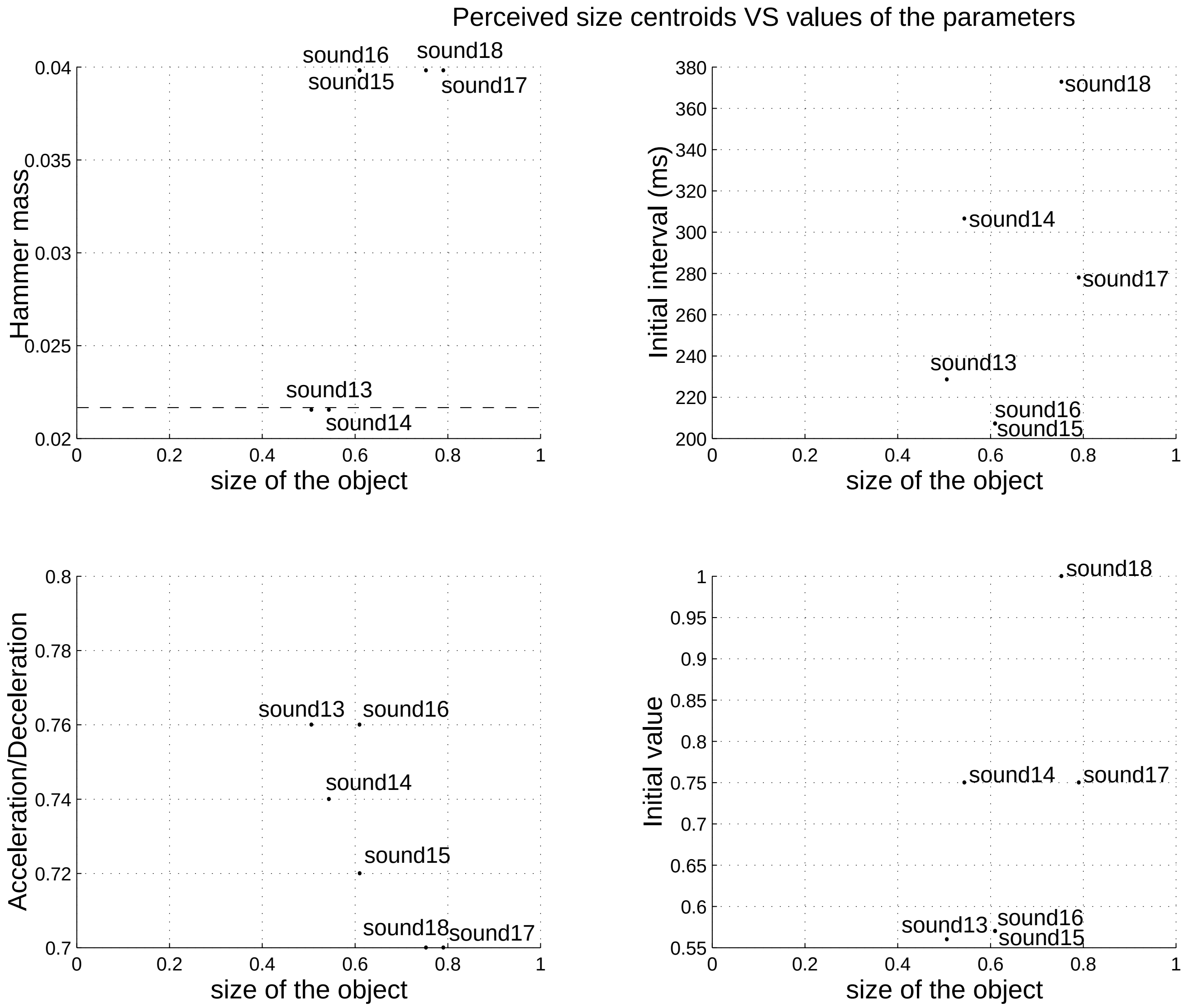


Figure 4.13: *Main experiment. Centroids of the scaling of the perceived size versus the parameters values, sorted by parameters. The perceptual scaling values are normalized between 0 and 1. The parameters values are only those of the second parameterization. Stimuli parameters with the same value are highlighted with horizontal dashed lines.*

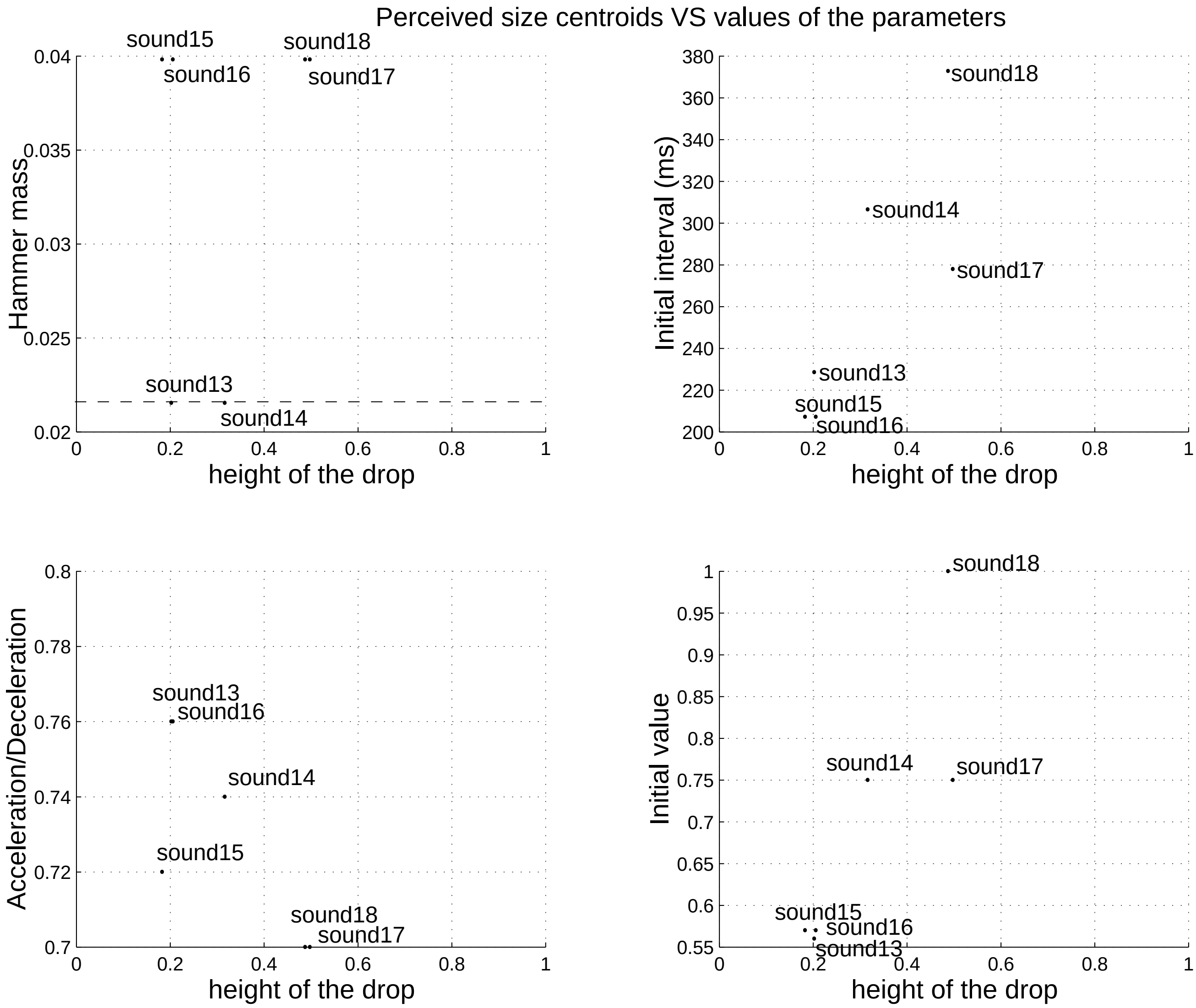


Figure 4.14: *Main experiment. Centroids of the scaling of the perceived height versus the parameters values, sorted by parameters. The perceptual scaling values are normalized between 0 and 1. The parameters values are only those of the second parameterization. Stimuli parameters with the same value are highlighted with horizontal dashed lines.*

In Figure 4.15, as we did for the pilot probe, we report the results of the questionnaire with cumulative participant responses displayed per question.

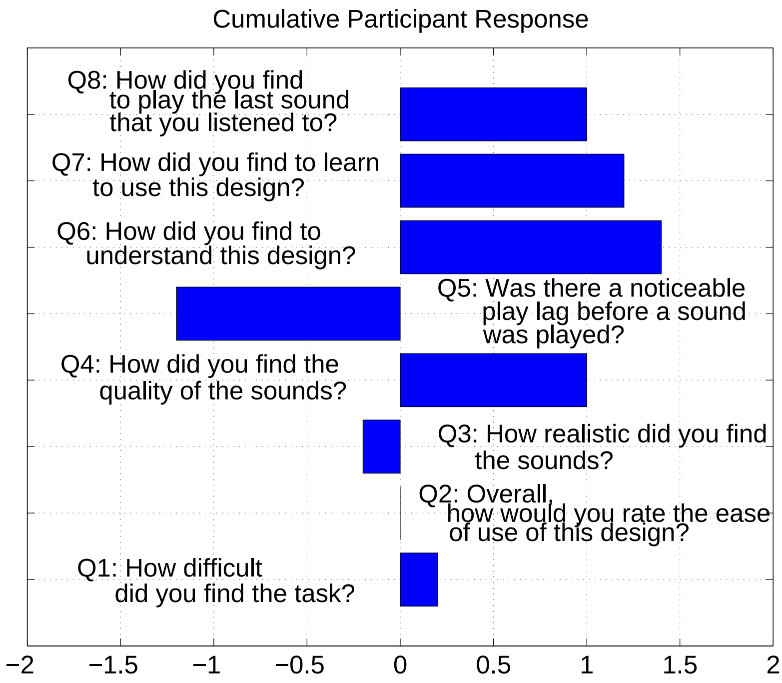


Figure 4.15: *Main experiment. Results of the questionnaire filled out during the debriefing phase: Cumulative Participant Response.*

We can see that the task was found to be non trivial (question 1) and the ease of use of the application was rated only on average (question 2). The participants judged the sounds to be of high quality (question 4), but not so realistic (question 3). 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 objects and the presence of a "buzz tail" at the end of the two mode sound objects. The slight delay when playing an audio file resulted in this experiment to be very noticeable to the users (question 5). This fact is evident by the verbal protocol, as well, since many participants were irritated by it. As what the added questions concern, the application was judged to be easy to understand (question 6) and to learn (question 7) and, also, to play back a sound (question 8), despite the slight delay.

The rich verbal protocol returned several interesting results during the experiment. Many participants found a problem with the scales and on "deciding the scale whether to start with big or small sounds". Some participants found that it was "much easier to judge size over height". The "use of speed of repetition as characteristic of height" was found to be helpful in classifying the height of a sound. Another common problem was that the "metallic zips distract/confuse the classification of sounds", which referred to the 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 particularly 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. He 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 Figure 4.8.

4.3 Experiment: perceived elasticity of the impact/bounce event versus perceived force "throwing" the object.

Examining real and model generated sounds of impacts and bounces of objects made with different materials, the previous experiments reported in section 4.2 concerned the relationship between perceived height of the object drop and perceived size of dropped objects. In this section, we introduce our further investigations, which continue the previous work, by analyzing perceptually the impacts and bounces sounds from a different perspective, focusing on the relationship between other two important perceptual dimensions: perceptual elasticity of the impact/bounce and perceptual force "throwing" the object, as well as looking at the judgments about the quality and the realism of the synthesized stimuli. With perceptual force "throwing" the object we mean both the perception of the object just dropped without any force and the perception of the object thrown by applying some force.

We will present the experiment conducted and the data collected, and we will comment the results obtained, comparing them with those from the previous experiments.

4.3.1 Participants

The participants were 10 volunteers who were studying at the University of Limerick. Four of them referred to have a musical training for about 10 years, while five referred to have less than 3 years of musical training, including two that said to have never practiced music. One subject referred to practice music for 5 years. Nobody referred to have hearing problems, and four required glasses.

4.3.2 Stimuli

Since we obtained quite good results in comparing recorded and synthetic sounds in the experiments presented in section 4.2, at this stage we preferred to involve only sounds synthesized with the sound models.

We decided to design them all with the PD-modules modeling impact interactions of two modal resonators [RF03], simplified returning only one mode, because it was the model that gave us less spread data in the aforementioned works. As we aimed at investigating the relationship between elasticity and force applied to the objects, we decided for objects of two materials with completely different elasticity properties: wood and rubber.

The stimuli included in the sounds set were 18, consisting of 11 sounds of wood and 7 of rubber. All of them were sounds of bouncing events, excluded 2 for each material who consisted in single impact events, that were *sound7_wood*, *sound11_wood*, *sound6_rubber* and *sound7_rubber*. We note that the wood sounds were privileged in number compared with the rubber sounds. Since the focus of this experiment was on bouncing events, we preferred to give more relevance, within the stimuli set, to sounds with a material less obviously connected with bouncing than rubber.

Moreover, in designing the stimuli set, we paid attention to change slightly no more than two parameters simultaneously, in order to be able to make some observations on the influence of the parameters' values on the scaling and estimation results. In particular, we worked on the following parameters of the model: elasticity of the contact, force damping, gravity force, strike velocity, frequency, decay time. In Table 4.4 we report the values of the parameters used in the PD-modules for synthesizing the stimuli. Just as an overview of the most important parameters that we mention, the elasticity parameter is the hardness of the impact and the lambda parameter is the force damping weight, accounting for the dissipation of energy during the contact. The alpha parameter, that we kept constant, is related to the local geometry around the contact area. The gravity force parameter represents the gravity force applied to the "striker", and the strike velocity is the velocity of the object during the contact, i.e. when the distance between the two colliding objects is zero. The decay time parameter controls one of the most important features for material perception [AR01], i.e. the decay time, and the frequency parameter controls the resonant frequencies of the resonator. For a reference on the main features of the models and a detailed meaning of the parameters, we refer to [RF03].

Short Name	Elasticity k	Damping λ	Gravity force	Strike velocity	Frequency (Hz)	Decay time (s)
sound1_wood	3162.28	21.5443	580	2290.87	1782.52	0.009999
sound2_wood	2486	21.5443	580	2290.87	1782.52	0.009999
sound3_wood	1665.3	21.5443	580	2290.87	1782.52	0.009999
sound4_wood	3161.6	21.5443	580	363.078	1782.52	0.009999
sound5_wood	3161.6	21.5443	580	144.544	1782.52	0.009999
sound6_wood	3161.6	21.5443	750	2290.87	1782.52	0.009999
sound7_wood	163.03	21.5443	580	2290.87	1782.52	0.009999
sound8_wood	3161.6	34.5443	580	2290.87	1782.52	0.009999
sound9_wood	3161.6	21.5443	580	2290.87	1349.45	0.009999
sound10_wood	3161.6	34.1455	580	2290.87	1349.45	0.009999
sound11_wood	224.64	21.5443	580	2290.87	1782.52	0.009999
sound1_rubber	3162.28	21.5443	580	2290.87	1378.11	0.00247
sound2_rubber	2486	21.5443	580	2290.87	1378.11	0.00247
sound3_rubber	3161.6	34.1455	580	2290.87	1378.11	0.00247
sound4_rubber	3161.6	21.5443	750	2290.87	1378.11	0.00247
sound5_rubber	3161.6	21.5443	580	463.516	1378.11	0.00247
sound6_rubber	163.03	21.5443	580	2290.87	1378.11	0.00247
sound7_rubber	163.03	21.5443	580	2290.87	1094.65	0.00247

Table 4.4: *Values of the parameters for the synthesized sounds.*

4.3.3 Procedure

The experiment was conducted in a quiet, but not acoustically isolated room of the Interaction Design Centre, at UL. The procedure was the same as the one applied in the previous studies: The participants had to navigate in the bi-dimensional plot of the Sonic Browser, where they were represented by the cursor and free to decide about the aura size, to listen to the stimuli through headphones, by placing the mouse on the objects representing the sounds, and to move the objects according to the two axes of the plot, i.e. perceptual elasticity of the impact/bounce and perceptual force "throwing" the object.

As in the previous experiments, the resulting data coordinates have been normalized between 0 and 1, for being able to compare the objects' locations estimated by each user, who preferred to arrange the sounds according to their own scales, and not paying attention to the screen boundaries.

Besides collecting data logging, we also recorded each session on video-tapes, in order to keep the data coming from the Thinking-Aloud Protocol.

After the scaling task, the subjects were asked to tag the sounds which they judged unrealistic and, in the debriefing phase that concluded each session, each participant filled out a 7-point Likert scale questionnaire, from a "poor" evaluation (0) to a "excellent" evaluation (6). The questionnaire was similar to that used in the previous experiments, but contrary to those experiments, we preferred to enhance the fact that each session comprised two tasks, i.e. scaling and tagging, by asking the subjects to evaluate the difficulty of the two tasks separately.

4.3.4 Results and Observations

In Figure 4.16 (a) and Figure 4.16 (b), we report the representation of the individual perceptual scaling and tagging information sorted by stimuli, and grouped according to the material: wood and rubber respectively. In each plot, we represent the perceptual scaling and the tagging information of all the users for a certain stimulus. The two axes represent the perceptual dimensions estimated, that are perceptual elasticity of the impact/bounce and perceptual force "throwing" the object. The 2-D plot identified by the two axes represents the perceptual space, where we report the locations of that certain stimulus for all the participants to the experiment. As we already said in subsection 4.3.3, we decided to normalize the data coordinates between 0 and 1. In the plot, we highlight the tagging information by means of different shapes. With black circles, we represent the 'realistic' judgments by users, while with crosses we represent the 'unrealistic' judgments by users.

In Figure 4.17 the data are summarized by means of a box plot.

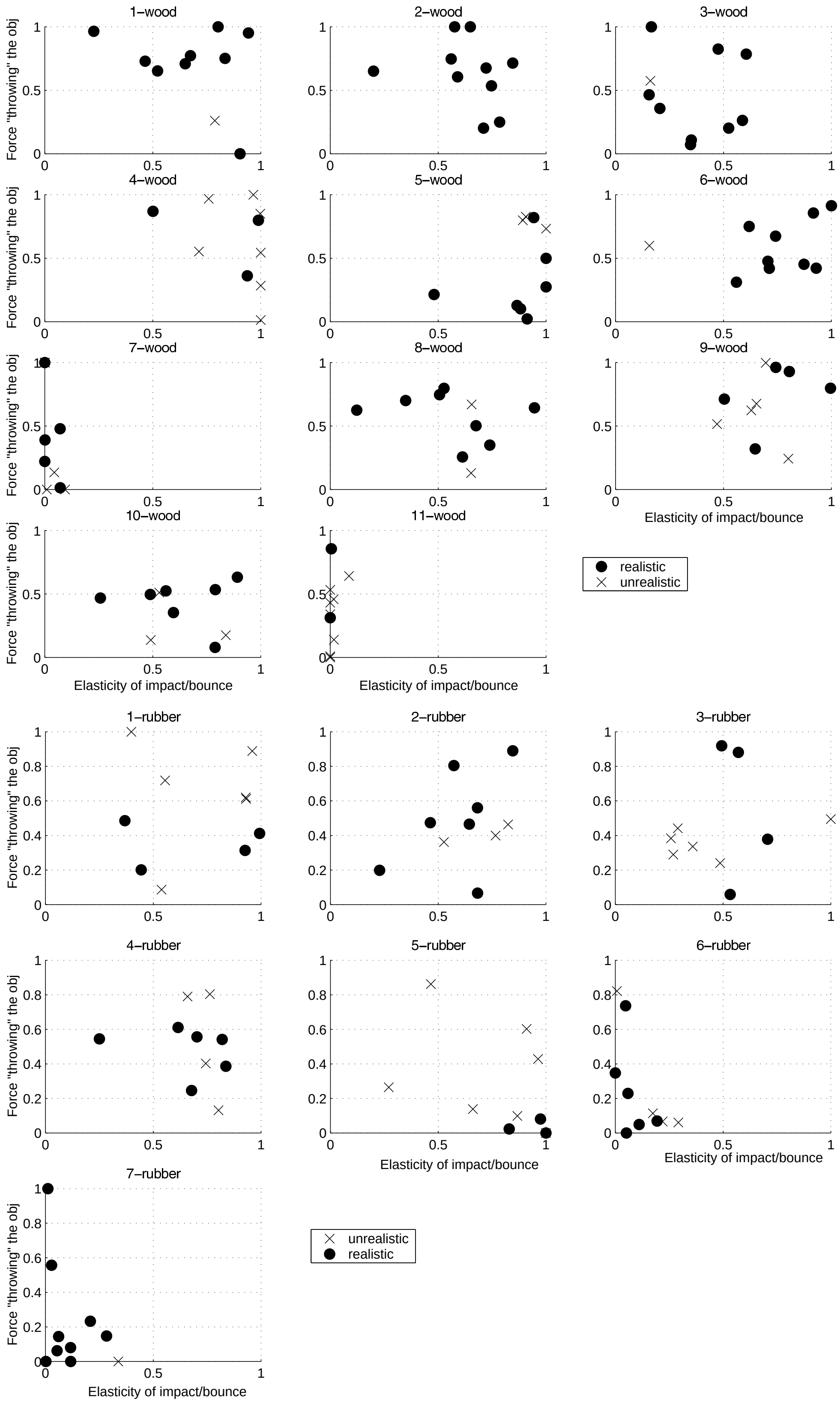
We can see that the sounds locations, according at least to one dimension, are slightly spread and some objects are placed in the 2-D plot uniformly in both the dimensions. In particular, the sounds *6-rubber* and *7-rubber* have only 2 outliers for the force axis, while *5-rubber* has 3 outliers for both the axes and *9-wood* has 4 and 5 outliers respectively for the elasticity and the force dimension.

Anyway, we can observe from the data collected that most of the users agreed in scaling the stimuli, at least in one dimension. In particular, a part from some outliers, the sounds *5-rubber*, *6-rubber*, *7-rubber* and *9-wood* were estimated uniformly in both the dimensions, while all the other sounds were judged uniformly in at least one dimension. We can observe that the sounds *6-rubber* and *7-rubber* are the only two stimuli of the rubber-set that consisted of a single impact event.

Looking at the tagging task, we can notice that there is one sound, *2-wood*, that was considered by all participants to be realistic, and 8 sounds were defined to be realistic by at least 7 users. It is interesting to underline that, among these 8 sounds, the sound *7-rubber* belongs to the best uniformly scaled stimuli.

The better results of the tagging task, achieved with this experiment rather than those conducted previously (section 4.2), could be due to the use of a sound set designed with the same sound model and of a model parameters' setting that is more value-centered and, maybe, because there were no real sounds in the stimuli set.

From the verbal protocol arose that the scaling task was difficult for some participants, because of the influence of other dimensions involved in the event, such as size or weight



(a) (b)

Figure 4.16: *Representation of the individual perceptual scaling and tagging information sorted by stimuli. (a) Only wood objects. (b) Only rubber objects. The perceptual scaling values are normalized between 0 and 1.*

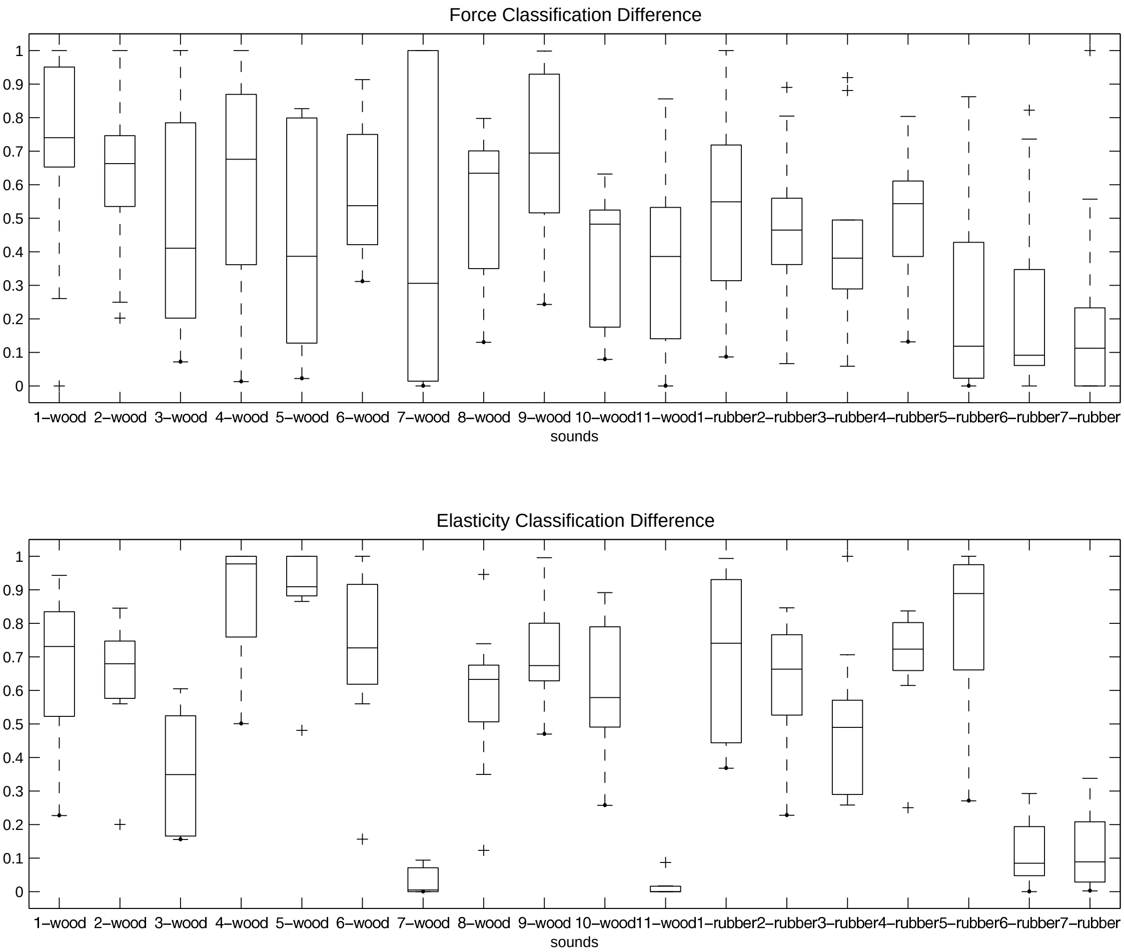


Figure 4.17: *Representation, by a box plot, of the perceptual scaling of the elasticity and force of the impact/bounce event sorted by stimuli.*

of the objects, loudness or height of the drop. In particular, five subjects referred to be influenced on their scaling task performances by a pitch variation in the sounds and two by the objects size.

For instance, a participant referred that "it's really hard to judge about force, because they sound like they are different sized objects. So you don't know whether the impact sound that they make is because they are thrown, or because they are bigger". Four subjects were "confused from them (some sounds) thrown from a different height". Some participants were biased by the loudness, the hardness of the dropping objects, their weight or their material. In particular, one subject reported to hear not only the dropping object but also the object where it is dropping on, its shape and material, by saying that "the object where has been dropped on sounds quite strange. It is a kind of complex thing being dropped on and there's something else vibrating as well " and, moreover, "it's like it has been dropped inside an object, where there is a kind of rim that is vibrating", and, as far as the material is concerning, "that's a completely different material, this one, where it is dropping on". Another participant noted that some dropping objects " hit on something that has a different density or dropped from a lower height".

As we already underlined, all the sounds were scaled quite uniformly, at least in one dimension, and, moreover, we can observe that the four sounds of single impact events, i.e. *7-wood*, *11-wood*, *6-rubber* and *7-rubber*, were all judged to have minimum elasticity, as it could be expected, and all were scaled very uniformly, probably because the scaling of the single impact events is not affected by the bouncing pattern. For example, in a bouncing sound "the second and the third (bounces) go higher, louder ... I think it's higher in elasticity" and "the sound of it seems to bounce for too long and the bouncing thing is very, very small ... too small for it". In addition, we can see that, among them, the rubber objects converge to the zero-elasticity point less than the wood objects, probably due to the characteristic elasticity of the materials involved. We can see this by looking at Figure 4.18, where we plot the centroid of the scaling of each stimulus. These centroids are computed by keeping all the values, including the outliers. Although the barycentres would be more accurate by excluding the outliers, the plot represented in Figure 4.18 can give a general view of the stimuli mean positions within the perceptual space.

Observing the locations of the single impact sounds, we can notice a distinction of their perceptual estimations from the rest of the stimuli set and, moreover, that the material characteristics of these events appear to be distinct on both the scaling dimensions. In fact, the wood objects with a single impact sounds are identified to have in average less elasticity and more force rather then the rubber objects.

In the plot of the centroid three more sounds, that are *4-wood*, *5-wood* and *5-rubber*, are judged to have a high level of elasticity, resulting to be on the right hand of the graph, quite separated from the other stimuli. We can observe that these three sounds were the only in the set to have the strike velocity parameter with a lower value, rather than the others, which could mean that the changing in the strike velocity in the model is perceived as an increase in elasticity.

As far as the other stimuli locations are concerned, we can see the complex relationships that connect the model parameters with the stimuli, although the centroid positions are computed by keeping also the outliers.

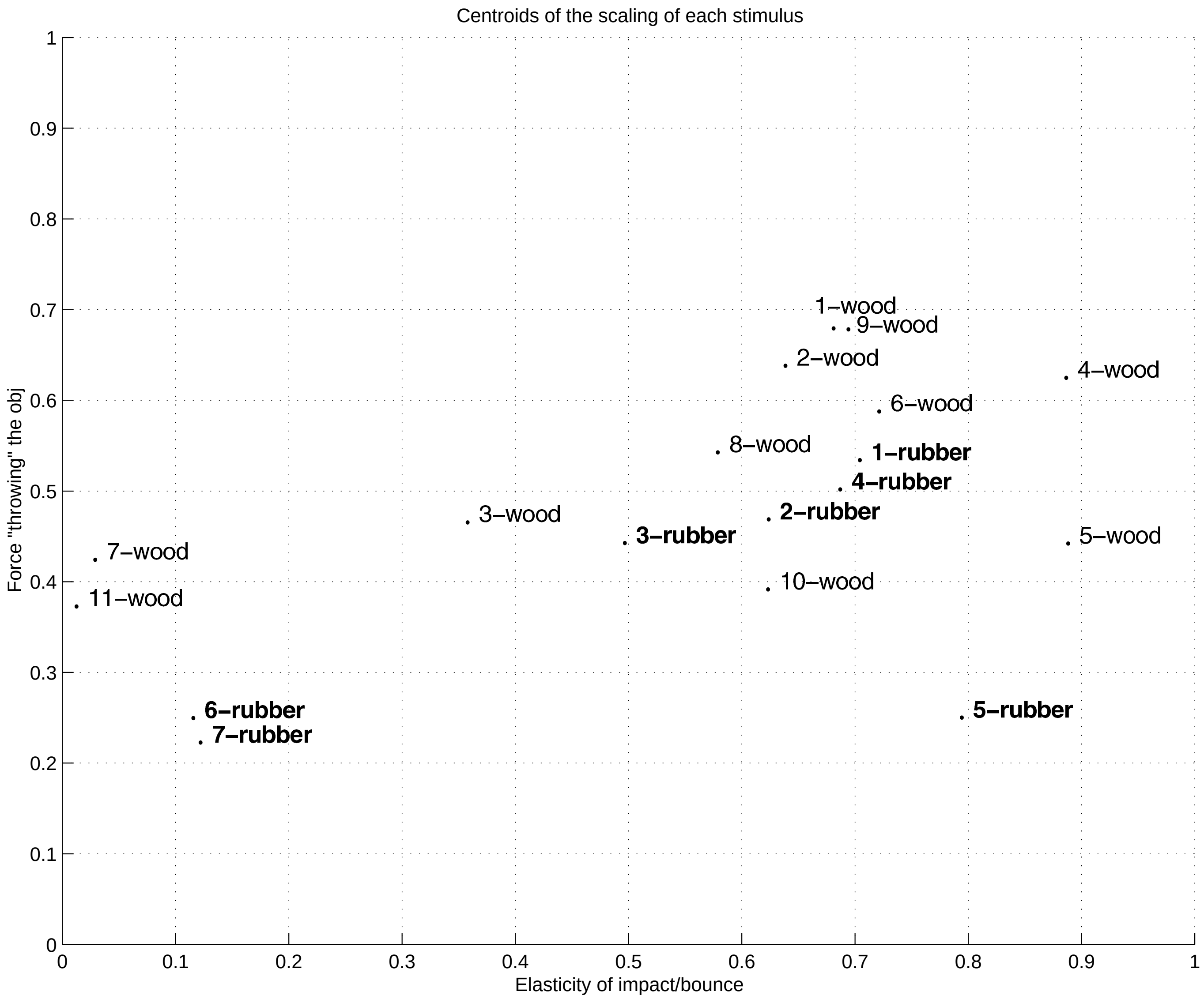


Figure 4.18: *Centroids of the scaling of each stimulus, sorted by material (Rubber objects written in bold). The perceptual scaling values are normalized between 0 and 1.*

In Figure 4.19 and Figure 4.20 we plot, sorted by parameters, the centroids of the scaling of the perceived elasticity and of the perceived force, respectively, versus the parameters values, for the wood objects. In Figure 4.21 and Figure 4.22, we report the same plots calculated for the rubber objects. In the plots, stimuli parameters with the same value are highlighted with horizontal dashed lines.

From the plots, we can see that the relationship between the parameters space and the perceptual space is complex and it is difficult to identify some clear trend in them, maybe due to the parameters values that were established during the synthesizing phase. Anyway, we try to point out some interesting behaviour.

We can see that there are three wood stimuli, i.e. the sounds *1-wood*, *2-wood* and *3-wood*, and three rubber stimuli, i.e. the sounds *1-rubber*, *2-rubber* and *6-rubber*, whose parameterization differs only in the elasticity parameter. All of them show both a different perceived elasticity and a different perceived force. It is evident also the pattern that we underlined previously, concerning the sounds *4-wood*, *5-wood* and *5-rubber* that have a low strike velocity parameter and show an high perceived elasticity.

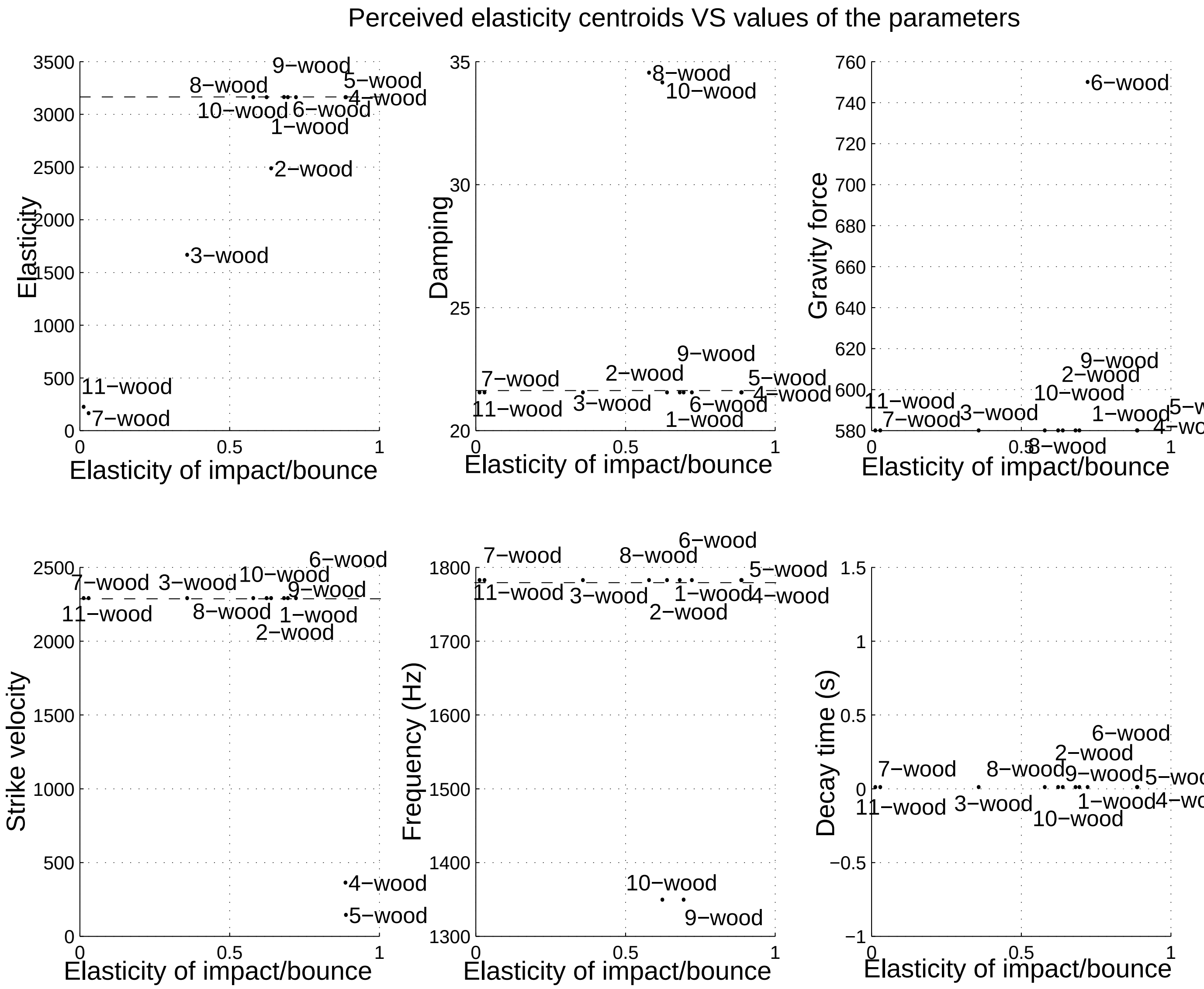


Figure 4.19: *Centroids of the scaling of the perceived elasticity of the wood objects versus the parameters values, sorted by parameters. The perceptual scaling values are normalized between 0 and 1. Stimuli parameters with the same value are highlighted with horizontal dashed lines.*

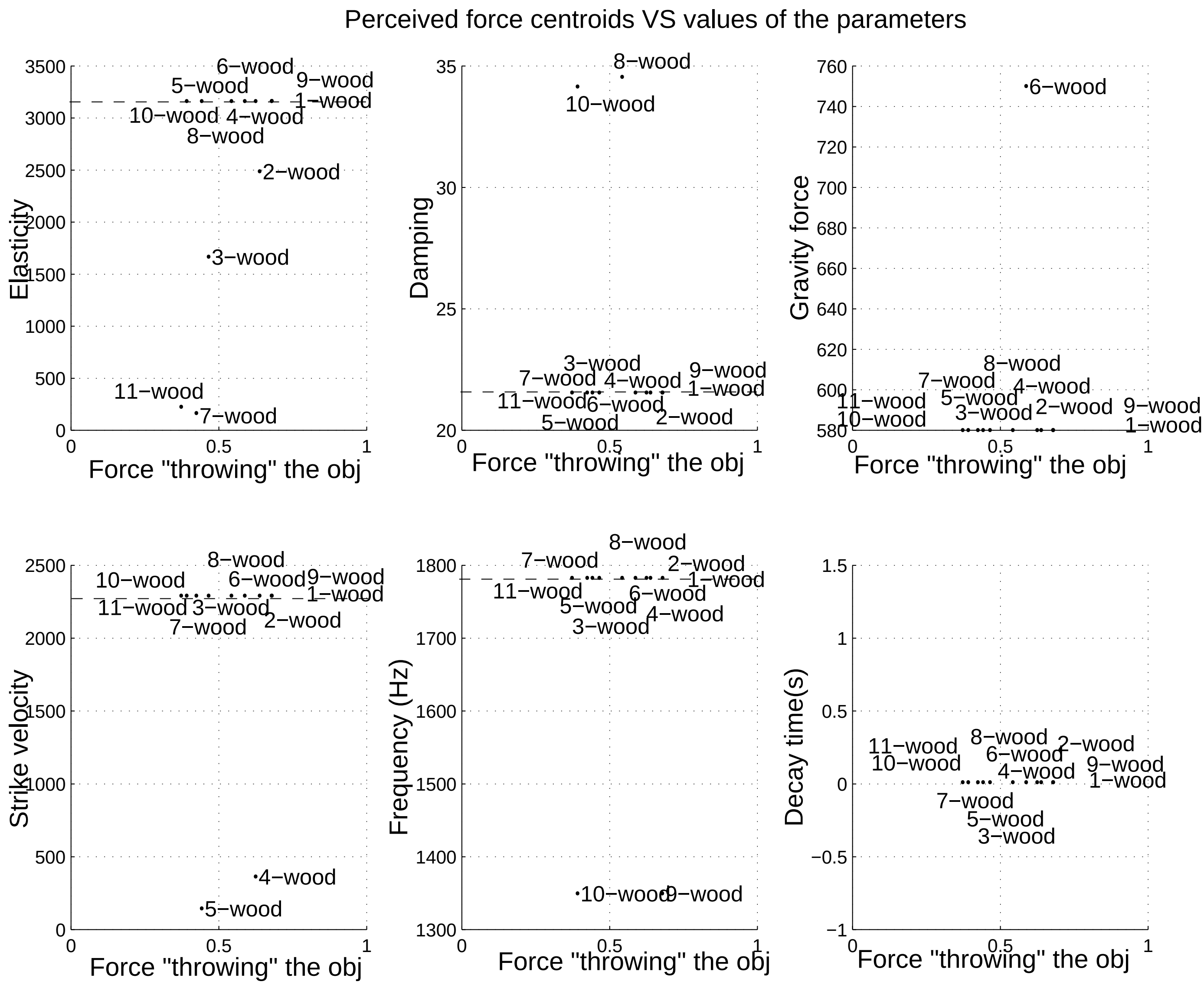


Figure 4.20: *Centroids of the scaling of the perceived force of the wood objects versus the parameters values, sorted by parameters. The perceptual scaling values are normalized between 0 and 1. Stimuli parameters with the same value are highlighted with horizontal dashed lines.*

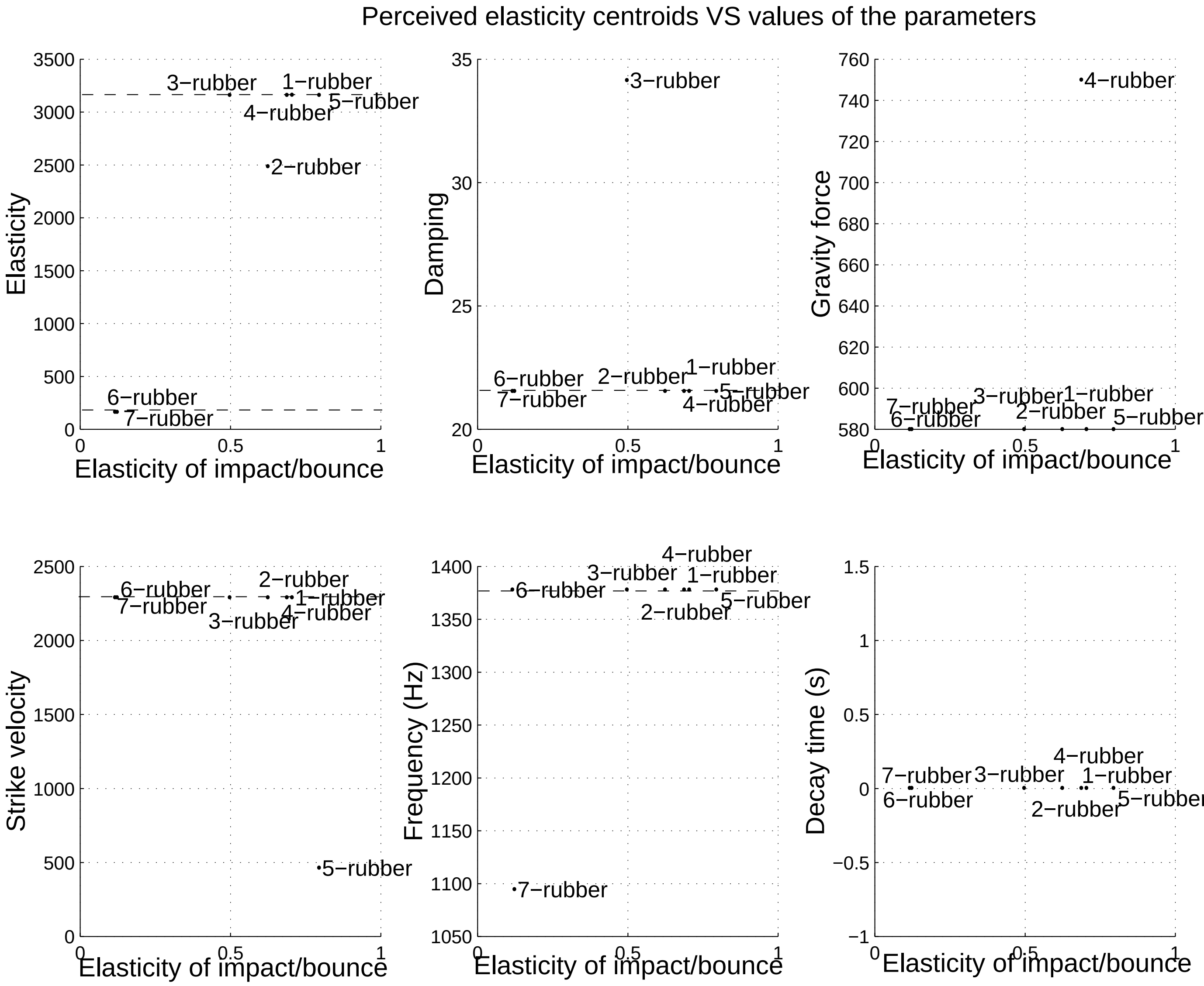


Figure 4.21: *Centroids of the scaling of the perceived elasticity of the rubber objects versus the parameters values, sorted by parameters. The perceptual scaling values are normalized between 0 and 1. Stimuli parameters with the same value are highlighted with horizontal dashed lines.*

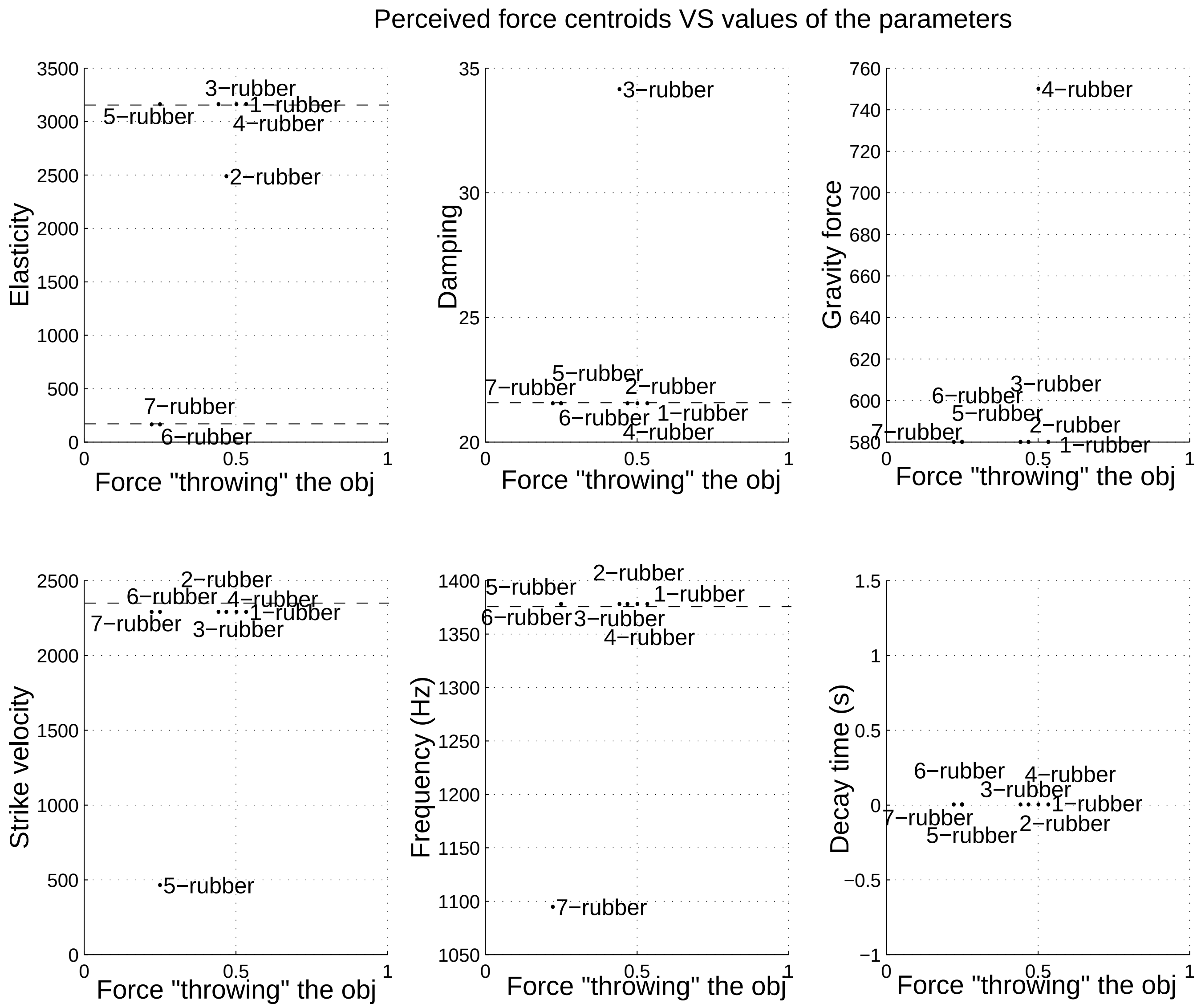


Figure 4.22: *Centroids of the scaling of the perceived force of the rubber objects versus the parameters values, sorted by parameters. The perceptual scaling values are normalized between 0 and 1. Stimuli parameters with the same value are highlighted with horizontal dashed lines.*

By the sessions task performances, it is clear the utility of the aura for conducting the scaling task. Usually, at the beginning of the experiment, the subjects took “the aura smaller. I will listen to them (the sounds) all together when they are organized a bit more. I slow down the aura”, while they used the aura for comparing the sounds and evaluating their judgments afterwards.

Moreover, it is interesting to notice that some participants preferred to judge the two dimensions together, while some others preferred to start by scaling one dimension and then moving to scale the other, because “it’s really hard to judge the two together”. In fact, “if I will try to think of the two things I will come confused”.

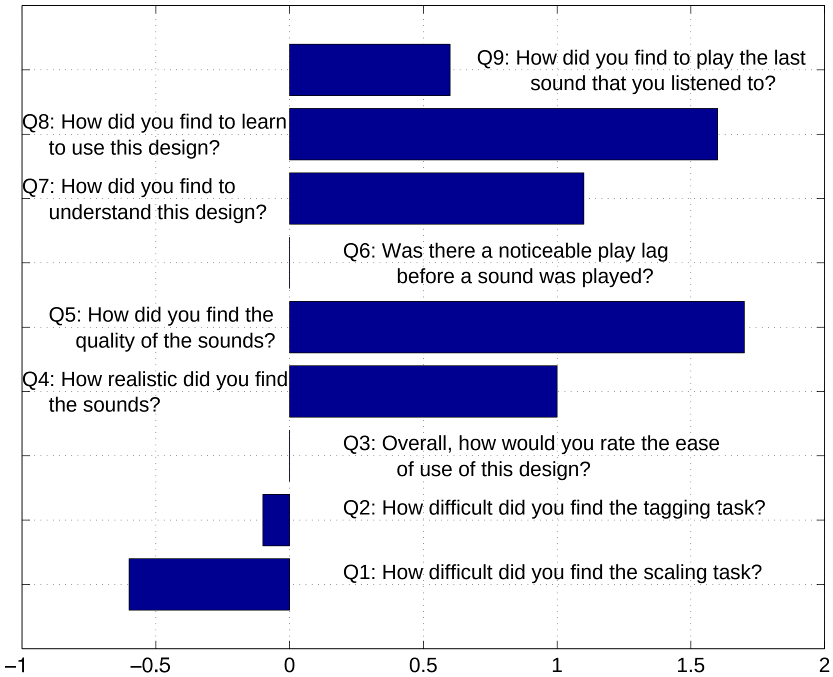


Figure 4.23: *Results of the questionnaire filled out during the debriefing phase: Cumulative Participant Response.*

In Figure 4.23 we report the results of the questionnaire filled out by each participant during the debriefing phase, by representing the cumulative participant response with a bar chart.

The most interesting result, arisen from it, is that, even if the users found both the tasks difficult, they judged the scaling task harder then the tagging task, probably because of the influence of other dimensions in the event perception, as resulting from the verbal protocol as well. Despite this difficulty, the sounds were scaled consistently, at least in one dimension.

Moreover, the sounds were judged to be realistic and of good quality and these positive results could be connected to those obtained from the tagging task, as we have already reported.

As far as the software application is concerned, the ease of use of the Sonic Browser was judged to be on average, the slight delay of up to .3 seconds for playing a sound was noticed but accepted, and it didn't affect the tasks performances, since the question about the difficulty in playing the last sound listened was answered positively. Finally, good evaluations were achieved for the questions regarding how subjects found the interface to be understandable and learnable.

4.4 Conclusions

As regards the experiments about perceived height versus perceived size, examining the results we can state that sounds convey information about the perceptual dimensions investigated even if the sounds are simplified, as we could see by using, in the experiments, also stimuli synthesized with simplified sound models with only one resonating mode. Apart from one case in the pilot probe, the unrealistic perception of sounds did not affect the possibility for the participants of being able to perceptually scale the perceptual dimensions we investigated. This illustrates that the "realism" of a sound does not affect the amount of information extracted by a participant. Moreover, synthetic sounds can be recognized as unrealistic events but their high-level parameters can still be extracted and evaluated, as it is stated by the technique of sound cartoonification. This issue was already discussed by Rath [RF03], focusing on sound objects in combination with cartoonification.

By analyzing the results concerning perceived elasticity versus perceived force throwing the object, we can see that the Sound Objects are judged to be quite realistic as well and, even if they are tagged as unrealistic, because they are cartoonification of the reality, they still convey information and the physical properties of the events are still perceived by the listeners.

In particular, we have seen that the participants distinguished quite clearly among materials and about event identity. We have noted a parameter, the strike velocity, that could be particularly involved in the elasticity scaling. Nevertheless, some other investigations are needed in order to confirm this hypothesis.

As in the other experiments, the results are affected by the influence of other dimensions, which weren't examined in this case. The sound objects could be located in a multidimensional physical space and in a multidimensional perceptual space, connected to each other by a complex relationship.

Nevertheless, the sound objects resulted to be more uniformly estimated rather than those of the previous experiments. Therefore, we can state that, by providing to users sound objects synthesized with the same model, the listeners could scale the stimuli more easily and clearly. In this way, the sounds could convey information to the listeners. Moreover, even if the sounds are judged to be not realistic, what it is important is not to introduce distractors, such as a buzz tail, that could turn the user listening attitude.

The experiments presented in this chapter show that, by using sounds of impact/bounce events synthesized with parameters accurately decided, an auditory display could provide quantifiable information of multi-dimensional data. In particular, we noted the ease for the users to recognize and estimate sounds even if recognized as unrealistic. Moreover, although the parameters involved in the sound synthesis are connected to each other in a complex relationship, it is possible to enhance some features rather than others in order to highlight certain aspects and to make those characteristics pop-out.

Chapter 5

Perceptual dimensions of the rolling process

Besides the impact/bounce event, another important sound event category, that is common in the everyday auditory environment, is represented by rolling sounds.

There are some works about rolling sounds synthesis [vdDKP01], but there are few studies about auditory perception of rolling sounds. Houben, Kohlrausch and Hermes in [HKH01] investigated the relationship between size and speed of rolling balls. They worked with both recorded sounds and manipulated sounds in order to evaluate the interaction of the two dimensions. As far as the recorded sounds perception is concerned, subjects were able to discriminate size and speed of the balls by listening to their rolling sounds, even if some of them found difficult to label the sounds with the right speed value. Anyway, the experiments showed an interaction effect of the two dimensions especially when both the parameters had high values, that is when the subjects listened to big balls rolling at high velocities. On the contrary, high speed values seemed to be more easily evaluated if the rolling balls were smaller.

> The interaction effect encountered in the previous experiment, when both size and velocity of a rolling ball are increased, may be caused by the fact that changes in these two physical parameters affect auditory cues which are used by the subjects, in a similar way [HKH01].

In order to study the auditory cues which affect the perception of speed and size of rolling balls, Houben et al. conducted auditory experiments with sounds manipulated by changing the temporal content of a sound with the spectral content of another one. They found that, while the size of rolling balls is identified by listeners according to the spectral content of the sounds and it is not affected by a variation of speed, the perception of speed is much more difficult if size is varied simultaneously and fast rolling balls are often confused with small balls according to their spectral content [HKH01].

In a more recent paper [HS02], Houben and Stoelinga focus on some temporal aspects of rolling sounds, by studying, through perceptual experiments, the influence of amplitude modulation on the perceived size and speed, by presenting to subjects recorded sounds of wooden balls rolling over a wooden plate. From the experiment, amplitude modulation resulted to be important for the perception of both the dimensions of the rolling process. Moreover, the authors suggest to apply amplitude modulation to synthesized rolling sounds in order to make them more natural.

In this context, we conducted a psychoacoustical experiment on rolling sound processes modeled by the EU-funded project "the Sounding Object" (SOb)[1]. The chapter was written with the precious help of the studies conducted by the student Germana Olivieri.

The experiment we conducted aims at finding some perceptual dimensions of sound objects which should be easy to be identified and quite uniformly estimated in order to convey data information. Instead of focusing on the auditory features of the sound waves, we investigate the perceptual dimensions of the sound processes and their relationship in the perceptual space. As we did for the other experiments conducted within our research activity, we used stimuli synthesized with sound models, in order to be able to control directly the physical dimensions of the process.

Contrary to chapter 4, we didn't use the Sonic Browser for conducting the experiment, as from informal experiments we noticed that the application was not suitable for those sound processes, probably due both to their length and their identity. Therefore, we preferred to develop a simple interface using the Matlab environment for conducting the tests and collecting the perceptual scaling data.

5.1 Experiment: perceived velocity versus perceived size of the rolling object

The experiment was preceded by a pilot probe in order to check the stimuli set, the experimental method applied and the type of data results achieved. In this section we will briefly present just the experiment and its main results and, only when necessary, we will mention the pilot probe.

5.1.1 Participants

The participants to the experiment were 12 volunteers with age ranging between 26 and 30, all students at the Department of Computer Science of the University of Verona. All of them reported to have no hearing problem.

5.1.2 Stimuli

The stimuli set consisted of 9 Sound Objects synthesized with the PD-modules modeling rolling processes. They were chosen among a set comprising 18 sounds, that in the pilot

[1] `http://www.soundobject.org`

Short Name	Velocity (m/s)	Amplification	Diameter (cm)	Surface depth (mm)
rot1	0.116188	63.0957	3.16228	0.0218776
rot2	**0.116188**	**162.181**	**3.16228**	**0.0218776**
rot3	0.116188	30.903	3.16228	0.0218776
rot4	0.116188	63.0957	2.16272	0.0218776
rot5	**0.116188**	**63.0957**	**2.16272**	**0.0549541**
rot6	**0.116188**	**63.0957**	**3.16228**	**0.0549541**
rot7	**0.116188**	**162.181**	**2.16272**	**0.0218776**
rot8	0.116188	30.903	2.16272	0.0218776
rot9	0.223888	30.903	2.16272	0.0218776
rot10	**0.223888**	**63.0957**	**2.16272**	**0.0218776**
rot11	**0.223888**	**162.181**	**2.16272**	**0.0218776**
rot12	0.223888	162.181	3.16228	0.0218776
rot13	0.223888	63.0957	3.16228	0.0218776
rot14	**0.223888**	**30.903**	**3.16228**	**0.0218776**
rot15	**0.661735**	**63.0957**	**3.16228**	**0.0218776**
rot16	**0.661735**	**63.0957**	**2.16272**	**0.0218776**
rot17	0.41929	63.0957	3.16228	0.0218776
rot18	0.41929	63.0957	2.16272	0.0218776

Table 5.1: *Values of the parameters for the synthesized sounds. (The sounds used in the main experiment written in bold).*

probe resulted to be too large for a listening experiment where we wanted to repeat each sound 3 times. Therefore, we selected 9 of them, according to their quality and the values of the parameters. In Table 5.1 we report the values of the parameters used for the synthesis of all the 18 sounds and we highlight in a bold font the 9 sounds used in the main experiment, that are reported in Table 5.2 as well. The velocity parameter controls the velocity of the rolling object, and the diameter parameter controls its size. The surface depth parameter deals with the roughness of the surface where the object is rolling on and the amplification is a constant multiplying the amplitude of the obtained signal in order to control the loudness of the sound. For a reference on the main features of the model and the meaning of the parameters, we refer to [RF03].

According to the observations from the experiments on the impact/bounce event (see chapter 4), we tried to synthesize this stimuli set paying attention to change slightly no more than two parameters simultaneously, in order to be able to make some observations on the influence of the parameters values on the scaling and estimation results.

	Short Name	**Velocity (m/s)**	**Amplification**	**Diameter (cm)**	**Surface depth (mm)**
sound1	*rot2*	0.116188	162.181	3.16228	0.0218776
sound2	*rot5*	0.116188	63.0957	2.16272	0.0549541
sound3	*rot6*	0.116188	63.0957	3.16228	0.0549541
sound4	*rot7*	0.116188	162.181	2.16272	0.0218776
sound5	*rot10*	0.223888	63.0957	2.16272	0.0218776
sound6	*rot11*	0.223888	162.181	2.16272	0.0218776
sound7	*rot14*	0.223888	30.903	3.16228	0.0218776
sound8	*rot15*	0.661735	63.0957	3.16228	0.0218776
sound9	*rot16*	0.661735	63.0957	2.16272	0.0218776

Table 5.2: *Values of the parameters for the synthesized sounds used in the main experiment.*

5.1.3 Procedure

The experiment was conducted with the method of category rating in a quiet, but not isolated room.

The subjects were asked to perform the test by means of an interface designed in the Matlab environment. They listened to sounds through closed headphones (Beyerdynamic DT-770) and they had to rate the perceived velocity and the perceived size of the rolling object respectively in the range 1-10 for the velocity and in the range 1-6 for the size.

All the sounds were repeated 3 times and presented in random order to the subjects who could listen to them as many times as they wanted. The experiment was conducted with the Thinking Aloud Protocol, as it provided good results in the experiments on the impact/bounce event (see chapter 4). Each test session was recorded on audio tape.

At the end of each session, there was a debriefing phase, where the subjects were asked to fill a 6-point Likert scale questionnaire, from a “poor” evaluation (1) to an “excellent” evaluation (6). The questions were about the ease of the required task performance, about the quality of the sounds and about their realism.

5.1.4 Results and Observations

In Figure 5.1 and Figure 5.2 we report, by means of histograms, the relative answers frequency respectively for the perceived size and the perceived velocity values, sorted by stimuli. Each bar of the histogram represents the relative frequency of a certain answer for a certain stimulus. The histogram is sorted by stimuli and the category values are represented by different colors, from white — the minimum, to black — the maximum, as explained by the legend in the plots. To clarify, if all the answers relative to a certain sound had all the same category value, i.e. all the participant rated the same sound with the same value all the times they listened to it, the relative answers frequency would be equal to 1. Therefore, the higher is the relative answers frequency, the more uniform are the judgments of the subjects.

Although the relative answers frequency for both the perceptual dimensions is not very high, we can see more clear judgments concerning the size — it reaches the 50% in two cases and the 56% in one case, while for the velocity all the answers frequencies are below the 40%. It is probably due to the several perceptual features that affect the velocity perception more than the size perception, such as the intensity of the sound and the roughness of the surface.

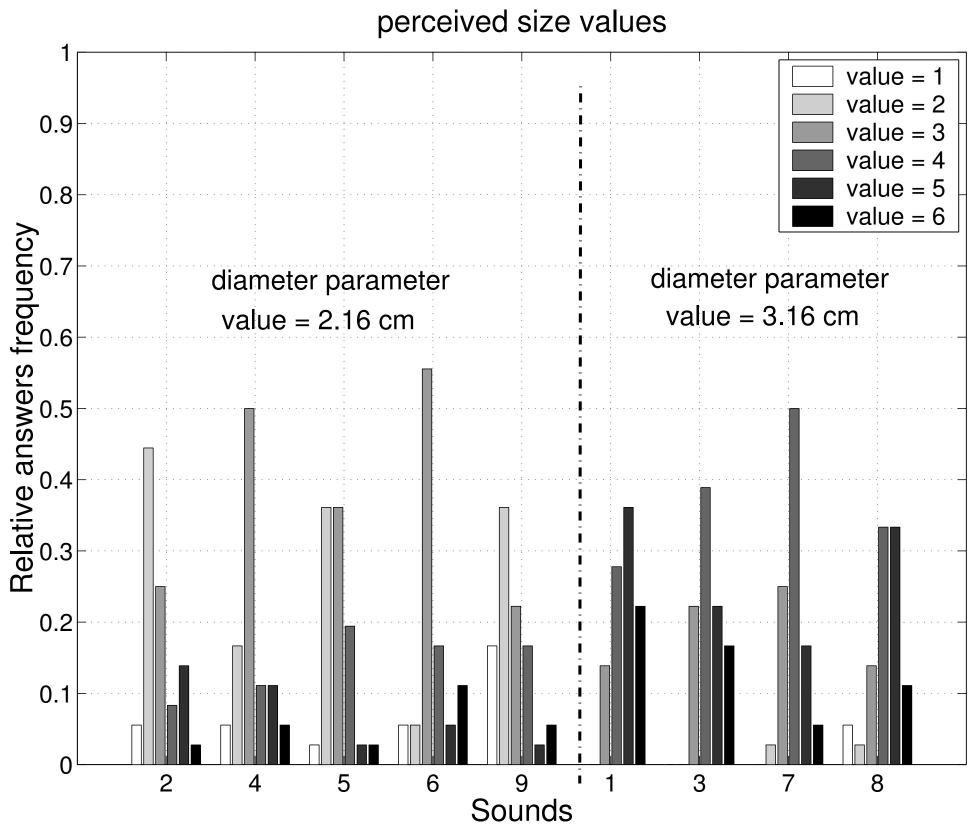


Figure 5.1: *Summary of the relative answers frequency of the perceived size sorted by stimuli. Each bar of the histogram represents the relative frequency of a certain answer for a certain stimulus. The histogram is sorted by stimuli and the category values are represented by different colors, from white — the minimum, to black — the maximum. The stimuli are divided according to the diameter parameter values.*

In the plot of the relative answers frequency of the perceived size (Figure 5.1), we divided the stimuli according to the diameter parameter values, while in the plot concerning the perceived velocity (Figure 5.2), we divided them according to the velocity parameter values.

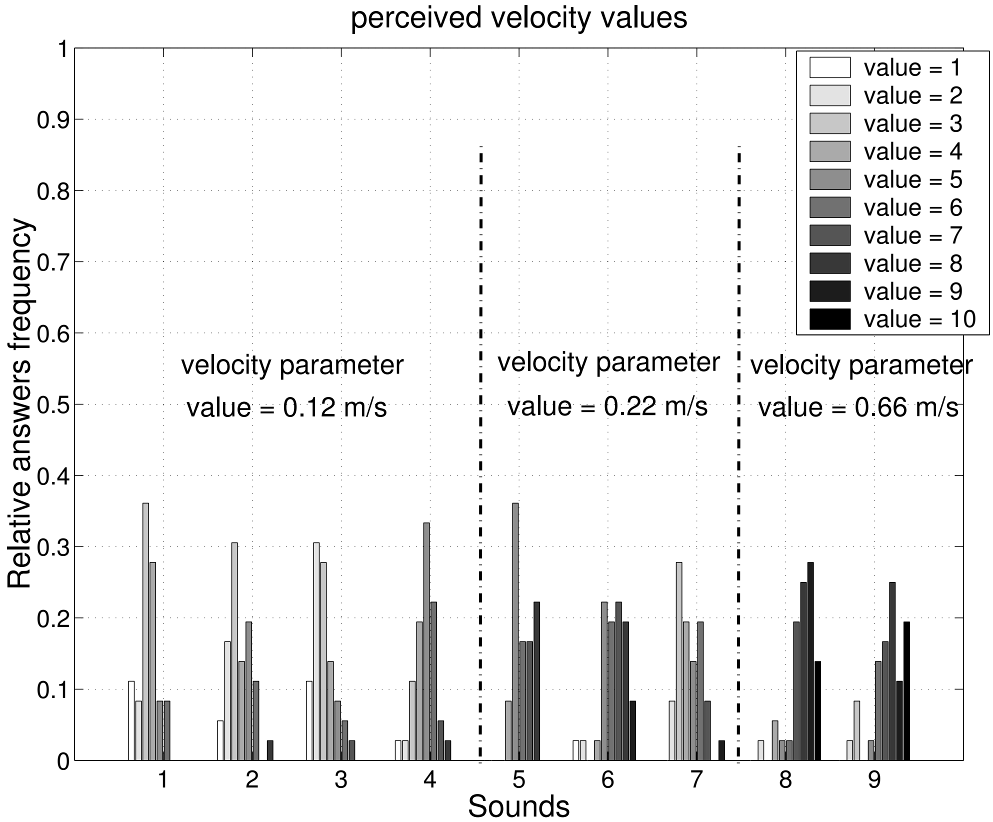


Figure 5.2: *Summary of the relative answers frequency of the perceived velocity sorted by stimuli. Each bar of the histogram represents the relative frequency of a certain answer for a certain stimulus. The histogram is sorted by stimuli and the category values are represented by different colors, from white — the minimum, to black — the maximum. The stimuli are divided according to the velocity parameter values.*

Therefore, we have two groups for the diameter parameter, i.e. 2.16 cm and 3.16 cm, while three groups for the velocity parameter, i.e. 0.12 m/s, 0.22 m/s and 0.66 m/s.

From the summary plot regarding the size perception (Figure 5.1) we can see that, for the stimuli synthesized with small diameter value — *sound2*, *sound4*, *sound5*, *sound6*, *sound9*, the answers frequencies are centered on the lowest estimated values, while, for the other stimuli group — *sound1*, *sound3*, *sound7*, *sound8*, they are centered on the highest range. This is more clear if we look at Figure 5.3 where we plot, by means of an histogram, the most frequent estimations relatively to the perceived size dimension. On the y-axis there are the category values and each bar of the histogram represents the most frequent estimations for each sound. In the plot, the stimuli are divided according to the diameter parameter values.

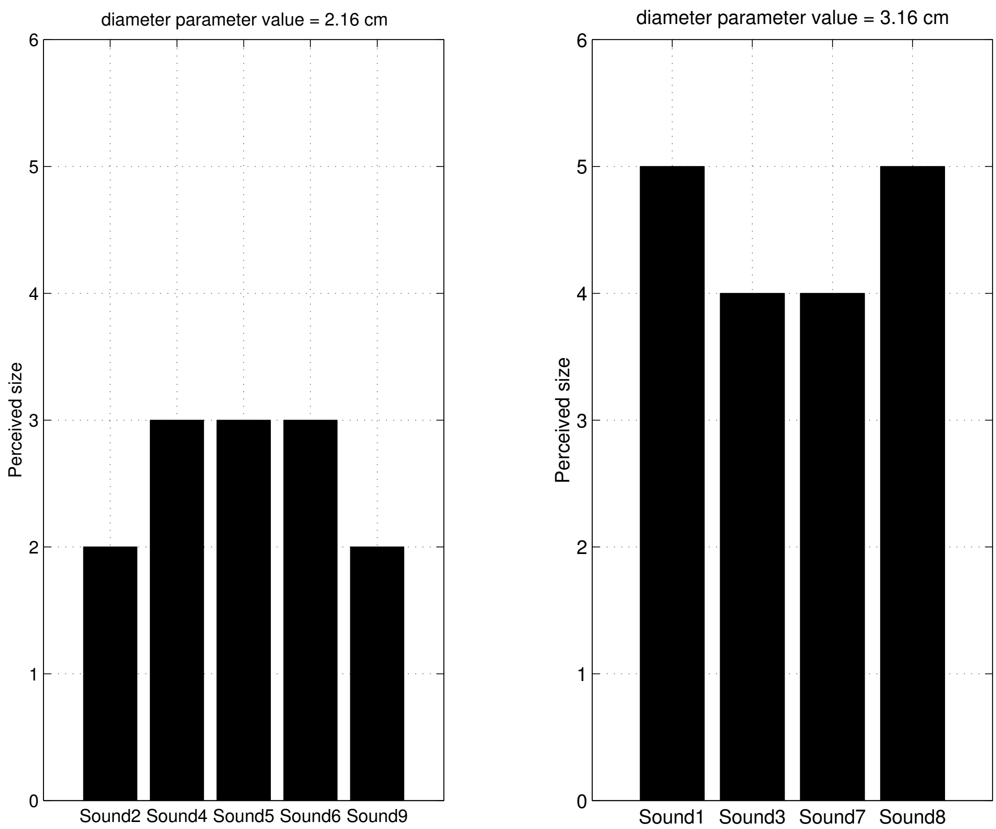


Figure 5.3: *Summary of most frequent estimations of the perceived size sorted by stimuli and divided according to the diameter parameter values.*

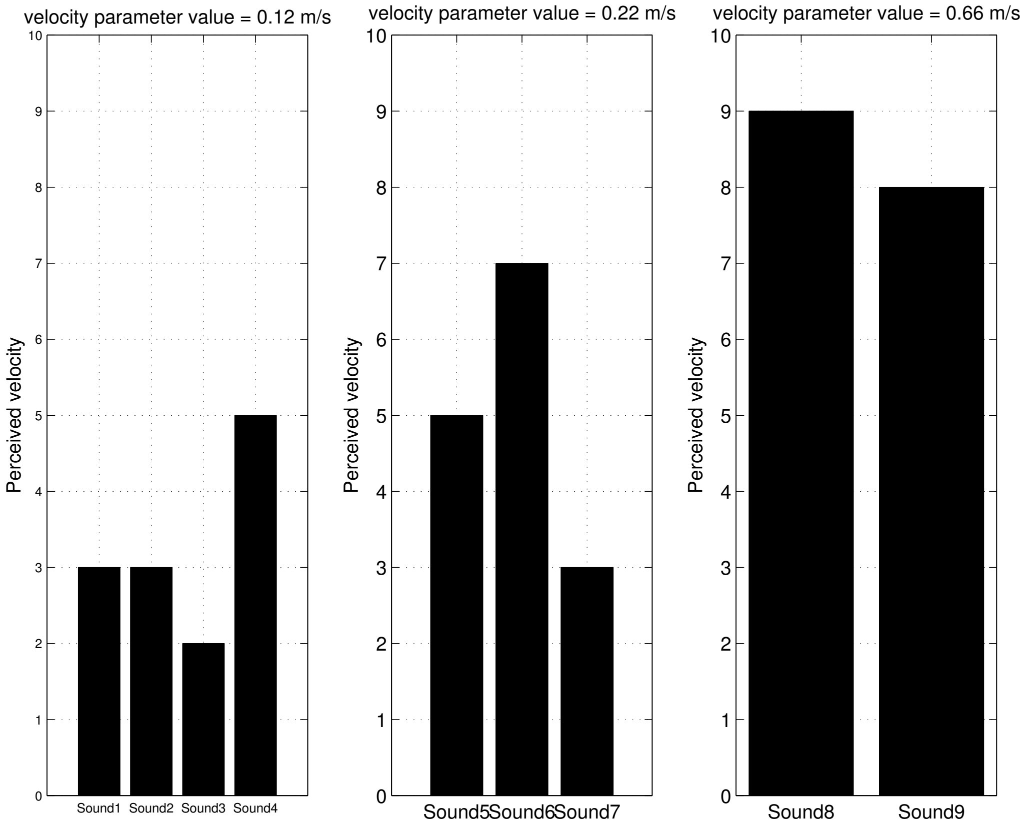


Figure 5.4: *Summary of most frequent estimations of the perceived velocity sorted by stimuli and divided according to the velocity parameter values.*

In Figure 5.1, we can see that only two stimuli, *sound5* and *sound8*, were judged with difficulty by the participants to the test, as there are two peaks in the relative answers frequency. In case of equal answer frequency, as maximum to be reported in Figure 5.3, we decided to consider the higher by default.

From Figure 5.3, it is clear that the sounds are able to provide category information quite well, since we can see the range of the perceptual dimension evaluations increasing with the parameter values. Moreover, for some sounds synthesized with the same diameter parameter value we can see even equal estimated values.

While the size dimension is easier to be estimated, the velocity dimension is affected, as we have already mentioned, by other parameters, especially by the amplification value. For instance, if we compare *sound1* and *sound2*, which have the same velocity parameter 0.12 m/s, we notice a spread estimation for the second stimulus, probably influenced by the change of both the amplification and the roughness values. *Sound4* represents an exception in the answers frequency plot, as its evaluation is shifted to higher values compared to the same synthesizing velocity value group.

It is more clear in Figure 5.4, where we plot, by means of an histogram, the most frequent estimations relatively to the perceived velocity dimension, as we did for the perceived size. On the y-axis there are the category values and each bar of the histogram represents the most frequent estimations for each sound. In the plot, the stimuli are divided according to the velocity parameter values.

As we already mentioned, the evaluation of *sound4* is shifted to higher values compared to the same synthesizing velocity value group. Probably, this fact is due to the amplification value which in this case is as high as for *sound1*, but the diameter value is lower.

As regards the second velocity group, comprising *sound5*, *sound6*, *sound7*, the estimated value range is shifted to higher values. In Figure 5.2, we can notice uniform judgments for *sound5*, while for *sound6* there is more uncertainty. A strange result regards *sound7*, whose answers frequency reaches its peak at the value 3, less than the previous results, as it is shown in Figure 5.4. This could be related to a clear decrease in the amplification value, which could have been interpreted by the listeners as a decrease in velocity.

Similar results are obtained in the third velocity group, where we can see uncertainty by the subjects in choosing a clear estimation peak. Anyway, the evaluation range is moved to the highest values, mirroring the velocity parameter setting used for synthesizing that group.

From the verbal protocol an interesting result arose. Most subjects spontaneously referred to the events they were listening to as to a rolling process. Some of them talked about an irregular shaped object such as an egg or a scraped ball. Many participants pointed out that they perceived a rolling objects moving on a not-straight path, such as a sinusoid or inside a sort of cone. Moreover, most of the times the subjects specified the materials involved in the interaction. The material perception of both the object and the surface will be investigated in further experiments.

It is interesting to look at the centroid position plot (Figure 5.5), where we represent the barycentres of the scaling of each stimulus within the perceptual space, identified by the two perceptual dimensions we investigated. In the plot, we highlight the different perceptual groups related to the parameter settings for the size and for the velocity. We can see that *sound7* is an outlier for the velocity scaling, since it belongs to the second velocity group but it has a low mean position for the perceptual velocity dimension. This exception could be related to the very low amplification parameter value, as we already mentioned. Another interesting observation is that *sound4* and *sound5* have the same mean position within the perceptual space and therefore, even if they have a different parameters setting, they are perceived to be the same sound object. We note that the parameters that change in these two sounds are amplification and velocity and, in particular, while *sound4* has an high amplification value and a low velocity value, *sound5* has a low amplification value but a velocity value that is higher than *sound4*.

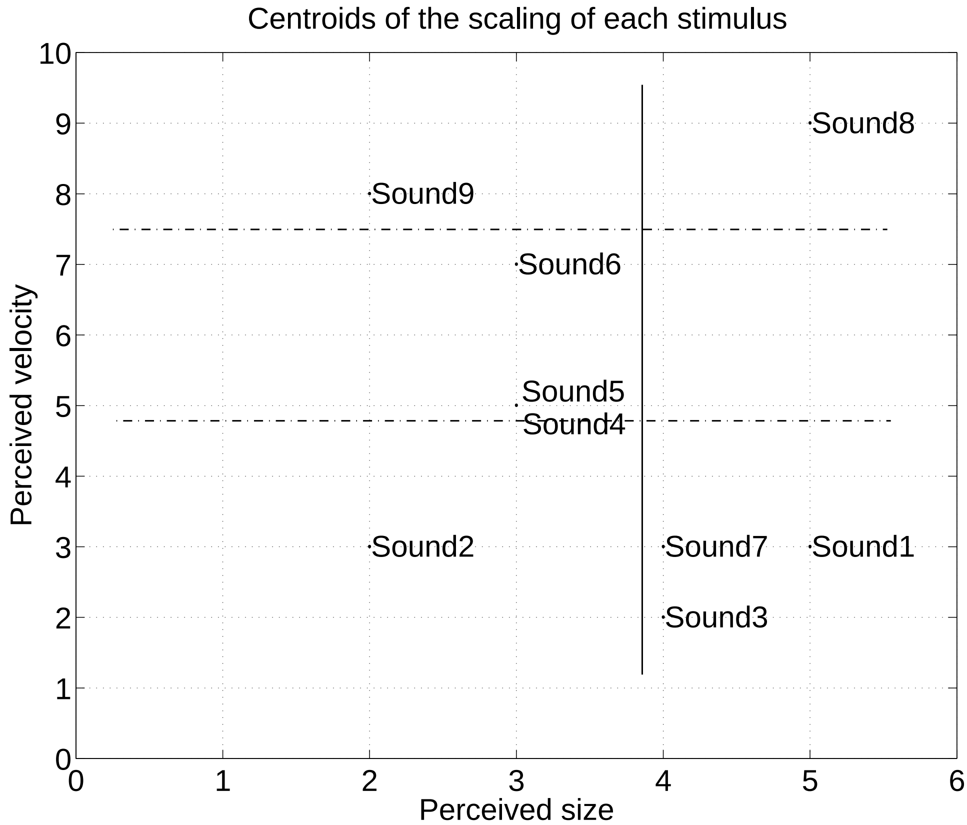


Figure 5.5: *Centroids of the scaling of each stimulus, where the different perceptual groups related to the size parameter setting are highlighted with a vertical solid line and the different perceptual groups related to the velocity parameter setting are highlighted with horizontal dash-dot lines.*

It is worth to look at the relationship between the perceptual scaling of the stimuli and their parameters setting. In Figure 5.6 and Figure 5.7 we plot, sorted by parameters, the centroids of the scaling of the perceived size and of the perceived velocity, respectively, versus the parameters values. Therefore, in the plots, we can compare the stimuli mean positions within the perceptual space and the stimuli positions within the parameters space. Stimuli parameters with the same value are highlighted with horizontal dashed lines.

We can notice a complex relationship between the perceptual space and the parameter space and it is quite hard to find some clear behaviour. The diameter parameter seems to influence quite strongly the perceived size, since there is an evident separation in the perceived size estimation according to the diameter values. The same happens for the velocity parameter considered with the perceived velocity. Moreover, the amplification parameter shows a slight relationship with the perceived velocity, since three stimuli, i.e. *sound7*, *sound5* and *sound6*, have increasing amplification values and increasing perceived velocity scaling. Anyway, the three sounds differ not only in the amplification value, but also in the diameter value, since *sound7* has an higher value than the other two.

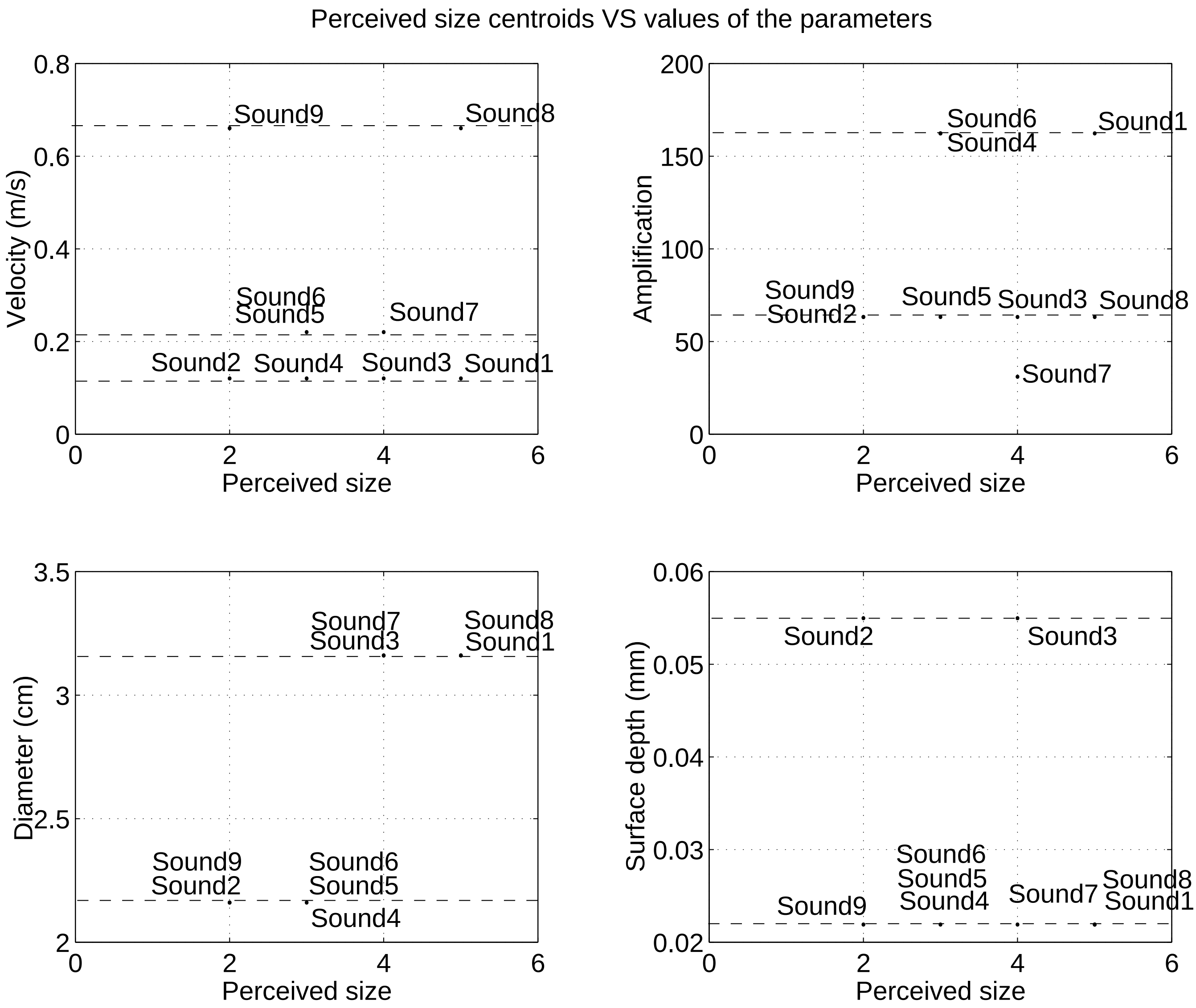


Figure 5.6: *Centroids of the scaling of the perceived size versus the parameters values, sorted by parameters. Stimuli parameters with the same value are highlighted with horizontal dashed lines.*

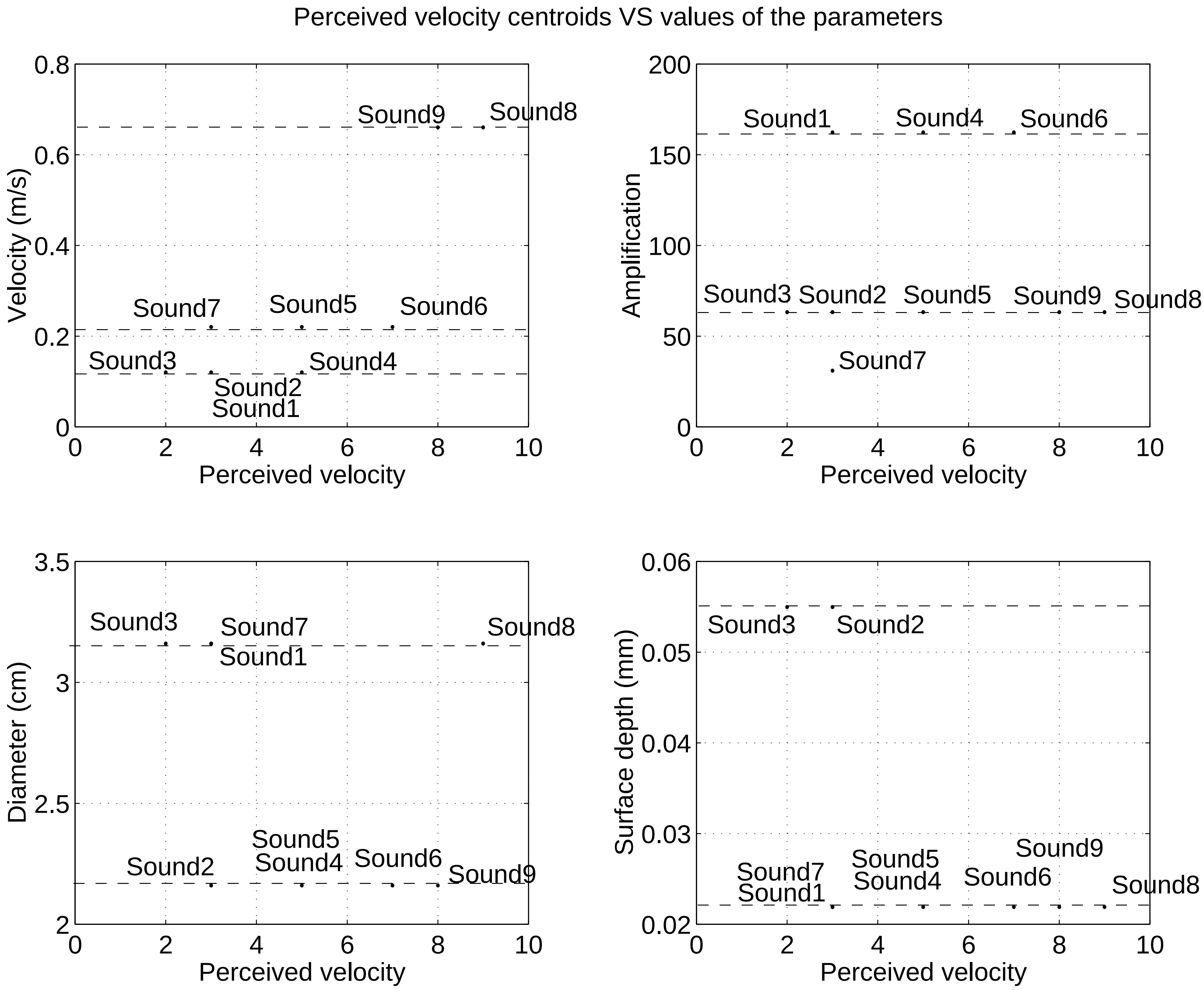


Figure 5.7: *Centroids of the scaling of the perceived height versus the parameters values, sorted by parameters. Stimuli parameters with the same value are highlighted with horizontal dashed lines.*

The debriefing phase gave the results reported in Figure 5.8 with cumulative participant responses represented through a bar chart, with -2.5 representing a negative result to the question and 2.5 a positive one, while 0 represents the average result.

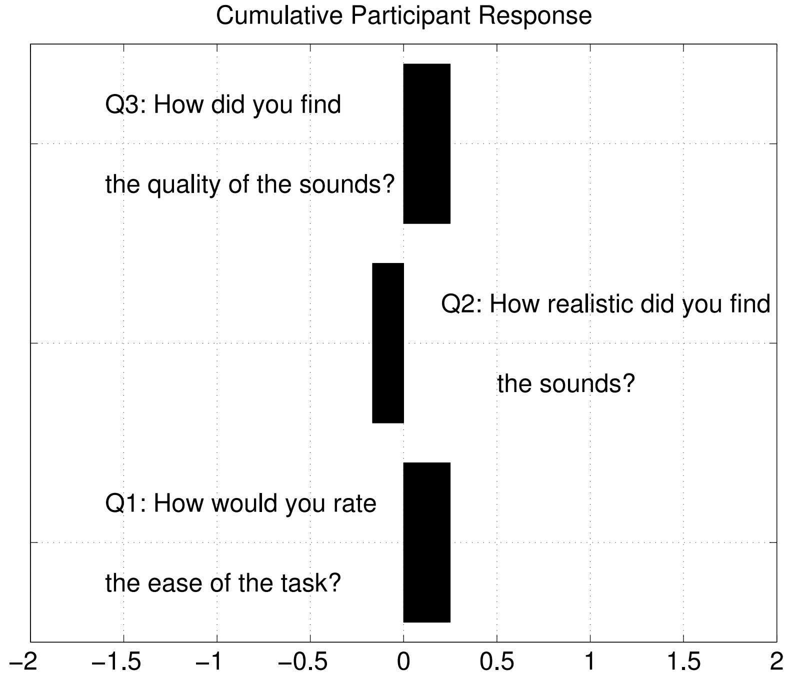


Figure 5.8: *Results of the questionnaire filled out during the debriefing phase: Cumulative Participant Response.*

We can see that the task was judged to be quite easy to be performed, and the sounds, even if they are considered not too realistic because of their cartoonification, are judged to have an acceptable quality. This experiment on the rolling process achieved similar results as those dealing with the impact/bounce event, since in both cases the cartoonification of the sounds does not affect the sounds capacity to provide information to users.

5.2 Conclusions

In this chapter we have investigated another sound object which is common in everyday life: the rolling process. We conducted the experiment with the sound objects we synthesized. Even if the rolling sounds are cartoonified, they are able to convey quantified information.

We investigated the bi-dimensional space created by perceptual size and perceptual velocity. The experimental results showed that perceptual size pops-out more than perceptual velocity, because perceptual velocity is influenced by more dimensions rather than the size dimension. In any case, both the perceptual dimensions were quite uniformly estimated, even if the sounds were judged to be not much realistic, for their cartoonified nature.

In this way, it is possible to provide data information by using cartoonified rolling sounds inside an auditory display and by highlighting the two dimensions of perceptual velocity and perceptual size which can be identified and evaluated quite well by listeners. Moreover, it would be probably possible to exploit the information conveyed by the perceptual dimensions of the object material and the surface material, since subjects pointed out many times the materials they heard. This aspect will be the topic for further experiments. However, this experiment shows that, if the parameters are set with accuracy, it is possible to provide data information with rolling sound processes, even if cartoonified.

Chapter 6

Conclusions

Human beings are able to recognize, through their auditory system, besides the proper dimensions of the auditory domain, other perceptual dimensions, such as geometric properties of objects, sound sources configuration, particular characteristics of sound events and processes, etc.

In order to auditorilly represent multi-dimensional data, it seems interesting to study some perceptual features able to provide information to listeners and it could be better if these perceptual dimensions are ecologically relevant, because they could provide users with information through a “direct input channel”, without the need of using knowledge or memory processes for gaining information from the auditory display.

Within the field of auditory display, there are various approaches which tried to solve the problem of representing multi-dimensional data, but some of them aim at presenting a solution to a precise practical problem in a pre-defined strict context, while the others try to study the general principles of auditory display and to present the rules and guidelines for an effective auditory design. On the contrary, few of them focus on the real problem of representing multi-dimensional data through auditory display from a more general perspective, trying to cover a variety of situations and scenarios. For these reasons, we investigated ecologically some features of sound events that were little or never studied in the auditory display field and we tried to find features useful for creating a “natural auditory environment” for the users of the auditory representation.

The underlying assumption is that auditory representations based on everyday sounds translate into environments that are more engaging for the users. Moreover, by using everyday sounds we think that the auditory display could be more general and therefore applicable to a variety of situations. On the contrary, unnatural sounds tend to be more distracting and irritating and, if an application uses abstract sounds, the mapping design should be more related to that precise data set in order to have a metaphorical relationship with the data it represents. In any case, everyday sounds seem to be more easily separated from other sounds when presented to users in a complex auditory scene. Furthermore, it seems that everyday sounds comprise naturally more perceptual dimensions as compared to abstract sounds, which, on the contrary, have to be synthesized on purpose and sometimes many perceptual dimensions make them more unnatural and difficult to be estimated.

From these starting points, our dissertation aimed at finding some interesting perceptual dimensions of everyday sounds able to ecologically convey information. We decided to use synthetic models of the sound sources and of the sound events/processes we wanted to investigate. An experimenter, by using physical models for generating the stimuli used in the experimental framework, is able to control directly each physical attribute [LCS00]. Since all these investigations were conducted for being, in the future, applied to auditory displays or multimodal applications which could enjoy their benefits, it could be more flexible to include sound models inside applications rather than pre-recorded sounds and, in particular, by using the sound models developed by the the EU-funded project "the Sounding Object" (SOb)[1], it could be computationally convenient. The aim of our dissertation was to investigate the perceptual features of sound sources and of auditory events/processes synthesized with these models able to convey information, and in particular quantifiable information, in order to auditorilly represent multi-dimensional data. Sound sources and auditory events have been recently considered effective in designing auditory display [HR99, HHR01].

Our main interest has been to cover the perceptual aspects of the topic, rather than the practical rules for the auditory design, widely studied, for instance, in [Kra94a, Kra94b, Bar97], etc., although our investigations on the perceptual dimensions of sound sources and auditory events/processes could define the basis for a new design of auditory displays for multi-dimensional data.

As for the perceptual dimensions of sound sources, we found that some object dimensions, which can be physically seen, can be "heard" as well, at least in certain conditions. We worked on the auditory perception of the shape of cubic and spherical resonators and on the relationship between pitch perception and size of the enclosure, by means of simulated 3-D resonators. Cubic and spherical resonators of the same volume are identified to have equal pitch under certain circumstances and, for certain dimensions, listeners are able to classify the enclosure's shape. In particular, during these studies, we identified an effect, the order effect, which affects the listening attitude and, therefore, can cause performance gaps between users.

Kubovy and Van Valkenburg, within their theory of indispensable attributes, define an attribute to be indispensable if and only if it is a prerequisite of perceptual numerosity [KV01]. They state that there are different indispensable attributes for the visual and the hearing channel. For the visual channel, space is an indispensable attribute. For the auditory channel, this role is played by frequency, that becomes an indispensable attribute for hearing numerosity. Time is an indispensable attribute for both vision and audition.

How can the size of the resonators, that we saw to be related to pitch, be considered within the theory of indispensable attributes? Is it a spatial attribute or a frequential attribute? If size would be a frequential attribute, it could be distributed over frequency and perceptual numerosity would be heard. On the contrary, if it would be a spatial attribute,

[1] `http://www.soundobject.org`

it could be distributed over space and perceptual numerosity would be seen. We think that from the visual point of view, the size of a resonator is a spatial attribute, but, from the auditory perspective, the size of a resonator is a frequential attribute. In fact, if the impulse responses of two resonators are played on one or two loudspeakers, they can be heard as two sounds only if they have different size. The case is similar to the example for pitch separation presented in [KV01], where a listener hears two sounds only if they are separated in frequency, regardless of whether they are played over one or two loudspeakers. As Valkenburg and Kubovy underline, "the theory of indispensable attributes does not claim that frequency separation is a sufficient condition for perceptual numerosity, but that frequency separation makes perceptual numerosity *possible* [VK03]". In our case, we have seen that size can be estimated under strict circumstances. Although it is a frequential attribute for the auditory channel, it is not easily used in auditory display design, because the perceptual numerosity is ensured only under strict requirements.

Before focusing on sound events, we worked on distance perception in the context of the validation of the distance rendering of a virtual resonating environment. We concluded that the synthetic resonator under analysis can efficiently render the distance cue and, therefore, it could be used in various auditory display applications.

As far as auditory perception of sound events/processes is concerned, we focused on the auditory cues of two particular sound events: the impact/bounce event and the rolling process. We found that, although some sounds are not easily scaled in the perceptual space by the listeners, they are estimated uniformly by most listeners, if they are synthesized accurately. Therefore, they can provide information.

The experiments conducted on sorting sound events/processes highlighted the ability of the human auditory system in estimating not only natural or realistic auditory events but cartoonified auditory events as well. In particular, we are sure that auditory displays with an accurate setting of the parameter values can provide multi-dimensional information, even to a non-trained user. The experiments were conducted without training, because we wanted to check the unbiased users performance. The subjects were allowed just to browse through the sounds for a while, in order to be able to start the organization of their own perceptual space. Even if we didn't test it appropriately, we think that, with a little training, it would be possible to render quantifiable data quite precisely, since we obtained quite good results with non-trained participants. We noted that most of the participants, before performing the task, preferred to group the sounds, that were randomly spread in the 2-D plot, according to certain mental models. Few subjects also preferred to start judging one dimension for all the sounds and then the other dimension and, only later, to estimate the dimensions together. With a little training, the "arrangement" step could be skipped and maybe even the few subjects who started the estimation of the two dimensions separately, could have scaled them together from the beginning.

It is important to note the better results obtained for sound events rather than for sound sources. In fact, the dimensions studied in the first part of our research activity were maybe difficult to be identified, recognized and evaluated, even if they are important features for a resonator. In particular, the shape of cavities is fundamental, for example, if we think to the vocal tract: A common listener is able to identify the utterances from its shape. As we mentioned in subsection 1.2.1, a speech sound can convey information about the changing geometry of the vocal tract. Even if both the vocal tract and the sound sources we used are resonators, in the case of the vocal tract, it is a monodimensional cavity, i.e. a tube of variable diameter, and the information are provided by the spectral envelope which is related to the shape of the enclosure, while the cavities we used are 3-D resonators, where the information are provided by the distribution of the resonances. The different performance of listeners in the case of the vocal tract and the case of 3-D resonators could be due to the human nature that is trained to extract information from monodimensional cavities, in order to be able to understand speech. From the experiments we conducted, we got good results, but under strict requirements and listening conditions. On the contrary, the auditory events/processes we worked on were immediate to be perceived — we could say that their features pop-out — and ecologically relevant.

Further experiments on the perception of sound events/processes will be interesting to be conducted. In particular, it would be interesting to focus on the relationship of the four dimensions of the impact/bounce event all together and on the perception of material in the rolling process both of the rolling object and the plane.

It is worth to mention that there are not similar studies in the literature covering perceptual experiments conducted with sound models with the aims we were pursuing. We know that there are other works regarding experiments on perceptual dimensions [MCR04], but they applied the technique of multidimensional scaling analysis. The stimuli set comprised sounds of impacted bars synthesized with a physical model. The subjects were asked to rate, according to any criteria that they considered perceptually salient, the perceived dissimilarity between all the possible pairs of stimuli. By applying a certain procedure [MCR04] to evaluate the appropriate dimensionality of the perceptual space, the authors could compare the perceptual and the physical space and analyze the results. In the case of the impact/bounce event, we decided to use the approach with the Sonic Browser in order to decrease the user fatigue and to make the experiment more involving by allowing the subjects to listen to the stimuli they wanted, how many times they found it necessary, and to create their own perceptual space. We already pointed out (chapter 4) the usefulness of the aura. On the other hand, for the rolling process, we used another procedure, performed with a Matlab interface, because of the nature of the stimuli. In any case, the stimuli set we used in both types of perceptual experiments were synthesized with sound models that are the combination of physics-based models and classical techniques of sound synthesis [RAB+03]. Even if our sound models are simpler than the physical models used in [MCR04], we showed that they are able to provide information and cartoonification doesn't reduce this capacity. On the contrary, the physical models of [MCR04] are quite complicate with many variable parameters, that, within the experimental framework, are reduced to two, while the remaining parameters are kept constant. It has been hard for us to find useful methods for investigating the desired perceptual features, both from the data collecting and from the data analysis point of view. We hope that in the future, with the experimental approach we used, there will be an established method for comparing physical parameters values of sound models and the scaling of perceptual parameters of the sound events they synthesize. In this dissertation, we have presented some experimental methods for analyzing the collected data, trying to find, in regard to perceptual scaling experiments on sound events and processes, useful representations and summary plots of the complex data gathered with such investigations.

The use of sound models was found to be useful for two main reasons. First, we were able to control directly the parameters for synthesizing the sounds that were used in the experiments. In this way, during our investigations, we were able to study the relationship between physical and perceptual dimensions and to generate stimuli that are difficult to be recorded, such as sound sources filtered by resonators, whose shape and size we could control. Second, since our research activity is oriented to representing multi-dimensional data within auditory displays with the goal of being exploited for the design of auditory display applications, we were able to investigate stimuli synthesized by sound models that could render the design both engaging and computationally convenient.

We already underlined in chapter 1 that we like to think of a multimodal display which could combine the visual with the auditory data presentation in order to get the advantages of both modalities. An example of the use of the sound events in a multimodal application is reported in appendix C. It doesn't want to show the benefits that an auditory representation of multi-dimensional data could have from the sounds events we synthesized, because this wasn't the purpose of this application. It wants just to show that our simulated sound events can cooperate with a graphical interface, that they can enhance some aspects of the scenario which is presented to the users, that they can engage the users and that they can attract the attention on certain features, which, if included in a proper auditory display application, could pop-out and be used for providing the desired information.

The novelties introduced by our work can be summarized by the following points:

- some perceptual experiments have been conducted by means of simulated 3-D resonators with cubic and spherical shape in order to study whether some object dimensions, that are commonly "seen", can auditorilly provide information. The investigated features, i.e. the size of the resonators and their shape, resulted to be usable only under strict requirements;

- an interesting perceptual effect, that is the order effect, has been found. It affects the listening attitude of the listeners and it can cause performance gaps between users;

- some experimental methods for analyzing the collected data of perceptual scaling experiments have been presented and studied. They focused on finding useful representations and summary plots of the complex data gathered with the experiments for comparing physical parameters of sound models, and scaling the perceptual parameters of the sound events they synthesize;

- some investigations about some perceptual dimensions of two sound event categories that are common in the everyday auditory environment, i.e. the impact/bounce event and the rolling process, have been conducted. Even if there is a complex relationship between the perceptual space and the space of the parameters, the investigations highlighted the perceptual dimensions that "pop-out" more or, at least, they found how to get more uniform perceptual scaling by means of a proper parameterization.

We hope that the investigations we have conducted both on the sound sources and on the auditory events and processes could provide interesting suggestions for future auditory display, characterized by multimodality, multi-dimensionality and an ecological approach, in order to "surround" the users with data information in the most natural and direct way, without requiring particular knowledge, expertise or memory.

Appendix A

Method of constant stimuli and Differential Limen estimation

The method of constant stimuli [Pur97] is one of the classical methods in psychophysics which attempt to describe the relationship between the measurable/physical aspects of a stimulus and an observer's perception of that stimulus.

If the experiment aims at establishing the Absolute Limen (AL), i.e. the minimum threshold for a stimulus condition to cause a certain perception, it consists in the repetitive presentation of a constant set of variable stimuli, one after each other. If the experiment's goal is to estimate the Difference Limen (DL), i.e. the minimum difference in two stimuli conditions, the stimuli are presented in pairs, where one is the standard stimulus.

The main characteristics of this method are: the randomized presentation of the stimuli set; the use of a reduced number of stimuli, approximately at equal distance one from the other on the physical continuum; the use of the appropriate stimuli set, in order to comprise both the extremes in the responses.

In our experiment, we are interested in estimating the Difference Limen (DL), using only two categories of responses ("higher" or "lower"). In general, the DL is half of the interval between the values of the variable stimuli judged positively 25% of the time and those judged positively 75%. Another important measure is the Point of Subjective Equivalence (PSE) that is the value of the variable stimulus judged equal to the standard stimulus 50% of the time. The third important measure in the classical psychophysical methods is the Constant Error (CE), that is the difference between the PSE and the objective point of equality and it is determined by finding the algebraic difference between the PSE and the standard stimulus value.

We calculated these measures by applying the method of interpolation with the least-squares regression line. The PSE is calculated with the following equation:

$$PSE = -\frac{a}{b} \tag{A.1}$$

where

$$b = \frac{n\left(\sum_{i=1}^{n} R_i z_i\right) - \left(\sum_{i=1}^{n} R_i\right)\left(\sum_{i=1}^{n} z_i\right)}{n\left(\sum_{i=1}^{n} R_i^2\right) - \left(\sum_{i=1}^{n} R_i\right)^2} \tag{A.2}$$

and

$$a = \bar{z} - b\bar{R} \tag{A.3}$$

in which n is the number of stimuli used, R_i are the values of the physical stimuli, and z_i are the z scores related to the probability p of positive responses. There are appropriate tables for calculating the z scores. We indicate with $\bar{z}$ and $\bar{R}$ respectively the mean of the z_i values and of the R_i values.

The CE and the DL are calculated with the following equations:

$$CE = PSE - St \tag{A.4}$$

$$DL = z_{75}\frac{1}{b} \tag{A.5}$$

where St is the value of the standard stimulus and z_{75} is the z score relative to a probability of 75%.

Appendix B

The Sonic Browser

The Sonic Browser is a software tool which was developed in the Interaction Design Centre at the University of Limerick in 1996 and which has been under improvement since then [BF01, BFTC02]. It allows the user to navigate a bi-dimensional multimedia space primarily through listening.

The Sonic Browser main purpose is managing huge catalogues of sound collections, a difficult problem in the sound designer community and among Foley artists. The development of sound models, whose parameters control the sound source characteristics, has lead to the use of the Sonic Browser for conducting psychophysical experiments, in order to test and validate the sounds produced by these sound models [OBF03]. The former case is the cataloguing scenario, while the latter case is the validation scenario. Within our research activity we used the Sonic Browser in the validation scenario.

In this appendix we briefly introduce the Sonic Browser, focusing on the validation scenario.

The Sonic Browser interface consists in a 2-D plot identified by two axes. It represents a 2-D space. Within this space there are sounds represented with coloured shapes.

While, in the validation scenario, the Sonic Browser interface consists only in the 2-D plot identified by two axes labeled according to the dimensions under investigation, in the cataloguing scenario, it includes also the control side, an area for controlling dynamically the application and for textually searching audio files within the database. It comprises:

- a slider for controlling dynamically the size of the aura;
- some visualization buttons for choosing which visualization to use for representing the audio data. There are currently four visualization mechanisms available, which are: the TreeMap visualization, the HyperTree visualization, the TouchGraph visualization and the SObGrid visualization, which can be seen in Figure B.1. In the validation scenario, the visualization used is the SObGrid. Since this appendix aims at just giving a general idea of the application we used for conducting our experiments, we suggest to refer to [BFO03] for further details on the characteristics of the visualizations and on the the Sonic Browser in general;

- the filter control mechanisms area, which offers the following functionalities: "and", "or", removal of the filter settings, saving of the filter settings and loading of the settings. A grey area displays the current filter settings;
- the status area that summarizes some status information about the application and the user.

In Figure B.1 we can see the Sonic Browser interface in the cataloguing scenario, i.e. the 2-D space with the SObGrid visualization on the right, and the control side on the left.

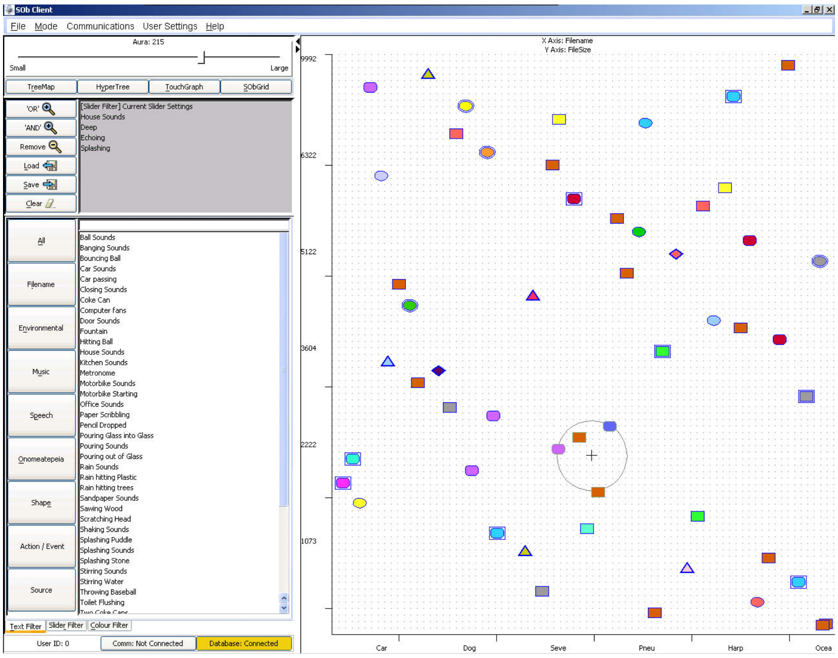


Figure B.1: *Interface of the Sonic Browser used for the cataloguing scenario. On the left hand, there is the control side for the dynamic control of the application. On the right hand, there is the sound space, with the SObGrid visualization, which can be browsed by the users. The sounds are represented with different shapes and colours. The circular area surrounding the cursor is the* aura*: The users listen simultaneously to all the sounds included in it.*

In the bi-dimensional plot of the Sonic Browser, the users are represented by the cursor, which is surrounded by a circular area called *the aura*. It can be seen in Figure B.1. The aura, which can be resized or turned off and on directly by the users, allows them to listen simultaneously to all the sounds it includes.

The users can listen to the sounds by moving the cursor on them and, if the aura is large enough, they can listen simultaneously to all the sounds it is surrounding. In the cataloguing scenario, the users just browse within the bi-dimensional space in order to find the sounds they are looking for, and they can tag the sounds they want to highlight. In the validation scenario they can move the stimuli in the perceptual space, identified by the axes labeled according to the perceptual dimensions. In this way, by dragging-and-dropping the shapes representing the stimuli, the participants to the scaling experiments can create their own perceptual space.

Another feature of the Sonic Browser has been used during the realism judgment phase in our experiment: the possibility of tagging some sounds, by changing the line style of the shapes which refer to those sounds.

Moreover, the Sonic Browser can easily collect the experimental data, both data logging for the object positioning in the 2-D perceptual space and the list of stimuli tagged by the participants to the experiment.

The Sonic Browser can be useful to our experiments for three main reasons. First, it allows the users to browse the sound space, providing fast and direct access to the sounds and making the task more natural to them. Second, it lets the users move the sounds according to a bi-dimensional evaluation scale and, therefore, create their own perceptual space. Finally, there is the main feature of the Sonic Browser, that is the aura. It represents the heart of the Sonic Browser and differentiates it from other tools [SLH02]. Despite not being so important in the validation scenario as it is for the cataloguing scenario, it has demonstrated to be very useful to users, who applied it for comparing stimuli and previous estimations, especially with very short stimuli, as those we used for the experiments on the impact/bounce event (see chapter 4).

Appendix C

The Virtual Touch Machine

In this appendix we report an interesting example of the use of the sound events in a multimodal application.

This application, called the Virtual Touch Machine, is an "hybrid artefact" of the Living Exhibition "Re-Tracing the Past", which took place at the Hunt Museum in Limerick (Ireland) from the 9th to 19th June 2003, developed by the EU-funded project SHAPE[1] and with the collaboration of the EU-funded project "the Sounding Object" (SOb)[2].

> Hybrid artefacts exhibit physical and digital features and can exist in both physical and digital worlds. They combine interactive visual (computer graphical and video) and sonic (music, recordings and live sound) material with physically present manipulable devices [Sha].

The aim of the exhibition was to allow the visitors of the museum to actively interact in both physical and virtual spaces in order to learn interesting details about four mysterious objects of the Hunt Museum, by stimulating their creativity.

Within this big project, the SOb project, by means of a troubadour grant for the dissertation's author, collaborated in the setting of the Virtual Touch Machine (see Figure C.1). The users with this machine were able to explore the four mysterious objects closely and, in particular, their material. The users could interact with the 3-D models of the objects projected on a flat screen and rotate them by means of a "magic wand". The system was multimodal, since the wand could be used for tapping the object and hearing the sound according to the shape, the size and the material of the object selected.

In Figure C.1 (above) we can see the Virtual Touch Machine within the Living Exhibition, while in Figure C.1 (below) we can see just the virtual models of the four mysterious objects as they appeared on the flat screen.

The sound models had been developed with a modular structure. They were PD-patches running on the modules of the impact sound of the SOb catalog and they were driven by an higher level patch, aiming at identifying which object the users chose.

[1] Situating Hybrid Assemblies in Public Environments — `http://www.shape-dc.org`

[2] `http://www.soundobject.org`

This is a good example of interaction between visual and auditory display based on our models and our perceptual research. We have seen that the interaction of the two modalities can enhance the sense of presence of the users and it can provide information about the physical features of the objects, such as shape, size and material characteristics.

We don't mention the Virtual Touch Machine for showing the benefits that an auditory representation of multi-dimensional data could have from the sounds events we synthesized, because this wasn't the purpose of this application. We would like just to show that our simulated sound events can cooperate with a graphical interface, that they can enhance some aspects of the scenario which is presented to the users, they can involve the users and that they can attract the users attention on certain features, which, if included in a proper auditory display application, could pop-out and be used for providing the desired information.

Figure C.1: *The Virtual Touch Machine in the Living Exhibition (above). The virtual models of the objects as they appeared on the flat screen (below).*

Bibliography

[AKZ02] D. Arfib, F. Keiler, and U. Zölzer. Source-filter processing. In U. Zölzer, editor, *DAFX: Digital Audio Effects*, pages 299–372. John Wiley and Sons, Inc., New York, 2002.

[Ale02] P.M. Alexander. *Towards reconstructing meaning when text is communicated electronically*. PhD thesis, University of Pretoria, South Africa, 2002. Available at `http://upetd.up.ac.za/thesis/available/etd-08192002-155431/`.

[ANS] American National Standards Institute ANSI. American national psychoacoustical terminology. S3.20, New York: American Standards Association.

[AR01] F. Avanzini and D. Rocchesso. Controlling material properties in physical models of sounding objects. In *Proceedings of the International Computer Music Conference 2001*, pages 91–94, La Habana, Cuba, 2001.

[Bar97] S. Barrass. *Auditory Information Design*. PhD thesis, Australian National University, 1997.

[Beg99] D.R. Begault. Virtual acoustic displays for teleconferencing: Intelligibility advantage for "telephone grade" audio. *J. Audio Eng. Soc.*, 47(10):824–828, October 1999.

[Ber92] L.L. Beranek. Concert hall acoustics - 1992. *J. Acoust. Soc. Am.*, 92(1):1–39, 1992.

[BF01] E. Brazil and M. Fernström. Let your ears do the browsing - the Sonic Browser. *The Irish Scientist*, 2001.

[BFO03] E. Brazil, M. Fernström, and L. Ottaviani. Psychoacoustic validation and cataloguing of sonic objects: 2d browsing. In D. Rocchesso and F. Fontana, editors, *The Sounding Object*, pages 297–340. Mondo Estremo, 2003. Freely distributed under the GNU Free Documentation License. Available at `http://www.soundobject.org/SObBook/`.

[BFTC02] E. Brazil, M. Fernström, G. Tzanetakis, and P. Cook. Enhancing sonic browsing using audio information retrieval. In *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan, 2002.

[BGBG95] R.M. Baecker, J. Grudin, W. Buxton, and S. Greenberg. A historical and intellectual perspective. In *Readings in Human-Computer Interaction: Toward the Year 2000, Second Edition*, chapter 1. Morgan Kaufmann, San Fracisco, CA, 1995.

[BR00] M.T. Boren and J. Ramey. Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, 43(3):261–278, September 2000.

[Bre90] A. Bregman. *Auditory Scene Analysis.* MIT press, 1990.

[Bro00] G. Brown. Auditory scene analysis: Listening to several things at once. Available at `http://www.dcs.shef.ac.uk/research/groups/spandh/info/asa.html`, 2000.

[BSB02] D. Brock, J.L. Stroup, and J.A. Ballas. Using an auditory display to manage attention in dual task, multiscreen environment. In *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan, 2002. Available at `http://www.icad.org/websiteV2.0/Conferences/ICAD2002/proceedings/`.

[BWSM96] D.R. Begault, E.M. Wenzel, R. Shrum, and J. Miller. A virtual audio guidance and alert system for commercial aircraft operations. In *Proc. International Conference on Auditory Display - ICAD 1996*, pages 117–122, Palo Alto, CA, November 4-6 1996. Available at `http://human-factors.arc.nasa.gov/ihh/spatial/personnel/wenzel.html`.

[Cab99] D. Cabrera. The size of sound: Auditory volume reassessed. *MikroPolyphonie*, 5, 1999. Online journal. Available at `http://farben.latrobe.edu.au/mikropol/volume5/cabrera_d/Volume.html`.

[CAKP98] C. Carello, K. L. Anderson, and A. J. Kunkler-Peck. Perception of object length by sound. *Psychological Science*, 9(3):211–214, May 1998.

[Cas98] M.A. Casey. *Auditory Group Theory with Applications to Statistical Basis Methods for Structured Audio.* PhD thesis, Massachusetts Institute of Technology, 1998. Available at `http://xenia.media.mit.edu/~mkc/thesis/`.

[CBI98] G.A. Calvert, M.J. Brammer, and S.D. Iversen. Crossmodal identification. *Trends in Cognitive Sciences*, 2(7):247–253, July 1998. Available at `http://www.physiol.ox.ac.uk/Research_Groups/Multisensory/Papers/tics.pd\%f`.

[Chi01] L. Chittaro. Information visualization and its application to medicine. *Artificial Intelligence in Medicine*, 22(2):81–88, 2001.

[Col62] P.D. Coleman. Failure to localize the source distance of an unfamiliar sound. *J. Acoust. Soc. Am.*, 34:345–346, 1962.

[Djo00] P. Djoharian. Shape and material design in physical modeling sound synthesis. In *Proc. International Computer Music Conference 2000*, pages 38–45, Berlin, 2000.

[Dom01] F. Dombois. Using audification in planetary seismology. In *Proc. International Conference on Auditory Display - ICAD 2001*, pages 227–230, Espoo, Finland, July 29-August 1 2001.

[DS93] S.A. Van Duyne and J.O. Smith. Physical modeling with the 2-D digital waveguide mesh. In *Proc. Int. Computer Music Conf.*, pages 40–47, Tokyo, Japan, 1993.

[Ell98] D.P.W. Ellis. Modeling the auditory organization of speech - a summary and some comments. Draft of a chapter for 'Listening to speech: An auditory perspective", S. Greenberg and W. Ainsworth, eds., 1998.

[Eng63] D. Engelbart. A conceptual framework for the augmentation of man's intellect. In P. Howerton and D. Weeks, editors, *Vistas In Information Handling*, volume 1, pages 1–29. Spartan Books, Washington D.C., 1963.

[Eng71] T. Engen. Psychophysics. i. discrimination and detection. ii. scaling. In *Woodworth & Schlosberg's Experimental Psychology*, pages 11–86. J.K. Kling and L.A. Riggs (ed.), Methuen, London, 3 edition, 1971.

[FK94] W.T. Fitch and G. Kramer. Sonifying the body electric: Superiority of an auditory over a visual display in a complex, multivariate system. In Gregory Kramer, editor, *Auditory Display: Sonification, Audification and Auditory Interfaces*, pages 307–325. Addison-Wesley Publishing Company, Reading, MA, U.S.A., 1994. SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII.

[Fon03] F. Fontana. *Physics-based models for the acoustic representation of space in virtual environments.* PhD thesis, University of Verona, Italy, 2003.

[Fre90] D. J. Freed. Auditory correlates of perceived mallet hardness for a set of recorded percussive events. *J. Acoust. Soc. Am.*, 87(1):311–322, January 1990.

[FRO02] F. Fontana, D. Rocchesso, and L. Ottaviani. A structural approach to distance rendering in personal auditory displays. In *International Conference on Multimodal Interfaces, ICMI'02*, Pittsburgh, USA, 2002. Available at `http://profs.sci.univr.it/~ottavian/publications.html`.

[FWGK01] J.H. Flowers, L.E. Whitwer, D.C. Grafel, and C.A. Kotan. Sonification of daily weather records: issues of perception, attention and memory in design choices. In *Proc. International Conference on Auditory Display - ICAD 2001*, pages 222–226, Espoo, Finland, July 29-August 1 2001.

[Gar68] M.B. Gardner. Proximity image effect in sound localization. *J. Acoust. Soc. Am.*, 43:163, 1968.

[Gar98] W.G. Gardner. Reverberation algorithms. In M. Kahrs and K. Brandenburg, editors, *Applications of Digital Signal Processing to Audio and Acoustics*, pages 85–131. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[Gav88] W. W. Gaver. *Everyday listening and auditory icons*. PhD thesis, University of California, San Diego, 1988.

[Gav93a] W. W. Gaver. How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology*, 5(4):285–313, 1993.

[Gav93b] W. W. Gaver. What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, 5(1):1–29, 1993.

[GB03] M. Grassi and R. Burro. Impact sounds. In D. Rocchesso and F. Fontana, editors, *The Sounding Object*, pages 47–72. Mondo Estremo, 2003. Freely distributed under the GNU Free Documentation License. Available at `http://www.soundobject.org/SObBook/`.

[Gib86] J.J. Gibson. *The ecological approach to visual perception*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1986.

[Gol02] G. Goldin. What you see and what you get. *LAWEEKLY*, november 2002. Available at `http://www.laweekly.com/ink/02/52/a.php`.

[Gre00] C.D. Green. Introduction to: "perception: An introduction to the gestalt-theorie" by kurt koffka (1922). Available at `http://psychclassics.yorku.ca/Koffka/Perception/intro.htm`, 2000.

[HBE95] C.G. Healey, K.S Booth, and J.T. Enns. Visualizing real-time multivariate data using preattentive processing. *ACM Transactions on Modeling and Computer Simulation*, 5(3):190–221, 1995. Available at `http://www.csc.ncsu.edu/faculty/healey/abstract/pubs.html\#tomacs.95`.

[Hea96] C.G. Healey. *Effective visualization of large multidimensional datasets*. PhD thesis, University of British Columbia, 1996. Available at `http://www.csc.ncsu.edu/faculty/healey/abstract/pubs.html\#phd.96`.

[HHR01] T. Hermann, M.H. Hansen, and H. Ritter. Sonification of markov chain monte carlo simulations. In *Proc. International Conference on Auditory Display - ICAD 2001*, pages 208–216, Espoo, Finland, July 29-August 1 2001.

[HKH01] M.M.J. Houben, A. Kohlrausch, and D.J. Hermes. Auditory cues determining the perception of size and speed of rolling balls. In *2001 International Conference on Auditory Display*, pages 105–110, Espoo, Finland, 2001.

[HR99] T. Hermann and H. Ritter. Listen to your data: Model-based sonification for data analysis. In M.R. Syed, editor, *Advances in intelligent computing and multimedia systems*. Int. Inst. for Advanced Studies in System Research and Cybernetics, 1999.

[HS02] M.M.J. Houben and C.N.J. Stoelinga. Some temporal aspects of rolling sounds. In *Journées Design Sonore*, Paris, France, March 2002. Société Française d'Acoustique. Available at `http://www.confs.loa.espci.fr/ds2002/`.

[Jot99] J.-M. Jot. Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces. *Multimedia systems*, 7:55–69, 1999.

[Kel99] D. Keller. touch'n'go: Ecological models in composition. Master of fine arts, Simon Fraser University, Burnaby, BC, 1999. Available at `http://www.sfu.ca/sonic-studio/srs/EcoModelsComposition/Title.html`.

[KPK00] R. L. Klatzky, D. K. Pai, and E. P. Krotkov. Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environment*, 9(4):399–410, August 2000.

[KPT00] A. J. Kunkler-Peck and M. T. Turvey. Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1):279–294, 2000.

[Kra94a] G. Kramer. An introduction to auditory display. In Gregory Kramer, editor, *Auditory Display: Sonification, Audification and Auditory Interfaces*, pages 1–77. Addison-Wesley Publishing Company, Reading, MA, U.S.A., 1994. SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII.

[Kra94b] G. Kramer. Some organizing principles for representing data with sound. In Gregory Kramer, editor, *Auditory Display: Sonification, Audification and Auditory Interfaces*, pages 185–221. Addison-Wesley Publishing Company, Reading, MA, U.S.A., 1994. SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII.

[KV01] M. Kubovy and D. Van Valkenburg. Auditory and visual objects. *Cognition*, 80:97–126, 2001.

[LCS00] S. Lakatos, P.C. Cook, and G.P. Scavone. Selective attention to the parameters of a physically informed sonic model. *J. Acoust. Soc. Am.*, 107(5), 2000.

[Lev70] H. Levitt. Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.*, 49(2):467–477, 1970.

[LKPG98] J.M. Loomis, R.L. Klatzky, J.W. Philbeck, and R.G. Golledge. Assessing auditory distance perception using perceptually directed action. *Perception & Psychophysics*, 60(6):966–980, 1998.

[LLP91] X. Li, R. J. Logan, and R. E. Pastore. Perception of acoustic source characteristics: Walking sounds. *J. Acoust. Soc. Am.*, 90(6):3036–3049, December 1991.

[LMC97] S. Lakatos, S. McAdams, and R. Caussé. The representation of auditory source characteristics: simple geometric form. *Perception & Psychophysics*, 59(8):1180–1190, 1997.

[Lut89] R.A. Lutfi. Informational processing of complex sound. i: Intensity discrimination. *J. Acoust. Soc. Am.*, 86(3):934–944, 1989.

[Lut90] R.A. Lutfi. Informational processing of complex sound. ii: Cross-dimensional analysis. *J. Acoust. Soc. Am.*, 87(5):2141–2148, 1990.

[Lut92] R.A. Lutfi. Informational processing of complex sound. iii: Interference. *J. Acoust. Soc. Am.*, 91(6):3391–3401, 1992.

[Map] Minard's 1861 map of napolean's army's russian campaign. Web site: `http://it.coe.uga.edu/studio/seminars/visualization/minard.html`.

[MCR04] S. McAdams, A. Chaigne, and V. Roussarie. The psychomechanics of simulated sound sources: Material properties of impacted bars. *J. Acoust. Soc. Am.*, 115(3):1306–1320, 2004.

[Met96] E. Metois. *Musical Sound Information - Musical gestures and embedding synthesis*. PhD thesis, Massachusetts Institute of Technology, 1996. Available at `http://xenia.media.mit.edu/~metois/Phd/home.htm`.

[MG95] A.D. Milner and M.A. Goodale. *The visual brain in action*. Oxford University Press, Oxford, 1995.

[MI68] P.M. Morse and K.U. Ingard. *Theoretical Acoustics*. McGraw-Hill, New York, 1968.

[MK96] S.W. Mereu and R. Kazman. Audio enhanced 3d interfaces for visually impaired users. In *Proc. ACM SIGCHI '96*, Vancouver BC, Canada, April 13-18 1996. ACM Press, NY. Available at `http://www.acm.org/sigchi/chi96/proceedings/papers/Mereu/rnk-txt.htm`.

[MKS99] E. Milios, B. Kapralos, and S. Stergiopoulos. Sonification of range information for 3-D space perception. In *Proc. International Conference on Auditory Display - ICAD 1999*, Berlin, Germany, March 20 1999. Available at `http://www.cs.yorku.ca/~billk/BERLIN99.PDF`.

[MMG86] M.R. Moldover, J.B. Mehl, and M. Greenspan. Gas-filled spherical resonators: Theory and experiment. *Journal of the Ac. Soc. of America*, 79(2):253–272, 1986.

[Mun00] T. Munzner. *Interactive visualization of large graphs and networks.* PhD thesis, Staford University, 2000. Available at `http://graphics.stanford.edu/papers/munzner_thesis/`.

[MW01] M. Mellody and G.H. Wakefield. A tutorial example of Stimulus Sample Discrimination in perceptual evaluation of synthesized sounds: discrimination between original and re-synthesized singing. In *Proc. International Conference on Auditory Display - ICAD 2001*, Espoo, Finland, July 29-August 1, 2001.

[Myn95] E.D. Mynatt. *Transforming Graphical Interfaces into Auditory Interfaces.* PhD thesis, Georgia Institute of Technology, 1995.

[Nav77] D. Navon. Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9:353–383, 1977.

[Nem96] K. Nemire. Evaluating visual and auditory enhancements to a virtual-object manipulation task. In *Proceedings of SPIE: Stereoscopic Displays and Virtual Reality Systems III*, pages 249–260, 1996.

[Nor93] D.A. Norman. *Things that make us smart.* Addison Wesley, Reading, MA, 1993.

[NW01] M. Naguib and R.H. Wiley. Estimating the distance to a source of sound: mechanisms and adaptations for long-range communication. *Animal Behaviour*, 62(5):825–837, 2001. Available at `http://www.uni-bielefeld.de/biologie/vhf/NG/Naguibandwiley.pdf`.

[OBF03] L. Ottaviani, E. Brazil, and M. Fernström. Psychoacoustic experiments for validating sound objects in a 2-D space using the Sonic Browser. In *Proc. of the XIV Colloquium on Musical Informatics (XIV CIM 2003)*, pages 90–94, Florence, Italy, 2003. Available at `http://profs.sci.univr.it/~ottavian/publications.html`.

[OFR02] L. Ottaviani, F. Fontana, and D. Rocchesso. Recognition of distance cues from a virtual spatialization model. In *Proc. of the 5th Int. Conference on Digital Audio Effects (DAFx-02)*, pages 187–190, Hamburg, Germany, 2002. Available at `http://profs.sci.univr.it/~ottavian/publications.html`.

[OS89] A.V. Oppenheim and R.W. Scafer. *Discrete-Time Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, 1989.

[PMS80] D.R. Perrott, A. Musicant, and B. Schwethelm. The expanding image effect: The concept of tonal volume revisited. *J. of Auditory Research*, 20:43–55, 1980.

[Pur97] F. Purghè. *Metodi di Psicofisica e di Scaling Unidimensionale*. Bollati Boringhieri, 1997.

[RAB+03] M. Rath, F. Avanzini, N. Bernardini, G. Borin, F. Fontana, L. Ottaviani, and D. Rocchesso. An introductory catalog of computer-synthesized contact sounds, in real-time. In *Proc. of the XIV Colloquium on Musical Informatics (XIV CIM 2003)*, pages 103–108, Florence, Italy, 2003. Available at `http://www.soundobject.org/articles.html`.

[RD01] D. Rocchesso and P. Dutilleux. Generalization of a 3D resonator model for the simulation of spherical enclosures. *Applied Signal Processing*, 1:15–26, 2001.

[Rep87] B. H. Repp. The sound of two hands clapping: an exploratory study. *J. Acoust. Soc. Am.*, 81(4):1100–1109, April 1987.

[RF03] M. Rath and F. Fontana. High-level models: bouncing, breaking, rolling, crumpling, pouring. In D. Rocchesso and F. Fontana, editors, *The Sounding Object*, pages 173–204. Mondo Estremo, 2003. Freely distributed under the GNU Free Documentation License. Available at `http://www.soundobject.org/SObBook/`.

[RO01] D. Rocchesso and L. Ottaviani. Can one hear the volume of a shape? In *Proc. IEEE Workshop Appl. Sig. Proc. Audio and Ac.*, pages 115–118, Mohonk, NY, 2001. Available at `http://profs.sci.univr.it/~ottavian/publications.html`.

[Roc01] D. Rocchesso. Acoustic cues for 3-D shape information. In *Proc. International Conference on Auditory Display - ICAD 2001*, pages 175–180, Espoo, Finland, July 29-August 1 2001.

[Sch00] E.D. Scheirer. *Music-Listening Systems*. PhD thesis, Massachusetts Institute of Technology, June 2000.

[Sem01] S.K. Semwal. MoVE: Mobiltiy training in haptic virtual environment. Available at `http://www.cs.uccs.edu/~semwal/NSF2001PS.pdf`, 2001.

[Sha] EU-funded project Situating Hybrid Assemblies in Public Environments (SHAPE). Web site: `http://www.shape-dc.org`.

[SKS00] L. Shams, Y. Kamitani, and S. Shimojo. What you see is what you hear. *Nature*, 408:788, 2000.

[SLH02] G.P. Scavone, S. Lakatos, and C.R. Harbke. The Sonic Mapper: an interactive program for obtaining similarity ratings with auditory stimuli. In *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan, 2002.

[SMK98] K.M. Stanney, R.R. Mourant, and R.S. Kennedy. Human factors issues in virtual environments: a review of the literature. *Presence*, 7(4):327–351, August 1998.

[SML98] W. Schroeder, K. Martin, and B. Lorensen. *The Visualization Toolkit*. Prentice Hall, Upper Saddle River, NJ, 1998.

[Spe01] R. Spence. *Information Visualization*. ACM Press, Addison-Wesley, Essex, 2001.

[SRB87] R.D. Sorkin, D.E. Robinson, and B.G Berg. A detection-theoretic method for the analysis of visual and auditory displays. In *31st Annual Meeting of the Human Factors Society*, volume 2, pages 1184–1188, 1987.

[SS01] S. Shimojo and L. Shams. Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology*, 11:505–509, 2001. Available at `http://neuro.caltech.edu/cns286/shams_shimojo01.pdf`.

[TM04] M. Tory and T. Möller. Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics*, 10(1):72–84, January/February 2004.

[TSHK87] J. Tzelgov, R. Srebro, A. Henik, and A. Kushelevsky. Radiation detection by ear and by eye. *Human Factors*, 29(1):87–95, 1987.

[Tuf83] E.R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983.

[vdDKP01] K. van den Doel, P. G. Kry, and D. K. Pai. FoleyAutomatic: Physically-based sound effects for interactive simulation and animation. In *ACM SIGGRAPH 2001*, August 2001.

[Vic60] G.B. Vicario. L'effetto tunnel acustico [on auditory tunnel effect]. *Rivista di Psicologia*, 54:41–52, 1960.

[VK03] D. Van Valkenburg and M. Kubovy. In defense of the theory of indispensable attribute. *Cognition*, 87:225–233, 2003.

[War70] R. W. Warren. Perceptual restoration of missing speech sounds. *Science*, 167:392–393, 1970.

[War00] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, San Fracisco, CA, 2000.

[Wen92] E.M. Wenzel. Localization in virtual acoustic displays. *Presence*, 1(1):80–107, 1992. Available at `http://human-factors.arc.nasa.gov/ihh/spatial/personnel/wenzel.html`.

[Wer12] M. Wertheimer. Experimentelle studien über das sehen bewegung. *Zeitschrift für Psychologie*, 61:247–250, 1912.

[WR88] R. Wildes and W. Richards. Recovering material properties from sound. In W. Richards, editor, *Natural Computation*, pages 356–363. MIT Press, Cambridge, MA, 1988.

[WV84] W. H. Warren and R. R. Verbrugge. Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5):704–712, 1984.

[Zah01] P. Zahorik. Estimating sound source distance with and without vision. *Optom. Vis. Sci.*, 78:270–275, 2001.

[Zah02a] P. Zahorik. Assessing auditory distance perception using virtual acoustics. *J. Acoust. Soc. Am.*, 111(4):1832–1846, 2002.

[Zah02b] P. Zahorik. Auditory display of sound source distance. In *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan, 2002.