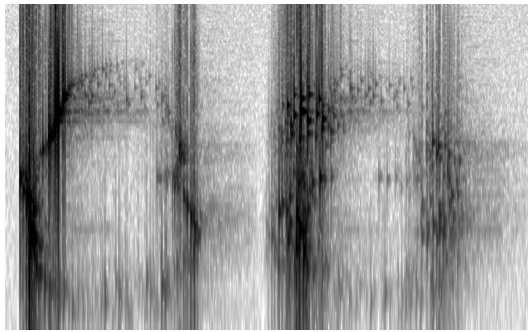


Listening-Mode-Centered Sonification Design for Data Exploration



Florian Grond

(genehmigte Fassung)

Der Technischen Fakultät der Universität Bielefeld
vorgelegt zur Erlangung des Titels

Dr. rer. nat.

Mai 2013

1. Reviewer: Dr. Thomas Hermann (thesis supervisor)

2. Reviewer: Dr. Gerold Baier (external reviewer, University College London)

Head of the examination board: Prof. Ipke Wachsmuth

Representative of lecturers: Dr. Frank Hegel

Day of the defense: December 5, 2013

Meinem Vater Kurt gewidmet

Acknowledgments

While working on this thesis, many friends and colleagues supported me through inspiring conversations and by sharing their thoughts, their ideas and their know how. I particularly like to thank my supervisor Thomas Hermann for the opportunity to conduct research with him in the Ambient Intelligence group, Till Bovermann for countless tips and support and collaboration with programming in SuperCollider particularly for the vowel class extension, Hans Diebner for continuing conversations and sharing ideas in the cross-section of arts and science, Trixi Drossard for the collaboration on the auditory graphs, Oliver Kramer for the collaboration on the monitoring application, Adriana Olmos for the opportunity to work on the audible sculpture project and Michael Ciarciello for his collaboration in this project, Hendrik Kösling and Nick Kasajanov for support during the eye-tracking evaluation of ancillary gesture sonification. Further, I like to thank Jean-François Denis and Robert Normandeau for literature recommendations and discussions.

I would like to thank the Ambient Intelligence Group at the Cognitive Interaction Technology - Center of Excellence (CITEC), Bielefeld university, for the scholarship which allowed me to work on this thesis, further the sonic interaction design (SID) COST action for a short term scientific exchange stipend, as well as the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) for the support through the audible sculpture project. During this thesis I could conduct research at McGill university in Montreal in 2008 and in 2010 for which I like to thank Prof. Marcelo Wanderley and Prof. Jeremy Cooperstock.

I would also like to acknowledge the work of all the individuals contributing to open source software used during this thesis, most of all SuperCollider and the always helpful and friendly SuperCollider mailing-list.

I would like to thank Alexis L. Emelianoff and Susanne Ackers for the proofreading of the manuscript. Last but not least I would like to thank Tamar Tembeck for her support and the proofreading of publications made in the context of this thesis.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 1.1 | Structural Organization of this Thesis | 6 |
| 2 | Listening, Interaction and Parameter-Mapping | 9 |
| 2.1 | From Sonification Definitions to Research Questions | 9 |
| 2.2 | Listening, a Mode of Perception with Intention | 13 |
| 2.2.1 | The Four Direct Listening Modes | 14 |
| 2.2.2 | Musical and Everyday Listening | 19 |
| 2.2.3 | Listening Modes and Monitoring | 24 |
| 2.2.4 | Listening and Sound Making | 25 |
| 2.2.5 | Recent Contributions to Listening Modes | 26 |
| 2.2.6 | Summary of Recent and Previous Contributions | 27 |
| 2.2.7 | Concluding Remarks | 29 |
| 2.3 | Parameter-Mapping, Connecting Data and Sound | 33 |
| 2.3.1 | The Fields of the Parameter-Mapping Diagram | 35 |
| 2.3.2 | Listening Intentions in the Parameter-Mapping Design Cycle | 38 |
| 2.4 | Parameter-Mapping in Vowel Synthesis | 40 |
| 2.4.1 | The Class <i>Vowel</i> , its Instance, and Control | 42 |
| 2.4.2 | The auxiliary Pseudo UGens | 44 |
| 2.4.3 | Ways to Use the Spectral Envelopes | 45 |
| 2.4.4 | Sample Sonification Applications | 47 |
| 3 | Mapping in the Sonification of Ancillary Gestures | 51 |
| 3.1 | Movement and Sonification | 51 |
| 3.1.1 | Ancillary Gestures | 52 |
| 3.1.2 | The Gestural Sonorous Object | 53 |
| 3.1.3 | Audio-Visual Displays | 54 |
| 3.1.4 | Research Questions | 56 |

CONTENTS

| | | |
|----------|--|------------|
| 3.2 | Annotating Sonified Ancillary Gestures | 57 |
| 3.2.1 | Motion Tracking Data | 57 |
| 3.2.2 | Data Preparation and Preprocessing | 58 |
| 3.2.3 | Sound Synthesis and Mapping | 58 |
| 3.2.4 | The Audio-Visual Display of Body Movements | 61 |
| 3.2.5 | Evaluation of the Display | 62 |
| 3.2.6 | Experimental Data Evaluation | 64 |
| 3.2.7 | Conclusion on the Annotation of Ancillary Gestures | 68 |
| 3.3 | Eye-tracking of Sonified Ancillary Gestures | 70 |
| 3.3.1 | Mapping and Sound Design | 71 |
| 3.3.2 | User Study | 74 |
| 3.3.3 | Evaluation of the Eye-tracking Data | 74 |
| 3.3.4 | Discussion | 81 |
| 3.4 | Summary and Outlook | 82 |
| 4 | Mapping and Interaction in Auditory Graphs | 87 |
| 4.1 | Literature Review on Auditory Graphs | 88 |
| 4.1.1 | Existing Applications | 90 |
| 4.1.2 | New Developments | 90 |
| 4.1.3 | Evaluating Auditory Graphs | 92 |
| 4.2 | Sonic Function | 93 |
| 4.2.1 | Mapping and Sound Design | 94 |
| 4.2.2 | User Study | 97 |
| 4.2.3 | Conclusions on <i>Sonic Function</i> | 102 |
| 4.3 | Singing Function | 104 |
| 4.3.1 | Mapping and Sound Design | 105 |
| 4.3.2 | Evaluation | 107 |
| 4.3.3 | Conclusions on <i>Singing Function</i> | 112 |
| 4.4 | Summary and Outlook | 113 |
| 5 | Mapping in Auditory Augmentations | 117 |
| 5.1 | Monitoring and Auditory Augmentation | 119 |
| 5.1.1 | Monitoring, Sonification and Evolutionary Optimization | 120 |
| 5.1.2 | Sound Design, Mapping, Audible Effects | 123 |
| 5.1.3 | Evaluation | 127 |
| 5.1.4 | Discussion | 129 |
| 5.2 | Audible Sculptures | 131 |
| 5.2.1 | 3D Data Preparation | 132 |

| | | |
|----------|--|------------|
| 5.2.2 | Sound Synthesis, Mapping, Participatory Design | 134 |
| 5.2.3 | Qualitative Evaluation | 141 |
| 5.2.4 | Discussion | 144 |
| 5.3 | Summary and Outlook | 145 |
| 6 | From Case studies to Guidelines | 149 |
| 6.1 | Listening Modes and Display Purposes | 150 |
| 6.2 | Repetition | 153 |
| 6.3 | Interaction | 156 |
| 6.4 | Complementary Modalities | 160 |
| 6.5 | Embodied Modalities | 163 |
| 6.6 | Parameter-Mapping and Sound Design | 165 |
| 6.7 | Concluding Remarks | 169 |
| 7 | Conclusion | 171 |
| | Glossary | 177 |
| | References | 179 |
| 8 | Appendix | 189 |
| 8.1 | Movement Annotation Plots | 189 |
| 8.2 | Gaze-density, Fixation Plots | 192 |

CONTENTS

1

Introduction

Through the ever growing amount of data and the desire to make them accessible to the user through the sense of listening, sonification, the representation of data by using sound has been subject of active research in the computer sciences and the field of HCI for the last 20 years. During this time, the field of sonification has diversified into different application areas: today, sound in auditory display informs the user about states and actions on the desktop and in mobile devices; sonification has been applied in monitoring applications, where sound can range from being informative to alarming; sonification has been used to give sensory feedback in order to close the action and perception loop; last but not least, sonifications have also been developed for exploratory data analysis, where sound is used to represent data with unknown structures for hypothesis building.

Coming from the computer sciences and HCI, the conceptualization of sonification has been mostly driven by application areas. On the other hand, the sonic arts who have always contributed to the community of auditory display have a genuine focus on sound. Despite this close interdisciplinary relation of communities of sound practitioners, a rich and sound- (or listening)-centered concept about sonification is still missing as a point of departure for a more application and task overarching approach towards design guidelines. Complementary to the useful organization along fields of applications, a conceptual framework that is proper to sound needs to abstract from applications and also to some degree from tasks, as both are not directly related to sound. I hence propose in this thesis to conceptualize sonifications along two poles where sound serves either a *normative* or a *descriptive* purpose.

In the beginning of auditory display research, a continuum between a symbolic and an analogic pole has been proposed by [Kramer \(1994a, page 21\)](#). In this continuum, *symbolic* stands for sounds that coincide with existing schemas and are more denotative, *analogic* stands for sounds that are informative through their connotative aspects

1. INTRODUCTION

(compare Worrall (2009, page 315)). The notions of *symbolic* and *analogic* illustrate the struggle to find apt descriptions of how the intention of the listener subjects audible phenomena to a process of meaning making and interpretation. Complementing the *analogic-symbolic* continuum with *descriptive* and *normative* purposes of displays is proposed in the light of the recently increased research interest in listening modes and intentions.

Similar to the terms *symbolic* and *analogic*, listening modes have been discussed in auditory display since the beginning usually in dichotomic terms which were either identified with the words *listening* and *hearing* or understood as *musical listening* and *everyday listening* as proposed by Gaver (1993a). More than 25 years earlier, four direct listening modes have been introduced by Schaeffer (1966) together with a 5th synthetic mode of *reduced listening* which leads to the well-known *sound object*. Interestingly, Schaeffer’s listening modes remained largely unnoticed by the auditory display community. Particularly the notion of *reduced listening* goes beyond the connotative and denotative poles of the continuum proposed by Kramer and justifies the new terms *descriptive* and *normative*. Recently, a new taxonomy of listening modes has been proposed by Tuuri and Erola (2012) that is motivated through an embodied cognition approach. The main contribution of their taxonomy is that it convincingly diversifies the connotative and denotative aspects of listening modes.

In the recently published sonification handbook, multimodal and interactive aspects in combination with sonification have been discussed as promising options to expand and advance the field by Hunt and Hermann (2011), who point out that there is a big need for a better theoretical foundation in order to systematically integrate these aspects. The main contribution of this thesis, as shown in Figure 1.1, is to address this need by providing alternative and complementary design guidelines with respect to existing approaches, all of which have been conceived before the recently increased research interest in listening modes. Gaver (1991) introduced the concept of affordances to the context of auditory display design, arguing that sound can provide the user with affordances that

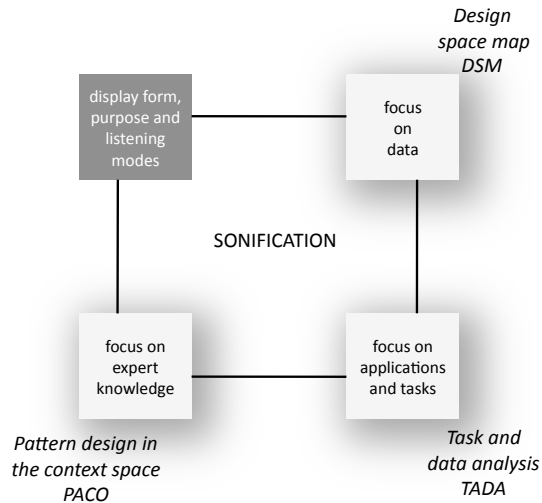


Figure 1.1: Design guideline contributions to the field of sonification. The field on the top left highlights the questions addressed in this thesis

complement visual user interfaces. Barrass (1997) introduced the *TADA* framework for auditory information design which is conceived to be *useful for the task and true to the data*. Frauenberger (2009) proposed a diversification of the task domain and developed *paco*, design patterns in context for auditory display design, which are mostly context and use case driven. *Paco* summarizes the expertise of the field based on a survey amongst practitioners and researchers. In both, the *TADA* framework and *paco*, auditory display design is usefully categorized and approached from the data, task and application domain. In the *TADA* framework, listening modes are mentioned but have back then not yet been diversified as they are today. With respect to *paco*, the diversified listening modes of today have the potential to provide a theoretical underpinning in addition to the expert knowledge from application domains. Frauenberger (2009, page 167) also identifies for future research the promising direction to apply *paco* on multimodal design, which as pointed out by Hunt and Hermann (2011) is in need of a theoretical foundation. The design space map (DSM) by deCampo (2007) connects sonification methods with data properties and puts an emphasis on the data structure by showing how it influences the form a sonification takes on. Although the DSM provides a tool for systematic reasoning by focusing on the data, it also makes indirect references to listening modes by referring to the form that the sonification takes on as the *sound object*. In the context of product sounds, contributions to the design process have been made by Jekosch (2005) and Hug (2009). The *TADA*, *paco* and *DSM* approach, and also the work by Hug, have partly touched aspects related to either multimodality or listening modes. None of the existing contributions to design frameworks integrates multimodality, and listening modes with a focus on exploratory data analysis, where sonification is conceived to support the understanding of complex data potentially helping to identify new structures therein. In order to structure this field the following questions are addressed in this thesis:

- How do natural listening modes and reduced listening relate to the proposed normative and descriptive display purposes?
- What is the relationship of multimodality and interaction with listening modes and display purposes?
- How can the potential of embodied cognition based listening modes be put to use for exploratory data sonification?
- How can listening modes and display purposes be connected to questions of aesthetics in the display?
- How do data complexity and PMSon relate to exploratory data analysis and listening modes?

1. INTRODUCTION

It can be argued that these questions are closely related to what can be summarized as the creative aspects of sonification. Admittedly, the answers to this questions in the form of guidelines cannot be as unambiguously structured as recipe-like instructions. Similar to the DSM, the guidelines will however help to engage in systematic reasoning, and provide a theoretical underpinning for what is currently expert knowledge. The guidelines will provide a genuine sound centered way to reflect on the role of sonification across applications and tasks and will apply accross sonification techniques in multimodal and interactive contexts. The guidelines are derived from theoretical considerations based on the phenomenology of listening and the experience from the practical sonification applications of this thesis.

With respect to the sonification method, I focus in the practical application prototypes in this thesis mostly on the popular technique of parameter-mapping sonifications (PMSon) which involves the mapping of data features to sound synthesis parameters. From Model-Based Sonifications (MBS) introduced by [Hermann and Ritter \(2005\)](#), I will also include data-sonograms, a technique which exhibits parameter-mapping components. In comparison with other sonification methods such as audification and MBS, where the sonic result is to a large degree determined through the method itself, PMSon leaves the designer of the sonification with various sonic decisions to be made. In order to arrive at a functional result, PMSon are usually created in an iterative process, which attempts to take the considerations of the data substrate, sound synthesis parameters as well as the perceived sonic result into account. Awareness about listening modes are a key ingredient in this iterative process and can be understood as part of what constitutes the sonification designer’s expertise. While for other sonification methods listening modes are relevant for the overall design of the auditory display system, for PMSon listening modes are directly tied to the design decisions related to the sound generating processes, which will be also discussed in the guidelines.

The investigation of sonification types with focus on listening modes needs to be based on concrete examples which are provided in this thesis through case studies in Sections 3, 4 and 5. As the theoretical Section 2.2 as well as Section 3.1 will motivate, the main aspects along which sonifications for exploring data can be organized are multimodality and interaction, data abstractness and data dimensionality. As we cannot sample this 4-dimensional space sufficiently dense, we will select and organize practical applications for subsequent analysis according to the selected main conceptual aspects of *multimodality* and *interaction* as depicted in Fig. 1.2. The practical applications are carefully selected to sample this space and give examples for sonification designs to be developed with a top-down listening mode-centered focus.

The bottom left corner shows an application for low interaction and a single modality as presented in Section 5.1 in the auditory augmentation-based monitoring application. On the bottom right corner, there are both gesture sonification applications discussed in Section 3.2 and Section 3.3 respectively. Both are conceived as fixed media audiovisual displays. They are hence high in multimodality but provide no means to interact with the data. On the left towards the top, there is an example of an auditory graph of mathematical functions as presented in Section 4.2

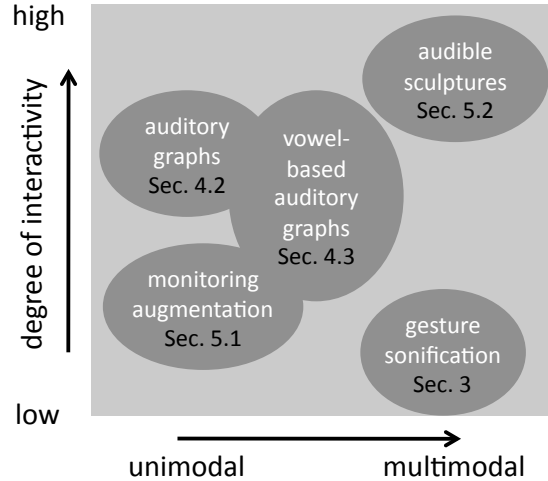


Figure 1.2: 2-dimensional space of conceptual aspects from which the practical examples in this thesis are sampled.

where an interactive auditory display offers the possibility for data exploration along one data dimension. The vowel-based auditory graph from Section 4.3, is placed towards the middle of the diagram in Fig. 1.2, because the use of vowel sounds appeals to the sound production capability of the user, which makes this example a case where a unimodal display can potentially evoke a multimodal experience. The audible sculptures from Section 5.2 are on the right and feature the multimodal aspects of listening and touch together with a high degree of interactivity: one part of the interactive display consists of sonic interaction and the other part consists of the rotation of a 3D object. Data abstractness and data dimensionality, although not depicted in this diagram, are also reflected in the case studies. The gesture sonifications, are for instance based on a high-dimensional data substrate that lends itself to the gestural and articulated character of sound. On the other hand, the auditory graphs translate fairly abstract and low-dimensional data into sounds. The high-dimensional data space of the monitoring application is purely abstract. The sonification of sculptures represents an example that is low in terms of abstractness, since sculptures represent concrete objects. However, this example is high with respect to the amount of data that are translated into sound. In all these examples, the translation of data into sounds was based on the method of parameter-mapping sonification.

1.1 Structural Organization of this Thesis

Chapter 2 contains all theoretical reflections which were developed in parallel to the practical applications and starts with a review of the development of listening modes. Section 2.1 gives a detailed introduction and motivation for the practical and theoretical scope of this thesis. This section starts with an overview of auditory display and sonification definitions and discusses the twofold hurdle that the information in the data has to overcome: The translation into sound and the interpretation through the act of listening. Section 2.2 continues by a review of the phenomenology of listening, which contrasts early concepts from auditory display by Gaver (1993a) with the preceding conceptual framework from *concrete music* by Schaeffer (1966). Recent contributions by Vickers (2012) as well as those by Tuuri and Eerola (2012), who proposed an embodied cognition centered approach are discussed with a focus on their role in multimodal and interactive auditory display systems. This section also includes reflections on temporal and persistent aspects in display as well as the relation of the sound object with sonification methods. Section 2.3 conceptualizes the PMSon design process as published in Grond and Berger (2011) in form of a diagram, which attempts to bridge the operationalizable and creative aspects of this sonification method. Different mapping topologies and their role in the practical chapters are discussed. Section 2.3 also includes a discussion of the relevance of listening modes during the PMSon design process. Section 2.4 presents an implementation of building blocks for the easy and flexible synthesis of formant like spectral contours published by Grond et al. (2011a), which was used in two applications in the practical part of this thesis. Various mapping possibilities of this synthesis class are discussed and sample sonifications based on this class are presented, which serve as prototypes for the practical chapters.

In Chapter 3, we develop an audiovisual display involving PMSon for multivariate motion tracking data of instrumentalists. In the introduction to this chapter a literature review of the mutual influence of auditory and visual stimuli is given. The first version of this multimodal display published in Grond et al. (2009) is evaluated through a free annotation of the movements consisting of the sonification and an abstract stick-figure visualization of the instrumentalists' movements. An alternative second sonification based on the articulation of the movements through vowels was evaluated by eye-tracking in order to study the influence of different parameter-mapping polarities on the visual perception of movements in this display. The chapter concludes by discussing the potential of sonification in audiovisual displays for exploratory data analysis of movements.

1.1 Structural Organization of this Thesis

In Chapter 4, we present two contributions to the field of auditory graphs (Grond et al. (2010) and Grond and Hermann (2012b)). In both, sonified curves of mathematical functions were conceived as a pedagogical aid for the blind and partially sighted. For the integration of several derivatives in one sound stream, we present the new concept of multi-parameter-mapping, which was developed in order to support gestalt formation of mathematically relevant features of the curve. The first sonification, which integrated the first derivative, was evaluated in a pedagogical context by blind students. The second sonification, which included the first and second derivative, was based on vowel synthesis and was tested for its perceptual contrast in a discrimination task. The conclusion of the chapter discusses the potential of multi-parameter-mappings, within the state-of-the-art in auditory graphs.

In Chapter 5, we develop sonifications that belong to the more recent field of auditory augmentation. The sonifications in this chapter belong to the category of MBS exhibiting a parameter-mapping aspect. The first application was developed to support the unobtrusive monitoring of algorithmic processes, in this instance for evolutionary optimizations. This application was evaluated qualitatively in a small user study and published in Grond et al. (2012). In a second application, I developed data-sonogram-inspired representations of 3D shapes in order to make sculptures perceivable for blind individuals. The interaction paradigm in this application is inspired by echolocation; the sonifications are convolutions with realtime impulses like finger clicks. The application was developed in a participatory design approach and evaluated qualitatively with blind subjects.

In Chapter 6, I develop listening-centered guidelines for the design of multimodal interactive sonifications with a focus on data exploration. These guidelines are based on the theoretical reflections from Chapter 2, the experience made during the design process and the quantitative and qualitative evaluations of sonification in the practical applications from Chapter 3, 4, and 5. These guidelines focus on the role of interaction, repetition, complementing and embodied modalities and the PMSon technique and offer support for systematic reasoning during the design process.

The thesis concludes with a summary of the contributions from the chapters and reviews how the proposed design guidelines help to address the questions raised in this introduction. The conclusion also gives an outlook how the guidelines can stimulate future research in the field of multimodal and interactive sonifications for exploratory data analysis. All sonifications are provided on a DVD and are numbered according to chapters and sections.

1. INTRODUCTION

2

Towards an Integration of Listening Modes, Interaction and Parameter-Mapping

2.1 From Sonification Definitions to Research Questions

The term sonification is used today in various contexts ranging from scientific applications to sound art and composition. Following an early definition by Kramer (1994a), sonification is understood in practical terms as *representing data with non-speech sound*. This states the primary purpose of sonification which is to adopt the role of a scientific display.

As straightforward as this definition seems, it hides a whole set of challenges in designing, composing or programming sonifications: The sonification community still struggles to define its relationship to practices in music or sound art. This is mostly due to the fact that music and sound art have a lot to say about sound but their goal and function is different from that of a scientific display. For a better definition of sonification which states the necessary requirements from a scientific display perspective, a taxonomy of sonification techniques was proposed by Hermann (2008). This detailed taxonomy stresses that sonification needs to be a systematic, reproducible, objective and transparent transformation of data into sound in order to establish a scientific display. Hermann's taxonomy also elaborates on questions of interaction, which are organized with respect to whether the user intervenes with the sonification, the data, or whether the user interacts in the world, which as a consequence influences the

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

sonification. Still, as pointed out by Hermann, this definition constitutes necessary requirements which does not mean that the resulting sound is automatically a functional i.e. information revealing sonification.

Sonification must be rooted in the data, which are algorithmically translated into a signal. However, what is perceived through listening is in essence just sound. After the conversion of data to sound, extracting information from it is the second crucial level of transformation. Interestingly, we find that information is here less a notion from the domain of informatics, see Shannon (1948), but rather refers to the etymological root of the word (from Latin *in-formare*), which literally means ‘*putting into form*’ as in an act of interpretation. This means that *what* kind of information we receive is not equal to what we perceive, but depends on *how* we perceive it. Hence despite the methodical first transformation into sound, the second challenge can be best described with a paraphrase of the saying: *the beauty is in the eye of the beholder*. Applied to sonification, the same sentence reads: *the information is in the ear of the listener*. This is not meant to promote a radical constructivist stance for sonification, which in my opinion would be unproductive, but rather wants to emphasize that information is contingent on its perception. There are different ways to look at one thing, revealing different aspects of it and so exist different listening intentions. Therefore information in sonifications depends on what possibilities we have to engage with the sound by extracting and constructing meaning¹, in brief how we can use the sound as a sign.

This aforementioned notion of beauty and information point towards the question of aesthetics. Aesthetics in sonification have been discussed by Vickers (2005) and Vickers and Hogg (2006) with an attempt to analyze and situate the aesthetics of sonification within musical practice (also compare Section 2.2.2.3). In sonification, aesthetics has not only something to do with pleasantness and the associative potential of sound but needs to be understood as information aesthetics (see Fishwick (2006)), a notion applied to sonification by Barrass and Vickers (2011). Their pragmatic approach is based on the TaDa (Task and Data) framework by Barrass (1997) and emphasizes functionality over representation. Whilst the function that a sonification attempts to fulfill is the angle from which its usefulness can be assessed, it is in a similar way distant from sound as the data. In the design guidelines presented at the end of this thesis, I will differentiate the function of aesthetics taking on different roles on a two-pole continuum of *descriptive* and *normative* tasks.

The information - being in the ear of the listener - emphasizes that sonification also needs to be understood from the phenomenology of listening in order to understand

¹For an early, auditory display related contribution to this discourse compare (Ballas, 1994, page 92).

2.1 From Sonification Definitions to Research Questions

how we use sounds as signs – how sound can point at and hence represent data. An exhaustive phenomenology of listening can be found in the *Traité Des Objets Musicaux* by Schaeffer (1966) first published in 1966. This seminal work has been made accessible to a wider community by Chion (1983) as a detailed commented index and was translated into English in 2009. A key concept of this body of work is the notion of the *sound object* and the ideal of *reduced listening*, both closely connected to a hierarchy of listening modes. *Reduced listening*, which will be described in detail in Section 2.2.1.1, is related to the acousmatic listening situation, which reframes sound apart from the audiovisual context. The concepts of Schaeffer have been adopted and reformulated to some degree by Chion (1998), who developed a phenomenological perspective for audio vision (Chion (1994)). In the context of auditory display, the mode of everyday listening has been introduced by Gaver (1993a). Gaver’s concepts dominated the phenomenological foundation of auditory display for many years. Recently Schaeffer’s listening modes have been discussed for monitoring sonifications by Vickers (2011) also including references to Chion and Gaver. Vickers (2012) extended Schaeffer’s taxonomy from acousmatic to direct listening situations. Tuuri and Eerola (2012) reformulated Schaeffer and Chion’s work with an embodied cognition approach for auditory display. These very recent contributions to the conceptual framework of listening modes demonstrate an increasing interest in this field. However they have not yet converged to a generally accepted taxonomy, and there are further terms from the phenomenological approach towards listening by Schaeffer that are worth being incorporated in the considerations for the design of interactive sonification applications.

When developing sonifications, reflecting on listening modes is what I call in this thesis high level considerations, equivalent to a top-down approach in the design process. In turn, low level or bottom-up considerations are those defined by Hermann (2008) in the beginning of this section. Low level considerations are related to the data to be sonified and the algorithmic processes of sound synthesis. With respect to the sonification method, this thesis has a focus on PMSon. In general, the challenge of PMSon can be described as integrating considerations about data and synthesis parameters, as well as the salience of perceptual dimensions and the high-level associative potential of the resulting sound. Worrall (2009), provides a detailed overview of the development of sonification¹ with a focus on PMSon, which I briefly recapitulate in Section 2.3, discussing earlier contributions to this sonification method. In the same section I discuss a scheme for the PMSon design cycle as it was presented by Grond and Berger (2011). In this design cycle, I also integrated the role of listening modes

¹Worrall (2009, page 313) also shows that high-level considerations played a role from the beginning of this research field and can be identified in early sonification / PMSon definitions.

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

and the sound object. By synthesizing these considerations in the PMSon diagram I attempt to integrate the top-down and bottom-up aspects which are in practice tightly entangled.

For the new trend of multimodal and interactive sonifications, listening modes have further the advantage to provide a genuinely auditory focus, thereby providing the necessary perspective from a top-down approach. Today the integration of multimodal and interactive aspects into design thinking is a field that has so far been merely structured as stated by Hunt and Hermann (2011, page 295):

“The challenge is huge; there are infinitely many possibilities, techniques, multimodal mixtures, tasks, etc. to be investigated. We are far away from a coherent theory of multimodal sonification-based interactive exploration.”

This thesis complements the design guidelines for sonifications, by taking interaction and multimodal aspects into account. The resulting guidelines proposed in Chapter 6 are grounded in the phenomenology of listening, the experience made during the implementation of the practical sonification applications and their quantitative or qualitative evaluations.

2.2 Listening, a Mode of Perception with Intention

The existence of different listening modes is reflected in most languages through various synonyms for the act of listening, describing different ways of perceiving the world through sound. In this section I will review various phenomenological approaches related to listening modes.

The first detailed descriptions of listening modes were given by Schaeffer (1966) in his seminal work *la Traité Des Object Musicaux*. Schaeffer (1910 - 1995) was a French composer and musicologist and is known as a major contributor to post war experimental and electronic music, in particular to the genre of *concrete music* (French: *musique concrète*). Concrete music starts from recorded concrete sound material and moves towards an abstract musical experience beyond the indexical aspects of the original material. One aspect of Schaeffer's theoretical work, namely its focus on listening intentions, offers interesting insights for a listening centered design framework. Similarly to sonification, concrete music evolved with new methods of sound production, which created a need for conceptualization of the new technologically-mediated perceptual phenomena. This led him to develop a taxonomy of sounds beyond the concept of musical notes, in order to have an apt phenomenological description for the seemingly limitless possibilities when working with a plethora of recorded sound material.

Despite his methodological and systematic approach, Schaeffer acknowledges in Hodgkinson (1986) that his practice of composing with recorded sounds is in essence bricolage. As discussed in more detail later in Section 2.2.2.2 and in Section 2.3.2, this insight is reminiscent of design cycles for auditory display: Although the world of sounds can be conceptualized, the creation of a composition or, for our concern, an auditory display involves an iterative process of trial and error. In this process guidance can only come from an interplay of sonic imaginations and the senses. This interplay can be methodological when appropriate descriptions of the auditory percepts as correlates of listening modes and intentions are at hand.

The reflection about listening intentions was rooted in the experience of sound coming from the radio. This *acousmatic* listening situation turned into a key concept in Schaeffer's theoretical framework. The ancient meaning of *acousmatic* describes a practice by a sect of disciples from Pythagoras, which would follow the masters teaching hidden behind a curtain. Acousmatic listening is conceived the opposite of direct listening, in which the sound is perceived as an *audiovisual complex* or more generally where sound sources are simply present. In principle, though more difficult, even in direct listening situations one can focus on the sound only. Similarly, in an acousmatic situation, it is not guaranteed that the sound is perceived for what it is but rather understood as a sign or index. Today *acousmatic* describes a situation that facilitates turning towards the sound itself and is not exclusively tied to the Pythagoras' teaching scenario.

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

2.2.1 The Four Direct Listening Modes

Understanding listening modes can help to identify if displays tend to favor one or another and can thereby help to create conditions that serve the purpose a sonification is made for. Schaeffer distinguishes 4 different direct listening modes as schematically depicted in Figure 2.1. They are ordered diagonally according to the abstract (3,4) versus concrete (1,2) juxtaposition: Abstract, “because the object is stripped down to qualities which describe perception (3) or constitute a language and express a meaning (4)”. Concrete, “because the causal references (1) and the raw sound data (2) are an inexhaustible given”. Further, the diagram is organized according to the objective (1,4) / subjective (2,3) juxtaposition, Objective, “because we turn towards the object of perception”. Subjective, “because we turn towards the activity of perceiving the object”, see Chion (1983, page 21).

Listening mode 1 - listening, French *écouter* - means using the sound in order to identify the source, the event or the cause. The sound takes on an indexical function and is objective insofar as we can identify what the sound is pointing at. In auditory display this is often understood as the physics of processes and energetic attributes of events, i.e. where and with what intensity things happen. The faculty of listening for indexes and causes is exploited in audification in the case of data substrates, whose oscillations originate from physical phenomena. MBS by Hermann and Ritter (2005) equally attempts to exploit this listening mode by rooting the cause of a sound event in invariant dynamical laws inspired from physics with respect to global data properties. However the index to data properties requires some conceptualization of the data space and the virtual physics of MBS, hence *hearing (3)* and also *comprehending (4)* are involved. An example would be to *hear (3)* and *comprehend (4)* two *pitch separated (3) clusters (4)* in a data-sonogram. This illustrates that despite their universality, the perception of

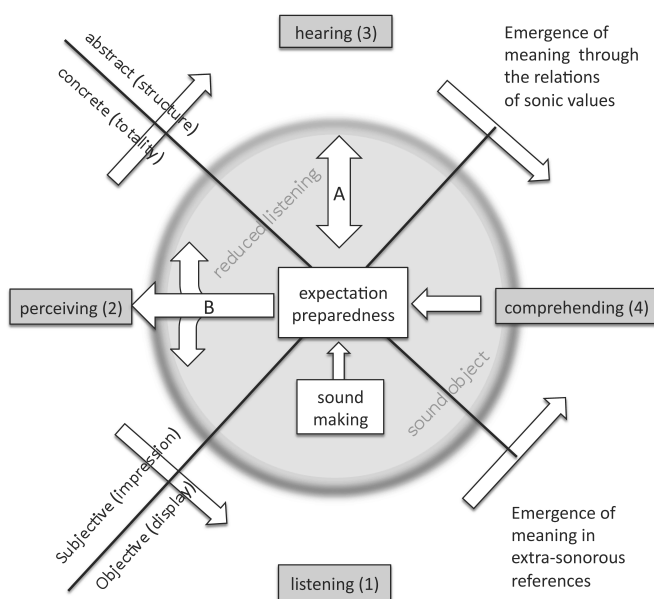


Figure 2.1: The four direct listening modes adapted from Chion (1983, page 21), see Section 2.2.4.

originate from physical phenomena. MBS by Hermann and Ritter (2005) equally attempts to exploit this listening mode by rooting the cause of a sound event in invariant dynamical laws inspired from physics with respect to global data properties. However the index to data properties requires some conceptualization of the data space and the virtual physics of MBS, hence *hearing (3)* and also *comprehending (4)* are involved. An example would be to *hear (3)* and *comprehend (4)* two *pitch separated (3) clusters (4)* in a data-sonogram. This illustrates that despite their universality, the perception of

2.2 Listening, a Mode of Perception with Intention

causalities in a physical sense, can require expertise and does not happen naturally. It needs to be stabilized through context. Auditory icons were equally conceived with respect to the causal aspects of *listening (1)*. *Listening mode 1* is not only about physical causality but more generally about indexes pointing from the perceived sound into the world. This is important for sonification in multimodal scenarios, where sounds potentially point towards other modalities.

In *listening mode 2 - perceiving -*, the original French term *ouïr* refers strongly to the sense of hearing as a whole. The lack of a similar unique term in English or German might explain why listening in auditory display is often reduced to dichotomic oppositions such as hearing versus listening, see Vickers (2011), or *musical listening* and *everyday listening*, see Gaver (1993a). Hence *listening mode 2* is more precisely translated as *perceiving through the sense of hearing*. In the English translation of Chion (1983), *listening mode 2* is translated in detail as *perceiving by ear*, which underlines the fact that this elementary level of perception informs us first and foremost that something is audibly present. This mode is very difficult to attain as it lacks an abstract language to represent it and the sound itself is on this level only sufficiently structured so that the sonic event can be distinguished from the background (compare Section 2.2.7.1). However each sonification is at first perceived on this level and might be trapped as such if the listener finds neither indices nor sonic values, nor meaning. By reducing the sound mostly to the raw outline of its audible presence and a very coarse categorization, it is the listening mode with an alarming function and has its place in monitoring situations.

In *listening mode 3 - hearing*, French *entendre*, we are not paying attention to the totality of the sound but rather to specific values of it (pitch, timbre, envelope, duration) which we choose deliberately. The deliberate choice of criteria is the reason why we can listen to environmental soundscapes as musical compositions. *Hearing (3)* is also responsible for why we often perceive a sonification as more or less musical. *Hearing (3)* is an interesting case as it is abstract and subjective, and illustrates a specific challenge in sonification. Being structured, this listening mode has the potential to reveal more information than just signaling an audible event to be present as in *perceiving (2)*. However, the subjective choice of sound properties that are abstracted from it make it uncertain if the information is decoded. If the sonification method that is used bears no indexical potential nor carries meaning, *hearing (3)* ensures that a structure is perceived, but being a subjective mode this structure is not stabilized through objective external references as is the case in the next listening mode.

Listening mode 4, comprehending, French *comprendre*, refers to extra-sonorous notions and meaning conveyed by signs, see Chion (1983, page 21). Hence the term

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

comprehending does not exclusively belong to synonyms for listening and hearing. It is an objective listening mode but differs from *listening (1)*, since the signs that are comprehended belong to an external frame of references, which can be either natural language or depending on the cultural context, a musical motif. For sonification this has the following consequences: *comprehending (4)* reflects the fact that semantic categories influence auditory perception. What we perceive is often anchored in a semantic / linguistic framework. Hence, once an object is identified or a meaning is decoded our attention shifts away from the sound itself, which could potentially be informative beyond semantic language categories. *Comprehending (4)* the meaning of sounds is the stabilizing basis for *auditory icons* as auditory signs, transitioning from *listening (1)* to *comprehending (4)*. *Earcons* as musical signs are situated in the transition from *hearing (3)* to *comprehending (4)*.

A central point that Schaeffer makes is that we tend to circle through these 4 listening modes (similar to the arrows in Figure 2.1), and that information emerges through changing between listening modes. This can be illustrated when reflecting on a learning process associated with sonification, or more generally speaking when acquiring auditory expertise. The doctor, who is in training and who diagnoses a patient's heartbeat through (acoustic) auscultation, can first only oscillate between *perceiving (2)*, that there is an extraordinary sound, and *hearing (3)*, that this sound has a recognizable structure. Then he or she can progress to *listening (1) for the cause* and will later recognize it by its name according to listening mode *comprehending (4)*, i.e. *diagnosing a specific heart murmur* for instance. This illustrates that without the subjective modes *perceiving (2)* and *hearing (3)* which are directed towards the perception of the sound itself, no auditory expertise can be acquired. However if a person has reached a certain level of expertise, he or she might first *comprehend (4)* a sound and then have to make an effort to *hear (3)* and *listen (1)* to it in order to differentiate the stage of a medical condition, also compare Chion (1983, page 22).

How can we think of these listening modes? A partly apt visual metaphor - with the restriction that it mostly addresses ambiguity in object recognition - are reversible or multi-stable images. Once we have perceived both alternatives we are reminded of the existence of the percept as a complement to the stimulus. The perceiving of one option is interestingly not experienced as deficient, which is clear if we remember that this option constitutes the experienced percept and not a subset of the stimulus. At the same time, multi-stable stimuli do not promote a radical subjectivist view, as most subjects agree on what the alternatives are, with the difference for sound that we are challenged in properly describing the alternatives. Multi-stable stimuli emphasize that their perception has an undeniable aspect of intention to it. Even more they provoke

2.2 Listening, a Mode of Perception with Intention

us to ask questions such as: which objective part of the stimulus (and for sound also the sonic context) is it that makes me perceive it as one or the other and why, i.e. how is this aspect of the stimulus related to my percept?

2.2.1.1 Reduced Listening, The Sound Object

The sound object is defined as the correlate to the *reduced listening* mode. This listening mode is meant to be a synthesis of the four natural listening modes. A reproduction of a diagram by Schaeffer depicting the relations between listening modes and the sound object is shown in Figure 2.2. The top part of the diagram shows the ordinary signifying listening modes with the two alternatives: meaning (left) and indexes (right). The bottom of the diagram shows the sound object as the correlate of reduced listening. *Reduced listening* is directed towards perception and is therefore first a synthesis of listening mode 2 and 3. *Reduced listening* captures from *perceiving* (2) the totality of the concrete given and from *hearing* (3) the potential to extract from this totality all perceivable structures.

It is an attempt to listen to the sound for its own sake beyond its abstract meaning and concrete indices. However it also compares to the objective listening modes *listening* (1) and *comprehending* (4) insofar as *reduced listening* constitutes the *sound object* as an objective unit. *Reduced listening* is therefore not a natural listening mode, on the contrary the listener has to consciously lift all conditioning and habituation.

The slightly paradoxical character of the sound object can be captured by the tension to reach an objective description of sound by overcoming the objective listening intentions of *listening* (1) and *comprehending* (4) and turning towards the subjective act of perception of *perceiving* (2) and *hearing* (3). However, by lifting individual experiences and conditioning, the sound object can claim more objectivity than indices and meaning which are the result of an individual's exposure to specific sounds and music- or language-specific terms. For the description of the sound object Schaeffer (1966) developed a meta-language, which is not purely abstract

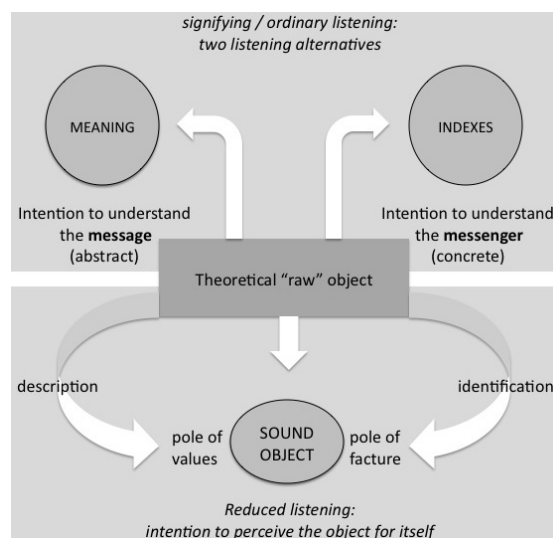


Figure 2.2: A reproduction of the table of listening intentions from Chion (1983, page 193)

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

and retains links with the four direct listening modes. As an example, one finds in Figure 2.2 on the right side of the sound object the pole of *factures*. *Facture* is a term that is part of the typological description of the sound object, and it is the qualitative perception of its energetic sustainment that leads to its identification. The term *facture* refers to the fact that each sound is made. We find it on the right side aligned with indices of ordinary listening. The terms in which *facture* is described refer however to abstract perceived sonic sustainment qualities and differ from an indexical or causal description of the sound. On the opposite side there is the pole of *values*, which stand for all possible perceived values such as pitch, timbre or duration. These values can establish a sonic reference system and are found on the left side under meanings of ordinary listening. This alignment corresponds to the fact that, as found in music sounds can establish a system of references similar to language and equally exploited in sonification. The pragmatic dimension of the terminology describing the sound object is further reflected in the categories of causalities that Schaeffer describes for the sound object, i.e. human, natural or mechanical, categories that can also be found in sound-scapes.

The primary goal of reduced listening is to be a synthesis of listening intentions that appreciates the totality of the sound and is not necessarily in strictest opposition to a specific listening mode. In fact, according to Schaeffer (1966, page 343), it is “*the swirl of intentions that creates connections or exchanges of information*”.

With respect to the relevance for sonification, one limitation of the theoretical frame work from Schaeffer (1966) needs to be kept in mind: this work focuses on the description of single percepts, however the concepts of listening modes and the intention of reduced listening is applicable beyond isolated percepts.

Chion (1994, page 25) reformulated the listening modes in three categories, *indexical* - understanding the messenger, *semantic* - understanding the message and *reduced listening* understanding the sound, the latter essentially fusing the modes *entendre* and *ouïr*, similar to the threefold organization of Figure 2.2 (also compare Chion (2009, page 471, 487, 489)).

The idea of the sound object, its status and construction as an ahistorical essence of sonic material has been critiqued by Kane (2007). A more recent critique of Schaeffer’s phenomenological method and his use of the term *intentional* is given by Kim (2010). For this thesis, I will use the concept of the *sound object* as a correlate to the *reduced listening* intention, which ideally attempts to free the sound from a first degree of signification. Its significance for sonification for exploratory data analysis lays in its attempt to lift habituation and conditioning, thereby circumventing the complexity reducing mechanisms of our perceptual and cognitive apparatus. It is a listening intention that oscillates between description and identification, see Figure 2.2. Reduced

2.2 Listening, a Mode of Perception with Intention

listening is also something that has likely been experienced by most listeners, when a certain sound or a short sonic sequence is repeated again and again like a broken record. In this commonly made experience, semantic and indexical listening intentions can get exhausted and the recording gets reduced to its sonic essence. In fact, when *reduced listening* was conceived, it has been studied in the closed groove experiment (compare Chion (1983, page 13)), where discs were deliberately manipulated in order to listen repeatedly to selected sections of the recording.

2.2.2 Musical and Everyday Listening

In auditory display and sonification, listening modes are mostly discussed according to those introduced by Gaver (1993b) in which he encourages an ecological research approach towards listening. Gaver (1993a) suggests the *everyday listening* mode as a complementary research agenda to *musical listening*. Interestingly, Gaver makes no reference to Schaeffer's theoretical work which precedes his own by about two decades. Both are interested in understanding sound complementary to psychoacoustics, i.e. to understand the perceptual essence of sound and how sound is used to signify events in the world. The difference in their approach is to some degree of practical nature: whilst Schaeffer is interested in listening modes with respect to sound objects as potential material for composition, Gaver discusses sound as an information carrier in auditory display. As a composer, Schaeffer was manipulating recorded sounds for musical purposes. Gaver, in turn, was interested in a bottom-up algorithmic synthesis of everyday sounds with the goal to gain access to salient meaningful synthesis parameters. Both share the conviction that a listener can change between different listening modes: music can be heard as interaction with the instrument and everyday sounds like traffic can be heard as music¹.

The pair of *musical* and *everyday listening* corresponds to the concept of *proximal* and *distal* sound stimuli. *Musical listening* is defined by Gaver (1993a) as listening to proximal stimuli - *the variations of air pressure near the ear*, a metaphor for listening to the sound properties themselves. Comparing with Schaeffer's terminology, *musical listening* can be identified as *hearing (3)*, i.e identifying specific sound properties, like timbre, pitch or duration.

The complementary mode of *everyday listening* corresponds to *distal stimuli*, which refers to the source of the stimulus distant from the ear, but equally relates to experience. Most of all, *everyday listening* suggests listening situations that are not *musical* with a focus on physical sound source properties. The research goal of *everyday listening*

¹For the latter, Gaver makes a reference to John Cage whose work is known for transgressing aesthetic boundaries, incorporating everyday sounds in musical pieces.

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

is therefore to relate the physics of the dynamics of the sound source with perceptually salient features of what we hear as outlined by Gaver (1993b). Compared to Schaeffer's listening modes, the ecological approach or *everyday listening* comprises the listening for indices and causes, with interference from the comprehension of extra-sonorous signs.

2.2.2.1 Things, Processes, Indices

This transgression between indexical listening and comprehending sounds becomes evident in Gaver's protocol-studies-based classification of sound material according to categories such as *vibrating objects*, *aerodynamic sounds* and *liquid sounds*, with possible further categories such as vocal sounds. Since identification errors tend not to transgress these boundaries, Gaver concludes that rather than context-based classifications, sound source categories are more likely to be generative. However, he gives a counterexample which transgresses sound source categories: the Mexican rainstick, which is often heard as pouring water (*liquid sounds*), but belongs in terms of the categories above to the subcategory *impacts* of the category *vibrating objects*. As far as object identification is concerned, physics on its own can become an unguided missile because physics teaches us a lot about the *relations* amongst objects but is less about *what* these things are. Gaver (1993b), discusses experimental evidence that physical processes of the sound source can tend to be interpreted as various objects and need a stabilizing context. Gaver (1993b) cites Vanderveer (1979) who found that subjects tended to identify the sounds in terms of the objects and events which caused them, describing their sensory qualities only when they could not identify the source events. This finding led to the hypothesis that interactions affect the temporal domain of sounds, and objects the frequency domain, which parallels Schaeffer's notion of *value* and *facture*. These examples underline that we do not always listen to the physics of the source but rather identify objects - *comprehending (4)*- depending on real or imaginary contexts. In fact, in *everyday listening* situations sound often carries a lot of contextual information, which can contribute to the stabilization of the perceived sound by limiting the plausible categories of objects. *Listening mode 1* from Schaeffer, is first and foremost conceptualized as indexical and causal and cannot not only be reduced to physical signal properties, making it a useful concept for multimodal contexts, where sound points towards other modalities.

2.2.2.2 Finding the Sonic Essence in Design Cycles

Gavers's practical contribution to auditory display (see Gaver (1989) and Gaver (1986)) is the concept and development of auditory icons. For an in-depth discussion of the development of this field see Brazil and Fernström (2011). Auditory icons are sound cartoons based on the simulation of the physical sound source properties that attempt to capture the essence of a real-world sound, without being concrete instances. In order to assess the connection between the physical simulation of source properties and the percept which it evokes, Gaver proposes an analysis and synthesis approach: the signal is first analyzed and re-synthesized based on the analysis, then evaluated simply by listening. Gaver points to various challenges in this approach for instance the confusion of the signal property with the percept, by stating: “*Insofar as it is based on an analysis of physics, it is liable to confuse source attributes that affect sounds with those that are actually heard.*” Gaver (1993b, page 16). He also acknowledges that everyday language might not be appropriate to attack this problem: “*... it is liable to confuse the effects of language use for the attributes of perceptual experience*” Gaver (1993b, page 16). Interestingly the challenge through language effects in the analysis and synthesis listening approach corresponds to Schaeffer's listening intention *comprendre* (4). Although Schaeffer is aware of these effects and developed the meta-language describing the sound object, the process of composition remains for him bricolage, i.e. a constant evaluation cycle through listening. We will encounter in Section 2.3.2 a similar challenge in the design process for PMSon, where the iterative process depends on the guidance of sound imagination and evaluation by listening.

2.2.2.3 Mapping Principles and Semiotics in Auditory Display

Gaver (1986, page 170) discusses three mapping principles for computer events and auditory icons: *nommic* mappings are the true physical model of the sound-producing event. *Metaphoric* mappings share some similarity between the process to be represented and the chosen sound. *Symbolic* mappings are the most arbitrary mapping scheme, where the relation between sound and the process has to be learned. Gaver notes that *nommic* mappings, understood as a caricature of physics, do not guarantee that an appropriate process or object is associated with them and hence also bear some arbitrariness. It is interesting to note that Gaver proposes the notions of *nommic*, the literary term *metaphorical* and *symbolic* for the different mappings, rather than using the semiotic terms *iconic*, *indexical* and *symbolic*. In fact, Gaver (1986, page 170) states that the semiotics in auditory display need to be worked out in more detail, which has also been noted by Oswald (2012) (on the problem of finding proper terms to categorize sounds

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

also compare Kramer (1994b, page 197) and for terms describing functions of sounds also Ballas (1994, page 81)).

Similarity, i.e. the iconic dimension of a sound, is often judged with respect to selected sound attributes. In fact, if we listen with different listening intentions to the same reproduced acoustic signal of a sufficiently complex sound, the two resulting sounds can be perceived as either similar (iconic) or different depending on what we have retained from the totality of the percept. Only if we attempt to lift habituation and conditioning contexts, more persistent and subtle similarities of the emerging sound object can be discovered. Further, the term *iconic* is a challenging concept for sounds if the percept is reduced to an object as the example of the rainstick has shown. This however extends the notion of iconicity to extra-sonorous systems like language and the semantic categories of objects. In this light, *auditory icons* are iconic only because of the stabilizing context, e.g. the desktop metaphor; in sonic essence they are only *auditory indices*. The iconic dimension of a sound itself is hence better judged through reduced listening and less with a direct listening mode.

Through indexical listening, sound always retains links with its source. For isolated sound objects in acousmatic conditions, these indices are related to the physicality of the sound source. In the rainstick for instance, the sound density is indicative for the amount of particles. Indices of non-everyday sounds can go beyond the material realm of physics, extending to situations and cultural experiences. Indexical listening is influenced and provoked by other modalities, in the audiovisual complex for instance, sound points to correlates in the visual substrate. What does indexicality mean for sonification? It is tempting to relate it to sonification methods and establish a hierarchy of indexicality as has been suggested by Vickers (2005). A hierarchical organization of sonification methods could be ordered from the most indexical being audification over MBS to PMSon. However, indexicality is first and foremost a listener category and hence examples of audification have been discussed by Grond and Hermann (2012a) where indexical listening might fail to relate to the data.

The symbolic mapping corresponds to earcons introduced by Blattner et al. (1989). For an overview of the field see McGookin and Brewster (2011). According to Schaeffer a single sound object cannot be symbolic since it cannot be arbitrarily linked to any possible meaning like words in language. Hence sound symbols or earcons can be generalized as a collection of percepts, beyond the notion of a strictly musical motif. Consisting of more than one percept, the tight indexical link to the sound source can be loosened through the combinatorial possibility of relations among the sonic values of the percepts. Even auditory icons can also consist of several percepts, the relation between them can create a context which stabilizes the overall meaning by helping sonic

2.2 Listening, a Mode of Perception with Intention

values to emerge such as time intervals between footsteps. Although symbols are by definition arbitrary associations between the sign and the signified, Chion (1983, page 84) quotes Schaeffer as: “*if the linguistic sign is arbitrary, the musical sign is not*”. This, even for the musical motif still tight connection between the sonic sign and its meaning is not only due to an indexical, i.e. source-related listening intention but is also related to embodied aspects of listening like the perception of tempo as motion for instance.

2.2.2.4 Comparing both Frameworks of Listening Modes

The most significant difference between the approaches is that Schaeffer’s phenomenological account of listening strictly tries to avoid any acoustic explanation of what we perceive. Schaeffer’s core conviction can be summarized as listening being everything but a passive mode of reception. This is why his links with semiotics are cautious but precise, insofar as the semiotic modality of a sound - its sign nature - is tied to the listening intention and not to the signal.

Gaver’s approach is motivated by the desire to operationalize sound for HCI. The ecological research approach towards *everyday listening* is hence the wish to influence perceptually salient features by controlling the physics of the sound source. This is the reason why his preliminary semiotics of sounds in HCI are tied to the mapping in the design process. The challenges that Gaver describes for the ecological approach, which can be captured as the tension between listening to sound source properties versus object recognition, indicate that Schaeffer’s conceptualization of listening modes provide a general structure which can differentiate perceptual responses that fall into the *everyday listening* category with Gaver.

It is however not trivial to identify or equate the listening modes with each other. Gaver’s dichotomy of the musical as opposed to the everyday for instance captures one insight from Schaeffer, which is that in traditional music in its absolute form, non-signifying meaning emerges that is at the cusp between *hearing* (3) and *comprehending* (4) (compare Chion (1983, page 28, 92)). Further, the categories of Gaver’s protocol studies are reminiscent of the categories of causes proposed by Schaeffer for the *facture* of the sound object.

Three aspects challenge the development of a general semiotics of sounds for sonification. First, there is the hybrid nature of sound as being somewhere between physics and language and not reducible to either or both¹. Second, sounds in sonification can range from single percepts over collections of percepts to sound streams. Third, sound is

¹Compare Chion (1983, page 84) citing Schaeffer: “*dragging music by hook or by crook from physicist determinism to linguistic structuralism.*”

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

perceived with or appeals to other sensory modalities, which influence the sign modality that a sound adopts. Important for sonification in multimodal contexts is that many sound qualities that would be attributed to *musical listening* or in Schaeffer's terms *hearing (3)* can become audible indices, because as stated in Section 2.2.2.3 indexical listening is a listening intention and not a sound source property.

2.2.3 Listening Modes and Monitoring

As one of the application prototypes, I develop in Section 5.1 a monitoring application for optimization algorithms. In the context of monitoring, Schaeffer's and Gaver's terms of listening modes have been discussed to some extent by Vickers (2011, page 5) in the context of monitoring. Vickers categorizes monitoring into three types of tasks and discusses how they relate to receiving/perceiving information from an auditory display:

Direct auditory display: The information to be monitored is the main focus of attention and does not allow for parallel activities. The person who is monitoring is pulling information from the display. Vickers describes the related mode of auditory perception as listening.

Peripheral auditory display: The monitoring person has his or her attention focused on a primary task whilst required information relating to another task or goal is presented (pushed) on a peripheral display. Vickers describes the related mode of perception as hearing.

Serendipitous-peripheral: Attention is focused on a primary task whilst information that is useful but not required is presented on a peripheral display and is monitored indirectly. Vickers describes this situation as the display is nudging the user, however he does not propose any specific listening mode.

Listening (1) and *comprehending (4)* summarized as *everyday listening* applies for the direct monitoring situation. According to Schaeffer, *hearing (3)* isolates specific characteristics from the sound, which is unlikely if the sound is in the background. In fact the peripheral display has features of an alarm (compare Vickers (2011, page 3)). This corresponds with the description of *perceiving (2)*, *ouïr* given by Chion (1983) as *being struck by sounds*.

As far as the serendipitous monitoring is concerned, a balance between the alarming aspects of *listening mode perceiving (2)* possibly, mixed with mode *listening (1)*, might be a proper description of the corresponding listening mode. The sound is concrete but still has to find what it objectively refers to by answering the question: what's going on. What strikes the listener can however be related to audible features that belong to *hearing (3)*. Any change in pitch or duration in pulsed streams, for instance can create the nudge, see Section 5.1.1.4.

2.2 Listening, a Mode of Perception with Intention

Since monitoring is often happening in a soundscape related setting, a listening taxonomy proposed by Truax (2001, page 21 and 24) and discussed by Tuuri (2011) must be considered, which proposes three levels of attention in listening: listening-in-search, listening-in-readiness and background listening. These modes match in their description well with the three monitoring tasks mentioned above.

2.2.4 Listening and Sound Making

For all interactive sonification prototypes as in Chapter 4 and 5, it is interesting to consider that listening to sounds in Schaeffer’s phenomenological approaches is strongly tied to making sounds. Schaeffer introduces the term *musicianly listening*, see Chion (1983, page 39), which must not be confused with Gaver’s *musical listening* (corresponding to arrow A in Figure 2.1). *Musicianly listening* is closely related to reduced listening and corresponds to arrow B in Figure 2.1. It refers to the listening mode of practicing musicians, who when making sound do not first and foremost express musical content but attend to the sound in all its facets in order to understand and improve the sound quality by altering the way they interact with the instrument. Although the term *musicianly listening* makes one think of music, it is a listening intention towards any audible action or interaction, fostered through an attitude of preparedness and expectation in Figure 2.1, also compare Chion (1983, page 36).

This review of listening in sound making would not be complete without the notion of *ergo audition* from Chion (1998, pages 84-85), which describes the perception of sounds while making them. In *ergo audition*, Chion interestingly differentiates both, the causal listening of the closed loop experience which has also been elaborated by Hunt and Hermann (2011), as well as the need to dissociate listening from making in learning processes such as in instruments similar to the *musicianly listening* mode. Related to the latter, Hermann and Ritter (2004) proposed in Gaver’s vein the term *analytical everyday listening*, where the listening focus is on accessing details about the object by actively querying sounds via interaction.

With respect to interactive sonifications, Schaeffer’s distinction between the abstract instrument’s timbre and the timbre resulting from the manner how the instrument is played, see Chion (1983, page 53) also needs to be considered. Through the process of *musicianly listening*, the instrumentalist – or for our purpose the user – acquires her skills by learning to distinguish between both. The importance for interactive sonification lies in the fact that depending on the purpose of the sonification, the focus is either on closing a feedback loop thereby directing the focus on interaction or to focus onto the data which would correspond to the instrument’s timbre. The later requires to be able to abstract from the sonic effects of interaction.

2.2.5 Recent Contributions to Listening Modes

Whilst listening modes have seen only few contributions in the beginning of auditory display research, interest in this field has increased recently. The most recent contribution by Vickers (2012) is also a review of his previous publications related to the aesthetics of sonification and indexicality. Vickers extends the four listening modes by Schaeffer with modes that take direct listening into account, thereby leaving the acousmatic context. This extension is mostly referring to the work on audio-vision by Chion (1994), in which he proposes causal listening, as indexical listening pointing to a visual cause. Vickers adds four direct listening modes which are also organized according to the objective / subjective and abstract / concrete grid. He proposes acronyms of three letters (A) acousmatic or (D) direct, (A) abstract or (C) concrete and (S) subjective or (O) objective. The direct listening experience includes in this scheme interactivity, which means that the cause is not necessarily visible but present.

The DAO direct-abstract-objective mode (complement to *comprendre*) is what we encounter mostly in auditory display. Vickers suggests for this mode an interactive MBS. The problem is that understanding of what the sound of a MBS means with respect to extra-sonorous properties of the data requires conscious involvement of other listening modes. If a visualization of the data is present at the same time, this can turn into a direct-concrete-objective DCO listening, which Vickers equates with Chion's causal listening, the complement to *écouter*.

Interestingly Vickers suggests that direct-abstract-subjective listening or DAS, the complement to *entendre*, has little place in auditory display. Being abstract he states that in an interactive display “*sound is being made, but for its own sake*”, see Vickers (2012). The direct-concrete-subjective mode DCS, the complement to *ouïr*, is what Vickers describes as an unsuccessful direct auditory display. Sound is perceived but neither recognized nor connected to present causes.

Vickers' extension provides a useful structure to think about listening modes in multimodal auditory display. Interestingly, the direct listening mode makes no distinction between complementing modalities, which can be visual, interactive, or haptic. It hence provides a genuine listening-centered way of thinking about multimodal auditory display.

Vickers also reviews Tuuri et al. (2007), which is part of the PhD thesis of Tuuri (2011), in which he makes several contributions to an embodied cognition-centered conceptualization of listening modes for auditory display. Tuuri's approach to auditory display contains an in-depth theoretical discussion of multimodality in HCI with an emphasis on embodied cognition (compare Tuuri et al. (2009)), where the authors argue for a non-symbolic *amodal* completion of unimodal stimuli. The arguments in

2.2 Listening, a Mode of Perception with Intention

this paper challenge a simplistic idea of multimodality based on the number of sensory input channels. The concept of *amodal* completion of percepts also challenges the direct significations of sounds in auditory display. This thinking is in line with the idea of a gestural sonorous object by Godøy (2006), which will be discussed in Chapter 3.

Tuuri and Eerola (2012) revised the scheme of listening modes and propose three hierarchically modes with various subdivisions. Despite the hierarchical organization, the authors point at the fact that this scheme is not meant as a bottom-up approach but that instead each mode can be activated independently and simultaneously.

The *experiential mode* contains the most basic listening modes which are the *reflexive* and *kinesthetic* mode, the later emphasizing the bodily basis of meaning-creation. The third mode is connotative, subdivided into the *action-sound-object* (individual sonic interaction), *action-sound-intersubjectivity* (sonic interactions with others) and *action-sound-habit* (cultural coupling). These *action-sound-couplings* emphasize that sound is always tied to actions that cause it and parallels the thoughts in listening and sound making in Section 2.2.4.

The *denotative mode*, encapsulates the indexical aspect of listening, which is subdivided in four categories that span from source orientation to the orientation towards the context. The subdivisions are *causal*, *empathetic*, *functional*, and *semantic*. The functional field has been highlighted by Vickers as useful for auditory display. This listening mode also allows for the most symbolic use of sound insofar as any kind of sound can be perceived as purposive in an interaction context. As mentioned by Tuuri and Eerola (2012), a *functional* listening intention means that the context provides affordances for how the sound is used.

The *reflective modes* include *reduced listening* which is a conscious turn towards the sound, as well as *critical listening* which involves aesthetic judgment of the sound. Both modes are, as Tuuri et al. point out, premeditated listening modes and involve higher cognitive structures.

2.2.6 Summary of Recent and Previous Contributions

It is interesting to see how the first accounts of listening modes by Schaeffer and later Gaver relate to each other and where Gaver's modes are located in Schaeffer's abstract / concrete and subjective / objective grid. Vickers identifies *everyday listening* as being primarily objective and hence equates it with *écouter* and *comprendre*. Tuuri and Eerola (2012) emphasize the similarity of *musical listening* with *reduced listening* as turning towards the sound. It can however be argued that *reduced listening* might even have its place in Gaver's quest to find sonic essence in analysis and synthesis design cycles which are evaluated by listening, as discussed in Section 2.2.2.2.

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

It is not straightforward to relate *ouïr* to the experiential modes presented in Tuuri and Eerola (2012). Schaeffer’s historic phenomenological framework seems not to be comparable on that level with an embodied cognition centered approach. The main question is, as Tuuri has phrased it in personal communication, if the reflexive aspect qualifies as an intention.

It is interesting to note that the subjective abstract listening mode *hearing (3)* (which is – as its etymological root of *entendre* suggests – pure and simple listening intention) tends to disappear in both the HCI centered approach in Vickers (2012) and the embodied cognition-centered approach in Tuuri and Eerola (2012). I interpret this observation as a sign for the difficulty to relate functional sounds from auditory display, with musical practices, which according to Schaeffer emerges at the transition from *hearing (3)* to *comprehending (4)* (compare Chion (1983, page 73)). Some aspects of musical practices can however be captured from an embodied perspective as *amodal* as suggested in Tuuri (2011). *Reduced listening* being at first a synthesis of *hearing (3)* and *perceiving (2)* (compare Section 2.2.1.1) also preserves *hearing (3)* as a more reflective listening mode.

Combinations of the listening modes from Tuuri and Eerola (2012) can be understood as specific listening intentions described by Chion and Schaeffer. The *kinesthetic* basis would for instance project into the *empathetic* and the *semantic* listening, thereby enabling to listen for intentions of the messenger. The action-sound-object in combination with *functional* listening mode, captures well what Chion named *ergo audition* in the closed loop. The particular role of *ergo audition* becomes evident in contrast to action-sound-intersubjectivity and denotative *empathetic, or causal* listening. This pair of combined listening intentions can be found in the experience that a sound that is caused by oneself stands out less compared to sounds made by others (compare Chion (1998, pages 84-85)). Also *musicianly listening* can be understood as an activation of the *connotative action-sound-object*, the *causal* and the *reduced listening*.

Although Vickers (2012) cites Chion, he does not elaborate how the four listening modes he adds relate to Chion’s term of *ergo audition*. The direct-abstract-subjective listening mode, where sound is made for its own sake, comes close to the notion of *musicianly listening*. Although not pointing towards data in particular, it potentially still serves an interesting role in auditory display. The nature of the relationship between reflective listening intentions and the action oriented embodied listening modes – and its relevance for interactive sonification – is maybe well described through the German word *aufhören*¹, which illustrates that the focus can be shifted towards perception to

¹*Aufhören* means to deliberately discontinue an activity, the root of the word is *to hear*. The Duden (2001, article: hören) suggests *aufhorchend von etwas ablassen, (to discontinue something in order to listen)* as the origin of the word. Although this interruption might be due to sounds that do not originate from the interaction, it illustrates the shift from the closed action and perception loop towards the reflective oscillation between description and identification.

2.2 Listening, a Mode of Perception with Intention

the point that action is interrupted. The effort of *musicianly listening* is hence to maintain or initiate the action, despite the focus on perception, or to keep the focus on perception, despite the need to cause the sound through action.

Vickers ironically points out that the spellchecker of the text editor he uses always insists to replace *sonification* with *signification*, taking that as a sign to exclude the non-signifying *reduced listening* as a functional mode for auditory display. This is somewhat in contradiction with other auditory display literature. In the design-space map (DSM) by deCampo (2007), for instance, the goal is to find ways to design potential sound objects, which are the correlates of *reduced listening*. As already mentioned in Section 2.2.1.1, I propose to think of *reduced listening* as an iterative nested approach which helps to discover indices in a sonic structure beyond the first level of signification. This approach towards *reduced listening* is in line with the pragmatic terminology describing the sound object. I will discuss in the conclusion of this thesis that the functionality, or efficacy, of *reduced listening* depends on the purpose for which the display is designed.

As far as the embodied cognition-centered approach is concerned, it must be noted that the fact that the reflective modes involve higher cognitive resources can be understood as offering higher levels of abstraction from first level significations. As Tuuri and Eerola (2012) state, *reduced listening* is a highly premeditated activity. In this sense it becomes potentially detached from the immediate meanings of the connotative and denotative embodied levels, which offers the possibility to engage with the unknown, for instance when exploring sonifications of complex data. Here reflective listening modes turn back towards the structure of sounds, which carry no immediate meaning, such as those provided to the listener through the connotative and denotative modes in the embodied cognition centered approach.

2.2.7 Concluding Remarks

2.2.7.1 Between Phenomenology and the Signal

When investigating the sense of listening, science is interested in understanding sounds with respect to quantifiable signal properties and their behavioral, neurobiological or psycholinguistic correlates. The sonic arts focus on the phenomenology of listening from a philosophical perspective. With respect to sonification, both the scientific and the phenomenological approach towards listening contribute in a complementary way: The main advantage of scientific findings is that they can be in some cases operationalized. In other cases they can provide upper and lower limits of some perceptual dimensions within which the auditory display needs to be designed. A good overview of these

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

findings and their implications for auditory display is given in Carlile (2011). In many cases perceptual laws apply to simple stimuli like pure tones, however, the more complex and therefore potentially engaging and informative the sound becomes the more difficult it is to operationalize these findings.

Psychoacoustic research has done a good job in establishing correlates between the signal and what we perceive. Further, synthesis techniques have dramatically advanced since the first auditory icons by Gaver and the modeling of the physics of instruments has led to astonishing results in simulating their sounds. However it remains true that we can analyze and describe many signal properties of perceptual relevance, but this does not mean that these sound features are actually heard, i. e. consciously perceived, meaning that we can hear a bell without registering its pitch.

With the progress in mapping the auditory cortex and auditory pathways, listening modes will find their neural correlate and hence a scientific explanation. In fact, the taxonomy of listening modes proposed by Tuuri and Eerola (2012) is supported with ample references to neurobiological literature. It can however be observed, that the arts and phenomenology are sensitive to effects and phenomena before their existence is proven or their mechanisms are scientifically explained. To name just a few: the neurobiological finding in the form of voice-selective areas in the human auditory cortex in 2000 by Belin et al. (2000), provides evidence for the high saliency of the vocal sounds, which constitute its own category in soundscape compositions. Similarly the psycholinguistic studies by Guastavino and Cheminee (2003) provide evidence for the audible categories in a soundscape. It is beyond the scope of this thesis to discuss in detail listening modes with their neurobiological correlates. However, it is worth mentioning that recent fMRI based studies by Lewis et al. (2012) start to shed light on the neurological basis of the scene-like or object-like status of sounds in auditory scenes, which might be partly identified with *listening mode perceiving (2)*. Interestingly, Lewis et al. report a dependence on top-down task demands. This dependence supports the view that while sonic substrates might have the potential to *evoke* sets of naturally appropriate listening intentions, specific listening intentions among these sets need to be *invoked* from top-down. Within the taxonomy of Tuuri and Eerola (2012) the hierarchical order suggests that the higher the listening mode is located in this vertical scheme the more they can and need to be consciously *invoked* such as the premeditated reflective listening modes.

2.2.7.2 Persistent and Timebased Media

Perception as processing the sensory input always happens in time. In this sense even still images or plots change as we study them. However, I want to maintain the distinction between time-based and persistent media, for the following reasons:

Timebased media whether visual or sonic bring forth phenomena that necessarily need to be reified as objects or reduced to indices or values in order to retain something once they are gone. A biological reason for this tendency is that of memory constraints, creating the necessity to reduce the totality of the sensory input. On the contrary, phenomena experienced in persistent media offer the possibility of exploration. Hence they have a higher potential to maintain their plethora of possible meanings and structures while they are perceived. One reason that the concept of the sound object and reduced listening is difficult to grasp (and for which it is usually critiqued), is exactly because it attempts to make the ephemeral phenomenon of sound accessible as a persistent object which can be explored and studied in its totality.

For sonification design, this distinction has implications for the task and display purpose. For closed action and perception loops with the purpose to achieve a well-defined goal, the reduction of the sound to features matters less, as long as they are perceived and help optimize the action in order to achieve the goal. Sound is meant to make us do something repetitively and further is meant to condition us to do it well through augmented perception. For a historic overview on sound as a corrective means see Schoon (2012).

Exploratory data analysis is on the opposite side of display purposes. Here the sound is meant to make us think, and Kepler (1967, page 134) comes to mind with his wish that *“The intellect articulates what the ear would naturally have to say”*. Since for exploratory data analysis, we do not know a priori what we are searching for in the data, we also should not have a biased listening intention which reduces the totality of the sound to specific aspects. In order to allow the sound to make us think and discover, we need to widen our listening intentions towards *reduced listening*.

In design, form is usually discussed with respect to the functionality of the designed artifact. This is commonly known as *form follows function* dating back to the Bauhaus tradition, also adopted for sonification by Barrass and Vickers (2011). In multimodal sonifications with an exploratory purpose, the form corresponds to the auditory display system and the functionality depends on whether this display allows the user to *invoke* an appropriate listening intention.

For design aspects, it is important to remember that while the sound object must by no means be confused with the sound signal, it was sound technology which fostered its conception, by allowing to reproduce exactly the same acoustic signal that gives

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

access to various perceivable aspects of the sound object. In that sense, technology does not only extend our senses in a cybernetic sense, but can also enable us to turn towards and reflect about our percepts, and later identify aspects of it with sources or in the context of sonification with data.

Ironically, the slightly transcendental notion of reduced listening and its correlate – the sound object – has a conceptual similarity with MBS, which proposes a particularly objective and rigorous translation of data into sound. Physics or science in general, approach time-based phenomena often by reconstructing time series in phase space and thereby treating the problem from a geometric perspective. This particular kind of objectification is akin to the idea of the sound object, which equally attempts to get hold of the fleeting nature of sound. MBS shows its particular strength in proposing physics inspired invariant translations of geometric data relations into sonic phenomena. The sound object in turn tries to reveal what is persistent beyond the bias of our time-based perceptual apparatus. This structural relation between the sound object and MBS, the latter of which was to a large degree conceived for exploratory data analysis, should remind us that creating the opportunity to *invoke* the right listening intention is important for the design of auditory display systems for exploration purposes.

Appreciating the complexity of sound objects and trying to use them to represent complex data is a movement away from encoding data complexity in temporally extending data driven scores (compare Grond and Hermann (2012a) for a discussion). In this popular approach which is close to data driven music, the surplus of each individual sound is on the one hand reduced through the musical context according to the listening mode *hearing* (3), which is a subjective process. On the other hand listening modes proposed by Tuuri, help us to understand that the listener relates to the sounds through embodied knowledge. Both are however likely not to be indexical to the abstract data substrate.

The thoughts elaborated here complement the rigorous definitions of sonification as a scientific display for which one would demand a certain degree of abstraction¹ and objectivity. Following the abstract and concrete versus objective and subjective grid of listening modes proposed by Schaeffer, one is left for sonification with the fusion of the subjective listening modes of *hearing* (3) and *perceiving* (2), which brings forth the sound object, whose structure can become indexical (*listening* (1)) to the data in the introspective oscillation between description and identification.

¹For the question of abstraction in auditory display, also compare Ballas (1994, page 80)

2.3 Parameter-Mapping, Connecting Data and Sound

Parameter-Mapping Sonification (PMSon) involves the mapping of data features to synthesis parameters, which can be *physical* (frequency, amplitude), *psychophysical* (pitch, loudness) or *perceptually coherent complexes* (timbre, rhythm) as phrased by Worrall (2010). Since sound has various perceptual dimensions, PMSon is at least in principle well suited for displaying multivariate data. PMSon complements audification and MBS, the latter of which can contain a parameter-mapping element. Guidelines on identifying the method of choice from either audification, MBS or PMSon, were introduced in the Design Space Map by deCampo (2007). Unlike audification and MBS, PMSon is challenging from a sound design perspective because of the wide range of mapping decisions, which provides enormous opportunities to create and tune appropriate auditory display for a particular desired purpose. Nonetheless, PMSon is often the option of choice, for instance, when the data rate is too low for audification or when a particular data dimension needs to be translated into a time structure in the display.

A detailed account of the development of PMSon has been given by Worrall (2009). The multivariate data PMSon was first discussed within the seminal book *Auditory Display* by (Bly, 1994, page 406); Scaletti (1994, page 224) presented in the same book the concept of the *n*th order mapping, meaning that each dimension in a multivariate data set is mapped to a distinct synthesis parameter. Ever since these early contributions, the question of how to justify concrete mapping decisions, and potential problems have often been addressed, for instance by Walker and Kramer (2005) and some rules have been derived by reviewing various attempts in the field of auditory graphs by Flowers (2005) (also compare Section 4.1). However, the challenges of parameter-mapping which is generally referred to as *the mapping problem* are not resolved. The challenges of mapping justifications can be identified as questions related to mutual dependence of psychophysical perceptual parameters as well as the metaphorical dimension of their interpretation (compare Vogt and Höldrich (2010)) and their polarity and magnitude (compare Walker (2007)). In practice, particularly for complex mapping topologies both questions are mutually dependent on each other. In the view of these challenges, one can say that sound design for sonification is only a concern where PMSon is involved, since for audification and MBS methods the sonic outcome is to a large degree defined through the data themselves or the sonification model.

The mapping problem has recently been addressed by Worrall (2010) where the author suggests as a remedy an embodied cognition approach towards parameter-mapping, based on a review of related embodied cognition literature. Worrall (2011), extends this approach more concretely by proposing a gesture encoded sound model.

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

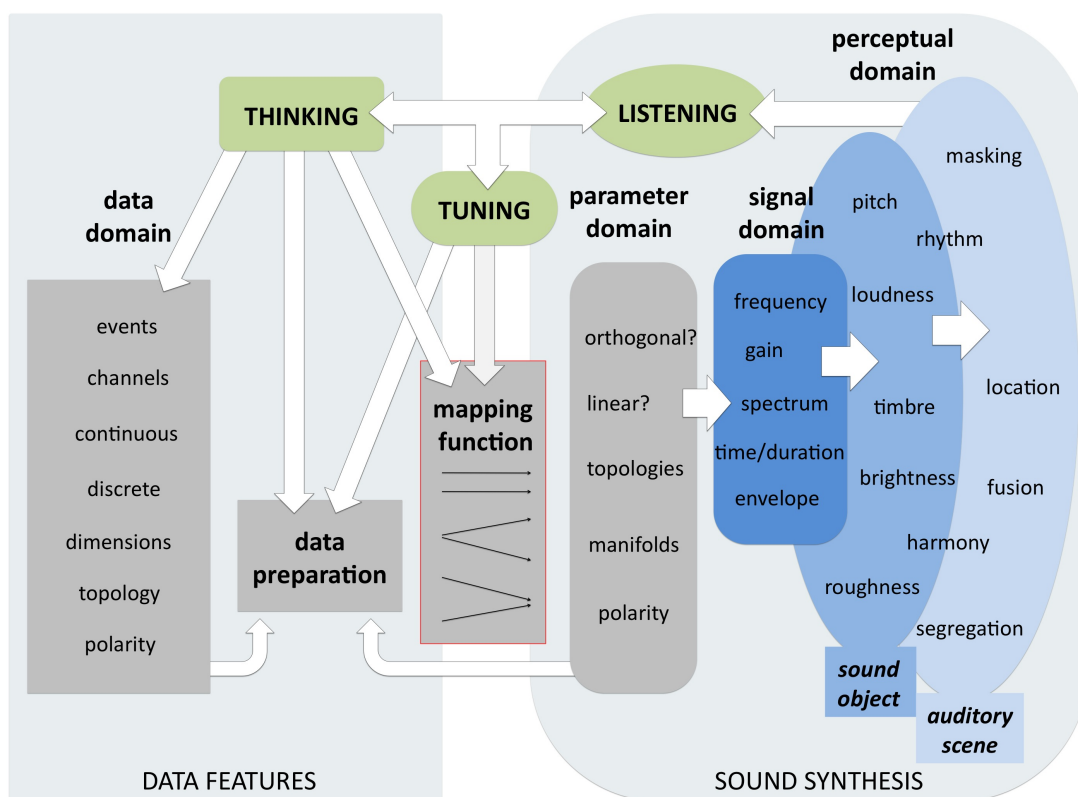


Figure 2.3: Map for a general design process of PMSon from Grond and Berger (2011) progressing from left to right. The diagram shows the data domain on the left and the signal and perceptual domain on the right.

While this model still waits to be fully implemented, the parameter-mapping sonifications in Chapter 3 and 4 have been similarly inspired by an embodied cognition approach.

Future developments might lead to generalizable mappings, for now it remains that the iterative process of PMSon involving trial and error is akin to bricolage, sharing similarities with composition (compare Section 2.2). The challenges and considerations that need to be made in PMSon can however be broken down systematically and are shown in the diagram of Figure 2.3, taken from Grond and Berger (2011). This diagram is divided between data and numerical control in grey, and sound and auditory factors in blue. Rectangular fields can be handled with rigorous objectivity. Oval fields involve human perception and impose a subjective component to the process. As stated in Grond and Berger (2011), designing a PMSon “*involves the interplay of, and the conscious intervention in both the data and the signal domains. Integrating both worlds is key in creating an effective sonification.*”

2.3.1 The Fields of the Parameter-Mapping Diagram

Since the fields in this diagram have already been discussed in depth in Grond and Berger (2011), I will give here specific examples that refer to the practical applications in the following chapters.

2.3.1.1 Data Domain

The *data domain* contains properties directly related to the data such as *channels* or *dimensions* and whether the data are *continuous* or *discrete*, aspects, which have been discussed in deCampo (2007). The movement data for the sonification of ancillary gestures from Chapter 3 for instance, provide a densely sampled continuous variable which translates into smoothly interpolating auditory streams. A similar case are the smooth function data in Chapter 4. The data type of both applications in Chapter 5 represent discrete data samples in an either high-dimensional search space or in 3D cartesian coordinates. The mappings in this chapter are part of an MBS approach, and the resulting sound clouds share a similarity with granular synthesis.

2.3.1.2 Data Preparation

The need to prepare or preprocess the data can serve various purposes. In Chapter 3, a principle component analysis (PCA) was used in order to reduce data dimensionality. Data preparation can also contain filtering as in Chapter 3 in order to smooth out artifacts. In Chapter 3 and 4, the data preparation consisted in deriving complementary data such as derivatives or joint angles. In the second part of Chapter 5, features of the data had to be extracted depending on the listening position with respect to the 3D dataset. Data preparation can also include the statistical analysis of the data which can be included in the mapping function as shown in the sonification examples for vowel synthesis in Section 2.4.4.

2.3.1.3 The Mapping Function and Topology

Although all fields in the diagram play an important role, it is the mapping function and the mapping topology that touch the heart of the sound design problem. Here decisions need to bridge the objective rigor of the data domain with the anticipated results in the perceptual domain, because data are mapped to signal parameters but the sonic result can only be judged by ear. This challenge becomes evident when comparing the formalization approaches by Hermann (2002) and Rohrhuber (2010). Hermann's approach suggests defining transfer functions and practically consists of a table which lists the associations of data features with signal parameters that come

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

close to perceptually approximatively linear units (e.g., MIDI notes, dBV). Rohrhuber in turn defines the sonification variable \mathring{S} , which has the advantage of tightening the data and the synthesis domain with a mathematical formalism, making it possible to distinguish data time t and signal time \mathring{t} . However \mathring{S} addresses exclusively the signal domain. The problem of formalizing the transition from data to sound is the inevitable identification of synthesis parameters with perceptual dimensions. A possibility to address the entanglement of perceptual features and the signal, is to experiment with various mapping topologies.

One-to-one mappings can strictly only be mappings to the signal domain, since the correlating perceptual feature is generally not independent. It does however often work well for mapping to signal properties related to pitch or frequency. Occasionally, timbre as spectral contour or brightness allows for smooth transitions as in vowels, see Chapter 3 and 4. This mapping topology was applied across all practical applications.

Kramer (1993, page 200) introduced the mapping of one data feature to several synthesis parameters, also known as *one-to-many* or *divergent mapping*, which takes into account that several perceptual dimensions can depend on one factor, e.g. the energy input to a resonating object. This mapping topology was applied in Chapter 3 and Chapter 5. The multi-parameter-mapping approach from Chapter 4 constitutes a hybrid between *one-to-one* and *one-to-many* mappings. In Section 2.4 an example is given where the data is mapped over various ranges in the signal domain with good effects in the perceptual domain, effectively providing a mixture between *one-to-one* and *divergent mapping*.

Many-to-one or *convergent mappings*, have been discussed by Hunt and Wanderley (2002), and Hunt et al. (2002) in the context of new musical interfaces. This mapping is implicitly always present through the perceptual interdependence of sound synthesis parameters. It has been applied in the second part of Chapter 5 in order to exploit the salience of one auditory aspect for a combination of two data features. The nature of convergent mappings can be best described by the fact that several sound synthesis units with individual parameters can contribute to a single percept.

The mapping topologies provide a useful concept in the design process. They are however not completely separate categories but overlap, which I will discuss in more detail in Section 2.4. As far as the psychophysical aspects of perception are concerned, we can rely on a good deal of empirical psychoacoustic formulae, which can be used as proactive corrections in the mapping function. In the practical applications, psychoacoustic amplitude compensation has been applied in order to approximate equal-loudness contours. Further in Chapter 5, the *Mel* scale was used to account for the nonlinear relation of frequency and perceived height of a tone.

2.3.1.4 Parameter and Signal Domain

Since parameters influence signal generating processes, the *parameter domain* and the *signal domain* in Figure 2.3 are strongly interlinked. The reason why they are two separate fields is that parameters still relate to data (grey), and the signal provides the substrate for the perceptual domain (blue). This is also the place where the Information Sound Space (ISS) from the TADA framework (Barrass, 1997, page 89) fits in. The ISS is a three dimensional spatial organization of auditory relations. During the design process these relations, potentially exceeding 3 dimensions, need to be matched with the data characteristics. In the parameter and the signal domain, one also needs to find meaningful relations between the data topology and the spatial characteristics of the sonic display. Similar to the perceptual qualities of the sound, its organization in space also needs to reflect data properties like circular, polar or continuous distinctions as left/right, front/back or elevation cues.

2.3.1.5 Fields of Listening, Human Activity

Listening, thinking and tuning are fields of human activity that involve perception or the reflection about it. This is why I wish to discuss here the *sound object* and further the *auditory scene* analysis introduced by Bregman (1994). While the sound object stands for the intentional aspect of listening, and the related design intentions, principles of Auditory Scene Analysis (ASA) contribute to the systematic design framework by helping to analyze how a mixture of sound objects interact and mutually influence each other. Sound objects are embedded in an auditory scene which influences what perceptual dimensions are available as mapping targets. The sonifications in Chapter 3 bring forth distinguishable sound objects. However, since they are not completely isolated percepts, considerations from ASA apply. Similarly, the monitoring application in Section 5.1 can result in an auditory stream with a mixture of sound objects. Since I experienced that the sound stream in the monitoring application reduced the number of mapping targets, I developed an interface that allows to monitor the state of the process as isolated sound objects. In the sonification of 3D shapes in Section 5.2, the amount of data that had to be sonified in a short amount of time had a strong influence onto the available mapping targets, and forced me to develop a synthesis scheme with a small spectral footprint in order to develop sufficient perceptual contrast for a differentiable and structured sound object.

Interaction as *tuning* is an inherent element in the PMSon design process. Various aspects of it have been investigated by Bovermann et al. (2008a,b) by facilitating it through tangible interfaces or by Hermann et al. (2007) by operationalizing some aspects within a numerical optimization algorithm. With respect to the applications in

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

this thesis different mappings have been explored and for some parameter dependent variants the tuning can be left to the user. In exploratory data analysis applications, this can be a hierarchical framework of mappings as in Chapter 4. A similar approach providing a mapping that gives an overview and one that gives more details can be found in the monitoring application in Chapter 5. In Section 5.2 a participatory design approach was used in order to narrow down the mapping options in collaboration with an expert from the target user group.

2.3.2 Listening Intentions in the Parameter-Mapping Design Cycle

As the discussion of the fields in the PMSon diagram has shown, the design process includes creative as well as operationalizable aspects. I want to review here some thoughts from Grond and Berger (2011), which argue for focusing on listening intentions within this framework. These thoughts parallel the concerns raised by Gaver in the analysis-synthesis-listening design cycle, see Section 2.2.2.2.

The arrangement in the diagram reflects the typical challenge a PMSon designer is faced with, which is the problem of representation in creative processes: an experienced sound designer has a preconception of the designed sound *before* it is generated and listened at. This is the reason why we chose the first field in the perceptual domain to be the sound object as an intentional turn towards the sonic substrate, followed by the more operationalizable insights from ASA which finds their phenomenological complement in soundscapes. This imagined result is a projection into the *perceptual domain*, which serves as a strong guiding principle, however the sound has to be heard for further evaluation and refinement. The descriptive framework of the sound object and a sensibility towards the listening modes which foster awareness about indices, meanings and sonic values demystifies the notion of *inexplicable intuition* guiding the design process.

The relationship between listening, thinking and tuning shares strong similarities with *musicianly listening*, in that it is a listening mode with a concrete purpose but one that requires some de-conditioning from first level connotations or causal and semantic denotations a sound may evoke. Since the sonification designers know what to listen for in the sound, a common pitfall in designing a PMSon is convincing oneself that the information in the data is clearly audible. However, the information is not necessarily present to the naive listener. Presupposing that a more neutral listening intention such as *reduced listening* can be cultivated, tuning the mapping can potentially avoid this trap. But only if the perceived sounds can be properly conceptualized through a critical perspective on the act of listening. The participatory design approach in Section 5.2 together with a skilled listener from the target user group also proved to be very helpful.

2.3 Parameter-Mapping, Connecting Data and Sound

Particularly for complex mapping topologies the salience of the sonic result for convergent or divergent mapping dimensions needs to be reassessed. In this situation *critical listening* (asking if all data features are mapped) and *reduced listening* (asking if they are salient in the mixture), are absolutely necessary in the iterative design cycle. Admittedly, *reduced listening* occasionally encounters already established principles from ASA, as it happened to me when designing the dense sound clouds in Section 5.2. The focus on whether the resulting sound provides a substrate for an intentional unit like the sound object helps to assess whether the generated sound is potentially meaningful. Here I experienced that the temporal extension of a sonification plays a crucial role. In this situation, principles of ASA can provide lower boundaries, meaning that if too many data points are rendered at once, little can be distinguished. On the other hand if the sonification becomes too long, it is increasingly difficult to engage in *reduced listening* and ordinary listening modes tend to set in which inevitably changes the information content of the sonification.

2.4 Parameter-Mapping in Vowel Synthesis

Sonic material where parameter-mapping is involved in its synthesis and which also appeals to the embodied listening intentions are vowel sounds. This section presents building blocks for the synthesis of vowel sounds in the programming language SuperCollider (SC3), which has been developed with Till Bovermann and is published in Grond et al. (2011a). Vowels are from a synthesis perspective a fairly accessible type of sound with highly characteristic spectral contours. Vowels are an essential part of human speech and have a natural sonic appeal and high perceptual salience. Vowels engage the listener on many levels and stand out noticeably from mixtures of sounds, which makes them for instance a distinct category in soundscapes.

On the sensorial level their salience is due to voice-selective regions in the human auditory cortex, which were found by Belin et al. (2000). Our capacity to differentiate vowel sounds from early on has been recently shown in newborns by Moon et al. (2013), giving evidence that language specific vowels must have been already perceived and distinguished in utero. The perception of vowels also engages the embodied aspect of listening. Hutchins et al. (2010) and Hutchins and Peretz (2011) find in studies with subjects affected by congenital amusia, that it is easier to match the pitch of one's own voice rather than that of an artificial one. This suggests that the perception of vowels is not only based on the extraction of certain auditory features, but is equally coupled with the motor control necessary for its production as an articulated vocal motor code. This connection of the vocal motor code with sound production can also be found in the extension of the gestural sonorous object by Godøy (2006), see Section 3.1.2. In brief, vowel sounds are very salient as an acoustic signal but also strongly engage the embodied dimension of their production, thereby literally making us hear with our vocal folds and vocal tract. This summarizes in a nutshell the notion of amodal couplings where unimodal input (sound as voice) leads to multimodal couplings (here listening and the activation of motor control for production) as discussed by Tuuri et al. (2009).

A growing number of sonifications originating from various contexts can be found in the literature, which exploit the appeal of vowel sounds. Ben-Tal et al. (2002) used vowel synthesis in stock-market and oceanographic data. Cassidy et al. (2004) used vowel synthesis to support the diagnosis of colon tissue. Hermann et al. (2006a) explored vowel-based sonification for the diagnostics of EEG signals. In the context of human motion display for golf movements, vowels were used by Kleiman-Weiner and Berger (2006).

In this section, I present classes implemented in the sound synthesis environment SuperCollider 3 (SC3) which allow for a convenient and yet flexible synthesis and control of vowel sounds. This implementation was used in this thesis in two applications, in

2.4 Parameter-Mapping in Vowel Synthesis

Section 3.3 and in Section 4.3. The primary goal of these classes is to control the spectral contour from vowels. While natural vowel sounds involve more than spectral contours, recognizable vowel sounds can easily be synthesized with these classes. Most of all, we want to provide with the set of synthesis building-blocks convenient access to the spectral envelopes that constitute a vowel and to get control over spectral contours in timbre space. However, this class also allows controlling all parameters of the vowel constituting formants individually.

The presented classes open access to the following sound design possibilities and advantages for sonification: Vowels offer a continuous and well-controllable dimension in timbre space partly orthogonal to pitch. Vowel sounds refer to ourselves and to our embodied knowledge to make vocal utterances. If the resulting sonification is utterable, it is easy to remember. As mentioned by Hermann et al. (2006a), vowel-based sonification can be directly communicated by mimicking the sound without a complex metalanguage describing the heard sound. The vowel class abstracts from the synthesis level and offers with vowel transitions a single parameter control for an otherwise more complex one-to-many mapping scheme.

As stated in the beginning of this section, PMSon involves the mapping of data to physical, psychophysical or perceptually coherent complexes. These three different mapping targets demonstrate the intricate relation that mapping establishes between the data, the signal and the sound. The resulting consequence for mapping topologies is that they are not necessarily distinct categories and partially overlap. This can be illustrated with formant based vowel synthesis: mapping data features to vowel transition can be understood as *one-to-one* mappings where the mapping target is timbre, as spectral contour. On the level of the individual formants however, it can be understood as the mapping of data features in a *one-to-many* topology to several formants, with individual transfer functions. If the formant filters are applied to an unpitched source the perceived location in the frequency spectrum is a complex interplay between the center frequencies of the dominant lower formants, making this mapping akin to the *many-to-one* topology.

Implementations of sonifications are often difficult to reuse. Functional frameworks, like the sonification sandbox¹, try to address this problem. However, they often do not allow to implement sonification methods like data-sonograms Hermann and Ritter (1999), or the application of flexible mapping functions as in Hermann et al. (2008). We believe that small but flexible building blocks are the best to form a basis for sonification developments, because many of the requirements mentioned above need a sound synthesis environment that can be flexibly scripted. The building blocks allow

¹http://sonify.psych.gatech.edu/research/sonification_sandbox

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

for the flexible control of both, the low-level synthesis parameters and the spectral contour, including the change in brightness and the controlled transition from one vowel to another. The *Vowel* class contains documentation with many examples, is publicly available and can be conveniently installed as the SC3 extension known as *Quark*.

2.4.1 The Class *Vowel*, its Instance, and Control

Sounds that are perceived as vowels have a characteristic spectral contour known as formants (consult Fant (1960) as a standard textbook for formant definition). On the signal level these formants have a *center frequency* (f_c), *bandwidth* (Δf) and *gain* (g). Two formants usually suffice to distinguish the five vowels known in German as *Vokaldreieck* $[i:], [e:], [a:], [o:], [u:]$ as in *bee, bear, bar, bot, boot*. Sets of five formants are used for more natural sounding results, a compilation of which are provided by the Csound manual, covering five vowels in the five registers *bass, tenor, countertenor, alto, soprano*¹. The class *Vowel* contains a library of formants as a class variable, that is initialized with the formants from the Csound manual. The library has hierarchically ordered entries, allowing for complete high to low-level access: `Vowel.formLib.at(\a)` returns a multilevel dictionary with all parameters for the vowel $[a:]$ across all registers. `Vowel.formLib.at(\vowel, \bass)` returns the 15 parameters for the chosen vowel and register as a dictionary holding arrays of f_c , Δf , and g of all formants. `Vowel.formLib.at(\register, \vowel, \freq)` returns the array of frequencies only. Δf and g can be accessed with the corresponding key `\bw` and `\amp`. The inclusion of this formant library is meant to provide parameter combinations where the sonic result is familiar. If one finds a parameter combination with a particular timbre when designing sounds, the class provides a method to save it in a specified file providing a specific name and register. These entries can be loaded and added to the library based on the Csound manual.

The member variables *freqs*, *dBs*, and *widths* of a *Vowel* instance automatically copy the entries of the library, when an instance of *Vowel* is created: `Vowel(\v, \reg)`, with `\v` being the vowel and `\reg` the register. The multi-channel expansion paradigm from SC3 applies to all arguments. Individual formant combinations from within the parameter space of the library can be composed:

```
Vowel.compose([\v1,... \vN,], [\reg1,... \regN,], [\w1,... \wN,]),
```

returning a linear combination according to the weights. Formants can also be defined independently by specifying the parameters manually:

¹www.csounds.com/manual/html/MiscFormants.html

2.4 Parameter-Mapping in Vowel Synthesis

`Vowel.basicNew([f_{c1}, \dots, f_{cN}], [bw_1, \dots, bw_N], [g_1, \dots, g_N]).`

All formant parameters can be directly set through the member variables *freqs*, *dBs*, and *widths*, or their complements *midinotes*, *amps*, and *rqs* returned by methods of the same name. Formants can be added or removed using the *addFormant* and *removeFormant* method.

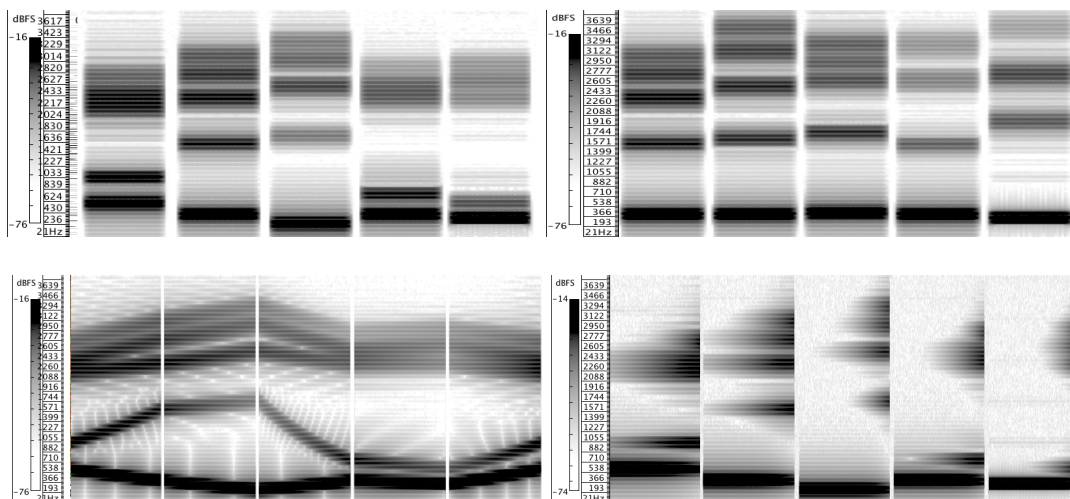


Figure 2.4: **top left:** spectrogram for the vowels [a:] [e:] [i:] [o:] [u:] for the register bass at 70 Hz. **top right:** spectrogram for the vowel [e:] at 70 Hz across the registers *bass*, *tenor*, *countertenor*, *alto*, *soprano*. **bottom left:** spectrogram for the blending between two vowels [a: e:], [e: i:], [i: o:], [o: u:], [u: a:] for the register bass at 70 Hz. **bottom right:** spectrogram for the brighten method applied to all the vowels [a:] [e:] [i:] [o:] [u:] for the register bass at 70 Hz, all from Grond et al. (2011a).

In vowel based synthesis, the transition across vowels is a salient mapping dimension, which has been extensively used in Hermann et al. (2006a). Vowel transition is accessible with the *blend* method; a spectrogram is shown in Figure 2.4. Blending between two vowels $v1$ and $v2$ is implemented as a linear interpolation between the parameter sets *midinotes*, *dBs*, and *widths*. `v1.blend(v2, frac)` morphs from $v1$ where *frac* is 0 to $v2$ where *frac* is 1. All three aspects of a formant (*midinotes*, *dBs*, and *widths*) can be morphed individually, the linear interpolation is done in MIDI notes, but the member variable holds the corresponding frequencies in *Hz*.

The class also allows to systematically change the brightness of a vowel sound by raising the gain of the higher formants through 3 methods: *brightenLin*(b, ref) changes the gain of a formant i by adding to the level of this formant a value that is based on the linear equation: $g_{new,i} = g_i + b \log(f_i) + N_i$. $N_i = (g_{new,i} - g_{ref})$ compensates the change in gain of all formants by adjusting it to the previous gain of the reference formant with index *ref*, b spans from negative to positive real numbers with 0 leaving

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

the formants unchanged. Positive values raise higher formants and negative values lower them.

brightenRel(b, ref) changes the formant gain according to: $g_{new,i} = b g_i + N_i$. All gains are multiplied with b and compensated through the term N_i as in *brightenLin*. b can take on all positive real numbers: 1 leaves the formants unchanged, values > 1 lower the gain of the higher formants and values < 1 brighten the sound. If b is set to 0 all formants are of equal level (0 dB). A spectrogram for this method is shown in Figure 2.4.

The difficulty with *brightenLin()* and *brightenRel()* is that the overall gain can become very high. The method *brightenExp* obeys the equation: $amp_{new,i} = amp_i^b \cdot N$ where the factor $N = \frac{\sum amp_{new,i}}{\sum amp_i}$ corrects the gain of all formants so that the sum of all amplitudes remains constant; this is an approximate compensation in order to achieve constant loudness, yielding satisfying sonic results.

2.4.2 The auxiliary Pseudo UGens

Vowel handles only the formant parameters. It reveals its full potential in combination with the *pseudo unit generator (UGen) Formants* and *BPFStack*, which are designed for additive and subtractive synthesis respectively. *Pseudo UGens* are classes implemented in *sclang* containing the methods *ar*. They instantiate and return a collection of *UGens*. The *Pseudo UGens* discussed here dynamically wrap around each formant of a *Vowel* instance the correct amount of unit generators.

Formants is based on the *UGen Formant*. It generates a set of harmonics based on a given fundamental around a centre frequency. *Formants* takes the arguments *baseFreq*, *vowel*, *freqMods*, *ampMods*, *widthMods* and *unfold*. The first argument - an instance of *Vowel* - assigns the formant parameters of as many *Formant UGens*, as the vowel instance holds formants. The *unfold* flag allows us either to return the sum of all or an array of individual *Formant UGens*, which can then be manually distributed over the display. Additive synthesis of a *soprano [o:]* vowel sound with a 200 Hz fundamental can be realized as: `{Formants(200, Vowel(\o, \soprano))}.play`. The arguments *freqMods*, *ampMods*, *widthMods* allow controlling the spectral flux for each formant individually and are set to 1 by default. They can be either a single modulator (*SinOsc.kr*) uniformly applied to all 5 formants, or an array of modulators modulating each formant individually. Additive synthesis of a vowel sound gives already recognizable sonic results, more natural sounding vowels however are usually achieved with subtractive synthesis through the pseudo *UGen BPFStack*.

In an independent source filter model as described by Klatt (1980) the source is typically modeling the vocal fold and filters model the resonances of the vocal tract.

2.4 Parameter-Mapping in Vowel Synthesis

This can be realized with *BPFStack* which is in its structure analogous to *Formants*. The basic *UGen* of *BPFStack* is a bandpass filter. The argument list differs only in the first argument, where in this case the sound source is passed. Voiced and unvoiced vowels can be realized through *BPFStack* depending on the sonic characteristic of the *in* signal. A vowel with a pronounced pitch of 200 *Hz* would be:

```
{BPFStack(Impulse.ar(200), Vowel(\a, \soprano))}.play.
```

Unvoiced vowels can be realized as:

```
{BPFStack(WhiteNoise.ar(), Vowel(\a, \soprano))}.play.
```

The transition between both provides a salient parameter-mapping dimension as used by Hermann et al. (2006a). It must be noted that with this transition the perceived height in the spectrum for unvoiced vowel sounds depends on the vowel used.

Since the SC3 server is often used without *sclang*, for instance when using the server together with other programming languages, it is desirable to include many vowel related parameters and methods in the synthesis definitions (*SynthDef*). This means that the sound design can be conveniently made in SC3, but the actual application only requires the compiled *SynthDef* to be invoked on the server. Most methods from the *Vowel* class can be used without the SC3 language. One or several instances of vowels can be constructed through the *compose* method within a *SynthDef*. Also the blending and brightening methods can be used within *SynthDef*, since none of them contain flow control statements that would not be properly executed in the DSP chain after compilation. If however *sclang* is used an instance of *Vowel* has the convenient method *addControls* which creates control busses within the *SynthDef*. The *asKeyValuePair* message distributes the data structure of a *Vowel* holding the formant information to these created control busses.

2.4.3 Ways to Use the Spectral Envelopes

Some synthesis possibilities discussed below require the amplitudes of partials as arrays. The amplitude of any frequency under the spectral envelope is accessible through the method *ampAt*, which also takes ranges of frequencies. For each formant the transition steepness can be modeled with an exponent as a function of the distance to the centre frequency of the formant. The convenience method *plot*, which is based on *ampAt* renders a visual display of the spectral envelope using the exponent arguments to control transition steepness. Resulting plots are shown in Figure 2.5. Extracting partials under spectral envelopes gives in combination with the *UGens Klang / Klank* or *DynKlang / DynKlank* very flexible but computationally more expensive synthesis options:

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

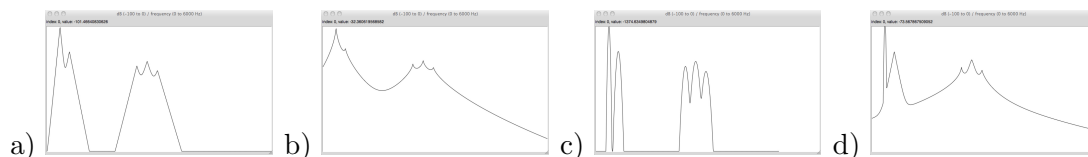


Figure 2.5: Typical spectral envelopes as *frequency / dB* value pairs from Grond et al. (2011a). These graphs are generated through the convenience method *plot* based on *ampAt*.

DynKlang holds a bank of sine oscillators, which can be dynamically changed after launching the sound synthesis process. Arrays of amplitudes for arbitrary selections of frequencies can be extracted

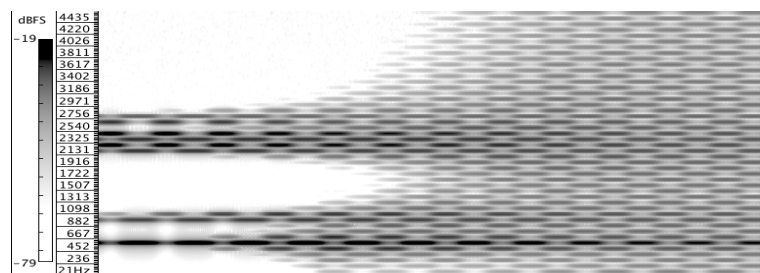


Figure 2.6: Spectrogram from Grond et al. (2011a) of the alternation of even and odd harmonics for *[a:] tenor* at 70 Hz together with a decrease in the transition steepness of the formants.

through *ampAt* and can be used as an argument for *DynKlang*. This allows changing dynamically between even and odd harmonics. The transition steepness of the formants can also be conveniently manipulated as shown in Figure 2.6. By using *DynKlang* the overall gain of all oscillators can be limited by applying the *normalizeSum* method onto the amplitude array.

In a source filter model, pitch mostly depends on the excitatory signal. In sonic interaction design a new trend of augmented auditory objects has been set by Bovermann et al. (2010) where contact sounds are treated in near real time through filters. Augmented auditory objects pair the appeal of causal sound-source oriented listening, with perceivable sonic values such as location in the frequency spectrum. Since contact sounds are mostly unpitched, this dimension remains however inaccessible for sound design. This restriction can be partly overcome using *DynKlang*, which simulates the resonant modes of an object. *DynKlang* configures its resonators through an array of frequencies, amplitudes and ring-times.

This argument can create a perceivable pitch (as a tonic mass with a spectral contour, beyond single frequency filtering) even if the excitatory signal has no pronounced pitch itself, such as in attack or friction sounds. A prototype of an auditory augmentation using surface friction as sound source is shown in Figure 2.7. The arguments of the resonant filters are extracted from vowels using *ampAt*. In the spectrogram of Figure 2.7, the formants as well as the partials of the resonators can be recognized.

2.4.4 Sample Sonification Applications

2.4.4.1 A sonic colormap

Based on the *Vowel* class, a *one-to-many* mapping sonification is presented in the following example. The dataset for this sonification is the z time series of the Roessler (1976) attractor which is a nonlinear dynamical system exhibiting

chaotic behavior. In such a system small deviations can grow exponentially, a sonification should hence make even small variations noticeable at any range they occur.

The spiking nature of the z variable of this system requires a logarithmic scaling of the mapping. For the given dataset, a salient mapping over a wide range is particularly important when looking at the data distribution:

10 % of the data points are found in the lower 0.015% range of the amplitude; 50 % are in the lower 0.037% of the amplitude range; 90% of the data points are within the lower 2.5 % range of the amplitude. In order to make deviations stand out across different amplitude ranges a *one to many* mapping approach was used to control various spectral synthesis parameters using *Vowel* and *BPFStack*, as presented in Grond et al. (2011a):

- the 0 to 30 percentile is mapped to an Δ gain of 90 dB,
- the 20 to 50 percentile fades between unvoiced and voiced,
- the 40 to 70 percentile blends between the vowels a and i ,
- the 60 to 90 percentile changes the pitch from 82 Hz to 116 Hz,
- the 80 to 100 percentile brightens the vowel.

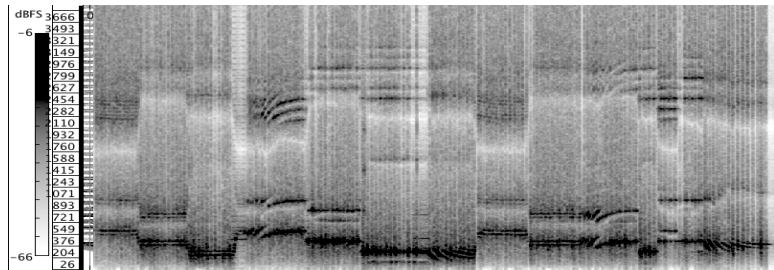


Figure 2.7: Spectrogram from Grond et al. (2011a) showing the transition of partials and formants of an auditory augmentation of interaction sounds creating noticeably pitched vowel sounds.

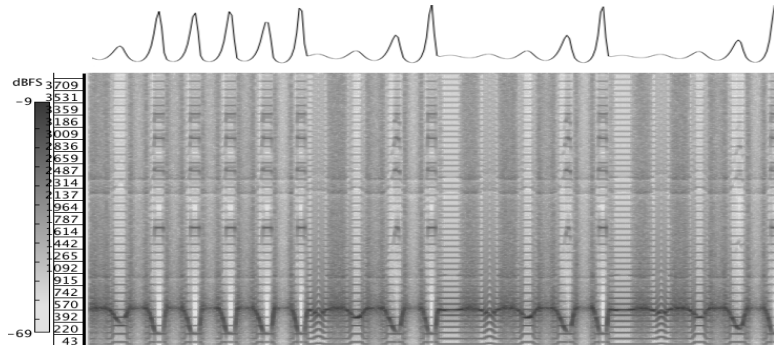


Figure 2.8: Spectrogram of the *one to many* mapping from Grond et al. (2011a), the logarithmic time series of the z variable is on top.

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

Since ranges overlap, the evolution of the sonic aspects controlled through the parameters build a single sound stream. Figure 2.8 shows the spectrogram of the resulting complex sound. The unvoiced-voiced transition is visible as the emerging partials. The vowel transition can be best seen in the changing first formant. The changing location of the partials corresponds to the pitch variation and the brightening of the sound lifts the gain of the higher partials. Note that the unvoiced-voiced transition was fully completed before the pitch change, which was preceded by the vowel transition. This order was chosen because only for pitched or sung vowels the vowel transition does not alter the perceived height in the spectrum too much. By using the methods of the *Vowel* class together with *BPFStack* this mapping was easy to implement. The range of most parameters such as the brightening of the vowel could be kept small, the perceived sum of all effects gave still a highly differentiable result. The sound was not necessarily utterable, but the mapping scheme helped to exploit the many salient features of vowel sounds. The unvoiced-voiced transition enhanced the auditory contrast by adding characteristics that were reminiscent of consonants.

2.4.4.2 Vocagram, a data sonogram with *Vowels*

The second sonification example based on vowel synthesis is a data-sonogram. This MBS, can be - apart from the data preparation step - implemented as an PMSon and hence offers the possibility to combine MBS with perceptual-based mapping by using the vowel timbre space. Figure 2.9 shows a screenshot of the graphical user interface. Any position within the dataset can be selected through a 2D Slider. Pushing the mouse button releases the radial virtual shock-wave and plays the sounds of each data point, when they are excited by the wave. The colors indicate the categories within the data-set, labels correspond to the vowels [a :] *blue*, [o :] *black*, [i :] *red*. The position of each data point relative to the center of the virtual shock wave was rendered in the stereo panorama.

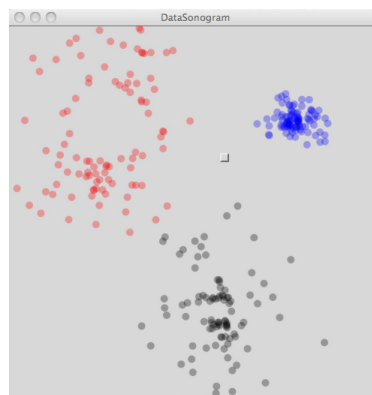


Figure 2.9: GUI of the data-vocagram from Grond et al. (2011a).

Figure 2.10 shows a spectrogram of the sounds of subsequent explorations of the dataset. Formant structures corresponding to the data clusters can be recognized. Distance was additionally mapped to decreasing pitch, which can be identified as glissandi-like movements of the partials. In most positions within the dataset, all data clusters and their positions could be well identified. Sometimes they came all in succession, which resulted in an utterable articulation with an unusual texture due to the many attacks. In many cases

the clusters were excited at the same time, giving the impression of two or three simultaneous voices. Here utterability was restricted by the complex data structure and not a question of sound design. However, the spectral gestalt of vowels was still useful to engage with the sounds and helped to separate the streams.

The data sonogram with vowel synthesis was also implemented as partitioned convolution kernels for interaction with contact sounds, similar to the augmented auditory objects from Figure 2.7. More precisely, the sonification

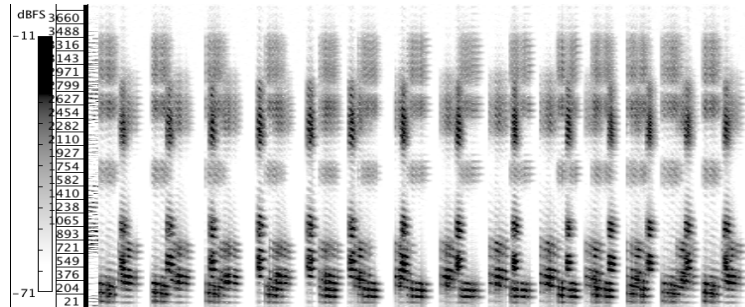


Figure 2.10: Combined spectrogram of the stereo channels taken from Grond et al. (2011a), showing the sounds of different shockwaves triggered in the data vocagram (approx. 30 sec).

of each cluster was distributed over four convolution kernels (giving 12 in total for 3 clusters) in order to be able to distinguish left and right as well as to some degree front and back. The sonic result depended on the interaction sounds but always gave the strong impression to engage the connotative listening mode of the object-sound coupling field as discussed in Section 2.2.5. I informally asked others to experiment with the interface with the sounds rendered through headphones. The feedback I got was often that it feels like a jaw harp, particularly when generating the excitation through tongue clicks.

2.4.4.3 Concluding Remarks

Although not all vowels had the same natural appeal, working with the *Vowel* class together with *Formants* and *BPFStack* and made it very easy to control perceptually salient synthesis parameters. This was due to the transparent abstraction from low-level synthesis units giving convenient high-level access to manipulate the spectral contour. Beyond the convenient synthesis of vowel sounds based on the formant data from the Csound manual, the very flexible design of these software building blocks also makes more subtle perceptual aspects accessible. Spectral flux, for instance, in the regions of individual formants could be easily realized. This offers further potentially information carrying perceptual dimensions. The *mapping problem* from Section 2.3, however, demands from the PMSon designer to carefully organize these mutually dependent perceptual dimensions, as demonstrated in the first *one-to-many* mapping example of the sonic colormap in Section 2.4.4.1.

2. LISTENING, INTERACTION AND PARAMETER-MAPPING

3

Mapping in the Sonification of Ancillary Gestures

This chapter presents two display prototypes for the sonification of ancillary gestures. The basis of the sonifications are data from clarinetists' postures captured with a VICON motion tracking system. Before meaningfully sonifying posture data, data selection and data reduction are important preliminary steps. Different data reduction approaches are applied in this chapter. While sonification can be a powerful means for representing the dynamics of the posture data, i.e their temporal structure, it remains difficult to interpret the sounds with respect to the actual postural configuration or concrete movement they represent. This is why the sonifications were developed and studied with a simple complementary visual display featuring an information content based on the same data as the sonification. In two different applications, we studied the interplay between the sonification of the movements and its visual counterpart.

The work on the first part presented in this chapter was conducted in Montreal in the Input Devices and Music Interaction Laboratory (IDMIL) under the supervision of Prof. Marcelo Wanderley and Dr. Vincent Verfaillie. It was completed in Bielefeld in the Ambient Intelligence group and the results were presented at the gesture workshop 2009 in Grond et al. (2009). The second part was conducted in Bielefeld in collaboration with Dr. Hendrik Kössling and Nick Kasajanov.

3.1 Movement and Sonification

Advanced recording and simulation possibilities create an ever increasing amount of 3D movement data, which are most often investigated through 3D visualizations of moving

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

points or models. This approach is self evident, since the human visual and cognitive system seems highly adapted to perceive and interpret human motion.

Sonification offers a complementary inspection technique, which is beneficial for several reasons: sonification is ideal for representing dynamic patterns in multivariate data sets with complex information. Sound requires neither a particular orientation of the user nor directed attention. The omnipresent nature of sound has the potential to direct visual attention. In many applications such as skill learning the eyes are already occupied with a specific task. Here sonification can offer a channel to deliver additional or complementary information. In the display type presented in this chapter, where audiovisual content is created, the interplay of both senses needs to be taken into account.

Movement sonification has been mostly applied to improve or facilitate movement and motor control in sports, early sonifications can be found with Effenberg (2005). A specific application has been developed for an audio-only environment in the system *Acoumotion* by Hermann et al. (2006b). Hummel et al. (2010) applied sonification to improve the practice of German wheel performers. Eriksson and Bresin (2010) designed a sonification system in order to improve running technique. Godbout and Boyd (2010) presented a corrective sonic feedback for speed skating which was applied successfully for correcting faulty movement sequences. Sonification was successfully applied to synchronization of rowers by Schaffert et al. (2012).

It is common to many of these examples that the data representing the movement are scalar to low-dimensional time series often based on real-time accelerometer readings or similar sensors. An exception is the sonification of swimmer movements based on multivariate data with the attempt to make them available in real-time, see Hermann et al. (2012). Most of these movements like running, rowing and swimming are repetitive coordinated body movements, and hence the power of sonification lies in the possibility to detect deviations relative to a repeating sound pattern, hence the sonic information does not need to be assessed in absolute terms. A good overview and source for further reading for the field of movement sonification in sports is given by Höner et al. (2011).

3.1.1 Ancillary Gestures

Ancillary gestures of instrumentalists are somewhat different compared to sport movements. The benefits of sonification are still of potential interest as a complementary element to the visual display, for exploratory gesture analysis. Here the temporal structure of the movement and the awareness of moving limbs relative to other concurrent movements are of interest.

For clarinet players, movements directly involved in the sound production, such as of lips and fingers, are effective gestures¹. Other movements like weight transfer and body curvature are expressive movements or ancillary gestures, which are defined as those body movements which are not directly involved in the sound production. These types of movements are omnipresent in musical performance and have been discussed by Cadoz and Wanderley (2000). As shown by Wanderley et al. (2005) and Nusseck and Wanderley (2009), their importance lies in the fact that they tend to align with musical phrases in the score. These movements show consistent patterns across various levels of expressiveness (compare Wanderley (2002)) and are an integral part of the instrumentalist’s performance. What makes ancillary gestures different from most movements in sport is that other than repetitive coordinated body movements, they do not repeat in regular intervals. Although they are consistent with musical phrases, preceding postures are likely to differ and hence create a different sonic context in which they are embedded.

Early sonifications of ancillary gestures have been developed by Verfaillie et al. (2006) and Savard (2008) both working on clarinetists’ movements. Similar to Verfaillie et al. (2006), Goïna and Polotti (2009) employs Rissetts infinite glissandi or Shepard tones in their mapping approach in order to represent continuous directed movements. Winters and Wanderley (2012) give an overview over expressive movement sonification which also discuss the results from Grond et al. (2009). They argue that the peculiarity of these kinds of movements is that they are “*not strictly goal-oriented and highly idiosyncratic*” and that they “*should be evaluated based upon their capacity to convey information that is relevant to visual perception and the relationship of movement, performer and music.*”

3.1.2 The Gestural Sonorous Object

Godøy (2006) proposed the term of the *gestural-sonorous object* based on Schaeffer’s notion of the sound object, extending it towards gestures and movement. This concept is of interest in this chapter for the following reasons. Godøy argues that the term known as *motor equivalence* from the field of motor control suggests thinking of movement on a similarly abstract level as Schaeffer’s sound object. Many sound objects, which are sonically different, are produced through the same movement intention on different instruments. This can be perceived as one schema through a listening intention on the perceiver’s side that matches a perceived movement intention. Godøy states that the movement intention is tied to motor imagery, which is based on the implicit

¹ Effective gestures have been sonified for skill training in the musical context like bowing the violin by Großhauser and Hermann (2010).

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

knowledge of biomechanics and motor control constraints. He concludes that we have kinematic and dynamic images of effort, chunking, and coarticulation, that resonate strongly in our images of sound¹.

Godøy gives an interesting overview of the meaning of the sound object, pointing to the fact that its status as an intentional unit helps us to segment the continuous stream of sensory impressions. The principles on which this segmentation is based are according to Schaeffer those of *articulation* and *stress*, where the first is the energetic event that interrupts and structures the perceptual stream, and the second refers to its sustainment. Godøy cites Schaeffer's notion of *articulation* as "*breaking up the sonorous continuum by successive distinct energetic events*". *Articulation* as the segmentation of the perceptual stream suggests a promising basis for the mapping of the sonification based on the energy of the movement.

Godøy further exemplifies *stress* through the place that vowels hold in speech. This example provides a conceptual basis for the second mapping that we applied, which is based on vowel synthesis². Godøy (2006) emphasizes that there is plenty of evidence that perception is more than an abstract processing of sensory data but rather always linked to some sort of simulated action. For the perception of speech-related sounds like vowels, this means that the perception is strongly linked to our own capacity of producing these sounds. Particularly with vocal sounds, the hearing intention is shaped through the heard intention, based on *stress* and *articulation* of the sound object (compare (Chion, 1983, page 27)) Related concepts with respect to vocal sounds have been discussed for the design of user interfaces by Tuuri (2010). The aspects of the *gestural-sonorous object* provided a guiding principle for the parameter-mapping design described later in this chapter.

3.1.3 Audio-Visual Displays

When developing sonification together with visualization, the interplay between both sensory modes in the display needs to be taken into account. Here the following phenomenological and psychological findings provide the context that informed the display design.

The audio-visual medium has been extensively covered by film studies. The notion of *synchresis* was coined by Chion (1994, page 58) (also compare Chion (2009, page 492)). This neologism is made up of the words *synchronous* and *synthesis* and describes

¹A simple example of this universal images of movement, articulation and effort is the animation of speech through a hand.

²The mapping approach is also inspired through the notion of the energymotion trajectory in the spectromorphology from Smalley (1997), to which Godøy refers in the development of the *gestural-sonorous object*, see Godøy (2006).

the irresistible connection that is spontaneously established between a short sound and a simultaneous short visual stimulus. This connection enforces a listening mode, which Chion (2009, page 471) termed *causal listening*. Recent psychological findings by Parise et al. (2012) support the causal connection of auditory events that correlate with visual ones, suggesting that the human neural apparatus in fact performs some sort of correlation between the sensory data.

Flückiger (2001, page 126) introduces the term of the *unidentified sound object*, which is a sound whose cause is neither resolved through context nor through a corresponding visual stimulus. In film, if sparingly used, a *unidentified sound object* can be the cause for discomforting emotions, due to the uncertainty it creates. Since the sonification of movements cannot be supported through a context of everyday experience, the concept of the *unidentified sound object* is mostly relevant with respect to the lack of a visual cause.

Some observations that have been made in films have equally been studied in psychology, one of which being the *ventriloquism effect*. This effect describes the phenomenon that we attribute sounds of speech to moving lips even if both, sound and lips, are not located in the same spot. In film for instance the speaking actor might enter the scene from one side while the words come from the center speaker. This effect of attributing the sound to a visual source has been termed “*the spatial magnetization of sound by image*”, see Chion (2009, page 491). For the audio-visual display of ancillary gestures in this chapter, we were interested to find out with eye-tracking if different vertical sound labeling of the limbs has the potential to influence the attention.

A review of the research field between auditory perception and vision by Spence and Soto-Faraco (2010), summarizes the interplay of audio-visual stimuli, which have the potential for *enhancement* of the sonic part, *extinction* of it, and for creating auditory *illusions*. An early example for *enhancement* is given by Sumby and Pollack (1954) where the visual perception of lip movements enhances auditory speech perception in noise. Here causal listening can be interpreted as searching for sensory correlates in order to find complementary information. The integration of complementary information can lead to auditory illusions, the best known example of which is the McGurk effect McGurk and MacDonald (1976), where paradox effects of perceived lip movement onto speech perception are found. In the sonification of ancillary gestures in an audio-visual display by Savard (2008, reported in personal communication) effects similar to the *ventriloquism effect*, have been experienced which will be elaborated later.

According to Chion (2008, page 563) the listening cinema is: *a bad listening cinema* by giving examples where sounds are misheard and wrongly attributed, which is only later corrected by binding the sound to its rightful - because visible - object in the

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

narrative. Similarly Flückiger (2001, page 143) states that in the combination of film and sound, aspects individually perceived in both senses can be either highlighted or narcotized. The narcotization of the sense of listening corresponds to inhibition, or extinction, which are possible psychophysical effects as discussed in Spence and Soto-Faraco (2010). The authors further discuss a body of literature on crossmodal perceptual load, and supramodal attention split, which is as the authors state still contradictory. Some accounts are given that suggest that vision can decrease auditory attention. The emphasis of the acousmatic condition as a prerequisite to approach the sound object gives indirect evidence that the visual sense can lead the focus away from listening.

For the given data substrate of changing postures, it is interesting to find in the field review by Spence and Soto-Faraco (2010), that dynamic stimuli facilitate stronger crossmodal grouping than static stimuli. Questions of temporal and spatial perception in multimodal settings involving auditory display has been studied by Guttman et al. (2005), with the finding that particularly in rhythmical multimodal stimuli, the auditory sense overrules vision.

Spence and Soto-Faraco (2010) elaborate that in a multimodal stimulus, it is difficult to assess the contribution of one sense to the combined percept, hence artificial situations are created in experimental setups, which induce inter-sensory conflict thereby studying how each sensory modality deals with discrepant cues about a given object property. The authors also point out that under everyday conditions our senses have to handle multiple sensory inputs and that deciding which visual stimuli should be bound with which auditory stimuli becomes a non-trivial problem.

3.1.4 Research Questions

In this chapter, two audio-visual display prototypes are investigated. Our goal was to find out, how an audiovisual display inspired by the *gestural-sonorous object* could be built for the movement of ancillary gestures, taking into account the intricate relation between sound and vision. Since the movement data were kinematic, we were mostly interested in the articulatory aspects of chunking and less in mental images of effort. For the first display prototype we wanted to find out if sonification helps the users to organize the segmentation of movements in a dynamic display of ancillary gestures.

As stated in the beginning, movement data are multivariate datasets and data reduction needs to be applied before they are mapped to sound. One question was therefore if global data preparation, specifically reduction, has a strong impact on the audio visual display. For the first prototype, we conceived two different data preparation steps. One is a direct mapping of marker velocities to sound, and the second involves a

principal component analysis of the movement data. We were interested to find out if subjects would have a preference for one of these two data-preprocessing steps. In order to investigate this question, we asked subjects to annotate the clarinetists' performance represented by various combinations of the resulting uni- and multimodal displays. The results were presented in Grond et al. (2009).

The second question was if different sonification mappings have the potential to guide the visual attention towards different aspects in the data. For the study of this question in combination with the visual display, we made use of eye-tracking in order to find out to which aspects of the movement data the user's focus is directed. We mapped the angular velocities to sound and presented it together with the visual display. In order to assess the audiovisual display we compared it with the eye-movements of a visual-only display of the ancillary gesture movements.

3.2 Annotating Sonified Ancillary Gestures

This part of the chapter describes the first sonification prototype of ancillary gestures. Two PMSon were developed, one based on the marker velocities, the other one on the principal components of the body postures. The effect of the combined audiovisual display was studied in a free segmentation task of the combined and separate stimulus. The sonification was based on VICON motion tracking data of clarinetists. Posture changes of clarinet players exhibit only ancillary gestures, since the effective gestures are mostly confined to finger and lip movements. All clarinetists played an excerpt of Brahms' Sonata for clarinet op. 120 no 1. The text in the following section is based on Grond et al. (2009).

3.2.1 Motion Tracking Data

The data of the clarinetists movements were captured in motion capturing sessions in previous projects at the Input Devices and Music Interaction Laboratory (IDMIL). All clarinet players were advanced instrumentalists. The movement data were recorded with a VICON 460 system. The standard plug-in-gait model used (Ferrari et al. (2008)) provides 38 marker positions. This set of markers gives a global description of the body posture.

From this set of data, we removed redundant channels. We decided to apply sonification to the posture information, which corresponds to the three data channels x , y and z , the Cartesian coordinates, and their derivatives. This choice was motivated because we wanted to sonify data which could potentially also be seen in a simple visual representation. The VICON dataset was reduced to markers shown in Table 3.1 some of them were combinations derived of originally measured data (for details compare Grond et al. (2009)).

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

| body part | left | middle | right |
|-----------|--------------------------------|--------------------------------|--------------------------------|
| head | front | back of the head | front |
| spine | | neck C7 T10 end of spine | |
| arms | shoulder elbow arm wrist | | shoulder elbow arm wrist |
| legs | hip knee ankle | | hip knee ankle |




Table 3.1: On the left: the reduced dataset of 18 markers. On the right: clarinetists in the VICON motion capturing system. The table and the images are taken from Grond et al. (2009).

3.2.2 Data Preparation and Preprocessing

Before the data reduction described above, the whole dataset was centered between both feet. This was accomplished for each time frame by moving the center of origin of the coordinate system to the point determined as the middle between the left and right toe. In almost all cases, with the exception of one recording, the toes of the clarinetists stayed in one place during the whole performance. Hence there was not much movement in the toe marker, which would have been discarded.

The whole movement represented through the dataset was dominated by the left and right as well as the back and forth movements; both are generally referred to as weight transfer. In the PCA transform, the global influence of the weight transfer became evident as the first and second components consisted almost exclusively of this movement aspects. In a similar way the weight transfer was present in the velocity sonification approach, when mapping marker speed to sound. Therefore we decided to remove this component from the data set as this would not allow identifying otherwise audible aspects in the movement sonification. For details about the moving center of mass removal, see Grond et al. (2009). After this data preparation step, all other movements were not overshadowed and could audibly stand out.

3.2.3 Sound Synthesis and Mapping

Based on the Design Space Map introduced by deCampo (2007) the recording rate of 100 Hz suggests as an appropriate strategy a continuous parameter-mapping sonification. The sound synthesis aspect in the sonification design was guided by the following considerations, which were influenced by the concept of the *gestural-sonorous object* of

Godøy (2006): the aim was that the continuous sonification should automatically lead to acoustic articulations which would allow for a perceptually natural segmentation of the sound, such that it can be meaningfully connected to the movement patterns.

In addition, the sonification of movements from different parts of the body between head and toes should be distinguishable. We wanted to design a single auditory stream, so that stream segregation would not be an interfering task for the test subjects when listening to the stimuli. We therefore chose to use a source filter model as a modular sonification unit, which could be applied to all markers or PCA components, and which would well blend together. This modular unit consisted of white noise filtered through a resonant filter based on Steiglitz (1994), which is implemented in SC3 as the *Ugen Resonz*.

$$s(t) = \sum_{i=1}^n H_{Resonz}(\eta_i(t); f_i, rq, g) \quad (3.1)$$

The resulting sonification $s(t)$ is the sum over all n sonified data-features, the 18 selected markers. H_{Resonz} stands for the resonant filter with the frequency f_i , the gain g , and the bandwidth, which is specified with the reciprocal q -value rq . As a source $\eta_i(t)$ a signal of white noise is fed to the filter. To address the frequency loudness dependency we used basic psychoacoustic amplitude compensation. For the details of the amplitude compensation we refer to the implementation details of the *Ugen AmpComp*. These resulting sounds of filtered noise merged nicely into one sound stream. However, the varying amplitudes of the different resonant frequencies f_i each representing one of the 18 markers could be audibly distinguished. This sound synthesis scheme was used for both the marker velocity and the PCA component sonification. The next section describes how the motion data are mapped to the synthesis parameters.

3.2.3.1 Velocity and PCA Sonification

In this mapping, the velocity of the 18 selected and derived markers described above was mapped to the sound parameters, as compiled in Figure 3.1. The mapping followed a one-to-one, as well as a one-to-many mapping paradigm.

The movement data contained some noise and hence they were smoothed with a rectangular window of 5 samples, before the velocity was calculated for each marker. Starting at the ankle, frequencies from 150 to 4000 Hz at the head markers were assigned to each marker, as audible labels on a log scale. The varying gain corresponded to the velocity and was mapped exponentially between 0.001 and 1. This mapping

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

was informed by the idea that the movement data would naturally express themselves following the notion of articulation from Godøy (2006). Additionally, the velocity of each marker was mapped to a center-frequencies modulation with $\pm 5\%$. This created a spectral flux which helped to draw auditory attention to the respective frequency band. Further, the articulation was enhanced by an exponential mapping of the rq of each resonant filter between 0.001 and 0.1.

The second mapping was based on the principal components of the data, which consisted of $3 \cdot 18 = 54$ features over a complete performance. The center of mass corresponding to the 54d vector of the mean posture has been subtracted. Then, we

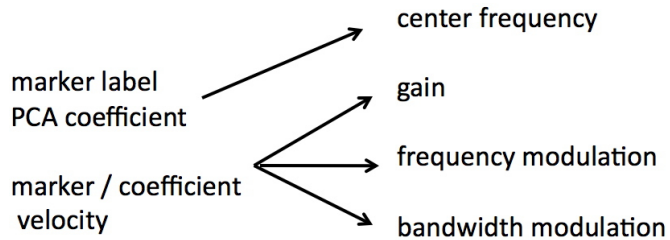


Figure 3.1: Mapping data features to signal parameters

computed the data set covariance matrix $C = \frac{1}{N} \sum_{\alpha} x^{\alpha} x^{\alpha T}$; the principal components hence represented the axis along coordinated activity. The first 6 coefficients with the largest eigenvalues of C covered approx. 85 % of the data set variance. Components with smaller eigenvalues than the first 6 represented very minute movement details; they could not be distinguished acoustically, and were also visually very difficult to identify. Similarly to the velocity mapping, the PCA coefficients of the principal components were mapped to signal parameters as compiled in Figure 3.1. The 6 coefficients were exponentially mapped between 300 - 2000 Hz and the frequencies and the range were chosen to yield an acoustically rich result, meaning that the articulation of each coefficient could be well perceived. The change of the coefficients was mapped in a similar way like the marker velocity to gain exponentially between 0.001 and 1. The frequencies of the filters were modulated with $\pm 5\%$ around the assigned center frequency controlled by the velocity of the time coefficients. The rq of the resonant filters corresponded to the principal components exponentially mapped between 0.001 and 0.1. Although both, the velocity and the PCA sonification were the result of the sum of eq. 3.1 over all sonified data features, their sonic quality was different due to the different durations of the sonic articulations of the data features.

3.2.3.2 Stimuli Selection for the Experiment

Particularly for the PCA sonification it was important to calculate the time coefficients for an extended period of time, so that the covariance matrix represented the longtime average. Otherwise each sonification would have represented only the unique moment of the selected time interval which would not have given comparable stimuli segments in a sense that would have represented consistently the same principal directions. Hence these sonifications were applied to the data set of the complete performance of 3 clarinetists. These performances were selected because they featured noticeably different movement patterns of ancillary gestures. Apart from a considerable amount of weight transfer movements, Clarinetist 1 and 3 exhibited various pronounced ancillary gesture patterns. On the contrary, clarinetist 2 was mostly standing still, making occasionally rapid movements with the elbows and the arm wrist.

3.2.4 The Audio-Visual Display of Body Movements

For the design of the audio visual display we considered the aforementioned multimodal effects, i.e that the percept is not merely a superposition of moving image plus sound. In the display, the user has to integrate the multiple sensory inputs, which is as discussed earlier a non trivial task. In order to bind both sensory modes into a combined percept, *causal listening* or the "searching" for a visual correlate can lead to unwanted results as the following experience shows: Savard (2008) constructed a framework for ancillary gesture sonification for clarinetists based on similar data. He reported in personal communication that in a first trial, subjects had to judge sonification videos from the clarinet players, in which finger movements were clearly visible. Due to this visual detail some test subjects wondered if the finger movements (effective gesture) were the cause of the sound. Interestingly, the sonification was based exclusively on ancillary gestures such as weight transfer and body curvature, however these movement types were visually less noticeable and were hence not attributed to the sound.

Through this experience on the interplay of sound and image, we asked whether different sonifications and visualizations change the way how the audio together with the visual modality can be efficiently integrated. We decided to give the test subjects an open task through which we wanted to find out if the presentation of ancillary movements changes how subjects interpret them in combinations of audio and visual displays.

The visual part of the display was an abstract depiction of the clarinetist. This abstraction is not only necessary in order to avoid a specific acoustic expectation in the subject, but also to exclude the representation of effective performer gestures, such as

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

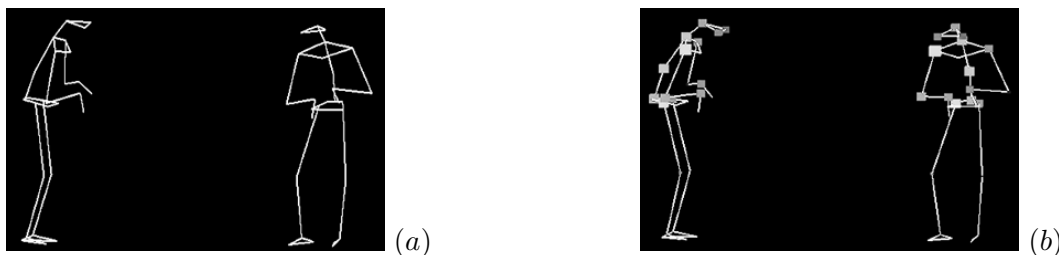


Figure 3.2: Abstract visual representation of the clarinetists: (a) shows a simple stickman (b) the enhanced stickman where the velocities of joints are additionally highlighted with glyphs in the form of red cubes. Both taken from Grond et al. (2009).

the movement of fingers and lips. The subjects were asked to rate the performance of all three clarinetists with respect to which sonification method matches best with the visual representation.

3.2.4.1 The Visual Display of Body Movements

The aforementioned challenges of audiovisual stimuli led us to develop a display which showed exclusively ancillary gestures¹. In Figure 3.2, the visual part of the display is shown. This consists of an abstract stick-figure in two positions: one in profile and one from the front; both omitting details such as finger and lip movements.

We conducted small and informal preliminary tests and we received good feedback for this display design with respect to consistency between sonification and image. Some participants still found it difficult to identify which part of the body was responsible for which aspect of the sound. This is why we added for the velocity sonification, a red glyph in the form of a cube to each sonified joint. These glyphs changed - according to the joints' velocity - its size and color (through the variation of the alpha channel). This addition gave a visually similar information, with respect to the time structure from the data, as it was mapped to the signal of the velocity sonification.

3.2.5 Evaluation of the Display

In a psychophysical experiment we studied how the audiovisual display would influence the perception of the ancillary movements.

We used subjects with a diverse background and designed an open task. Specifically we asked the subjects to identify and detect events in stimuli which stood out for them. The subjects were asked to perform this task for selected sequences of movements of the three clarinetists. All stimuli from the three performers were taken from time

¹The visualization in `SuperCollider3` was implemented in `SCgraph` Schmidt (2007)

3.2 Annotating Sonified Ancillary Gestures

intervals which all represented the same musical phrase, which consisting of 4 distinct melodic units. Each stimulus was presented to the test subjects 9 times. During each presentation, they were asked to attentively look at and listen to the display and to mark events per mouse click in real time. The subjects were told that it was up to them which and how many events they could mark in the segment. The test subjects were instructed that the first two runs are considered as test runs. They were also encouraged to repeat the mouse clicks that marked their selection of events consistently in the subsequent 7 runs. At the end of all trials, a questionnaire about the different stimuli was presented to the test subjects, which they were asked to fill out. In this questionnaire they were also asked about their musical experience.

The combination of the stimuli of the two visual representations (with and without cubes) and the two sonifications (velocity and PCA) gave in total 7 different stimuli, which were all presented to the subjects. Table 3.2 compiles the stimuli organized according to the visual and auditory aspects.

| id | sonification | visualization | acronym |
|----|--------------|----------------------|---------|
| 1 | velocity | - | A1 V0 |
| 2 | velocity | stick-figure | A1 V1 |
| 3 | velocity | stick-figure + cubes | A1 V2 |
| 4 | - | stick-figure | A0 V1 |
| 5 | - | stick-figure + cubes | A0 V2 |
| 6 | PCA | - | A2 V0 |
| 7 | PCA | stick-figure | A2 V1 |

Table 3.2: Table of stimuli, taken from Grond et al. (2009), showing the different combinations of visualizations and sonifications. In the text, the acronym will be used to denote the stimuli.

For the PCA-based sonification we decided to show only the simple stick man and omit the highlighting cubes, since their visual appearance changed according to the individual marker velocity those aspects would not have corresponded to the sonification. The order of the stimuli was randomized for each test subject; the randomization made sure that the 3 clarinetists were interleaved, and no performance pattern was shown twice in a row. The velocity-based sonification gave a very structured sound for the performances of clarinetists 1 and 3. The sonification appeared to be easy to connect with the visualization. The PCA-based sonification for the same selected regions gave a less structured impression and was hence more challenging to connect with the visualization. The situation was reversed in the case of clarinetist 2, where the PCA sonification gave more structured sounds.

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

3.2.6 Experimental Data Evaluation

12 subjects participated in the experiment, their age range was from 22 - 33. 11 were male, 1 female, and 7 of them were playing an instrument. We analyzed the data for each clarinetist individually, since the movement patterns differ strongly. This resulted in different sonifications for the velocity as well as for the PCA sonifications.

Average click frequency: As a first step in the evaluation we looked at the average number of clicks for each stimuli, that was given by all the subjects for the identified events. Figure 3.3 compiles the results. All the subsequent figures arrange the results for the different stimuli in the following order: the audio-only condition is shown on the left (A1V0, A2V0), the visual-only condition on the right (A0V1, A0V2), in the middle there is the combined stimuli (A1V1, A1V2, A2V1). Please refer to Table 3.2 for the meaning of the acronyms.

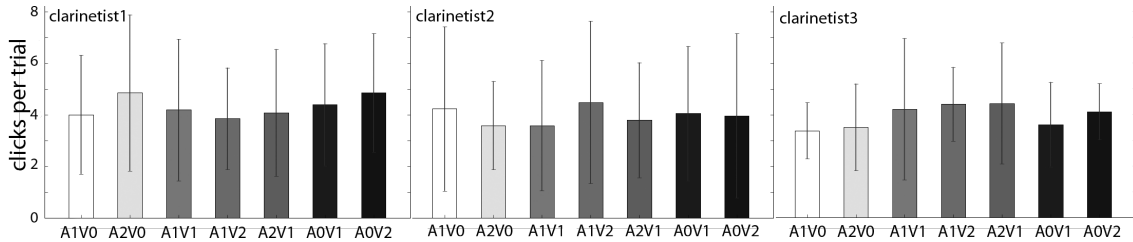


Figure 3.3: Average number of clicks per trial, taken from Grond et al. (2009).

As the large standard deviation shows, the click-frequency does not vary significantly across the stimuli. However it is interesting to observe that stimuli, which included a visual representation received the highest click frequency. Across all performances, the audio-only condition gave on average the lowest click frequency. Clarinetist 3 performed with very structured movements exhibiting pronounced ancillary gestures. Interestingly for this clarinetist, the standard deviation is smaller for most of the conditions when comparing it with those of the other two clarinetists. This comparison suggests that the intersubjective convergence in the perception of ancillary gestures is influenced by the ancillary gestures themselves, more precisely how pronounced they appear.

Kernel estimated click density: For a better comparison of the different click annotation patterns along the performance of the clarinetists, we computed kernel-estimated click densities and visualized the result. In Figure 3.4, some intervals of the click densities are depicted, which demonstrate that there are noticeable differences in the click density for various combinations of stimuli and suggested that either different events were perceived or the same event was annotated at different times. The three

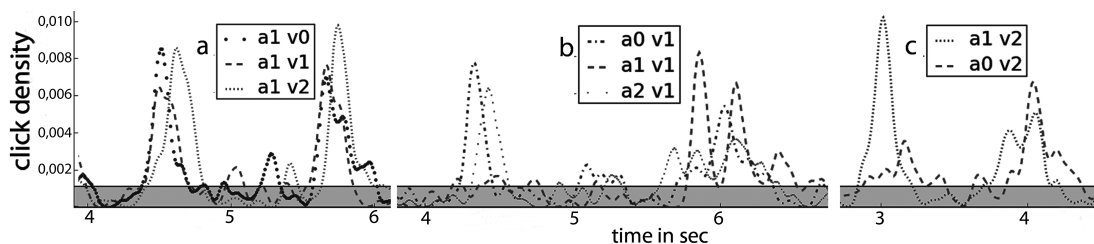


Figure 3.4: Kernel estimated click densities from Grond et al. (2009). a, b and c are selected intervals of clarinetist 1, 2 and 3 respectively. The grey horizontal bar indicates the average click density $\int = 1$.

depicted plots (a, b, c) are selected cases for patterns we found in the plots, which are compiled in the appendix 8.1.

In plot a , which compares the velocity sonification with the three different visualization types (no visualization, stick-man, stickman with glyphs), we see that multimodal conditions give a noticeable delay in the reaction of the subjects. It must be noted however, that we also found one moment with increased click densities in the performance where the visualization-only stimulus V2A0 made the subjects click faster than in the multimodal case V2A1 (cubes plus velocity sonification). A possible explanation for the delay might be that integrating two perceptual inputs increases the perceptual load and therefore causes the delayed response.

Selection b is a plot in which the simple visualization of the stick-figure is compared with the two different sonifications and the stimulus without sonification. In this comparison, interestingly the velocity sonification in clarinetist 2 made the subjects ignore an event (between second 4 and 5), which was however selected in A0V1 and A2V1. When comparing the last two conditions, the shape of their density peaks are very similar in this time region except the already discussed delay in the condition A2V1. This suggests that the PCA-based sonification (A2) seemed not to have overruled the visualization. In the time region around second 6 we see that the stimulus A1V1 made the subjects clearly annotate two successive events which we could identify as two quick arm-wrist movements. In the PCA-based sonification and the stimulus without sound the distinction between both events was less clear.

The plot c shows in the selection around second 3 a peak for the stimulus A1V2, i.e. the velocity sonification shown with the stickman visualization highlighted with glyphs. At this moment of the performance clarinetist 3 made a step, which was not noticeably annotated by clicks in the V2 only condition.

In addition to the qualitative evaluation of selected time segments, we were also interested if there is a statistically significant difference for the different stimuli. Table 3.3 compiles the results of the Kolmogorov-Smirnov (KS) test comparing the click-trains

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

| compared stimuli | clarinetist 1 | clarinetist 2 | clarinetist 3 |
|------------------|---------------|---------------|---------------|
| $A1V0 - A1V1$ | 0.233 | 0.032 | 0.011 |
| $A1V1 - A1V2$ | 0.010 | 0.493 | 0.302 |
| $A1V2 - A1V0$ | 0.180 | 0.010 | 0.078 |
| $A2V0 - A2V1$ | 0.226 | 0.086 | 0.248 |
| $A0V1 - A0V2$ | 0.144 | 0.105 | 0.441 |
| $A0V1 - A1V1$ | 0.618 | 0.018 | 0.127 |
| $A1V1 - A2V1$ | 0.223 | 0.007 | 0.360 |
| $A2V1 - A0V1$ | 0.487 | 0.508 | 0.368 |
| $A1V0 - A2V0$ | 0.086 | 0.001 | 0.012 |
| $A1V2 - A0V2$ | 0.006 | 0.006 | 0.104 |

Table 3.3: This table from Grond et al. (2009) compiles the values of the Kolmogorov-Smirnov test comparing different click-trains. Values below 5% are highlighted. The 3 stimuli pairs (line 1-3) compares $A1$ with $V0$ $V1$ $V2$. The 3 stimuli pairs (line 6-8) compares $V1$ with $A0$ $A1$ $A2$.

of the different stimuli. The most general finding in the last row is that for clarinetists 1 and 2, the velocity sonification $V2$ made a significant difference. Also for clarinetist 3, the value of 10% is low. Similarly in row 10, comparing the velocity sonification with the PCA-based sonification, both without visualization, made the subjects annotate significantly different events for clarinetist 2 and 3, even for clarinetist 1 the KS test returns a result that is close but not below a threshold of 5%. Across many stimuli pairs, the click distribution significantly differed particularly for clarinetist 2.

The qualitative analysis from Figure 3.4 reveals that also for clarinetist 1 and 3, different moments were annotated by clicks made by the subjects depending on the modality of the stimuli. We illustrated those differences in Figure 3.5 by plotting the click time versus the click number. For clarinetist 3 we compared the conditions $A1V0$, $A1V2$, $A0V2$. We chose to compare the enhanced visualization $V2$ because it was the one with a more visible connection with the velocity sonification. The plot shows that adding the velocity sonification $A1$ to the enhanced visualization $V2$ narrowed the distribution of succeeding clicks along the diagonal. The condition $A1V0$, i.e. the velocity sonification without visualization exhibits few outliers along a pronounced diagonal. In the condition $A0V2$, i.e. the enhanced visualization alone, there is less coherence in that subjects marked identified events in a similar order. The depiction of the condition $A1V2$, i.e. the combined audiovisual stimulus shows a coherence that lies in between the aforementioned stimuli. Comparing these three cases can be interpreted as decreasing intersubjective convergence in the perception of the stimuli $A1V0$, $A1V2$, $A0V2$. More precisely velocity sonification alone or together with the enhanced visual display made the test-subjects select similar events in a similar order.

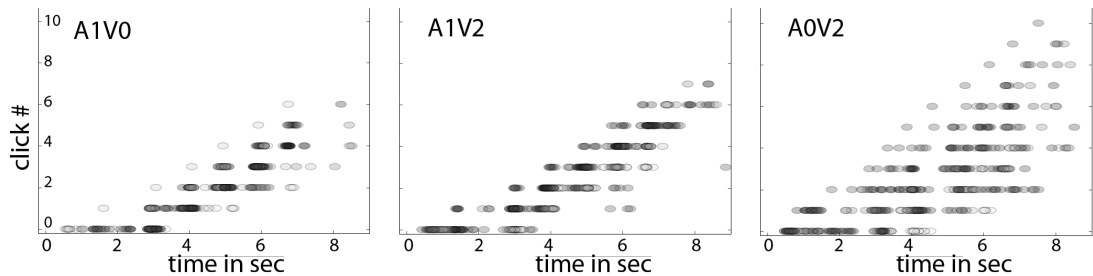


Figure 3.5: Clarinetist 3, click time versus click number from Grond et al. (2009). The outliers from the diagonal show the decreasing intersubjective convergence in the order of the stimuli A1V0, A1V2, A0V2.

3.2.6.1 Sonifying the Annotation

In order to get a better grasp on the click annotations we sonified the click trains as short filtered impulses and mapped the kernel-estimated click density to their center frequencies. The resulting sound files were merged with the visual or multimodal stimuli for all movement sequences. This exploratory evaluation supported the trends for clarinetist 3 that are depicted in Figure 3.5. In the A1V0 condition the unanimously annotated events were clearly audible, which could also be recognized, but were less pronounced in the A1V2 condition. The sonification of the clicks in the A0V2 condition suggested that the subjects were looking at various aspects in the display and that there was little convergence across the subjects, also the rhythmic aspect of the movement was less represented through the clicks in the visual only display. The advantage of sonifying the annotations consisted in the possibility to connect the annotations with the corresponding movement pattern in the visualization. This experimental evaluation approach also made it possible to get a sense for what kind of rhythmic structure in the sonification was perceived.

3.2.6.2 Subject Rating of the Stimuli

At the end of the experiment, we asked the participants to rate all the stimuli with respect to how helpful they were for the given task. The rating options were between 1 (difficult) and 5 (easy) with respect of consistently selecting events in the display. Figure 3.6 shows the compiled results. The audiovisual conditions A1V2 and A1V1 both involving velocity sonification were rated as first and second. The PCA based sonifications A2V1 and A2V0 rank among the least preferred together with the simple stick-figure visualization A0V1. This is consistent with the findings in Figure 3.4 in the selected time interval b , where the PCA-based sonification made it difficult to consistently select events.

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

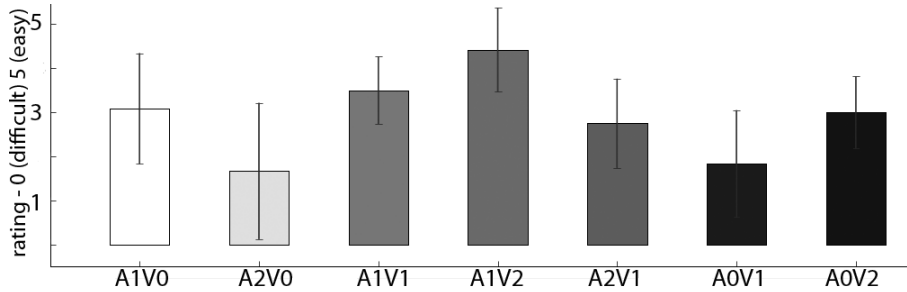


Figure 3.6: Self rating of test subjects from Grond et al. (2009). x shows the stimuli acronym according to Table 3.2, y the mean \pm standard error.

3.2.7 Conclusion on the Annotation of Ancillary Gestures

In the free annotation task we evaluated a velocity and a PCA-based sonification for ancillary gesture sonifications alone and in combination with a visualization of the movement. Although the open task we have designed did not test specifically the reaction time, we found cases suggesting that audiovisual displays make the user react more slowly in annotating identified events. The overall quality of the stimuli depended on the actual movement of the clarinetists. It is hence difficult to arrive at general conclusions across all possible ancillary gestures. The following findings are based on the qualitative evaluation of the three different performers that provided the data substrate for the sonification and the visualization.

In Figure 3.5, we have found an example where the sonification-only display directed the focus of the participants towards similar events, which were annotated in a more unanimous way among test subjects with respect to their temporal order. When comparing the velocity-based sonification with the PCA-based sonification, we found examples depicted in Figure 3.4 of indirect evidence that the velocity-based sonification is more efficient in highlighting otherwise overseen aspects of the gestures. The velocity based sonification was also preferred by the test-subjects over the PCA-based sonification and visualization-only modes. The coordinated directions of movement extracted through the PCA does not seem to correspond well with what subjects perceive in the visualization of the stick-figure, which we believe is the reason for the difficulties in connecting these two display modes. Another drawback of the PCA-based approach is the fact that it extracts for each clarinetist different coordinated directions of movement of their idiosyncratic patterns. As a result, the way in which a PCA-based sonification connects with the visual display has to be figured out by the test subjects for each clarinetist anew. Our experience from the PCA-based approach is instructive for the design of sonifications for multimodal displays. It is in line with the idea of *causal listening*, meaning that we tend to look for a visual correlate that explains our acoustic

3.2 Annotating Sonified Ancillary Gestures

impressions. If a correlate is found, the sound seems to originate from the image and is not perceived in isolation. If no correlate can be found in the visual display, as it seems to have been the case in the PCA-based sonification, no integrated perception emerges. This lack of integration is reminiscent to the unidentified sound object discussed by Flückiger (2001, page 126).

From all the three performers the movement pattern of clarinetist 3 was the most structured and also the most obviously influenced by the musical phrase. In this case it was interesting to find that subjects annotated different events for the A0V2 stimulus where the visualization alone was highlighting data aspects similar to the sonification, compared to the more unanimous selection in the A1V0 and A1V2 case. This suggests that sonification can potentially guide attention to information in the visual display, which is not necessarily in the focus of attention.

The main purpose of sonification in this audiovisual setting consists in guiding attention rather than adding information. Complementarity in information displays, takes on a particular meaning for multimodal cases: if sonification can be used in a complementary manner depends on whether the senses can integrate the additional information in the display to a meaningfully combined percept.

3.3 Eye-tracking of Sonified Ancillary Gestures

The free annotation task that was discussed in the previous section provided some qualitative insight into how sonification contributes to unimodal or multimodal displays of body movements and how it influences the perception of events in the stream of ancillary gestures. Asking the subjects to annotate events by mouse clicks indicated that some events were perceived, and the experimental analysis of sonifying the annotations made it possible to identify events in the visual display that were likely to be the correlate to the perceived sound. Although the task was deliberately very open, selected events can none the less still have been influenced by the requirement to annotate consistently. Since several audible events were present at the same time or in rapid succession, it is not possible to determine which sound was most prominent or which was the basis for searching a visual correlate.

I therefore decided to study multimodal sonification of ancillary gestures with eye-tracking, in order to better understand where subjects look when they hear at the same time a sonic representation of the movement. Eye-tracking does not inform us what subjects are listening to. However, by showing directly where subjects try to find visual correlates to the sonification, eye-tracking can give indirect hints about where the attention is directed. By conducting an experiment without a specific task, subjects can also freely explore the audiovisual stimuli and their eye movement is not influenced by a specific goal that occupies perceptual or cognitive resources.

Since the only data that we gathered from the subjects are eye-movement data, the sonification-only condition was omitted, because it would not have led to any meaningful data acquisition due to the lack of a visual component. The second audiovisual prototype differed in many ways from the first and hence both are not directly comparable. While the sound design in the aforementioned sonification resulted in noticeable spectral centroids with some spectral flux, we adapted the sonification design in order to use two different audible features (pitch and timbre) as labels for the joints. These new explorations for the sound design were motivated by the search for richer sounds with more mapping targets. The choice of vowel-like spectral contours – realized through the vowel synthesis class which is described in detail in Section 2.4 – was motivated through the ideas of the *gestural-sonorous object* described by Godøy (2006) and discussed in the introduction to this chapter. In the mapping of vowel-based sounds to the markers we were also interested in testing if the pitch polarity with respect to high and low in the visualization would have an influence on the visual perception of the display.

A further aspect that we changed in the data preparation was that we sonified the angular velocity and not the absolute velocity of the markers as we did in the previous display prototype. The reason for this choice was that we had learned from

the comparison between the absolute marker velocity and the PCA-based sonification that it is important to have comparable information in the visual and the audible display in order to merge the percepts. The absolute marker velocity is a superposition of the velocity of all joints to the end effector, a hand or leg for instance. Therefore, the absolute velocity of a hand that holds still can be big if there is a lot of movement in the joint of the shoulder or the elbow. Although one can argue that it is in fact the movement of the end effector that we perceive, the sonification of the angular velocity provided an alternative to highlight independent contributions to the movement. This was particularly interesting in combination with the enhanced visualization in which the angular velocity was additionally displayed with the animated glyph.

The stimuli of the previous sonification prototype were selected from different performers along one musical motives. This led to movements of very different expressivity. In the second prototype two different movement segments from two clarinetists were selected from sections of the movement data with varied movement patterns. This resulted in four different movement sequences.

3.3.1 Mapping and Sound Design

As mentioned above and in the introduction to this chapter, the sonification was based on vowel sounds taking advantage of the synthesis class introduced in Section 2.4. In Table 3.4 the audible labels consisting of sound with a distinct spectral envelope and pitch are listed together with the joints of the stick-figure to which they were mapped. Two different mappings with respect to the audible labels were conceived. The first mapping *s1* followed the natural association of low pitches with markers at the bottom of the stick-figure and high pitches at the top. Low pitches went along with low registers which also increased to the top, from *bass* at the ankle and knees to *soprano* at the neck and the front. The second mapping *s2* reversed the association of pitch and register with high being at the bottom and low at the top. The interval of the pitches were four semitones spanning two octaves from MIDI note 40 to 64. This range of two octaves was chosen in order to keep the fundamental frequencies fairly low and the partial frequencies narrowly spaced in order to well support the spectral contour.

The vowel sounds were synthesized as subtractive synthesis using the *BPFStack* unit described in Section 2.4. Figure 3.7 shows the SC3 synthesis definition. The filters were applied to a source that crossfades between two signals: the first is the pitched sound of a sawtooth signal, and the second is the same sawtooth multiplied by pink noise. The second source had a less pronounced to vanishing pitch. This transition between pitched and unpitched sounds was the first mapping target. The second mapping target was the brightening of the vowel sounds using a method that kept the sum of the gain

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

| joint angle | mapping s1 | mapping s2 |
|-------------|---------------------------------------|---------------------------------------|
| | MIDI Hz vowel register | MIDI Hz vowel register |
| front | 64 329.628 [e:] soprano | 40 82.407 [i:] bass |
| neck | 60 261.626 [a:] soprano | 44 103.826 [e:] bass |
| back | 52 261.626 [a:] alto | 52 103.826 [e:] counter |
| shoulder | 56 207.652 [u:] alto | 48 130.813 [a:] tenor |
| elbow | 52 164.814 [o:] tenor | 52 164.814 [o:] tenor |
| hip | 48 130.813 [a:] tenor | 56 207.652 [u:] alto |
| knee | 44 103.826 [e:] bass | 60 261.626 [a:] soprano |
| ankle | 40 82.407 [i:] bass | 64 329.628 [e:] soprano |

Table 3.4: Mapping joint angles of the clarinetist to audible markers.

of all formats constant in order to approximate constant loudness when changing this parameter (compare the paragraph on brightening in Section 2.4.1). The brightness change is implemented by adding the *addControls* method to the Vowel instance, which provides the necessary control busses for the formant parameter manipulation of the synthesis units. The third mapping target was simply the gain of the overall signal.

```

{lfreq=200, pan=0.0, gain=0.01, lg=0.1, xfade= -1}
var sig, source, sourceXfade;
source = Saw.ar(freq.lag(lg));
sourceXfade = XFade2.ar( PinkNoise.ar(1) * source, source, xfade.lag(lg) );
sig = Pan2.ar(
    BPFStack.ar(sourceXfade,
        Vowel(\i, \bass).addControls(i,\kr,0.05),
        pan.lag(lg), gain.lag(lg) )
    );
Out.ar(0, sig);
}.play

```

Figure 3.7: SC3 synthesis definition for the vowel-based ancillary gesture sonification

In a one-to-many mapping approach, the joint velocity was mapped to all three mapping targets. Over the velocities of all joints, the global 90, 95, and 99 percentiles were computed. The absolute velocity $|v|$ was linearly mapped to the 90 percentile for the transition between the voiced and unvoiced source. With increasing speed of the movement the sound was brightened by mapping $|v|$ up to the 95 percentile to the parameter of the brightening method, that lifts the higher formants in the spectrum. Finally, $|v|$ was mapped up to the 99 percentile to the gain from -30 to 0 dB.

This one-to-many mapping was inspired from the first example in 2.4.4. The idea is to articulate joint angle speed without a threshold below which the sound needs to be artificially muted. This would be the case if $|v|$ were only mapped to the gain of a pitched sound. Because even if pitched sounds are of low volume they tend to be heard as chords if other pitched sounds are present. If there are too many pitched sounds

3.3 Eye-tracking of Sonified Ancillary Gestures

at the same time an unpleasant saw-tooth like sound is perceived. It is therefore necessary to find possibilities to allow the sound to move to the background and to minimize interference with other audible labels that might be present. Particularly the first mapping changed noticeably the presence of the sound by altering its locatability in the pitch-field. By introducing the crossfade with the unpitched source, low volume sounds gave room for other elements that were simultaneously present in the auditory display. Comparing it with the descriptive terms of the sound object, this corresponds to a transition from a *complex mass* to a *tonic mass*, for a definition of both see Chion (2009, page 472 and 496). Since both extremes were shaped with the same spectral contour the transition was perceptually smooth. With increasing speed, the brightness additionally augmented the presence and articulation of the joint angle velocities.

3.3.1.1 Preparing the Stimuli

Using the sonification described above, different stimuli were prepared based on 4 clarinetist movements each of 20 seconds in length. For the movement sequences, moments in the clarinetist’s performance were selected which exhibited noticeable articulations.

| | no sound | sonification 1 | sonification 2 |
|-------|----------|----------------|----------------|
| lines | <i>L</i> | <i>Ls1</i> | Ls2 |
| cubes | C | <i>Cs1</i> | Cs2 |

Table 3.5: Rendering modes of the clarinetist movements. The abbreviations are explained in the text, **bold** and *italics* indicate the grouping of the rendering modes in the experiment.

Each movement was rendered in 6 different modes, which are compiled in Table 3.5. The first capital letters stand for the visualization: L is the movement of the stick-figure as line drawing with no sound. C stands for the stick-figure with glyphs in the form of cubes highlighting the joint angle movements. The further elements s1 and s2 stand for the sonification according to the mappings in Table 3.4.

Compared with Figure 3.2, from both visualizations with and without glyphs the left depiction of the clarinetist in profile was used with white cubes as glyphs, since for a meaningful eye-tracking experiment only one visual instance of a clarinetist can be displayed. For the experiment, we split the stimuli into two groups, in order to avoid exposing the subjects to different mapping polarities of the audible labels within one session. The two groups A and B consisted of (*L*, *Ls1*, *Cs1*) and (**C**, **Cs2**, **Ls2**). Each group contained one visualization-only stimuli and two audiovisual stimuli. The order of the two sessions A and B was balanced. Similar to the experiment in Section 3.2 the stimuli were randomized and interleaved such that the same movement sequence was never presented in succession.

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

3.3.2 User Study

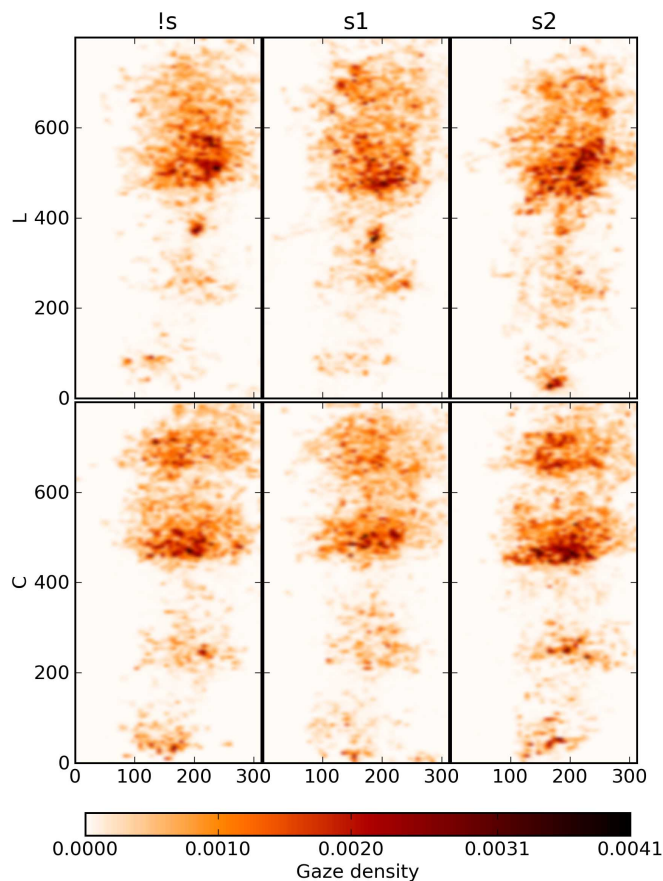


Figure 3.8: Gaze distribution density for movement 1.

tions. The visualization was presented on a 52 x 32 cm LCD screen. The subjects sat in front of the screen at a distance of approximately 120 cm. The sonifications of the audiovisual stimuli were rendered through AKG K271 headphones. Both eyes were tracked at a rate of 500 Hz. The data analysis was based on the dominant eye. The experimental procedure was implemented in the software V-designer 2005, which controlled the eye-tracking PC with the system Eyelink2 from SR-Research.

3.3.3 Evaluation of the Eye-tracking Data

The plots of the results are only partially shown in this section, the rest together with a detailed discussion is compiled in the appendix of this thesis in 8.2. Since 8 subjects do not suffice for valid statistics, standard deviations are shown in the plots for approximate orientation.

In order to get a qualitative user response about the designed interface we conducted an experiment with 8 subjects based on the recommendations of Nielsen (2000) (5 male and 3 female, with an age range from 24 to 49). Each subject was asked to participate in sessions A and B. Each session consisted of the 3 stimuli modes for 4 movements. Each stimuli were presented with 4 repetitions with randomized order, overall resulting in 48 stimuli presentations. After 24 stimuli, subjects were asked to take a break of 5 minutes. A drift correction was performed between all stimuli presenta-

3.3 Eye-tracking of Sonified Ancillary Gestures

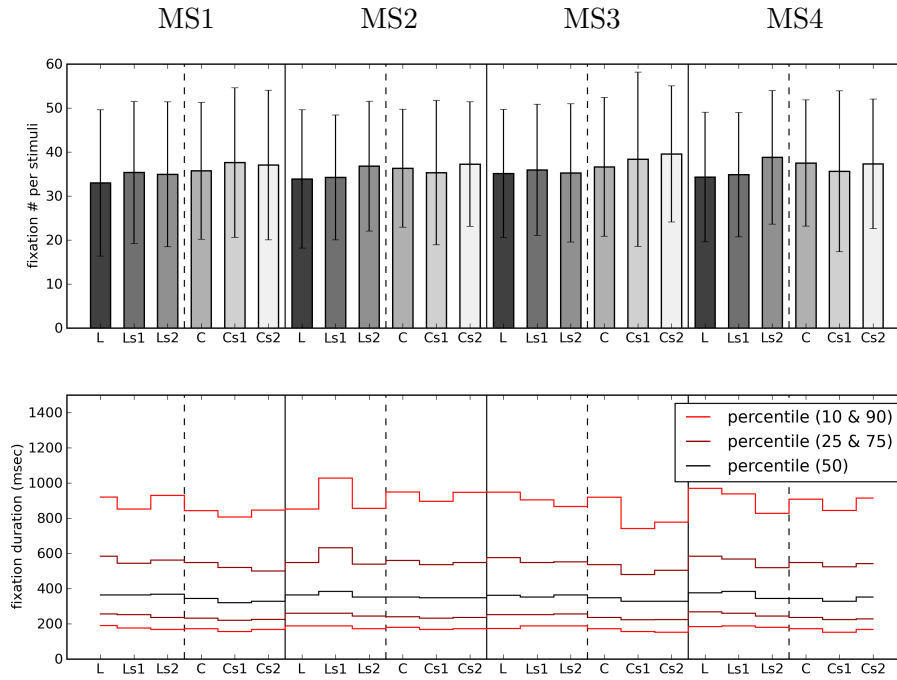


Figure 3.9: Number of Fixations \pm standard deviation on top. Percentiles of the duration of fixations at the bottom. Movement sequences MS1 to MS4 are ordered from left to right. For a description see 3.3.3

Figure 3.8 shows a heat map of the gaze density for all 6 stimuli modes. The axes of the heat map correspond to extension of the movement within the display, 768 pixels high and 300 pixels wide. The top row shows the gaze at the stick-figure. The bottom row shows the gaze at the stick-figure augmented with the cubes. The left column depicts the mode without sonification. In the middle there is *s1* and on the right *s2*. Figure 3.8 shows that subjects were mostly looking at the upper body for a detailed description of this plot and all other movement sequences see Section 8.2 in the appendix of this thesis.

For all 4 movements, the global average for the number of fixations per trial is shown in Figure 3.9 at the top. This plot is essentially the same for the number of saccades. From the fixation data, outliers were removed according to the following criteria: Fixations had to fall within the section of the display of 768 pixels height and 300 pixels width that contained the movement of the stick-figure. Given that there were about 5 back and fourth movements (each direction approx. 2 seconds), fixations lasting longer than 3 seconds were discarded as moments when subjects did not pay attention to the display. The middle in Figure 3.9 shows for all 4 movements the 10, 25, 50, 75 and 90 percentile for the duration of all fixations. The movements are ordered from left to right and are separated by vertical solid lines. Within the same

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

movement, the stick-figure stimuli are separated from those with cubes through dashed vertical lines. In all three plots no deviation or trend across the different stimuli can be observed.

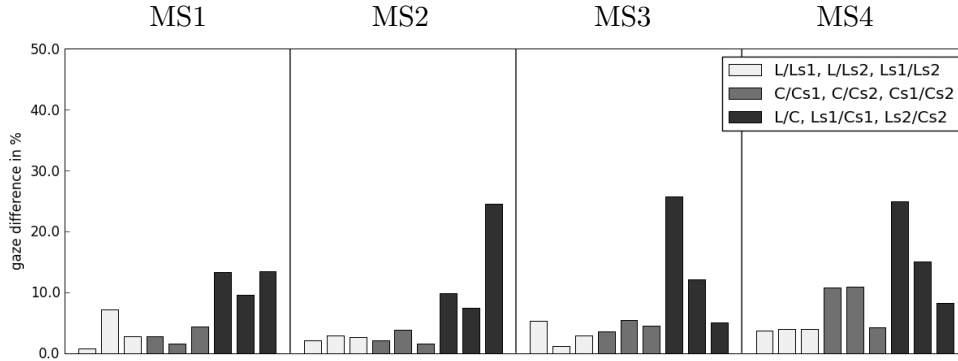


Figure 3.10: Amount of time with vertically differing gaze distribution for selected stimuli pairs. Movement sequences MS1 to MS4 are ordered from left to right.

In order to find out if the vertical mapping polarity, $s1$ and $s2$, caused different vertical gaze distributions for certain moments during the stimuli presentation, the vertical gaze distribution was compared based on the fixations for time frames of an 8 msec interval of selected stimuli pairs by the KS test. Figure 3.10 shows the amount of time for which the p-value of the KS test dropped below 5%. This figure is organized along the movement sequences from left 1 to right 4, separated through vertical lines. The colors encode sets of compared stimuli as shown in the legend, the bars in the plot correspond to the order in the legend from left to right. Although the comparison of the conditions leads to different results for the different movement sequences, a strong increase can be observed for each movement in at least one of the comparisons of stimuli with constant or no sonifications (dark bars on the right) (L/C , $Ls1/Cs1$, $Ls2/Cs2$). For movement 4, adding sonifications $s1$ and $s2$ had a noticeable influence for visualizations with cubes (C) in that it leads to double the time with different vertical gaze distribution (approx. 10%) compared to the stick-figure only visualization (L) (approx. 5%). This suggests that sonifications have a stronger effect if we can identify a clear cause in the display, which conforms with the written reports of most subjects that the cubes and the sounds were appreciated as mutual matches. However, comparing the same bars in movement sequence 1 in Figure 3.10, we find the opposite, i.e. a stronger influence for (L) compared to (C) for the sonification $s2$.

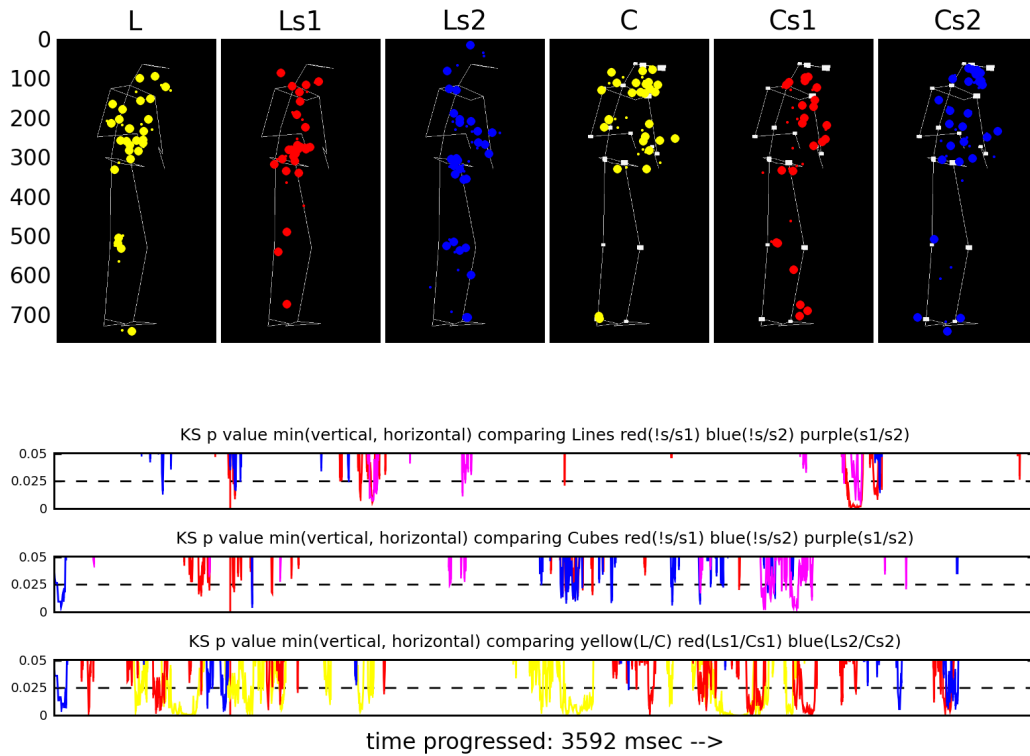


Figure 3.11: Selected timeframe of the interactive display for dynamic data inspection.

3.3.3.1 Vertical and Horizontal Gaze Distribution along the Time Axis

For a better assessment of the eye-movement over time the horizontal and vertical fixations over the audiovisual display are depicted in Figure 8.5 in the appendix to this thesis. The horizontal eye movement aligns with the back and forth movement of the clarinetist and shows in the plot as the up and down wave. The joint angle s of the stick-figure were mostly organized vertically corresponding to the audible labels. Some small differences can be seen with respect to the vertical gaze distribution. It is however difficult to interpret how these differences can be attributed to the movements. Therefore we developed means to interactively inspect and qualitatively evaluate the subjects' gaze for selected time frames.

3.3.3.2 Interactive Inspection along the Axis

In order to better understand the gaze distribution at a given point in time, I rendered for time intervals of 16 milli seconds the positions of the eye-tracking trajectories of all subjects for all the 6 different stimuli as depicted in Figure 3.11.

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

Below the movement visualizations including the fixations as larger circles and the eye-trajectories as small dots, the timeline of the stimuli (20 seconds) is depicted showing the p -values of the KS test for each time frame. Using the KS test, I compared the vertical gaze distribution between two different stimuli based on the fixations, as an indicator for deviating visual attention for a given time frame, similar to the comparison in Figure 3.10. Values below a confidence threshold of $p = 5.0\%$ are plotted, the $p = 2.5\%$ threshold is indicated as a dashed horizontal line. The plots for each timeframe were concatenated to a movie of the length of the stimuli. The sonifications $s1$ and $s2$ were added to two copies of the data inspection movie. This made it possible to simultaneously listen and observe the movement together with the eye-tracking data.

The KS plots give a convenient overview over the timeline and make it easy to navigate to timeframes where the gaze distribution is likely to differ in the vertical dimension, despite the small number of subjects. Around those time frames the movie can then be played back with 60 frames per seconds and the fixations over the Clarinetists' movement can be qualitatively assessed. As already found in Figure 3.10 the fixations differ in more timeframes, when comparing stimuli that differ in the visual condition (C) and (L). The results of this qualitative evaluation are reported in the appendix of this thesis together with a discussion of the gaze density plots in Section 8.2.

3.3.3.3 Post Experimental Questionnaire

After the eye-tracking experiment subjects filled in a questionnaire, in which they were asked to rate how engaging they found the different stimuli. Figure 3.12 compiles the answers as bar charts with standard deviation and parallel coordinates for each subject. The division indicated through the vertical solid line separates the two balanced groups A and B, after which the subjects were asked to rate. This means that the rating cannot be compared across all stimuli but only within the sessions A or B. All stimuli with sound are higher rated than the visualization only. Only the condition C was rated noticeably low. This is most likely because the cubes added to the display explicit information about the joint angle changes, the relevance of which was not obvious to the subjects without the sonification.

For each stimuli mode, subjects were asked to describe briefly what they observed. Two subjects mentioned explicitly the mapping polarity, one correctly identifying upper body movements with higher sounds for $s1$, another reporting wrongly the opposite for a $s1$ mapping.

As visible in the parallel coordinates in Figure 3.12, one subject disliked the sound and rated all display conditions as very low. Another subject also reported that sounds related to the head were unpleasant in the $s2$ sonification.

3.3 Eye-tracking of Sonified Ancillary Gestures

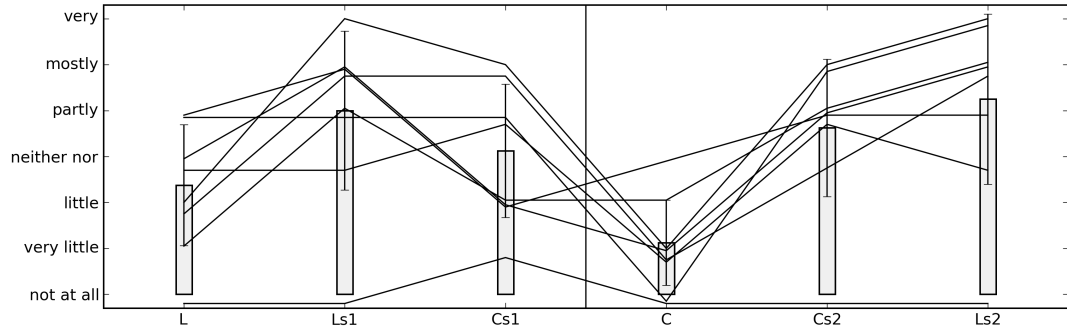


Figure 3.12: Subject rating how engaging they found the audio-visual display with parallel coordinates. The labels on the x-axis indicate the stimuli.

Most answers to this open question referred to the details of the visualization and the relation to the sound heard. Five subjects reported that the highlighting cubes caught their attention, while two subjects found the cubes irritating. For the visualization only stimulus **C**, four subjects mentioned the cubes and one reported them as supporting the movement. For **L**, four subjects found the movements to be uncontrolled or difficult to follow. For this condition subjects were mostly referring to the movement per se rather than visual details. For **Cs1**, one subject mentioned that movements of the legs came like a surprise.

When looking at all audiovisual conditions, 6 subjects reported at least once that the sound was helpful or a good fit with the visualization. Interestingly one subject tried to remember the cubes from other stimuli and tried to transfer this experience to connect it with the sound. For a **Ls2** stimulus one subject reported that it was better than without sound but still not a good fit. For the same stimulus another subject reported that there is little in common between the visual and the sound.

After this set of questions, subjects were asked to briefly describe whether the sounds had reminded them of something familiar. In this section subjects listed various sonic associations some of them referring to cinematic settings (like film effects or light saber) some associations were natural sound sources like bees some were referring to machines. None of them referred to vocal sounds. However, many instrumental associations were made, such as saxophone, accordion, tuning fork, singing bowl and vibrato. One reported that movement and intensity changes are reminiscent to electro motors. Two subjects made a negative judgement (annoying) about the sounds one a positive (relaxing) and one mixed.

Next, subjects were asked to describe the sound attributes and how they relate to the visual representation. Three subjects reported on the relation between pitch and

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

different positions on the body. One of them reported the correct mapping for *s1* and **s2**. One subject associated pitch with the movement direction, one with the movement intensity. One subject reported *s1* to have no connection and **s2** to have a good connection with the visualization. Two subjects reported the sound to be pleasant, one judged high pitches as unpleasant and low pitches more pleasant.

For timbre, two subjects made a positive judgment for **s2** describing it as good. Three subjects mentioned a relation between timbre and the part of the body that was moving. For one subject timber was not coupled to the visualization. The subject that described polarity mapping correctly, characterized the timbre development in a way that was reminiscent to the brightness mapping used. A second subject described the timbre mapping correctly for *s1*.

Four subjects described the volume as related to the parts or the sum of the parts of the body that move. The volume was judged by one subject as having no connection with the movement.

In the last section of the questionnaire, subjects were asked to express their preference between the following stimuli pairs (*L*, **C**), (*Ls1*, **Ls2**) and (*Cs1*, **Cs2**). This pairing was chosen to potentially learn more about the different mappings. For preference selection subjects could play back again the stimuli.

For the (*L*, **C**) pair, 6 subjects preferred the stick-figure with cubes, stating that it attracts attention or that it fits the movement. Two subjects preferred the stick-figure only and reported that they found the cubes distracting.

For the audiovisual stimuli pairs, subjects were asked to indicate which stimuli in the pairs had a better connection between sound and image. Two subjects reported that the cubes **C** emphasized details and that the stick-figure *L* helps to perceive the overall movement. One subject referred to the cubes as speakers in the visualization stimulus **C**, and found those helpful to concentrate on the movement. When comparing (*Ls1*, **Ls2**), four subjects preferred *s1* and two preferred **s2**, two subjects did not indicate a clear preference. When comparing (*Cs1*, **Cs2**), two subjects preferred *s1* and three preferred **s2**, three did not indicate a clear preference. Subjects were also asked to describe how the compared stimuli differ. Only one subject described that in the preferred condition the sounds related to the higher parts of the movement were lower. Interestingly this subject, who also described the mapping polarity and the timbre development correctly, did not play an instrument.

3.3.4 Discussion

As shown in Figure 3.9 the various audiovisual or visual only display forms had no influence on the number of fixations or their length averaging over the total time of the stimuli. The heat map plots in Figure 3.8 show that the direction of the gaze focuses mostly on the upper body, where most visual details were present. The upper body has also shown the pronounced back and forth movement, which could be clearly identified as the horizontal eye-movements in the plots of the fixations in Figure 8.5 in the appendix. This back and forth movement was visually prominent in the display but not explicitly sonified through the angular velocities.

The vertical eye movement in Figure 8.5 has shown a more complex and less directly interpretable pattern. Based on the comparisons of the vertical fixations and their distributions for each time frame in Figure 3.10, adding sonifications *s1* and **s2** had a small impact on the fixation distribution for the 4th movement sequence (*C/Cs1*, **C/Cs2**) compared to (*L/Ls1*, **L/Ls2**) where the sounds had a more explicitly visible correlate. This would suggest that in order to influence the gaze, sound needs identifiable visual causes. This difference could however not be identified in the data inspection movie, as discussed in the appendix.

An opposite case was found for the first movement sequence, where sonification **s2** had a stronger influence in *L/Ls2* comparing it with **C/Cs2**. This corresponds to the findings in Figure 8.4 as discussed in the appendix, and shows that the highlighting cubes could equally well overshadow the influence of the sonification.

The biggest impact could be observed when comparing stick-figure stimuli with those enhanced with cubes by keeping the audio constant (*L/C*, **Ls1/Cs1**, *Ls2/Cs2*). This shows that the cubes provide a strong visual attractor. Although only 2 subjects reported them explicitly as irritating, cubes in the visualization only stimulus were generally perceived as less useful (compare **C** in Figure 3.12). This is, however, in contradiction with the direct comparison (*L*, **C**), where 6 subjects preferred the cubes over the stick-figure. This contradiction might reflect the overall mixed results and is maybe explained through difference between the appreciation of salient visual details and their interference with the perception of the overall movement.

The global gaze distributions in Figure 8.4 in the appendix indicate that the polarity of the sonification influences the gaze for selected stimuli, when averaging over time. Some explicit moments could be identified through interactive data inspection for two cases, where the attention was shifted downwards for the mapping **s2** with higher pitches at the feet. For clearer results, the amount of 8 subjects was too small. The fact that there were multiple sensory inputs competing for attention also made it difficult

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

to identify a clearer influence of the movement sonifications and their polarity in the audiovisual display.

The answers to the questions related to the qualitative description of the sound properties show that polarity is observed by some but only one reported the correct orientation. This is most likely because in an audiovisual display, sounds serve first and foremost as events that are correlated to visual causes, and less for their qualities and based on which their vertical mapping was organized. This supposed lack of auditory focus would be an example for the narcotization of sound by a moving image and might also explain why there was no clear preference (four *Ls1*, two **Ls2**, two no preference, two *Cs1*, three **Cs2**, three no preference) for one or the other mapping when comparing them directly. Nonetheless, given that three subjects mentioned pitch and three mentioned timbre as related to the parts of the body that were moving the combination of both gives recognizable sonic labels, that allowed the subjects to connect the sounds with the audiovisual display.

The confusion of the pitch mapping polarity might also be due to the mutual influence between pitch and timbre, which was further influenced by brightening the sound. Also changes in brightness seemed to address the embodied perception of sound and were attributed to upwards movements. The intensity changes were attributed by most subjects to the movement intensity. The description of the perceived changes in intensity and timbre suggests that an audible articulation of the movement similar to gestural-sonorous-objects was achieved. The negative judgement of the sound by two subjects is most likely due to moments when many pitched sounds were played at the same time which lead in two occasions to the unpleasant sawtooth like sounds. This means that it would be desirable from a sonic perspective to find ways to reduce the data so that the information of the movement can be mapped to a smaller number of streams the separation of which would then be less pitch dependent. As a consequence the spectral parameters became perceptually more independent to properly encode further features of the movement. This sonification-motivated data reduction in turn would mean for the visualization to move away from more explicit visual causes.

3.4 Summary and Outlook

This chapter presented two prototypes for audiovisual displays of ancillary gestures of clarinetists with varied approaches towards sound synthesis and the data preparation in order to explore options to create an audiovisual display which combines both sensory modalities in a meaningful way. The possibilities to sonify movement sequences in combination with a visual display are manifold and require the intervention in the

data and in the sound synthesis domain. This was the reason for the exploration of different options in the design space. Since both display prototypes were different with respect to the sonification of data features and sound synthesis choices, they cannot be directly compared. However some more general conclusions about audiovisual displays of ancillary gesture movements can be drawn. The experiences made with respect to the data preparation, the evaluation of the combined audio-visual display and the sound synthesis approaches are discussed in the sequel.

The data preparation step tried to address the following question: how can the dimensionality of the movement data be reduced such that the movement sonification remains interpretable together with its visual representation. Three different data preparation steps were explored. In the first prototype, either the marker velocity or movement features based on the principal components were used. The sonification of the second display prototype was based on the velocity of the joint angles. The marker velocity corresponds in certain cases directly to the intensity of the visually perceivable movement and its causes, for instance when the knee is bent. In other cases the marker velocity translates the superposition of the activity of various joints or even the whole body. This was the reason, why the weight transfer movement was removed in the first display prototype, since it was overshadowing more minute aspects of the gestures. The principal components of the movement, that was applied in the first prototype had two somewhat related shortcomings, which was on the one hand the balance between global body-movements and movement details, and second the difficult visual identifiability of movements along principal components that were along coordinated movement dimensions with smaller variance. Further, PCA had a tendency to highlight the idiosyncratic aspects of the movement for each instrumentalist, so that sonifications across instrumentalists could hardly be compared. The joint angle velocity was complementary to the marker velocity, in that it would indicate movements of joints in the wrists and feet independently of the movement of the arm or leg respectively. In the case of the joint angle at the knee, angle changes and the velocity of the marker at the knee had a similar visual effect.

For the design of future display prototypes, I conclude for the data reduction that it might be worthwhile to segment the markers or joint angles on the body hierarchically. Based on this segmentation, the dominant principal components of the trunk and head can be sonified, followed by the dominant components of subsets like those of arms or legs. During the development of the first prototype I experimented with related approaches but abandoned them eventually due to the lack of reasonable scalings of the data subsets so that the PCA for left and right limbs can be computed independently but scale with respect to a common magnitude of coordinated movement dimensions.

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

Experimentation with data reduction, will need to find solutions how the representation of the overall movement can be balanced with movement details.

The visual display was developed together with the sonifications so that the meaning, i.e. the attribution of the audible articulations to moving parts of the body, could be understood through a visual representation. The first display prototype was qualitatively evaluated by a free annotation task, in which subjects were asked to consistently identify events of their interest in repeated presentation of the clarinetists movement. The results were that sonifications have a twofold influence in combinations with the visual display. First, they have the potential to direct the attention to otherwise barely noticed events. Second, sonifications can support the temporal organization of the stimuli and helped to raise the awareness of moving limbs relative to other concurrent movements, which we found most explicitly in the movement sequence 3 with well articulated features. In the same movement, the repeated annotation of events became more consistent, as shown in Figure 3.5¹. Further, subjects reported that they preferred audiovisual displays where the visual component provided easy to identify causes, such as the cubes highlighting the movement (compare Figure 3.6). Audio-visual displays where the sonification was based on a PCA was least preferred, this can be interpreted as being due to the less apparent visual causes and corresponds to the notion of unidentified sound objects introduced by Flückiger (2001, page 126). In the second prototype, two sonifications exploring different mapping polarities were qualitatively evaluated by eye-tracking in order to find out if the mapping polarity of the sonification influences how subjects visually perceive the display. With respect to the number of eye fixations, no global effect could be found when adding sonification to the visual display. However, in the gaze density plots averaging over time in Figure 3.10 some influences of the mapping polarity could be found as discussed in detail in Appendix 8.2 and Section 3.3.4.

The experience from the feedback of the subjects with respect to audiovisual displays can be summarized as that they like to be able to identify causes in the display. These causes, if too explicit, can however be perceived as distracting from the overall movement. For future audiovisual display prototypes of movement sonifications it might be worth considering to temporally separate both sensory modalities in the display. As a possible eye-tracking experiment, it would be interesting to study if a repeatedly alternating presentation of a sonified and visualized short movement sequence, alters the way how we look at the movement. This might be beneficial for the act of listening by allowing for more reflective modes and for the act of looking at the movement by removing the reflexive need to identify the visual correlate as a cause for the perceived

¹This is reminiscent of the findings by Guttman et al. (2005) that audible rhythms overrule vision.

sound. Such an experiment might shed light on whether the temporal organization of several *gestural-sonorous objects* can be transferred from listening to a visual movement display. In such a temporal separation different auditory mappings could be explored for their potential to direct the focus towards different aspects of the movement.

Two different sound synthesis approaches were developed for the display prototypes. The first movement sonification was based on audible labels with noise passed through resonance filters and a small degree of spectral flux. For the second movement sonification, audible labels consisting of a vowel-like spectral contours using the *Vowel* class from Section 2.4 were used. The gain of both sonifications was controlled through changes in the data features either by marker velocities, principal components or the joint angle velocities. When comparing both approaches I experienced the following tradeoff: The sufficiently narrowly tuned resonant filter created only a small spectral footprint which helped to make the labels distinguishable, but also provided only few other mapping targets of little saliency. The vowel-based sonification on the other hand, provided pitch (as label and as transition between a complex and a tonic mass), spectral contour and brightness as mapping targets but created a much thicker sound with a large spectral footprint. This tradeoff points at the importance of the data preparation step and how it is linked with parameter-mapping and possible future avenues.

If the dimensionality of the movement data can be reduced through principal component analysis in a hierarchical manner as described above, the body movement can be represented in few subsets, each of which could be mapped to a distinct vowel-like spectral contour. Then, mapping targets such as the transition of pitched to unpitched and vowel brightness can be used for the 2nd and 3rd principal components of the same data subset. By additionally exploiting the stereo panorama, 3 auditory streams of different spectral contours might suffice to represent the principal components of the legs, the arms and the trunk plus head of the performer. With only three different clearly identifiable pitches and spectral registers, movement sonifications could potentially also be understood without visualizations, by just knowing how they relate to the parts of the body. With only three spectral contours the separability of these auditory streams will also be less pitch dependent so that these mapping targets can be used for other data features. For an emphasis of the *gestural-sonorous object*, mapping to brightness could for instance more appropriately be mapped to movements towards postures that can be associated with brighter sounds. Another interesting avenue would be to find appropriate mapping targets for the second derivative in the sonification as investigated in Chapter 4, so that images of effort can be partly captured through the kinematic data set. For future work this might lead to temporally evolving sonically rich gestural-sonorous objects for different parts of the body beyond highlighting and sonifying specific movement aspects.

3. MAPPING IN THE SONIFICATION OF ANCILLARY GESTURES

4

Mapping and Interaction in Auditory Graphs

Auditory graphs are the auditory equivalent of visual data representations such as plots, graphs and charts. Auditory graphs play an important role as an assistive technology for visually impaired or blind users. For a general overview of this field, please consult Edwards (2011).

The two applications presented in this chapter are sonification prototypes for mathematical functions as a teaching aid. Mathematics teaching material for the blind and partially-sighted is generally tactile, using Braille notation or reliefs. When it comes to function analysis, this form of notation has some limitations. Appropriate teaching material is limited because of its involved method of production. Due to the fact that the blind and partially sighted often possess heightened auditory capacities, there have been efforts to develop auditory displays and more specifically auditory graphs for teaching mathematics. Almost all auditory graph implementations are examples of PMSon. The only exceptions are scatterplots as data-sonograms which belong to MBS (see Section 2.4.4.2).

During the development of the first prototype, *Sonic Function*, we collaborated with Trixi Drossard, who has a background as math teacher for the partially sighted and the blind. She evaluated the prototype together with blind pupils in a teaching context. The results were published in Grond et al. (2010). The pedagogical observations and qualitative aspects from this study are published in Droßard et al. (2012). The second prototype *Singing Function*, was published by Grond and Hermann (2012b) and expanded on the sound design possibilities by using vowel based synthesis. The content of the following chapter is based on these publications. This chapter gives first an overview of the literature of the field, followed by a description of the new approach

with implementation and evaluation details. The chapter ends with a conclusion based on the experience made with *Sonic Function* and *Singing Function* and an outlook for future work integrating the discussion of listening intentions.

4.1 Literature Review on Auditory Graphs

This literature review is based on two summaries, which have been published by the author in Grond and Hermann (2012b) and Grond and Berger (2011). Mansur et al. (1985) presented the first system for the creation of computer-generated sounds for the representation of two-dimensional plots. The objective of this system was to provide the blind and visually impaired with a means of understanding line graphs in a manner similar to sighted users by mapping ordinate values to pitch and varying it continuously along the x-axis. In a human factors experiment, the authors compared their prototype system with tactile-graph methods. According to Mansur et al. (1985), mathematical concepts such as symmetry, monotonicity, and the slopes of lines could be well determined.

Bonebright (2005) suggested an agenda for auditory graphs after conducting longitudinal research. The main items of the agenda were effectiveness, role of memory and attention and longitudinal studies of learning. For *Sonic Function*, as well as for *Singing Function*, we consider the role of memory to be very important. We believe that this factor depends highly on the sound design and whether sounds can evoke familiar listening experiences.

Stockman et al. (2005) give an overview of the field of auditory graphs and also propose the formation of a research agenda. Of particular interest for *Singing Function* are comparisons, understanding and recall of sonified charts. Further, in Stockman et al. (2005) the multiple views paradigm is introduced which inspired our work in *Singing Function*. This paradigm calls for intuitive means of representing the same information using different levels of detail. Interactivity as a way to explore auditory graphs was also addressed as an important point of the proposed research agenda.

Harrar and Stockman (2007) address the need to study the effects of changing presentation parameters on the users ability to gain an overview or to identify specific graph characteristics. Among other aspects, they discuss the influence of playback speed in which the auditory graph is presented. Speed and continuity were presentation parameters over which users had control in the interactive settings of both prototypes presented in this chapter. For the constant playback in the vowel based *Singing Function* we oriented our choice on which speed can be articulated with ease.

4.1 Literature Review on Auditory Graphs

Hetzler and Tardiff (2006) presented a study contrasting discrete and continuous auditory graphs. This study gives evidence that both representation modes serve different purposes. During the sound design process in *Sonic Function* Grond et al. (2010) we noticed that discrete presentation compared to a continuous one helps to gauge intervals. However, a discrete mode of presentation is conceptually in contradiction with the notion of a steady function and an x-axis populated with dense real numbers. This contradiction motivated us to look for sonifications whose understanding do not exclusively rely on the interpretation of intervals of function values along small increments of the x-axis.

The most complete overview of the state of the art is given by Flowers (2005). From a PMSon perspective, this account is particularly interesting as it reports not only successful mapping strategies but also those which failed. As already reviewed in Grond and Berger (2011), the mapping strategies which were successful are: pitch coding of numeric value, manipulating loudness changes in pitch mapped stream as contextual cues and signal critical events, choosing distinct timbres to minimize stream confusions and unwanted grouping, and sequential comparisons of sonified data. Approaches that failed were either due to the complex nature of loudness perception which cannot be used to represent an important continuous variable, or due to grouping if the simultaneous presentation of some continuous variables were of similar timbres, or if too many simultaneous continuous variables were presented at the same time using pitch mapping. Flowers also suggests that we need to know more about the effects of stream timbre and patterning on perceptual grouping, and the representation of multiple variables in a single auditory stream. This very complete account by Flowers lists many issues that we addressed by *Sonic Function* and even more with the improved version *Singing Function*.

Nees and Walker (2007) proposed a first conceptual model of auditory graph comprehension. The authors present an extensive review of the research field, the necessity of which they motivate by pointing out that: “*Auditory graph design and implementation often has been subject to criticisms of arbitrary or a-theoretical decision-making processes in both research and application.*” We agree, as indeed many design decisions in auditory graphs are mostly based on intuition. For the conceptual model Nees and Walker (2007) discuss the relevant literature on basic auditory perception, which according to them supports a number of common design practices. The general design process diagram for PMSon presented in this thesis in Figure 2.3 provides a supporting structure for future decision making. Nees and Walker (2007) mention that occasionally spectral characteristics such as brightness have been used in auditory graphs: however they restrict their discussion to the still most popular and dominant approach of mapping function values to frequency/pitch. Auditory tick-marks and the indication of

4. MAPPING AND INTERACTION IN AUDITORY GRAPHS

quadrant transition are also discussed in Nees and Walker (2007) as auditory context in an auditory graph.

4.1.1 Existing Applications

Although not all questions in the field of auditory graphs are resolved, some prototypes exist already. The most complete is the java program *MathTrax* by Shelton et al. (2006)¹. *MathTrax* is an application with a user interface tailored to typical requirements for the blind and partially sighted: It works together with screen readers and features shortcuts and hot-keys for efficient navigation. *MathTrax* also provides descriptive information about the mathematical functions to be analyzed. Several functions can be sonified at the same time, which can also be multiple derivatives of one function. Despite the clearly impressive features of *MathTrax*, the sound design relies exclusively on the dominant principle of pitch/frequency mapping. This poses a problem, if the user wants to listen to several functions at the same time. In this case each sound stream exhibits a differently varying pitch requiring highly analytic listening skills to decode the information, a problematic design choice as identified by Flowers (2005).

4.1.2 New Developments

The results from the field of auditory graphs provide a good basis for function sonifications of one variable $f(x)$. We believe however that there are still possibilities for further improvements. Rather than delivering additional information about the same function in separate auditory streams, the usefulness of auditory graphs can be extended by integrating information about the shape of $f(x)$ in one auditory stream. Further, auditory rich information can be more interesting to listen to and can provide perceptually distinguishable dimensions when the embodied aspects of listening intentions are addressed.

If we want to create an auditory display that provides “*means of understanding line graphs in a manner similar to sighted users*”, compare Mansur et al. (1985), we need to analyze first what contributes to the understanding of visual plots. How can principles of visual gestalt perception be tied to mathematical properties of the curve? The idea to use analytically established properties of an object in order to describe its shape can be traced back to Birkhoff (1933), where he defines the *aesthetic measure* of an object to be the ratio of order and complexity. Birkhoff applies his idea to prototypical objects of aesthetic interest like vases and describes the associated order through the analytical properties of the enveloping function that defines its shape. Shapes of a

¹The browser based application *MathTrax*: <http://prime.jsc.nasa.gov/mathtrax/>

4.1 Literature Review on Auditory Graphs

mathematical function, which facilitate our holistic understanding of it, are perceptual units and do not fall apart into constituent analytical elements like the function and its derivatives. Therefore, our goal for *Sonic Function* and *Singing Function* was to design a single auditory stream of articulated auditory variations according to function shapes. The *aesthetic measure* of Birkhoff describes static objects, the medium of sound however is time based. This connects the challenge to translate function shapes into sonic shapes in an interesting way with the notion of the sound object as discussed in Section 2.2.7.2. Hence, we developed articulated variations using textural and timbre based characteristics as sonic dimensions in order to translate the notion of function shapes into a sonic form. We hoped that this would allow to perceive and identify graph characteristics more independently of the time based progression along the x-axis.

As a new methodological contribution to the field beyond existing mapping strategies focusing on pitch, we introduced in Grond et al. (2010) a multi-parameter sonification of mathematical functions. In detail, we utilize the Taylor expansion of function f at location x in order to create unique sonic representations for $f(x)$ making the perception of the shape more independent from intervals as increments of x . We suggest to take advantage of the first m terms of the Taylor series ($f(x), f'(x), \dots, f^{(n)}(x)$) at location x as a *fingerprint* for the local characteristics of the function and map it to corresponding sonic features of a single sound stream.

For the concrete mapping of function values $f(x)$ and its derivatives to sonic properties, the main association of $f(x)$ to pitch can be maintained, as it seems natural and was reported being successful in Flowers (2005). Additionally we propose to reflect slope as pulse rate, and curvature $f''(x)$ as attack time of events if we consider for instance a discrete mapping.

In the first evaluation of *Sonic Function*, we restricted the multi-parameter-mapping approach to fit the subject group of pupils. Since the concept of higher derivatives are difficult to grasp and not part of the curriculum for pupils of the age of our test subjects, we included only the first derivative. In the second application, *Singing Function*, we implemented the described approach for all three derivatives by taking advantage of the vowel synthesis building block that was described in Section 2.4.

As stated in Hermann and Hunt (2005) interaction introduces new and exciting possibilities for a better understanding of sonification. For auditory graphs, this usually means to freely navigate along the x -axis. Interaction in auditory graphs mostly serves an exploratory purpose. Interaction also offers means to point into the graph in order to communicate perceived features in it. In the empirical study of *Sonic Function*, pupils used this interaction possibility in order to select specific points of interest such as turning points, saddle points or local optima. In the empirical study of *Singing Function*

4. MAPPING AND INTERACTION IN AUDITORY GRAPHS

we looked at the discrimination rate within a family of functions for interactive and non interactive sonifications.

Apart from the main concept of multi-parameter-mapping into one auditory stream, in Grond et al. (2010) we also contributed to the field by developing an auditory bounding box – an auditory context element within the conceptual model of auditory graph comprehension from Nees and Walker (2007).

4.1.3 Evaluating Auditory Graphs

The interdisciplinary aspects that need to be taken into account when developing auditory graphs imposes the challenge that there is no single methodology for their evaluation. Mansur et al. (1985) for instance, focus on the identification of mathematical properties like symmetry slope and monotonicity. On the other side, Bonebright (2005) asks sighted subjects to match the sound of the auditory graph with the visual representation of it.

Auditory graphs of mathematical functions often contain a mixture of explicit and implicit auditory display elements. In *MathTrax* for instance, the transition from positive to negative function values is indicated by white noise that is played in parallel with the sonification of the negative function values. Inspired through visual plot boundaries, we introduced in Grond et al. (2010) an acoustic bounding box, which is an equally explicit element. However, being rendered within the stereo panorama along the x -axis, it is at the same time implicit indicating the interaction progress. A short sound event similar in its function, are the auditory tick-marks along the x -axis indicating y -axis crossing. The evaluation of the explicit elements is mostly related to whether they can be perceived simultaneously with the auditory graph and if their meaning is understood. This was part of the qualitative evaluation of *Sonic Function*.

It is important that auditory graphs are tested in a realistic scenario with the target audience. This is why our blind colleague Trixi Droßard tested *Sonic Function* with blind and visually impaired math students. In this evaluation we were mostly interested if they could identify the explicit elements in the graph but also if they developed an understanding for shapes as they are found in extrema.

Many auditory graphs are difficult to reproduce exactly based on their description. As a result evaluations and experiments with auditory graphs hardly ever compare new developments with old ones. We hence decided to evaluate the multi-parameter-based sonification of *Singing Function* together with a pitch mapping version of it. This allowed to compare multi-parameter-mapping with this popular approach.

The evaluation of *Sonic Function* had shown us that the understanding of concepts like extrema and singularities influence the interpretation of the auditory graph.

Therefore we set up for *Singing Function* a psychophysical experiment focusing on the question if multi-parameter-mapping adds perceptual contrast by comparing members of a function family with similar shape with respect to $f(x)$ but different derivatives.

Since the multi-parameter-mapping in *Singing Function* included the second derivative – a mathematical concept that the students had not covered yet – the order in which the development took its course was somehow reversed. From an ethnographic top-down design approach however, it was useful to see that the simpler version *Sonic Function* was appreciated in practice and hence we tested in *Singing Function* the auditory contrast with sighted users but without a visualization of the graph. We believe both studies provide a good basis for future longitudinal evaluation steps with respect to usability and pedagogical questions.

4.2 Sonic Function

Based on the publication Grond et al. (2010), I describe in this section *Sonic Function*, giving details about its implementation, mapping and sound design. *Sonic Function* is implemented in python and Tcl/Tk in order to provide a minimal user interface. Open Sound Control (OSC) provides the protocol to send the parameters to the SC3 sound server scsynth. The keyboard serves as an input device, since it is a very familiar interface for the visually impaired. The user can interact with the program through the following keys:

The *up* and *down* arrow keys adjust the volume of the sonification. Although volume control is trivial, it is important to adjust it independently for *Sonic Function*, since screen readers, such as Jaws for Windows often run in parallel and hence volume levels need to be matched. The *left* and *right* arrow keys allow the user to navigate along the x-axis. If the keys are pressed constantly, the function is browsed in a continuous movement from left to right and right to left respectively. While navigating the function on the x -axis the sonification is presented on the corresponding position within the headphone stereo panorama. The consecutive keys on the bottom row of the characters x , c and v set the step size on the x -axis to $1/30$, $1/10$ and $1/6$ respectively. This allows the user to switch between a quick overview over the function and a more detailed inspection. The keys h , t , n and a for (German: Hochpunkt, Tiefpunkt, Nullstelle, Ordinatenabschnitt) set markers for maxima, minima, $f(x) = 0$ and $x = 0$ respectively. The interaction for placing markers is included since we want to evaluate the sonified function, by recording and analyzing user interaction. These markers together with the movement on the x -axis are registered in a protocol file. The number keys 1 to 6 correspond to 6 different test functions. During the evaluation, these keys are operated by Trixi Drossard in order to test the understanding of features across different functions.

4. MAPPING AND INTERACTION IN AUDITORY GRAPHS

4.2.1 Mapping and Sound Design

The sound design has two goals. One is to provide appropriate feedback for the interaction through a discrete sonification of $f(x)$. The second goal is to create a continuous sonification which represents the dense real values on the x-axis. The auditory graph can be roughly divided into two parts, the first mostly dealing with implicit sonic information about the function, its shape and the interaction feedback. The second part deals with the more explicit aspects of the auditory graph, i.e. the indication of positive and negative function values as well as the auditory bounding box.

4.2.1.1 Interaction Feedback and Function Shape

For each step, the user makes along the x-axis, the discrete sonification is rendered in the respective position on the stereo panorama. SC3 code for the SynthDef of the discrete sonification is given in Figure 4.1.

```
SynthDef(\discrete,  
  {|note, pan, delay, duration, vol, cutfreq|  
    var freq, harm, amp, ring, srcEnv, noise, noteEnv, source, klank, sig;  
    freq = note.midicps;  
    harm = {|i| i+1}!9;  
    amp = ({|i| i+1}!9).reverse.normalizeSum;  
    ring = {|i| i+1}!9;  
    srcEnv = EnvGen.ar(Env.new([0,0,1,0],[delay, 0, duration],-3), 1, doneAction:0);  
    noise = EnvGen.ar(Env.new([0,0,1,0],[delay, 0, duration/20],-3),1, doneAction:0);  
    noteEnv = EnvGen.ar(Env.new([0,1,0],[0.0,delay + 0.5],-3), 1.0, doneAction:2);  
    source = srcEnv + (attackNoise * ClipNoise.ar(1));  
    klank = Klank.ar([harm, amp, ring], source, freq);  
    sig = LPF.ar(klank, freq * cutfreq) * noteEnv;  
    OffsetOut.ar(0, Pan2.ar(sig * AmpComp.kr( freq, 40.midicps ),pan, vol))  
  }).send(s)
```

Figure 4.1: The SC3 synthesis definition for the discrete sonification.

The discrete sonification produces a sound reminiscent of a synthetic instrument. The basis of this subtractive synthesis approach is the Ugen `Klank.ar`, which is an assembly of overtones realized as `Ringz.ar` filters, which are implemented based on Steiglitz (1994). These partials can be identified in Figure 4.3 annotated as A. The base frequency can be set by the argument `note` and the overtones have all decaying gain corresponding to the array of amplitudes in `amp`. The lowest MIDI note to which function values $f(x)$ is mapped is 30 and the highest 77. This corresponds to a frequency range covering approx. 46.25 to 698.46 Hz, covering about 4 octaves. This considerably low range was chosen in order to have enough overhead in the spectrum for the 9 overtones, so that each of the sounds has a similarly perceivable pitch and richness. The excitation of the `Klank.ar` filter is an attack/decay envelope with a noise component

in the attack phase. The resulting sound is shaped through an envelope which also has an attack/decay characteristic.

The sound is played after a small delay. This allows the continuous sonification to ramp to the target frequency. The discrete sonification of $f(x)$ is played back within the stereo panorama corresponding to the actual position on the x axis within the bounding box. The gain of the signal is adjusted through basic psychoacoustic amplitude compensation `AmpComp.kr` according to the base frequency.

The continuous sonification resembles the discrete sonification with respect to spectral characteristics. Differently to the discrete sonification, it is implemented as additive synthesis using the unit generator `Klang.ar`. The corresponding SC3 SynthDef is depicted in Figure 4.2.

```
SynthDef(\continuous,
  {|note, pan, vol, cutoffreq, modf, moda|
  var freq, harm, amp, ring, srcEnv, noise, noteEnv, source, klang, sig;
  freq = note.midicps;
  harm = {|i| i+1}!9;
  amp = ({|i| i+1}!9).reverse.normalizeSum;
  phase = ({|i| i+1}!9).reverse.normalize;
  klang = DynKlang.ar(`[harm, amp, phase], freq.lag(0.1) );
  sig = LPF.ar(klang, freq * cutoffreq) * SinOsc.kr(modf, 0, moda, 1);
  OffsetOut.ar(0, Pan2.ar(sig * AmpComp.kr( freq, 40.midicps ),pan, vol))
  }).send(s)
```

Figure 4.2: The SC3 synthesis definition for the continuous sonification.

In addition to the $f(x)$ pitch mapping, the continuous sonification is also the carrier of the information about the derivative. $f(x)/dx$ is mapped to a low frequency amplitude modulation. The modulation of the amplitude approaches 0 if the derivative approached 0. If the slope of the function was high a noticeable modulation of the amplitude was setting in. This amplitude modulation corresponds to the detail `SinOsc.ar(modf,0,moda,1)` in Figure 4.2. In Figure 4.3, this amplitude modulation can be found in the split of the lower frequencies, in regions with a high slope, which is annotated as B.

4.2.1.2 Auditory Context Information

There are further audible elements in *Sonic Function* which are part of the auditory context: the differentiation between positive and negative $f(x)$, the acoustic bounding box and auditory tick-marks.

The difference between positive and negative $f(x)$ is indicated through the brightness of the sound through a lowpass filter, `LPF.ar`. The cutoff frequency is set to two different values - 5 or 2.5 times the base frequency in order to synthesize sounds of

4. MAPPING AND INTERACTION IN AUDITORY GRAPHS

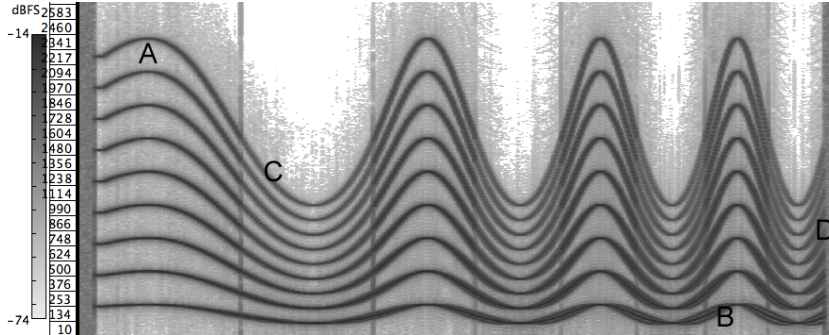


Figure 4.3: Spectrogram of the combined stereo channels for the continuous sonification of the auditory graph for function of equation 4.5. $f(x)$ is navigated from left to right along the x-axis. For a description of the visible mapping features see text.

two different brightness modes. The brighter sound indicates positive function values. In Figure 4.3, the attenuated higher partials can be found at the regions around the minima, annotated as C.

The acoustic bounding box indicates whenever the interaction on the x -axis either reaches the negative x limit on the far left or the positive x -limit on the far right. These limits are indicated through noise played back on the corresponding stereo channels. The acoustic bounding box further indicates if the $f(x)$ exceeds negative or positive limits. For the upper and the lower limit of the bounding box, noise is sent through a band pass filter `BPF.ar` with a low (200 Hz) and a high (5000 Hz) center frequency, respectively. The noise source is played back on the actual x position within the stereo panorama. For all the functions the bounding box is set from -10 to $+10$ in x and -5 to $+5$ for $f(x)$. In Figure 4.3, the auditory bounding box limits on the left and right can be seen as white noise, annotated on the right as D.

The acoustic bounding box appeared to be instructive at singularities. When navigating the function $f(x) = 1/x$ from left to right for instance, the auditory graph first exhibits a descending frequency. When the graph hits the lower limit, the bounding box is played back. At the singularity, the center-frequency of the bounding box changes to high. Finally, the function is audible again at high frequencies, descending towards the right.

Tick-marks on the x -axis are indicated as audible clicks synthesized through short envelopes over an additive synthesis of 4 overtones with a base frequency of 1.000 Hz. The high frequency range is used in order to distinguish the tick-marks from the lower frequency range of the discrete and continuous sonification. The tick-marks are played within the stereo panorama according to their position on the x -axis. The tick-marks at the centre of the bounding box $x = 0$ have an elevated base frequency of 1.600 Hz.

4.2.2 User Study

In order to test *Sonic Function* with our targeted user group in a teaching situation our collaborator Droßard conducted a user study with 14 students from the Carl-Strehl-Schule in Marburg, Germany. Fourteen students participated in the study (7 female, 7 male) blind and partially sighted with an age range from 17-19. From the participants seven were stated to be blind, four to be partially sighted, and three high-grade partially sighted. For further demographic details of the participants see Grond et al. (2010).

4.2.2.1 Qualitative Evaluation of Test Functions

The evaluation was conducted with each student individually in a quiet room. The sound was rendered through regular headphones. Droßard instructed the students how to use the program *Sonic Function*. The students were encouraged to ask everything they needed to know in order to make meaning of the sounds and the interaction possibilities of *Sonic Function*. The instructor made sure that all sonic features were understood, which were relevant for the tasks.

The participants reported verbally what kind of features they encountered when browsing a selected function. For the evaluation and transcription of the verbal feedback, compare (Droßard, 2010, page 111). When they encountered values for $x = 0$ or $f(x) = 0$, minima, or maxima and reported their findings to the conductor, Droßard set a marker through the keys on the keyboard thereby creating an entry in the log file. Further, the students were asked to describe with words the shape of the function that they had heard. In Figure 4.5 the trajectory of the interaction along the x -axis for the function from Equation 4.5 are depicted for three participants.

After the session each participant was asked for feedback about *Sonic Function*. Further, they were asked to describe what kind of learning type they are (visual, auditory or haptic). According to their description, ten students were visual learners and four auditory learners. In this context, visual learners generally prefer haptic representation for instructions.



Figure 4.4: Photo from the *Sonic Function* user study from Grond et al. (2010): the test subject sits in the foreground on the right following the instructions by coauthor Droßard in the background on the left.

4. MAPPING AND INTERACTION IN AUDITORY GRAPHS

The participants were presented with the functions from Equation 4.1 - Equation 4.6. The choice was made by Droßard with a focus primarily on pedagogical aspects of function analysis and the respective knowledge of the age group of the participants. The functions were presented to all participants in the same order. This unbalanced choice, with respect to the stimuli presentation, was mostly due to pedagogical reasons.

$$f_1(x) = \sqrt[3]{4} (x + 1)^2 - 2 \quad (4.1)$$

$$f_2(x) = 2x + 3 \quad (4.2)$$

$$f_3(x) = x^2 + 1 \quad (4.3)$$

$$f_4(x) = 0.5/x \quad (4.4)$$

$$f_5(x) = 1.5 \cdot \sin((0.2x + 3)^2) \quad (4.5)$$

$$f_6(x) = 1/\sin(x) \quad (4.6)$$

The main purpose to include Equation 4.1 is to introduce the test-subjects to all the audible features of *Sonic Function*. All x and $f(x)$ values of the parabola cover all quadrants within the bounding box. The participant can hence hear the tick-mark for $x = 0$, the different timbre for negative and positive function values $f(x) = 0$. At the minimum at $x = -1$, the low frequency oscillation vanishes, due to the derivative $df_1(-1)/dx = 0$. The purpose of Equation 4.2 is to verify if the point for $x = 0$ is understood by the test subjects as well as the audible brightness transition of $f(x) = 0$ at $x = -3$. The identification of the minimum and the position with $x = 0$ is the task in the third function, the symmetric parabola from Equation 4.3. Equation 4.4 helps to evaluate whether test-subjects are able to make sense of an audibly represented singularity. Equation 4.5 is used because it is an interesting test case of how the extrema identification depends on the curvature i.e. the acoustic contrast around $df(x)/dx = 0$, when navigating along x . Equation 4.6 is included to find out how minima and maxima - located at $\pi/2 \cdot m$ with $m \in \{-5, -3, -1, 1, 3, 5\}$ - are perceived between singularities.

The test case functions together with the recorded markers for $f(x) = 0$ and $x = 0$ are depicted in Figure 4.6 on the left, the markers for minima and maxima in the same figure on the right.

4.2.2.2 Discussion of Figure 4.5 and Figure 4.6

In Figure 4.5, we find the interaction patterns of different qualities for 3 participants. All three participants navigate the function at almost constant speed back and forth at least once. The interaction plot from the top shows the data of a participant who

4.2 Sonic Function

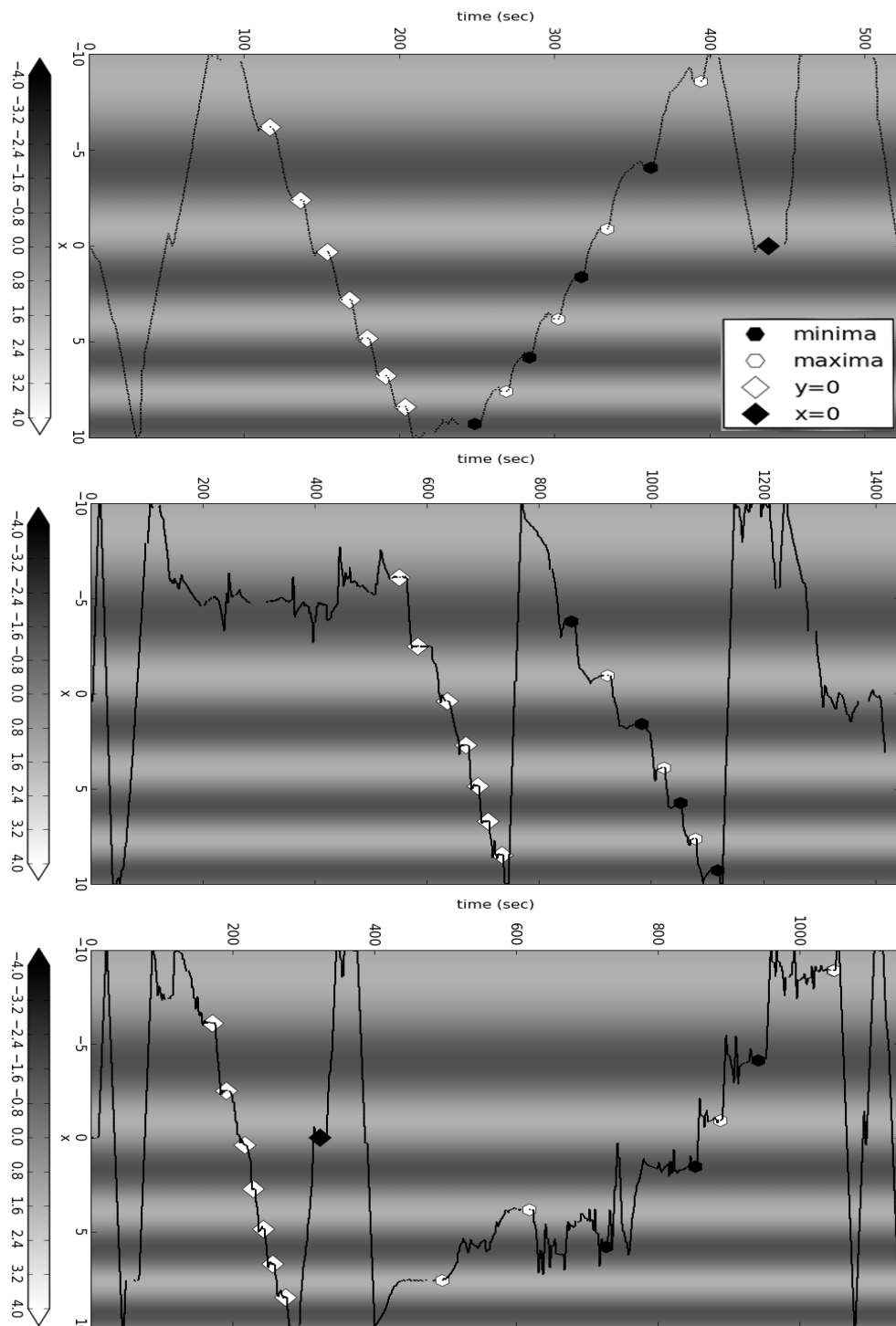


Figure 4.5: Example of typical explorations of $f_5(x)$ from Grond et al. (2010) and Droßard et al. (2012). The function values $f_5(x)$ are encoded in grey. The participant started in the middle and explored the function to both limits of the bounding box. Then $f(x) = 0$, further extrema and finally, $x = 0$ were marked.

4. MAPPING AND INTERACTION IN AUDITORY GRAPHS

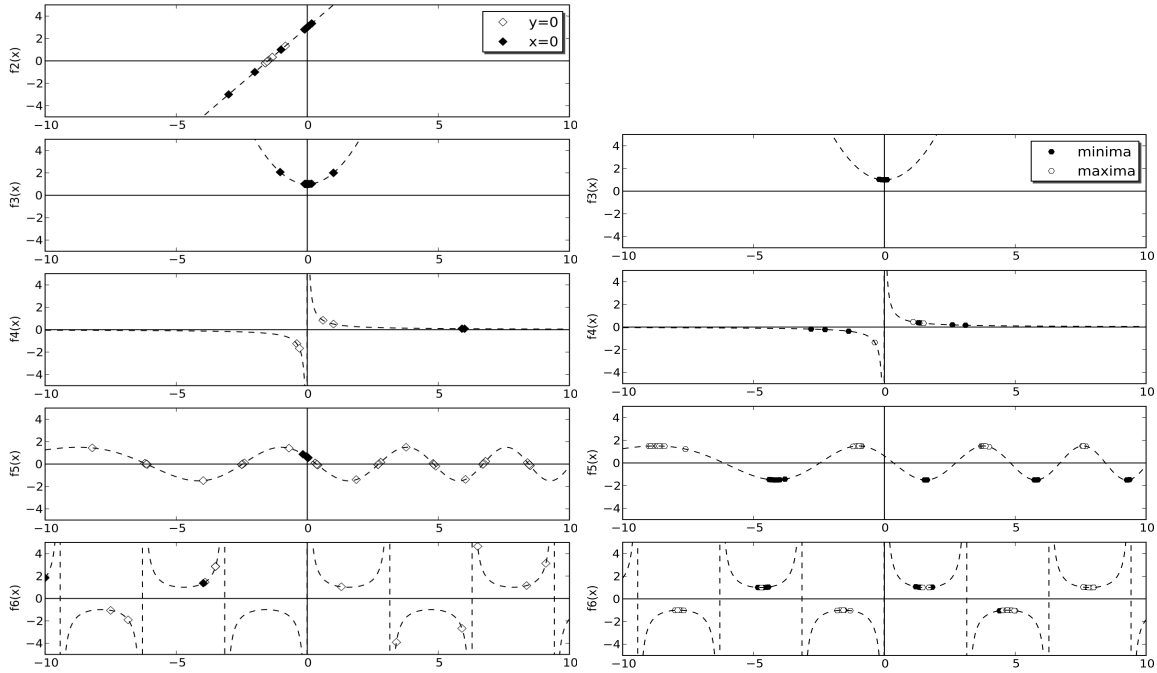


Figure 4.6: Plot of the test functions with the $f(x) = 0$ and $x = 0$ markers on the left. On the right are the same functions with the markers for $f(x)/dx = 0$. Both plots are taken from Grond et al. (2010)

picks up clearly on the audible cues and finds the requested features almost always at the moment when they occurred. We find back and forth navigating for confirming the audible feature mostly for extrema with low curvature. The interaction plot in the middle shows more insecurity when detecting the extrema. The explicit audible cues are however still easy to identify as is also the case for the interaction plot at the bottom. When comparing it with the other two, we find in this last plot the most insecurity with lots of back and forth navigation in order to identify the extrema.

In Figure 4.6, we find the markers of the participants for $x = 0$ and $f(x) = 0$. For $f_2(x)$ most of the markers are placed around $x = 0$ and $f(x) = 0$. There are outliers for $x = 0$ indicating that ordinary tick-marks on the x-axis seem to be mistaken for tick-mark at $x = 0$. Similar problems can be found in $f_3(x)$. Here again, most of the markers are placed correctly around $x = 0$, except for two outliers are found. In $f_4(x)$ none of the concepts $x = 0$ or $f(x) = 0$ are explicitly present. This poses apparently a conceptual challenge, some markers were placed, where $f(x)$ approaches 0. Some markers are placed where the function exhibits the strongest curvature. The plot for $f_5(x)$ shows how the participants become familiar with the auditory graph and how most succeed in identifying the tested positions. The function $f_6(x)$ is similar to $f_4(x)$, as neither $x = 0$ nor $f(x) = 0$ are present and participants show no particular pattern in the positioning of the markers.

Markers for minima and maxima are compiled in Figure 4.6 on the right: The minimum in $f_3(x)$ is identified by all participants. In $f_4(x)$ extrema are wrongly identified where the function approaches the x-axis. This seems to be confusing in a similar way as is the testing for $x = 0$ and $f(x) = 0$. In $f_5(x)$ the broader distribution of the markers at extrema with lower curvature are noteworthy. In $f_6(x)$ extrema were well identified, however, the concept of minima and maxima was confused.

For a detailed listing of the more exploratory statistics of the markers, see Grond et al. (2010).

4.2.2.3 Curvature and Perceptual Contrast

In contrast to the qualitative aspects, the function $f_5(x)$ is a particular case, because here a hypothesis can be formulated and tested. We suspect that the exact location of extrema depends on the perceptual contrast, i. e. the curvature at the extrema. This is why we take a closer look on function $f_5(x)$ and compute $\frac{df_5(x)}{dx}$ and $\frac{d^2 f_5(x)}{dx^2}$ as given in Equation 4.7 and Equation 4.8.

$$\frac{df_5(x)}{dx} = \frac{3}{25} (15 + x) \cos \left(\left(3 + \frac{x}{5} \right)^2 \right) \quad (4.7)$$

$$\frac{d^2 f_5(x)}{dx^2} = \frac{-3}{625} \left(-25 \cos \left(\left(3 + \frac{x}{5} \right)^2 \right) + 2 (15 + x)^2 \sin \left(\left(3 + \frac{x}{5} \right)^2 \right) \right) \quad (4.8)$$

We obtain values for x within the interval from -10 to 10 by using numerical methods¹ to solve the equation $\frac{df_5(x)}{dx} = 0$. The values for x together with the corresponding curvature are compiled in Table 4.1.

| | | $df_5(x)/dx = 0$ | $d^2 f_5(x_i)/dx^2$ |
|-------|-----|------------------|---------------------|
| x_1 | max | -8.733 | -0.377 |
| x_2 | min | -4.146 | 1.131 |
| x_3 | max | -0.988 | -1.885 |
| x_4 | min | 1.580 | 2.639 |
| x_5 | max | 3.799 | -3.393 |
| x_6 | min | 5.784 | 4.147 |
| x_7 | max | 7.594 | -4.901 |
| x_8 | min | 9.270 | 5.655 |

Table 4.1: Extrema and curvature values for $f_5(x)$, taken from Grond et al. (2010)

¹Such as damped Newton's Method, as implemented in the software package Mathematica.

4. MAPPING AND INTERACTION IN AUDITORY GRAPHS

| | correlation coefficient | p-value |
|-------------------------------------|-------------------------|---------|
| $ curvature $ versus σ | -0.7381 | 0.0366 |
| $ curvature $ versus $\hat{\sigma}$ | -0.6905 | 0.0580 |

Table 4.2: Results from the Spearman rank correlation test from Grond et al. (2010)

In Table 4.1, the decreasing standard-deviation of the extrema along the x-axis seems to correspond to the increasing absolute curvature of the extrema. Figure 4.7 from Grond et al. (2010) depicts a correlation plot of the absolute value of the curvature against the standard-deviation σ . Further, the absolute value of the curvature is also plotted against the standard deviation of $(x_i^k - x_0^k)$ denoted as and $\hat{\sigma}$. The term x_0^k denotes the exact position of the extremum.

Table 4.2 shows the Spearman rank correlation coefficient together with the two-sided p-value. The null hypothesis for the test is that the two data sets are uncorrelated. Only the correlation with σ is below, when accepting a threshold for significance

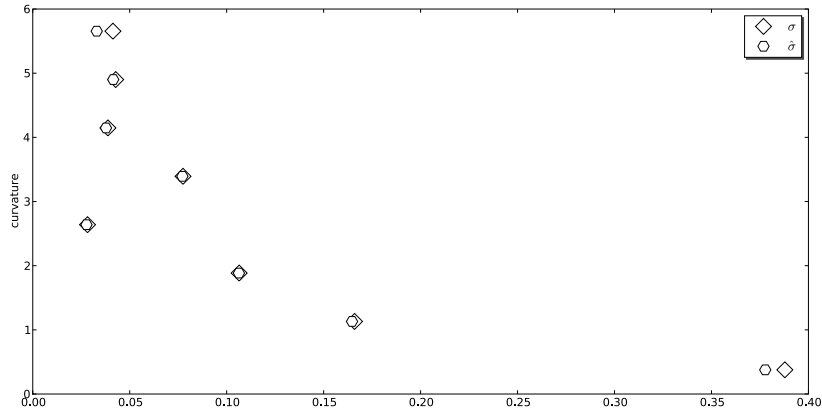


Figure 4.7: Curvature versus the σ and $\hat{\sigma}$ from Grond et al. (2010). Low curvature leads to a broader distribution of clicks around the function extrema.

of 5%. Given however the low amount of data (8 extrema), we think that it is reasonable to assume that the precision with which extrema are spotted depends on curvature and the resulting acoustic contrast.

4.2.3 Conclusions on *Sonic Function*

Sonic Function was presented to students in order to teach them about aspects from function analysis like extrema and axes crossing. This preliminary feedback on multi-parameter-mapping in auditory graphs in a real-world teaching situation is useful in many ways. Looking at the results from the analysis of the markers set by the participants, we find that not all mathematical concepts were properly understood. Nonetheless, spontaneous feedback from the subjects was mostly positive: 9 out of 14 said they would like to use *Sonic Function* in physics or math courses (see Droßard et al. (2012)).

The account of the feedback from the subjects (Droßard, 2010, page 121) leads to the following overall conclusion. If subjects could name and identify a sonic quality, such as brightness, it was appreciated as informative. Otherwise it was perceived as something that needs to be learned to become an informative aspect of the auditory graph.

Two different categories of auditory information were available in *Sonic Function*. Students generally performed well in terms of precisely spotting explicit auditory information, such as tick marks and transitions between positive and negative function values. Implicit auditory information as in extrema posed two challenges: In function $f_6(x)$, we find that students confuse the definition of minima and maxima, although they perform well with respect to this question in $f_5(x)$. This might be explained by the fact that the pitch of the maxima was lower than the one for minima and that all extrema were separated by singularities. Hence maxima and minima have partially not been related to the sign of the curvature at the extrema but to the absolute function value at $df_6/dx = 0$.

As shown in the case of $f_5(x)$, the precision with which extrema can be localized depends on the acoustic contrast (the curvature around the extremum). Interestingly, none of the participants reported explicitly the increase of frequency while navigating $f_5(x)$ along the x-axis. Since the functions were explored interactively, the navigation along the x-axis was not necessarily constant. Hence, the frequency of oscillations between the extrema could not be properly interpreted by the participants.

Sonic Function encourages us to proceed according to the new Taylor-series-based multi-parameter-mapping concept. Due to the unpredictability of how the x -axis is explored, a constant rate in the progression of the function along the x -axis cannot be assumed, which makes it difficult to decode information such as curvature. We hope that the integration of further derivatives in the same auditory stream enhances the sonic fingerprint and helps to address this challenge. Further, integrating information from the auditory context in the sonification, such as the transition from negative to positive function values as timbre filter, is promising, if solutions can be found that are easier to recognized. This makes space in the auditory context for other kinds of information such as an auditory bounding box.

It is difficult to measure the usefulness of *Sonic Function* in a quantitative manner, which is mostly due to the varied sight capabilities of our test subjects. Additionally, the students were not previously familiar with the idea of auditory graphs. The pedagogical setting was interesting because it shows that auditory graphs are in principle appreciated by the target user group. However, the limited a-priori knowledge about function analysis made it difficult to assess if a certain concept is not understood or if the auditory feature was not heard. We plan to make future evaluations with subjects who are either already familiar with function analysis, or design the evaluation such that it does not depend on the knowledge of specific mathematical concepts.

4.3 Singing Function

In this section, I present *Singing Function* published in Grond and Hermann (2012b), a further development of the multi-parameter-mapping approach that we took in *Sonic Function*. Let us reconsider the visual interpretation of the graph from the introduction of this chapter, which at first sight, seems to be nothing but a line connecting $(x, f(x))$ value pairs. At a closer look, these $(x, f(x))$ value pairs are the basis for the perception of various *gestalts*, without which we would not be able to visually interpret a function graph and distinguish it from others.

The findings from Mansur et al. (1985) show that auditory graphs allow us to recognize symmetry, slope and monotonicity. These are obviously important characteristics of a graph, however these audible features do not suffice to provide the same holistic understanding like the perception of visual *gestalts* in graphs.

Important features of a function which need to be translated into auditory graphs are the variations in slope and curvature, which are characteristic for turning and saddle points. Another important characteristic that we can see is whether a function exhibits high or low curvature at extrema, making it look more or less pointy. The fact that the whole visual graph is static and persistent makes these features visually accessible and hence they can be interpreted.

For auditory graphs however, the situation is different. Curvature for instance can be understood only indirectly. When moving along the x-axis, the only audible feature is how fast an extremum is reached and left behind. The listener encounters even more difficulties, when an auditory graph is explored interactively. Whether a feature of the graph becomes audible as a *gestalt* or not depends on navigation-speed as discussed in Harrar and Stockman (2007).

For *Sonic Function* by Grond et al. (2010), we have already introduced the new methodological contribution to the field, by going beyond the dominant pitch mapping-based strategies and by introducing *multi-parameter* sonification. The main focus in this new approach is to utilize the Taylor expansion of function f at location x in order to develop stationary sonic representation that captures the local characteristics of the function shape.

The mapping strategy in *Sonic Function* employed only $f(x)$ and $f'(x)$. For *Singing Function*, we also include $f''(x)$ using a vowel-based synthesis approach taking advantage of the rich and fairly well controllable perceptual space of vowel sounds.

As already discussed, vowel based sound synthesis has been explored for sonification in various cases. The intuitive understanding of vowels is the benefit which has already been pointed out by others, see Section 2.4. This is of particular importance in pedagogical contexts, because the vocal articulation of a graph can be captured without

already knowing what it means. The possibility to discriminate vowels from each other makes it easy to learn sonic fingerprints. From a sound synthesis perspective, vowels give access to a continuous dimension in timbre space, which is to some degree orthogonal to pitch and loudness. Most important, discrimination does not only depend on the perception of the vowel but also on the possibility of its reproduction, as discussed in Section 2.4.

In *Sonic Function*, the auditory graph contained some explicit elements of auditory information such as tick-marks. In *Singing Function* we focus less on those elements and address the following questions as formulated in Grond and Hermann (2012b):

- Can we create an auditory graph which brings forth a *sonic gestalt* which helps to distinguish features of mathematical functions beyond those of slope, monotonicity and symmetry?
- How sensitive are these *sonic gestalts* with respect to the interactive exploration of the auditory graph? Can these *sonic gestalts* help to identify function characteristics, can they make them independent of a uniform scan along the x -axis?
- How can an evaluation for these *sonic gestalts* take the already popular pitch-based mapping strategy into account, and compare the two?
- Will subjects report to use their embodied knowledge to interpret a vocally articulated sound?

4.3.1 Mapping and Sound Design

For the multi-parameter-mapping of the function value $f(x)$ and its first two derivatives $f'(x)$, $f''(x)$ we pursued the scheme from Grond and Hermann (2012b). The hierarchical combination of the mappings for (f) , (f, f') and (f, f', f'') , resulted in three different sonification types (ST). This can be seen as an example of the multiple-views paradigm from Stockman et al. (2005). Although we discussed the shapes as perceptual units, each view can be thought as adding perceivable details to the graph:

- **Function value $f(x)$:** The main association of $f(x)$ to pitch was maintained for two reasons: First it seems natural to associate low function values with low pitch. Second, this mapping on its own without further derivatives gives a classic pitch-based sonification which we can compare with the proposed Taylor series expansion mapping. Additionally a pitch variation with a constant spectral envelope keeps the sound comparable with respect to the sonic vowel qualities. For the frequency range we selected one octave from 110 to 220 Hz. Two reasons let us

4. MAPPING AND INTERACTION IN AUDITORY GRAPHS

to choose this low range: First this frequency range is typical for human speech. Second, the low fundamental gives a narrowly spaced overtone series. This narrow support of the spectral envelope helps to recognize the formant shapes.

- **Slope $f'(x)$:** For the slope, the absolute value of the first derivative was mapped to the transition of vowel [a:] to [i:] both in the bass register, which was chosen because of the rather deep fundamental frequencies. The transition between the spectral envelopes was realized with the blend method described in Section 2.4. The following consideration supported the choice: the first two formant center frequencies go roughly in antiparallel directions when moving from the vowel [a:] the two vowels [i:]. The vowel [e:] is approximately in between of both. This relation between [a:] [e:] and [i:] can be seen in Figure 2.4 on the top left, which makes the contrast along this vowel scale particularly perceptually salient. As argued in Grond and Hermann (2012b), the change in the format centre frequencies corresponds to the increasingly closed vocal tract during the transition from [a:] to [i:]. The alternative transition [a:] [o:] [u:] which also goes with an increasingly closed vocal tract, shows less variation in the formant center frequencies and hence exhibits less perceptual contrast. We hoped that the high perceptual contrast in the transition from [a:] to [i:] could potentially stimulate the listeners involvement.
- **Curvature $f''(x)$:** From the second derivative we took the absolute value and mapped it to the brightness of the vowel sound. The brightening was realized through the corresponding methods in the Vowel class (compare method *brightenRel(b, ref)* in Section 2.4.1 on page 43). As an auditory result, high curvature sounded brighter and a bit louder than low curvature, which metaphorically reflects more energy that could be associated with the second derivative being acceleration. Perceptually, the brightness value shifted the spectral centroid but the pitch of $f(x)$ was still recognizable.

As a result of this mapping, a recognizable auditory gestalt was building up, when moving along the x -axis across turning points, saddle points or local optima. Since we did not want to include explicit auditory information we chose to map the absolute value for $f'(x)$ and $f''(x)$. We will later discuss the sonic effects of the proposed mapping with Figure 4.9.

4.3.2 Evaluation

We used a function family depending on one parameter a for the evaluation of *Singing-Function*. The functions of this family exhibit a similar shape in $f_a(x)$ (see Equation 4.9). In its derivatives $f'_a(x)$ (Equation 4.10) and $f''_a(x)$ (Equation 4.11), however different shapes develop depending on the variation of a .

$$f_a(x) = \frac{\tanh(a \sin(x))}{\tanh(a)} \quad (4.9)$$

$$f'_a(x) = \frac{a \cos(x) b}{\tanh(a)} \quad (4.10)$$

$$f''_a(x) = \frac{-(2a^2 \cos^2(x) \tanh(a \sin(x))b + a \sin(x)b)}{\tanh(a)} \quad (4.11)$$

$$b = \operatorname{sech}^2(a \sin(x))$$

For small a , functions from this family are equivalent to $f(x) = \sin(x)$. As amplitude increases through a , $\tanh()$ acts like a wave shaping function. For very big values of a , $f_a(x)$ approaches a rectangle shape. 4 instances of $f_a(x)$ with the parameter $a = \{0.1, 1.0, 1.5, 2.0\}$ were the basis of our experiment, which are depicted together with their derivatives in Figure 4.8.

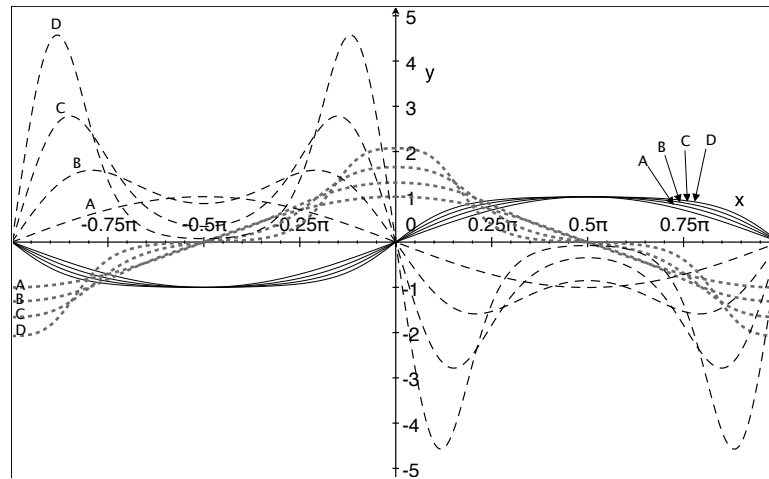


Figure 4.8: Plot of the curves from the functions in eqn. 1, 2 and 3. from Grond and Hermann (2012b). The straight line represents $f_a(x)$, the thick grey dashed line $f'_a(x)$, the thin black dashed line $f''_a(x)$. A, B, C, D identify the parameters $a = \{0.1, 1.0, 1.5, 2.0\}$.

Figure 4.9 depicts the spectrograms of the corresponding sonifications showing the functions with the biggest contrast $a = \{0.1, 2.0\}$. The mapping scheme for the sonification types $ST(f, f', f'')$ can be recognized in the spectrogram. The pitch mapping $ST(f)$ in the top row, features constant positions of the formants. The up and down movement of the fundamental and the partials is equivalent to $f(x)$ in Figure 4.8.

4. MAPPING AND INTERACTION IN AUDITORY GRAPHS

In the middle Figure 4.9 depicts $ST(f, f')$ with varying formants according to the first derivative. The first formant drops with increasing slope; the higher formants rise resulting in the earlier discussed transition from [a:] to [i:].

At the bottom of Figure 4.9 there is $ST(f, f', f'')$. The increasing brightness according to the change in curvature can be seen in the dark higher formants at the beginning, the middle and the end of the spectrogram. Compare the positions with brighter sounds on the x -axis with the high curvature in Figure 4.8.

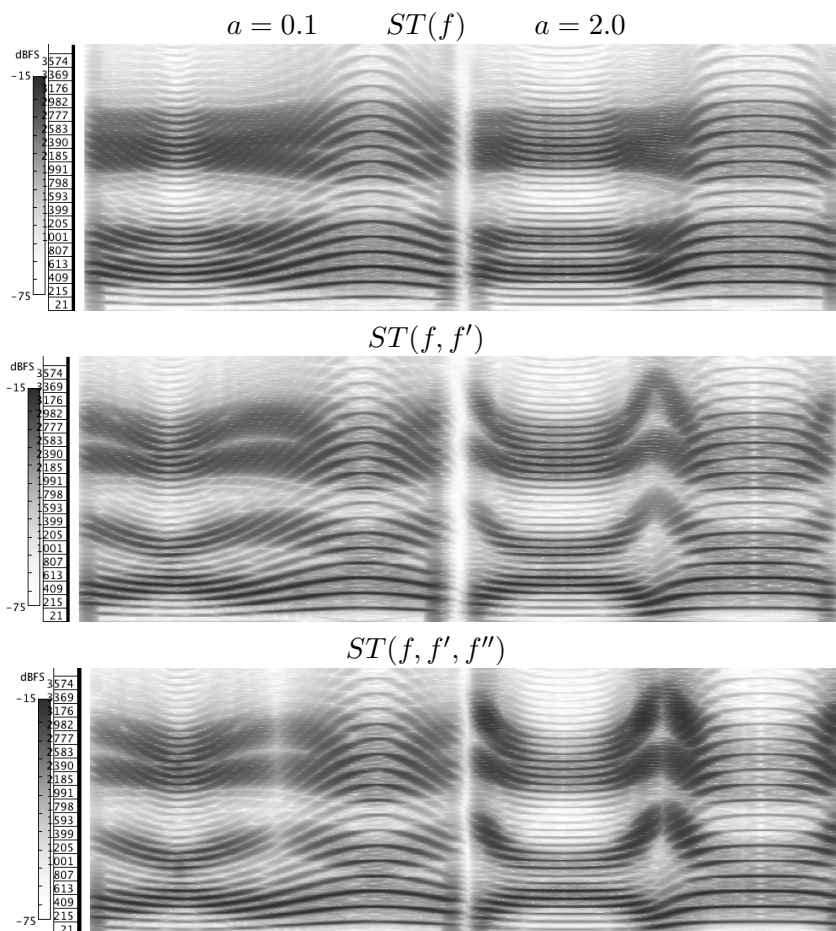


Figure 4.9: Spectrograms from Grond and Hermann (2012b) of the spectral contour of the sonified functions with the parameter a set to the biggest contrast for each sonification.

4.3.2.1 Experiment

In order to evaluate the perceptual contrast of the different ST, we asked subjects to rate whether a given pair of sonifications from the same ST appeared to them as same or different. Each combination of identical pairs $p_{i=j}$ 4 times and 2 times to both different function pairs $p_{i>j}$ and $p_{i<j}$ was presented, giving 120 pairs to rate for the 3

ST. After 60 ratings, subjects were given a break. The whole experiment lasted about 40 minutes. The bias in the frequency of occurrence between identical and different pairs had a ratio of 2:3. However, subjects had a tendency to rate more pairs as the same mostly for the simple pitch mapping $ST(f)$. We conducted the experiment with 10 participants (4 female, 6 male, age range 27 to 40). The experiment consisted of two parts (A and B) balanced across all subjects. In part A, subjects were rating the playback of two sonifications, each 1.5 seconds and rating this pair as *SAME* or *DIFFERENT*. In part B, subjects rated sonification pairs which were interactively explored through two sliders in a GUI arranged side by side. When moving the slider, subjects could hear the sonifications at the corresponding x in the stereo panorama. Measures were implemented in the GUI, which prevented subjects from adopting simple problem solving strategies, for example listening to the end positions of the slider or other visually recognizable positions. We asked the subjects in a post test questionnaire which strategies they used to distinguish the sounds. We also wanted to know which mode (playback or interactive) they preferred. For further details of the experiment see Grond and Hermann (2012b).

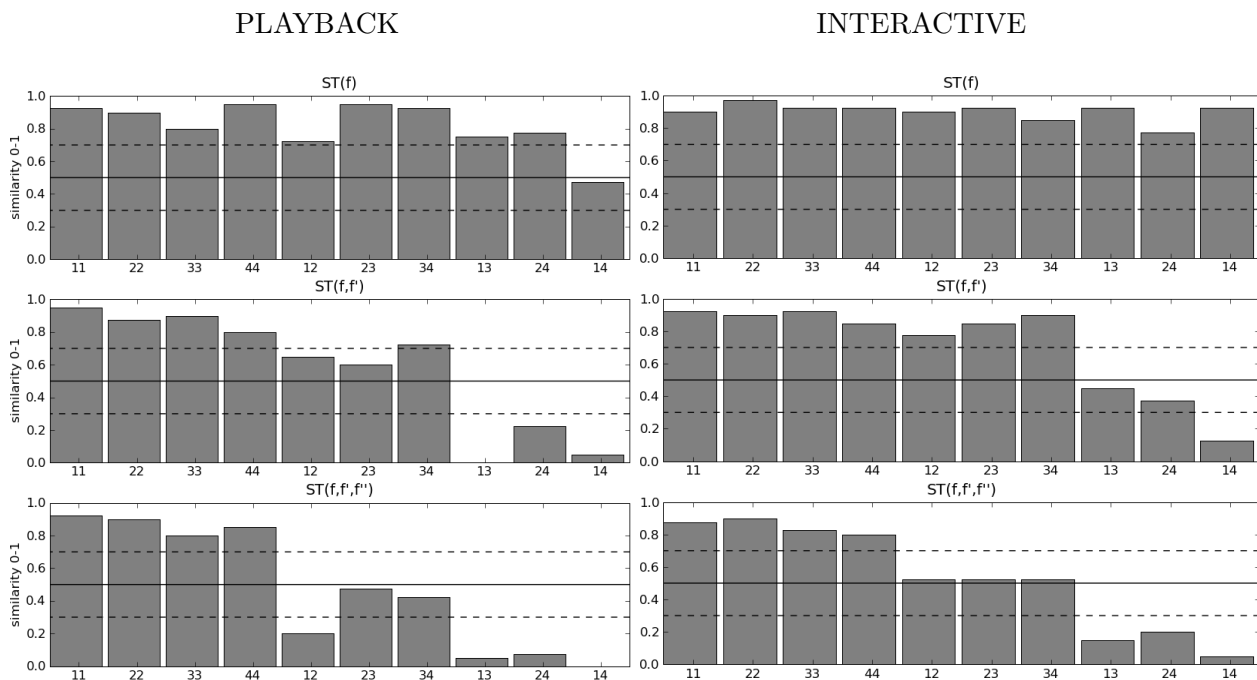


Figure 4.10: This plot from Grond and Hermann (2012b) shows the similarity of all pairs as rated by the subjects. On the left are the result for the playback and on the right for the interactive exploration. The order of the pairs across all charts is as follows: The first four vertical bars represent the cases $p_{i=j}$. The following bars are $p_{i>j}$ and $p_{i<j}$ pairs with increasing contrast p_{12} , p_{23} , p_{34} , p_{13} , p_{24} . The last pair p_{14} exhibits the biggest contrast.

4. MAPPING AND INTERACTION IN AUDITORY GRAPHS

4.3.2.2 Results

Figure 4.10 summarizes the results from the experiment. The left side shows the playback of the sonification, the right side shows the interactive exploration. In the plots, the vertical axis represents a scalar value for similarity (0-1). For a given sonification pair, this value is the mean rating of all subjects.

The first four vertical bars represent the cases $p_{i=j}$, i.e. the same stimuli in one pair was compared by the subjects. The following bars are $p_{i>j}$ and $p_{i<j}$ pairs. Difference increases with respect to parameter a from left to right. The null-hypothesis, which is that the sound contains no information is represented by the horizontal line at 0.5. A mean rating value of 0.5 is equivalent to random guessing. The 95% confidence interval is the area between the two dashed horizontal lines. All average values in above and below means that the null-hypothesis can be rejected.

One interesting result is that the identical pairs are never unanimously rated as identical. All of them however are above the confidence interval of 5%. Only two different pairs - both in the non-interactive mode in $ST(f, f')$ p_{13} and in $ST(f, f', f'')$ p_{14} - are unanimously rated as different.

A general trend can be found when comparing playback and interactive sonifications: The same sonification strategy is more often rated as identical in the interactive mode. For $ST(f)$ - the simple pitch mapping - all sonification pairs are rated as identical above the confidence interval of 5% showing, that $ST(f)$ does not provide enough perceptual contrast in the interactive mode. Even in the playback scenario $ST(f)$ performed poorly, however for the pair $p_{i>j}$ with the highest contrast ratings are equivalent to random guessing.

Enough perceptual contrast for $ST(f, f')$ can only be found for the 3 pairs with the biggest difference - p_{13} , p_{24} , p_{14} - in the playback mode and only for p_{14} in the interactive mode.

In $ST(f, f', f'')$ the similarity for small differences - p_{12} , p_{23} , p_{34} - is pushed down further being below the upper border of the confidence interval but still random guessing. The only exception for the playback mode is p_{12} rated as significantly different. In both the playback and the interactive mode, only the big contrasts - p_{13} , p_{24} and p_{14} - are rated as significantly different.

| | sonification 1 | sonification 2 | sonification 3 |
|------------|----------------|----------------|----------------|
| time (sec) | 15.33 | 19.48 | 17.15 |
| % | 79 | 100 | 88 |

Table 4.3: Total interaction time. Percentage is based on sonification 2

We additionally logged the interaction data of the subjects. For the 3 different $ST(f, f', f'')$, the results are shown in Figure 4.4 as plots of interaction time versus slider position. This plot shows that more time was spent at the slider ends. The total time spent for all $ST(f, f', f'')$ is shown in Table 4.3, the percentage in the second row is based on the exploration time of sonification2.

4.3.2.3 User Feedback

The playback was preferred by 7 subjects. Two subjects preferred the interactive mode. One subject was indifferent towards both modes.

Subjects were repeatedly surprised about how many pairs sounded to be the same. The feedback that we got from the subjects with respect to the listening strategy they used to rate the stimuli pairs was diverse. In the interactive mode, subjects said they tried to listen to the auditory graphs at differ-

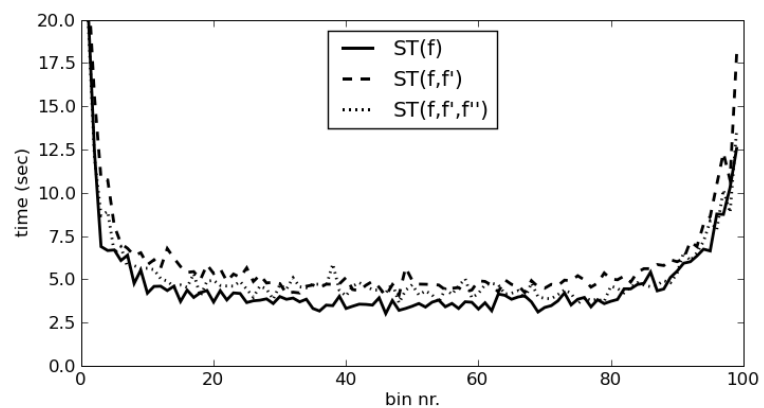


Figure 4.11: Interaction time from Grond and Hermann (2012b) for all 3 sonification modes. The times were collected in 100 bins across the horizontal slider axis.

ent speeds, sliding forwards and backwards. Other strategies they reported was listening to particular sections, most often near the slider extremes (see Figure 4.11), but also the central region. Subjects said they tried to remember particular parts of the sounds. Sliding across both sonifications with similar speed was mentioned to be difficult. In the playback mode subjects said they tried to remember the sounds at the beginning and the end. One subject reported trying to imagine the curve. One of two subjects, who tried to imagine the sound as spoken utterances, said that she felt her lips silently reproducing the sounds in order to be able to compare them. Many subjects closed their eyes in order to concentrate better on sonic details. Subjects also reported that they compared the rhythm and its distribution in the stereo panorama.

4.3.3 Conclusions on *Singing Function*

The evaluation of *Singing Function* provides the following answers to the questions raised on page 105. Multi-parameter-mapping adds perceptual contrast for functions from the same function family but with different parameters. The contrast increases with the number of derivatives that are included in the sonification. Although the evaluation did not test whether the meaning of these articulated vowel sounds was understood with respect to function analysis, the added function-shape-specific perceptual contrast can be considered a prerequisite to distinguish features of mathematical functions.

The increased perceptual contrast was observed in the playback as well as in the interactive presentation mode. However, in the playback mode it is still easier to recognize differences despite the fact that the *sonic gestalts* provide more contrast. Subjects reported that it is difficult to explore the x -axis with constant speed, a problem that we identified in *Sonic Function*. Here, the multi-parameter-mapping strategy could not lift the tight connection of the perceived *sonic gestalt* to its temporal evolution.

The hierarchical order of the sonification types $ST(f,)$, $ST(f, f')$, and $ST(f, f', f'')$ allowed comparing multi-parameter-mapping with the already popular pitch based method. It also provides an example of a multiple-views paradigm as proposed by Stockman et al. (2005).

The feedback of one subject suggests that the sound engaged at least for her the embodied aspects of listening modes. This evocative potential of vowel sounds can maybe be enhanced by telling users to listen for specific vowel transitions, or to try to mimic the pitch contour by singing.

The differences in the total interaction time across the three STs might be due to the following reasons: For $ST(f, f')$, subjects interacted with the graph longest, because differences were noticeable but were smaller compared to $ST(f, f', f'')$. In the simple pitch mapping approach $ST(f)$, pairs of stimuli had so little contrast that subjects arrived soon at the decision to rate the pairs as similar.

The time spent on the slider position as shown in Figure 4.11 suggests that subjects find it difficult to situate sounds in an unstructured homogenous interface. This observation compares with the direct listening modes of Vickers, compare Section 2.2.5. Structural elements with explicit auditory information such as acoustic tick-marks might help from an auditory display perspective. In *Singing Function*, we excluded them, because we focused on *sonic gestalts* and not on explicit elements.

For the given task of rating the difference of sonified function pairs, the playback mode was preferred by the majority of subjects over the interactive exploration. Interestingly, subjects occasionally closed their eyes in the playback mode, which can be

interpreted as an effort to focus closely on the sound. This shows that the question how to advantageously integrate interactivity in exploratory auditory graphs is still a challenge. Beyond the interference with the temporal evolution, interaction also occupies cognitive and perceptive resources, promoting a certain listening mode such as *ergo audition*, and possibly reducing sounds to crude features that serve mostly the purpose of providing feedback for the interaction rather than pointing towards the function data.

4.4 Summary and Outlook

Based on the discussion of current research in auditory graphs, this chapter introduced a novel multi-parameter-mapping sonification concept for auditory graphs based on the publications of Grond et al. (2010) and Grond and Hermann (2012b). This multi-parameter-mapping approach takes advantage of the Taylor series expansion of the function displayed in the auditory graph and maps the derivatives to audible features of a single auditory stream. Multi-parameter-mapping offers not only a sonification method for auditory graphs, it is a conceptual approach to create sonically richer auditory information by encoding shape or context specific features into a stationary timbre. For auditory graphs, stationary sonic representations were created in order to make the perception of graph shapes more independent from interaction. Although the stationary representations created a bigger function shape related contrast for the sonifications, the interaction dependent temporal evolution when exploring it along the x -axis still had a noticeable influence on distinguishing function shapes.

In the application *Sonic Function*, we applied this concept for the function value $f(x)$ and the first derivative $f'(x)$, where the function value was mapped to pitch and the first derivative was mapped to a low frequency oscillation envelope of the resulting sound. *Sonic Function* was evaluated in a pedagogical context. Although the pupils were challenged with mathematical concepts, the overall design of the function sonification was appreciated. In the qualitative evaluation, we found that the exact identification of extrema still depended on the sonic contrast i.e. the curvature around extrema, meaning that it did not become independent from exploration along the x -axis.

In the application *Singing Function*, we applied the multi-parameter-mapping to the function value and the first two derivatives, thereby creating 3 sonification types $ST(f)$, $ST(f, f')$, $ST(f, f', f'')$, each of them a distinct level of the multiple-view paradigm proposed by Stockman et al. (2005). In this prototype we mapped the first two derivatives to vowel transition and vowel brightness respectively. For the function value we kept the established pitch mapping approach.

4. MAPPING AND INTERACTION IN AUDITORY GRAPHS

Evaluating *Singing Function*, we found that the proposed concept adds more perceptual contrast when comparing pairs of a sonified function family with similar graph shapes. When the sonified function was interactively explored, it was still more difficult to discriminate the sonified pairs. Through the novel concept however, the perceptual contrast was increased when comparing it to pitch mapping only. We also found that in the case of vowel based synthesis, approaching and passing along critical points leads to complex yet familiar *dynamic timbre patterns* which can possibly be better detected, classified and remembered. In combination with a further improved mapping to loudness variation, these dynamic timbre patterns could be even more articulated and provide the basis for genuine sound objects, which also engage the vocal motor code allowing subjects to remember complex shapes. In fact the mapping to brightness for high curvature noticeably influenced the loudness of the sonification and provided such an articulatory stress.

The results demonstrate the limits of the orthodox and dominating $ST(f)$. However, the simple pitch mapping sonification has its place as the first instructive step, when explaining to users the modular approach of multi-parameter-mapping, and how it relates to function analysis. The multiple-views paradigm gives a framework within which all three ST can be embedded in a hierarchically structured display. The user can listen for each view for distinct sonic features: pitch, spectral contour and brightness. The first two of them can potentially be reproduced by the listener.

Building on *Singing Function* as the basis of an auditory graph, there will be the need to include information about the polarity of $f'(x)$ and $f''(x)$, since subjects had confused minima and maxima in $f_6(x)$ in *Sonic Function*. It might be best to represent this polar quality of the derivatives with a dichotomic sound parameter e.g. by voiced or unvoiced vowels. Stability of pitch perception needs to be taken into account, which could be achieved by restricting the noise component of unvoiced vowels to the higher formants only. Another possibility is the introduction of spectral flux, which can be conveniently realized by moderately modulating the center frequency of selected formants.

According to the findings of Hutchins and Peretz (2011), which is that it is easier to match the pitch of one's own voice rather than that of an artificial one, it is an interesting option for future development to adjust the pitch and register to the range of the user. This can be easily implemented through the vowel class presented in Section 2.4.

By comparing *Singing Function* and *Sonic Function* the advantages of perceptually fairly independent parameters of vowel-based synthesis become apparent. With low-frequency amplitude modulation, the first derivative was mapped in *Sonic Function* to

a strongly time dependent audible feature. This makes it more difficult to integrate in a complementary discrete sonification. The spectral contour of vowels, however, remain noticeable even in short sounds.

Further, the synthesis of vowel sounds allows keeping the pitch constant and to vary only the vowel transition according to the first derivative. Similarly, only the second derivative could be displayed by changing the brightness only. By defining a default pitch and / or vowel for these settings, multi-parameter vowel-based synthesis offers a hierarchically ordered as well as an independent multiple-view paradigm for auditory graphs as demonstrated in the provided sound sample.

The perceived pitch depends to some degree on the spectral contour. Here compensations can be easily realized by changing the formant bandwidths. In the case of very open filters, timbre as spectral contour almost vanishes, and only pitch remains audible: this means that the degree to which spectral contours shape the source sound can be left to the user as an intuitively controllable parameter.

In the evaluation of the auditory contrast, we did not determine just noticeable differences, because those depended on the mapping range for $f'(x)$ to vowel transition and $f''(x)$ to brightness. Since both – the absolute slope and curvature of functions – can take on all positive real numbers, both mapping ranges can be conveniently represented by one exponentially mapped parameter, over which the user will learn to have intuitive control.

For future developments and with respect to input devices for the interaction and listening modes, it is instructive to remember that sound with interaction is likely to be heard as causal feedback for the interaction. Here simple input devices like the right and left arrow key as used in *Sonic Function* might be better suited for exploration than the slider interaction from *Singing Function* due to the following two reasons: keeping the key pressed leads to a continuous speed progress along the x -axis, which is difficult to achieve with the slider movement. Also the feedback of the sound as response to pressing the key is likely to be exhausted quickly, and attention to the sound is possible. The relative positioning of mouse interaction however involves proprioceptive resources the relation of which to the absolute stereo positioning of the sound along the x -axis is unclear.

When these challenges are addressed, the potential to include additional information about the function in the same auditory stream, which was achieved with the multi-parameter-mapping approach combined with vowels-based synthesis can be fully unlocked.

4. MAPPING AND INTERACTION IN AUDITORY GRAPHS

5

Mapping in Auditory Augmentations

In this chapter, I present mapping strategies in two sonification applications from the emerging trend of auditory augmentation introduced by Bovermann et al. (2010). In auditory augmentation, interaction sounds are recorded and additional information is imprinted onto these sounds through near-real-time signal processing, as already briefly described in Section 2.4.4. An advantage of this approach is that the physicality of the impact sound translates into the sonification. The sonic result has the potential to be perceived as natural since it integrates with the already existing soundscape of the environment. Since it is part of the naturally occurring interaction sounds, this approach can engage *ergo audition* and is hence potentially less obtrusive for the user.

The original augmentation scenario by Bovermann et al. (2010) is an auditory display that mapped weather data to near real time filtered copies of the impact sounds on a keyboard. In this thesis, I expand this approach to higher dimensional datasets exemplified in two sonification applications. High dimensional datasets can be integrated through the MBS method of the data-sonogram introduced by Hermann (2011). This MBS method includes parameter-mapping aspects, as discussed in Section 2.4.4.2. One monitoring application for optimization algorithms has been developed which is also based on the augmentation of keyboard strokes. Further, an auditory display for 3 dimensional shapes has been developed for blind users. In this display, impact sounds such as finger snapping, hand clapping or tongue clicking are augmented through real time convolution of the interaction sounds.

In both cases, the amount of data does not allow limiting the augmentation of the sound to a filtered near-real-time copy of the sound. The information in the sonification has to be distributed over a short time span in order to create sufficient perceptual

5. MAPPING IN AUDITORY AUGMENTATIONS

contrast for either distinguishing states in optimization runs or recognizing 3D shapes through sonic representations.

The distribution of information over time in auditory augmentation faces the following implementation challenges, where two different cases can be identified: In the monitoring application, the abundance of the search space means that there is no a priori knowledge about the state of the system to be monitored. However, the amount of data which has to be mapped to sound parameters can be dealt with in real time. For the representation of 3D shapes, all information is available up front, but the amount of data that needs to be mapped to sound features, and therefore the amount of synthesis processes is immense. Therefore, these sonifications need to be rendered offline.

The perceptual challenge in the monitoring case is that a high frequency of impact sounds muddles the auditory display by having too many temporally extending augmentations overlap in time. The sonification of 3D shapes posed perceptual challenges in the sound design, since the amount of information about the 3D shapes has to be mapped to spatial, temporal, and timbral sound parameters ensuring differentiable results. Additionally, listening skills and habits of the blind community had to be taken into account in the design process.

The auditory augmentations developed in this chapter have a peculiar place within established methods in auditory display. As far as an augmentation through near real time sonic overlays is concerned, these sounds are reminiscent of auditory icons if we take into account that the physical source properties of the interaction sounds translate into the signal. They are on the other hand related to earcons, as there is some arbitrariness with respect to how the data are mapped to mostly spectral sound features. Bovermann et al. (2010) defined augmentation as near-real-time sound treatment, which applies to some aspects of the monitoring situation. In this application the decreasing filter bandwidth evokes a transition between listening modes that leads from *ergo audition* of realistic interaction sounds towards Schaeffer's listening mode *hearing (3)* of identifiable spectral sonic values.

Mapping the data to time delays leads to augmentations which are not near-real-time any more. The resulting sonifications create in many instances several sound objects, which respond like an echo to the interaction sound. The physicality of the interaction sound still translates into the sonification. Additionally, through the delay, these scenarios can create suspense and preparedness. By decoupling action and perception, this approach has the potential to create favorable conditions for *musicianly listening*. Interestingly, this collection of sound objects can also adopt the role of sonic motives and are hence related to the idea of earcons (also compare Section 2.2.2.3).

The monitoring of optimization runs has been a collaborative research project together with Oliver Kramer and Thomas Hermann in which I designed and developed

the sonification application. It was first presented at the ICAD 2011 (International Conference on Auditory Display) by Grond et al. (2011b), receiving the best paper award. A further development of the application was later published by the authors in Grond et al. (2012). The content of this chapter is based on both publications, the structure of which is partly adopted, rephrased and extended in the following chapter.

The sonification of 3 dimensional shapes has been developed within the audible sculpture project. This project has been initiated by Adriana Olmos, Jeremy Cooperstock and others and has received a strategic innovation fund award from the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) McGill University. Within this project, I developed the sonic interaction design as well as the mapping strategies for the representation of the 3 dimensional shapes. The sound design was developed together with Michael Ciarciello, who contributed valuable expertise as a blind composer and instructor for the blind community.

By developing and evaluating display options for interactive sonification monitoring, we have contributed to auditory augmentation of typing sounds by addressing some shortcomings in existing displays, which have so far dealt with the representation of small sets of values only. With the auditory displays proposed in this chapter, states of high-dimensional processes become accessible for monitoring through auditory augmentation. The auditory display developed in the audible sculpture project extends methods from auditory image representation into 3D shapes. Both works are starting points to connect the established MBS approach of *data-sonograms* with the emerging field of auditory augmentation. Although we have focused on specific tasks in this work, the sonification strategies employed here are of more general applicability for the auditory augmentation of high-dimensional data.

In this chapter I will first discuss the background and implementation details of the monitoring application together with its qualitative evaluation, followed by the audible sculpture project and its evaluation.

5.1 Monitoring and Auditory Augmentation

Sonification is very suited to support monitoring in various application areas. Various examples illustrate this field of potential applications: Sonification has been used for the monitoring of stock markets by Janat and Childs (2004), for the monitoring of network traffic by Rubin (1998), for the monitoring of electrocardiograms by Ballora et al. (2004), and for the monitoring of EEG by Baier et al. (2007). In the recently published sonification handbook the chapter about monitoring by Vickers (2011) discusses theoretical considerations as well as implementation examples. A discussion of

5. MAPPING IN AUDITORY AUGMENTATIONS

listening modes for monitoring applications is given in the introduction of this thesis, see Section 2.2.3.

The human auditory system is particularly apt for monitoring tasks, various mechanisms studied in the field of auditory scene analysis (see Bregman (1994)) can be put to use for this purpose. The phenomenon of backgrounding sets in when a sound becomes steady and allows to focus on other sonic input such as other audible streams in sound mixtures. If a sound is in the background and its characteristic sonic features change, it can come back to our attention since the human auditory system readily notices transients. Backgrounding parallels the idea of *ergo audition*, i.e. we turn our conscious focus towards any sound only as much as necessary in order to gather information from it. For interaction sounds this would be the minimal effort we need to make to close the action and perception loop.

5.1.1 Monitoring, Sonification and Evolutionary Optimization

Evolutionary optimization seeking (ES), to which we apply sonification for monitoring purposes, is in essence a stepwise process. ES can last from a few seconds to several minutes or hours depending on the problem dimension and computing power. The dynamic evolution of the iterative steps, ranging from fast progression to slow movements in the search space, makes it an ideal candidate for monitoring through an auditory display.

5.1.1.1 Evolutionary Optimization Seeking

Since the sixties and seventies, ES has developed into powerful optimization heuristics (as a standard textbook confer Beyer and P. (2002)). ES algorithms are population-based, randomized search heuristics, which are inspired by the Darwinian principles of evolution. These include: inheritance, mutation, and selection of the fittest as proposed by Darwin (1859). Fogel (1966), and Schwefel (1995) were amongst the first to translate these principles into algorithms known today as evolutionary computation. This field has brought forth a diversified set of methods to efficiently solve optimization problems in high-dimensional parameter spaces.

5.1.1.2 The Principles of ES

For a better understanding of the sonic metaphors which we used in the sound design, we describe here in more detail the principles that govern iterations during the optimization procedure. Finding an optimal solution corresponds to stepwise approaching the global minimum of a cost function. This problem-dependent function is embedded

in a potentially high-dimensional search space. In black-box optimization procedures no derivatives are available to be used such as in a gradient descend method during the search process. In these non-trivial cases, ES employ stochastic methods containing a random element in order to explore the search space. The consecutive steps during an iteration consist of:

1. **Sampling:** Create a population of random points that cover the search space.
2. **Evaluation:** Evaluate their fitness, i.e. their cost function.
3. **Selection:** Discard a defined percentage with the poorest fitness values.
4. **Inheritance:** Produce offspring with the distribution σ around the fittest.

These steps (1 to 4) are repeated until the population of points converges towards the optimal solution. As an example of a converging optimization process, the development of the cost function and the σ vector in a 30-dimensional parameter space is shown in Figure 5.1.

5.1.1.3 Why Monitor ES?

The success of stochastic methods is to some degree dependent on parameter tuning, an important issue for successful optimization. A self-evident example of a parameter that has to adapt during an optimization run is the vector of the mutation strengths σ . This vector has to decrease as the points of the population approach the minimum. Otherwise the distribution of the offspring of step 4 continues to covers too wide of a range in the search space and never converges to the optimal solution. Although automatic parameter-tuning methods have been developed by Bartz-Beielstein et al. (2005) and Meyer-Nieberg and Beyer (2007), many practical problems remain hard to solve. Ideally, the practitioner can be kept in the loop through monitoring and hence intervene in real-time by controlling important parameters during the optimization process.

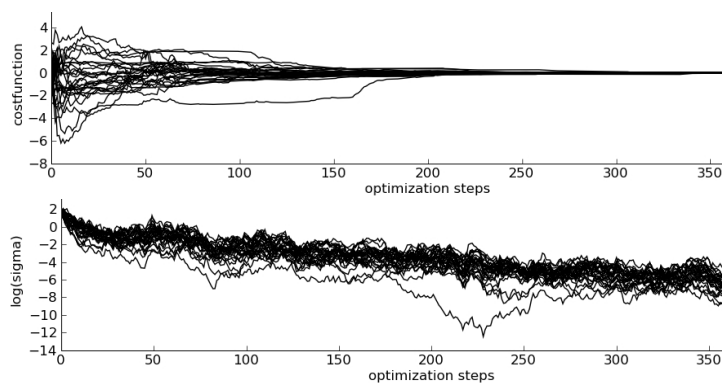


Figure 5.1: Cost function plots for an optimization run on top terminating after 359 iterations. This case is known to have its global minimum at the origin, i.e. 0 in all components. The development of the σ vector is shown at the bottom. Both plots are taken from Grond et al. (2012)

5. MAPPING IN AUDITORY AUGMENTATIONS

In order to turn the non-interactive algorithm through user interventions into a closed-loop interaction, more feedback during the run about actual details of the search process is needed. From an HCI perspective, current real-time displays are restricted to quite poor visual only interfaces consisting mostly of a flow of output numbers. Since many parameters, e.g. the cost function and the vector of sigmas, change rapidly at the same time, this challenge of designing an appropriate HCI can be well addressed by sonification.

In a first step, we applied sonification to ES in Grond et al. (2011b), where we experimented with signal parameters such as delays in order to exploit time structure as a perceptual dimension representing the convergence of the optimization. In a second step we refined the use of signal delays and conducted an evaluation of 3 different auditory display settings. In the sequel I will describe the sonification design which was finally published in Grond et al. (2012).

5.1.1.4 Auditory Augmentation in Monitoring

For a monitoring display, the particular challenge with respect to the sound design is to maximize the conveyed information and at the same time reduce obtrusiveness as much as possible. For an overview of monitoring-related tasks and listening modes see Section 2.2.3. From the three types of tasks which have been discussed by Vickers (2011) with respect to monitoring, I recapitulate here the first two from Grond et al. (2012), which allow to situate auditory augmentation in the context of ES monitoring:

- Direct auditory display (PULL characteristic): the information to be monitored is the main focus of attention and does not allow for parallel activities.
- Peripheral auditory display (PUSH characteristic): attention is focused on a primary task whilst required information relating to another task or goal is presented on a peripheral display.

Auditory augmentation as presented by Bovermann et al. (2010) allows to think of an auditory monitoring display as dynamically situating itself within the two PUSH and PULL characteristics. Whilst querying the optimization process through triggering a sound can be described as having a PULL characteristic, sonic information that is tightly integrated into an everyday activity can be informative but is not distracting and hence has a PUSH characteristic.

For the development of the auditory augmentation for ES, we were inspired by the application *Reim* from Bovermann et al. (2010) which augments keyboard interaction sounds with filtered copies. The configuration of the filters depends on external data.

The interaction scenario from *Reim* is particularly useful for this case since ES are usually performed on the computer workplace which naturally includes the interaction with input devices such as the keyboard.

Hence the soundscape at work is subtly enhanced (PUSH characteristic) rather than adding disturbing additional sounds to it. However, the state of the ES can always be queried (PULL characteristic) by interacting with the keyboard casing. As a result, the two monitoring categories listed above, which seem to mutually exclude each other at first sight, can be combined through auditory augmentation. It becomes the user's choice to engage with either display purpose depending on the available attention.

5.1.2 Sound Design, Mapping, Audible Effects

The objective for the sound design is to find a mapping that indicates the progress of the optimization, i.e. the state in the parameter space and the convergence. Since it is a matter of training if variations in the sound can be distinguished and sufficiently well attributed to the search space, we decided to keep a mapping that we found acceptable and pleasant after some sound design experiments, and focus on how the display could serve the two PUSH and PULL characteristics. The underlying principles of the mapping design are described in Grond et al. (2012). Its objectives are:

- The sonification should include hints about the mutation strengths for each dimension of the actual solution.
- The progress of the optimization should be noticeable on all scales, i.e. the strong variations at the beginning and the small adjustments during convergence towards the end of the optimization.
- The sonification should help to discern different areas within the parameter space, such that convergence to different local minima should be audibly distinguishable.

The iterative steps during an optimization run suggest some choices as mapping metaphors for the sound design. One range of values which must be mapped is the space of the cost function. This space has no a-priori range limit. For testing our sonification prototype we used Schwefel's or Griewank's multimodal function, see Schwefel (1995). For this function, the parameters in the solution space of interest are positive real numbers of different magnitudes. Another important feature from ES is the mutation strength i.e. the vector of σ which determines the distribution of the offspring. Unlike the cost function, the vector of σ has a lower limit for all elements, which is 0. The mapping from data features to sound synthesis parameters is compiled in Table 5.1. The delay mapping of up to 0.5 seconds was chosen to keep the sound objects with the longest possible duration plus the decay of the resonant filter around one second.

5. MAPPING IN AUDITORY AUGMENTATIONS

| data feature | | synthesis parameter |
|-----------------------------|---|-----------------------------------|
| $\log(\sigma[0.005, 0.5])$ | → | stereopan [left, right] |
| $\log(\sigma[0.0005, 0.5])$ | → | delay [0.0, 0.5] |
| $\log(\sigma[0.0001, 0.5])$ | → | filterwidth [0.5, 0.001] |
| $y_i[-2, 2]$ | → | $\text{freq}_i \cdot n$ [0.25, 4] |

Table 5.1: Mapping scheme of data features to sound parameters taken from Grond et al. (2012): value ranges are specified in square brackets.

5.1.2.1 Mapping the Cost Function and Mutation Strength

The elements of the cost function vector, each representing one dimension of the search space, were mapped to a two-pole resonant filter which is implemented in SC3 as the unit generator *Ringz*, compare Steiglitz (1994). The centre frequencies freq_n of the filters were equally distributed on a logarithmic scale between 100 and 2000 Hz. As shown in Table 5.1, the movement of the population of the fittest individuals along all these dimensions was mapped through a multiplier n for each frequency.

In a high-dimensional search space it is not reasonable to assume that each single dimension can be individually heard. Rather, the augmented sound being the sum of all filtered copies should give an impression of the overall position within the search space. However, transient movements in the search space were audible along individual dimensions y_i , due to the multiplier n for each frequency. The second vector to be mapped to sound is the mutation strength. This vector of σ components converges fast at the beginning in an ideal optimization run, and decays exponentially towards 0 when the global optimum is reached. It usually takes some time for the population to approach and settle at the optimum. This is when the change of the cost function values is minimal, a moment when the algorithm risks getting stuck in local minima due to small σ values.

Since the mutation strength is a good indicator for the optimization progress we decided to emphasize its presence in the auditory display through a one-to-many mapping strategy (including filter bandwidth, delay, and stereo panning.) These parameters were equally chosen in order to exploit perceptual features of human listening and to support their interpretation through the metaphoric aspect of the mapping. As discussed in Grond et al. (2012), one motivation of the mapping was to reduce masking effects when many different sounds are present. This is why we chose the delay and the distribution over the stereo panorama. As a metaphorical dimension of the mapping, the filter bandwidths reflect the vector of σ components determining the variance of Gaussian mutation. Here strong variations in the mutation result in a broad bandwidth. The three selected sound parameters were controlled with exponential mapping functions of different decays during optimization runs so that each effect would set in

5.1 Monitoring and Auditory Augmentation

during different phases. The spectrogram in Figure 5.4 shows both stereo channels and the effects of the mapping annotated and described in Section 5.1.2.4.

The mapping resulted in unique, spatially distributed *Ringz* filters during the course of the optimization process. Since an optimization run typically starts with strong mutation strengths, the sum of all filters let most of the spectral components pass through, hence the augmentation was almost undistinguishable from the original typing sound. At first the stereo panning gradually faded out during a run followed by the delay time. This gives the impression of an echo-like reflection in a big space, which continuously shrinks to the acoustic impression of a small room. The third effect to fade out is the filter bandwidth. As the bandwidth became smaller a tiny but noticeable ringing of the filters at their centre frequency occurred, in the last phase during convergence.

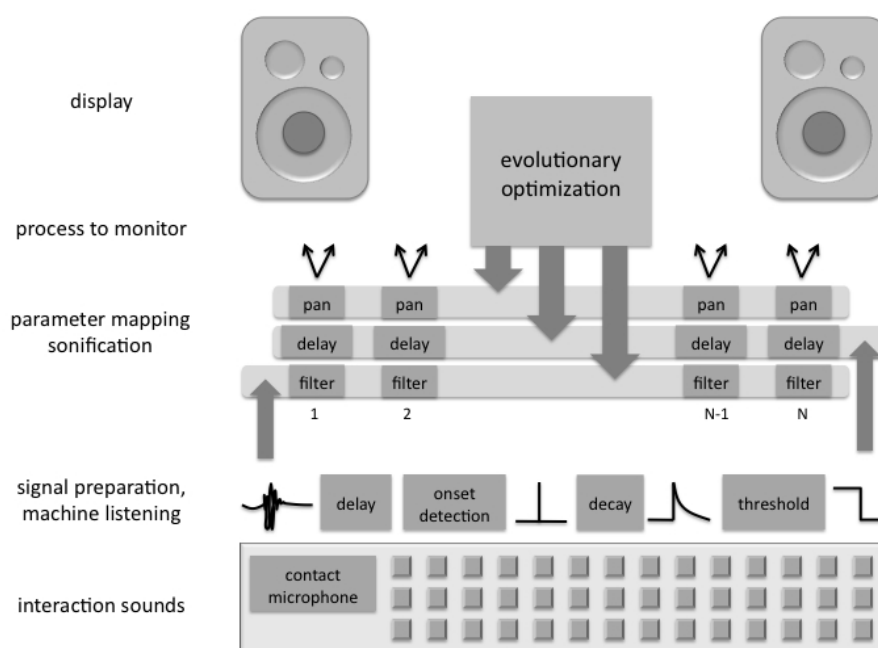


Figure 5.2: This diagram from Grond et al. (2012) depicts the link between interaction sounds with the keyboard and the auditory display through parameter-mapping of data from the ES. At the bottom the role of onset detection is schematically depicted.

5.1.2.2 Software and Hardware components

The sonification is implemented in SC3. The interaction sounds - typing onto the keyboard - are picked up by an AKG C411 PP condenser contact microphone, which was attached to an external USB keyboard in order to avoid unwanted sonic influences

5. MAPPING IN AUDITORY AUGMENTATIONS

from the laptop such as fan and hard-disk noise. The augmented sounds were displayed over Genelec 6010A active speakers. For further details see Grond et al. (2012).

5.1.2.3 Sonification Modes

The design of the mapping is based on preliminary sound design experiments and is kept fixed in the sequel, because we are mostly interested in the unobtrusiveness of the display. In order to minimize this factor we design three different display options all of them based on the parameter-mapping described above. The main difference between these options is the use of the delay mapping. One of the three mapping options includes machine listening on the contact sounds in order to control the delay. Figure 5.2 depicts a diagram of the auditory augmentation plus the machine listening element.

The augmentation with delay represents the mapping compiled in Table 5.1. In order to be able to distinguish the different dimensions of the optimization problem, the delay is introduced to distribute the sonic information in time. This mapping approach which organizes geometric data relations as temporal order in the sound domain is conceptually related to *data sonograms* of MBS. The disadvantage of this approach is the overlap of earlier augmentation sounds during fast typing. Additionally, the changing delay parameter creates occasionally unpleasant glissandi.

The augmentation without delay keeps the same mapping but all delay parameters are set to 0. Hence the progress of the optimization can be heard as a variation in the spectral contour. Individual dimensions of the search space could hardly be noticed due to masking effects. This setting, however, allowed to keep the augmentation non-disturbing when typing fast and there were no artifacts from the changing delay.

The third setting is augmentation with delay suppression, which is essentially a combination of the two from above. When fast typing sets in, the augmentation changes from *with delay* to the second variant *without delay*. We use machine listening techniques to detect typing activity, because we want the augmentation to be an autonomous program operating on the signal level only. Typing is monitored through onset detection, which is described in Stowell and Plumbley (2007) and implemented in SC3. An intelligent signal switch is created which is described in Grond et al. (2012). The state of this switch is at first in the variant *with delay* and gives enough time to augment the typing sound. After that, it fades into the augmentation *without delay* and remains in this state as long as typing continues. The interruption of typing changes the state of the switch back to the augmentation *with delay*.

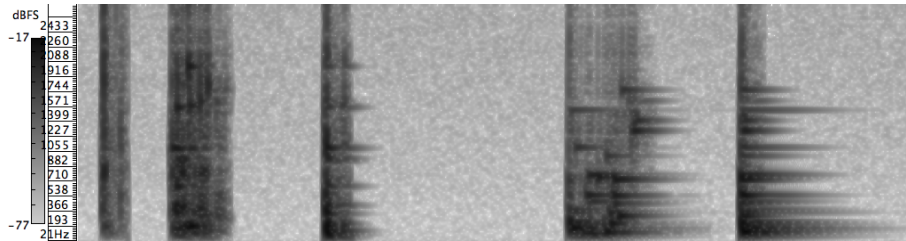


Figure 5.3: This spectrogram shows a down-mix of both stereo channels of the augmentation of two keystrokes, first with delay, then without delay. The filter bandwidth shows that the state of the optimization on the left features large sigmas. Stronger convergence is found in the optimization on the right.

5.1.2.4 Auditory Effects of the Mapping

In Figure 5.4 two annotated spectrograms of the stereo channels are shown featuring two optimization runs. The dashed vertical line indicates the border between the two runs. **A** annotates in the left channel high range frequency bands present in both runs. These bands are present but weaker in right channel. One high frequency band persists in the first run until the end. This indicates that the optimum that was reached in this run is far away from the origin with respect to this component.

Gaps in the right channel are annotated by **B**; they happen when the whole signal is mapped to the left channel for small σ values. A momentarily converging optimization manifests itself as noticeable and salient fluctuation in the stereo panorama. Occasionally, the algorithm leaves found minima and the augmented sound occupies again the whole stereo panorama.

C shows the end of an optimization, which has almost converged. The sound does not disappear completely from the right channel because the σ values are still not small enough. Nonetheless the run ended because it reached the hard limit of computed generations. In the first optimization run, though very small σ values were reached annotated by **B**, the optimization algorithm found only a local minimum.

5.1.3 Evaluation

We conducted a small user study to determine which of the display variants is preferred with respect to obtrusiveness of the display and information content. By using the prototype as mentioned above in Section 5.1.2.2 we conducted a study in a closed and quiet room. 14 subjects took part ranging from 27 to 38 years of age, 9 were male, 5 female. They had an academic background in informatics or psychology (PhD and postdoctoral researchers) with daily keyboard typing practice. In brief, the subjects

5. MAPPING IN AUDITORY AUGMENTATIONS

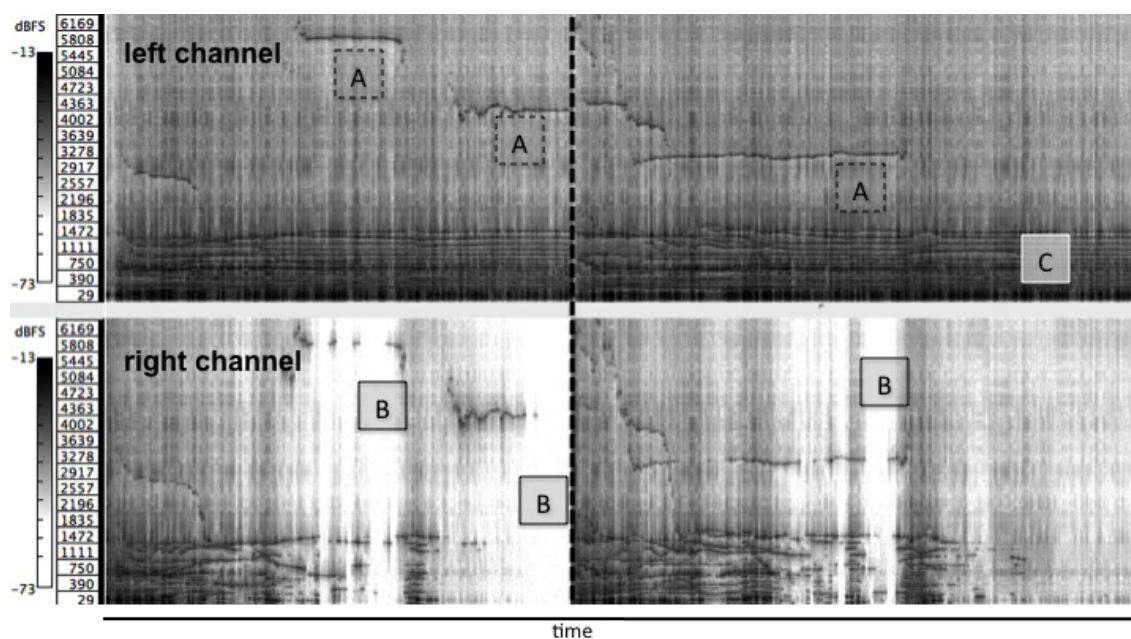


Figure 5.4: Spectrograms of two recorded optimizations runs from Grond et al. (2012). The effects of the mapping are annotated as **A**, **B**, **C** and are discussed in Section 5.1.2.4. The vertically dashed line separates the first optimization run (49 seconds) from the second (62 seconds).

were asked to explore the auditory augmentation of optimization runs. For further details of the study, please consult Grond et al. (2012).

| | |
|----|---|
| Q1 | Is the augmentation with delay informative? |
| Q2 | Is the augmentation with delay obtrusive? |
| Q3 | Is the augmentation without delay informative? |
| Q4 | Is the augmentation without delay obtrusive? |
| Q5 | Is the augmentation with delay more informative than without delay? |
| Q6 | Is the augmentation with delay more obtrusive than without delay? |

Table 5.2: Questions about informativity and obtrusiveness of the auditory display taken from Grond et al. (2012).

After the exploration part, the subjects were asked to fill in a questionnaire with 6 questions and one ranking task. The questions are compiled in Table 5.2. The last point of the questionnaire was to rank the 3 variants as either 1, 2, or 3 corresponding to least preferred, middle, and most preferred.

5.1.3.1 Results

The answers to the questions from Table 5.2 are compiled in Figure 5.5 depicting the average, the standard deviation and parallel coordinates for all subjects.

Comparing the answers about the informativity of the display options *with* (Q1) and *without* (Q3) delay, there is a tendency to find the display with delay more informative than the display without (the p-value for the *t*-test gives 0.0022). However, with respect to obtrusiveness of the display options *with* (Q2) and *without* (Q4) delay, the answers are in average indifferent (the p-value is 0.34). For the comparative question Q5, subjects preferred again *with* delay with respect to how informative it is, but judged less clearly the difference in obtrusiveness Q6 (the p-value is 0.068).

Figure 5.5 compiles the answers to the ranking task. We find that the mixed mode, (machine listening enhanced combination of the two variants) was significantly preferred over the *delay* variant (the p-value is $1.409 \cdot 10^{-7}$). There is a wide spread in the ratings for the variant *without delay*. The p-value for the preference of the version without delay over the one with delay is 0.12. The p-value for the preference of the mixed mode over the version without delay is 0.035.

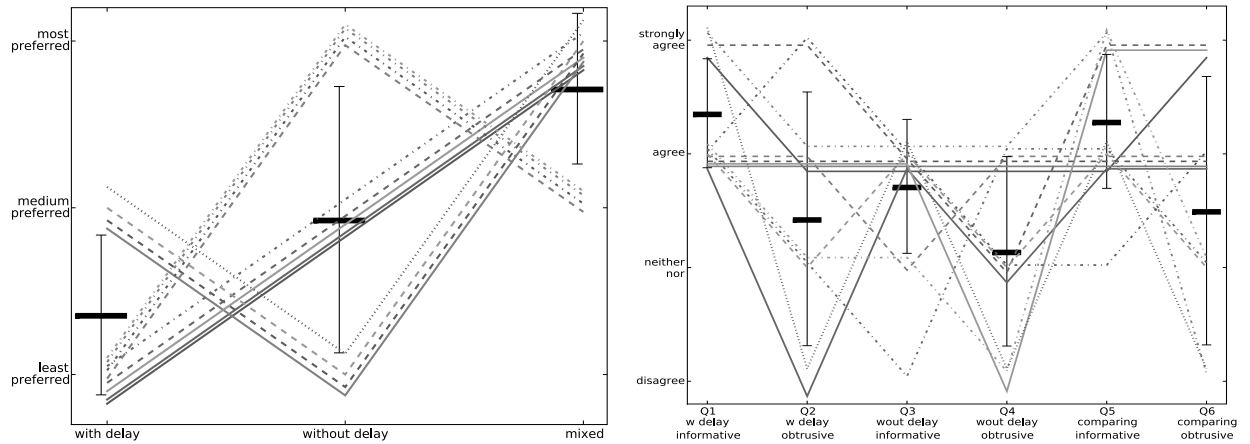


Figure 5.5: Box-plots of the ranking for the three display settings and answers to questions Q1 to Q6 both with parallel coordinates. Individual answer profiles can be distinguished by the small offset and subject-dependent line style.

5.1.4 Discussion

Based on the plots of Figure 5.5, we arrived at the following conclusions: We find a slight preference for the version *with* delay, with respect to questions (Q1, Q3, and Q5) about how informative the display is. This supports our design goal which is to

5. MAPPING IN AUDITORY AUGMENTATIONS

distribute auditory information in time in order to avoid masking of simultaneous sound layers within on augmentation.

With respect to the obtrusiveness of the display (Q2, Q4, and Q6), subjects were quite indifferent, which might indicate that the time of exposure to the interface was not long enough. This reason is further supported through the remarks by some subjects, who mentioned that obtrusiveness needs more exposure time for a better assessment.

The synthesis of all the answers to the questions Q1– Q6 did not indicate a significant preference for any of the modes. However, many subjects preferred the mode *with* delay the least, and some the one *without* delay the most. This could indicate that the delay makes the augmentation obtrusive despite the indecision from the answers to questions (Q1– Q6).

Interestingly, the mixed variant was preferred in the ranking task. It is possible that the option to switch between monitoring a background process to intentionally querying the state of the optimization contributed to the preference to this display option. The distinctly different auditory characteristics of the augmented sound seem to conform with the different listening intentions of both activities. It was interesting to find that one subject wanted to have more control when actively querying the state of the convergence. This subject did not want to wait until the onset triggered decay relaxed and preferred to have instead a key combination to switch into the delay augmentation for intentionally querying the optimization progress. For future developments, addressing this request might help to better match preparedness, action and the listening intention.

5.2 Audible Sculptures

The second practical application in this chapter is an auditory display for 3D shapes. It was designed for the blind and partially sighted. Following observations during a field study with blind participants within the ISAS project, we were motivated to make sculptures more accessible for our target audience by rendering them as audible shapes and forms¹.

As a concrete example, our goal was to convey the geometrical features of sculptures in Parc Mont-Royal, Montreal, Canada, through sonifications, so that blind individuals could gain a mental representation of their form through sound. Findings from Kim and Zatorre (2010) show that reliefs can be successfully identified based on sonifications of them, and suggest that a sonic mental representation of two-dimensional reliefs is possible. Despite the artistic field of application to sculptures, the methodological and systematic approach with respect to the data to sound mapping, contributes a method for the sonification of 3D shapes and forms in particular and to the development of auditory display for blind users in general.

The interaction paradigm which we adopted is inspired by echolocation, often used by blind individuals to facilitate orientation in space. For a detailed perspective on echolocation we refer to the review article by Kish (2003). When using this technique, echolocators either pay attention to the reflections of environmental sounds or snap their fingers or click their tongue and listen specifically to the reflections of these short impulses from nearby objects. Echolocation is hence a listening situation or habit that tries to extract information through acting in a given environment.

In the audible sculptures project, echolocation is used as an interaction paradigm to trigger sounds and engage with the sonification of the sculpture. The short impulses of snapping fingers are rendered like an echo by convolving them with the sonification of approximately 1 second in duration. This echo is specific for each listening angle around the sculpture, varying smoothly with small changes.

This echo-like response is composed by designing a systematic sonification of geometric features, which varies according to the perspective of the listener with respect to the sculpture. This sonification of geometric features is not based on the exact acoustic reflection of the sculpture, but rather mixes elements from the data-sonogram and parameter-mapping sonification of 2D images. The resulting sounds are hence like sonic fingerprints similar to earcons for specific perspectives on the sculpture.

¹I participated in the ISAS (In Situ Audio Services) project during its beginning in 2010, the results are published in El-Shimy et al. (2012), the field study was conducted by Adriana Olmos.

5. MAPPING IN AUDITORY AUGMENTATIONS

The sonic characteristic of the impulse sounds (sharp or dull finger clicks) can be heard in the convolved sound, which we hoped to increase the listeners' level of engagement through embodied listening modes of action sound couplings. The variability in the creation of impact sounds like finger snapping created an interesting tension between repetition and variation: no click produced by the listener can ever be repeated exactly, and this is always reflected in the slightly different sonic qualities of the resulting convolved sound.

Through the interaction paradigm that we use, the project is also situated in the field of auditory augmentation. In the project that was previously described in this chapter, the amount of data that had to be mapped to real-time sound synthesis consisted of approximately 60 floating point values that could easily be used as parameters for real time sound synthesis. When sonifying arbitrary free-form shapes, the amount of data on which the sonification is based consisted of roughly 1200 triangles for a given perspective. Hence the number of unit generators that have to be employed for sound synthesis is substantially bigger. Here, a real-time sonification is not feasible anymore. This is why I decided to render the sonifications offline. The sonifications last for about 1 second, and are rendered for 36 equidistant angles around the sculpture by keeping a fixed distance. These sounds were then used to augment interaction sounds through convolution.

Realizing the sonification as an offline-rendered convolution kernel is advantageous as the convolution algorithm scales linearly with respect to the kernel length. Hence the computational costs for the auditory augmentation are independent with respect to the sound processes involved when creating the sonification.

Convolution as a basis for data-sonograms was already discussed in Section 2.4.4.2, in which vowels based on formant synthesis were used to represent different data labels. In the sequel we follow this convolution-based approach but will explore different synthesis options, in order to create short and distinguishable sonic fingerprints.

5.2.1 3D Data Preparation

The sculptures are modeled with 3D mesh data. Regions of the model that are approximated as being flat are represented as triangles. The modeling of the sculptures was carried out by Adriana Olmos using Autodesk Maya 2013. The size of the objects modeled to fit proportionally a cube of 10 cm^3 . Then the standard mesh smooth and triangulation functions from Maya were used in order to obtain approximately a total of 2400 mesh triangles in the object. The smooth function was applied in order to represent sharp edges with surface normals pointing towards the listener. Figure 5.6 depicts a 3D mesh model next to a photo of the corresponding sculpture. For each

surface triangle we compute the following features, which are the basis for the mapping of data features to sound parameters:

Triangle area: The area of an individual triangle is normalized with respect to the sum of all triangles representing the surface of the sculpture.

Normal vector: The normal vector with respect to the plane defined by the triangle is computed such that it points outside of the sculpture. For each selected orientation of the sculpture towards the listener, the angle between this normal vector and the listening direction is computed and provides the value that is mapped.

Distance to the listener: For a selected orientation of the sculpture the distances of all triangles facing the listener are computed.

Center of the triangle: The center of the triangle is computed as the average of all three corner points defining the triangle. The z component of the centre is used to represent the height of the selected triangle. The horizontal position of a surface triangle with respect to the listener position is equally based on its center.

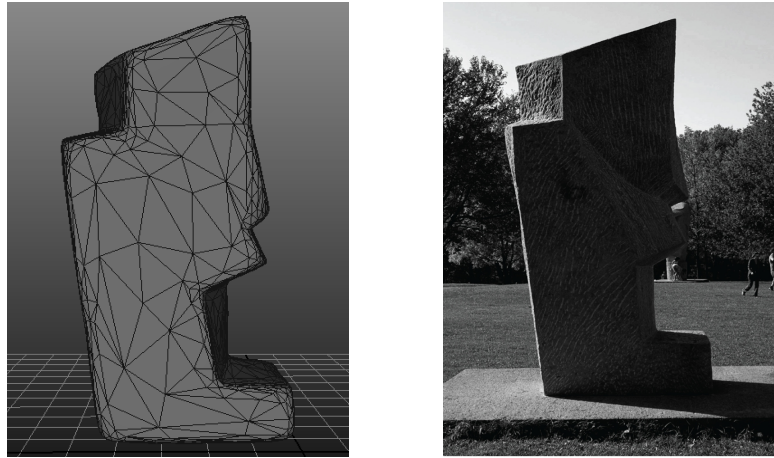


Figure 5.6: A 3D mesh modeling a sculpture on the left, and a photograph of the original sculpture from Parc Mount-Royal on the right.

The 3D data are resized so that their biggest extensions fit into a sphere with a radius of one unit. The center of this sphere is located at the origin $(0,0,0)$ of a Cartesian coordinate system. The listener is positioned at a height of $z = 0$, which corresponds to the middle of the vertical extension of the sculpture. Each sculpture is modeled with approximately 2400 triangles. For the sonification of a specific position towards the sculpture only those triangles are taken into account which have surface normals between pointing directly towards the listener and enclosing an angle of $\pi/2$.

5. MAPPING IN AUDITORY AUGMENTATIONS

5.2.2 Sound Synthesis, Mapping, Participatory Design

Since we want to use echolocation as interaction paradigm, the idea is to represent each surface triangle as a short sound event which does not model but alludes to a reflected impulse sound. This approach is similar to the data-sonogram, which is not strictly modeling the acoustic circumstances, but adopts for the data mapping acoustic inspired principles. This method essentially offers a translation of spatial structures from the data domain into temporal structures in the sound domain.

The short sound event that represents the reflection had an impulse-like envelope, long enough so that it could carry audible characteristics that could be located in the frequency range. Experimental sound design leads us to create a short impulse through an amplitude envelope and multiply it with noise. A code example of the synthesis definition for SC3 is given in Figure 5.7.

```
SynthDef(\panFormletReverbPulse,  
  { lamp, freq, attack, dur, pos, mix, room, tamp |  
    var src, srcReverb, excite, pan;  
    RandSeed.ir(trig: 1, seed: 56789);  
    excite = EnvGen.ar(Env.perc(0.0,0.001,1,-10), 1.0, doneAction: 0) * ClipNoise.ar(1);  
    src = Formlet.ar( excite, freq,  
      [0.001, 0.01, 0.1, 1].pow(1 - attack) * attack * 0.9 * dur,  
      [0.001, 0.01, 0.1, 1].pow(1 - attack) * dur,  
      ({|i| 4 -i}!4).normalizeSum *  
      AmpCompA.ir(freq, lowestFreq) ).sum;  
    srcReverb = FreeVerb.ar(src * amp, mix, room, tamp);  
    DetectSilence.ar(srcReverb, synthTresh, 0.01, doneAction: 2 );  
    pan = Pan2.ar( srcReverb, pos, 1.0);  
    OffsetOut.ar(0, pan);  
  }).load(s);
```

Figure 5.7: Synthesis definition for the sonification of one surface triangle. See text for further explanations.

`ClipNoise.ar` is implemented in SC3 through a signal random generator with values of either -1 or 1, producing the maximum energy for the least peak to peak amplitude. `RandSeed.ir` is used in order to ensure that the short noise bursts contain the same energy. This impulse is further filtered through a stack of 4 `Formlet.ar` filters. Similar to the *Ringz* filter used in the monitoring application it is a resonant filter with an impulse response of a sine wave shaped with an envelope, allowing to control attack and decay time. The reason to use four filters instead of one is to better tune the resulting sound with respect to the balance between broadband attack versus a distinct pitch. The resulting sound can be panned in the stereo panorama and additionally reverb can be added as a possible distance cue through `FreeVerb.ar`. Since the impulse response characteristic of the filter is the one of a sine wave, a psychoacoustic amplitude

compensation is implemented through `AmpCompA.ir` as shown in Figure 5.7. Through the unit generator `DeductSilence.ar` synthesis is stopped if the signal falls below a defined threshold.

The code part `[0.001, 0.01, 0.1, 1].pow(1-attack)` has the following function: If the attack argument is set to 0 the four filters have all different attack and decay times of exponential order. The factor 0.9 ensures that the attack is always shorter than the decay, which also avoids complications with the `DeductSilence.ar` Ugen. This results in a broadband impulse with a noticeable pitch component, according to `({|i| 4-i }!4).normalizeSum` i.e. the corresponding array of gains. As the attack parameter is reaching 1 the expression `1- attack` reached 0 and the sum of all 4 filters give a narrow-band and soft audible grain.

This synthesis definition is then used in non-realtime synthesis using the `Score` object in order to render a sonification for a selected perspective of the listener towards the sculpture. In the interactive application, this sonification is then convolved through `PartConv.ar` in real time by using one partitioned convolution engine for each of the stereo channels.

5.2.2.1 Two Mappings, Two Parameter Sets

I developed two mappings, both being complementary with respect to the organization of time and space in the data versus the sound domain. These two different approaches are inspired through existing methods as well as our experience from the ISAS project (see El-Shimy et al. (2012)) in which we used similar paradigms to organize time and space.

5.2.2.2 Organizing Space and Time

The first approach is based on a method for auditory image representations, introduced by Meijer (1992). This method is the basis for many sonifications of images; recent applications in experiments with blind participants can be found in Striem-Amit et al. (2012). In this method, images are scanned from left to right and the intensity of the grey values for each pixel is mapped to its corresponding position in the stereo panorama. The vertical position of the pixel is mapped to frequency. This method was also used by Kim and Zatorre (2010), in order to test, whether subjects can develop a mental model of 2-dimensional reliefs based on an auditory representation.

The second approach is based on the data-sonogram, which is an MBS technique introduced by Hermann (2011). In a data-sonogram a virtual shockwave is launched in the region of interest in data space. The simulated response to this impulse by

5. MAPPING IN AUDITORY AUGMENTATIONS

the data points maps the spatial data structure through an invariant principle to a temporal structure. Data-sonograms usually employ frequency or spectral mapping to distinguish data points with different category labels, see Section 2.4.4.2.

Both methods were conceived for data of different dimensionality. Image data points, for instance, have no depth compared to the surface mesh data of the sculpture or more generally high dimensional data point clouds. Scanning along one axis and mapping it onto the stereo panorama as in Meijer (1992), versus ordering data points according to distances in a spherical space as in Hermann (2011) have different effects in the sound domain. The method by Meijer (1992) renders positions along the scanning axis in succession. This means that time separates sounds which are separated in space with respect to their position in the stereo panorama. In the data-sonogram method by Hermann (2011), time separates sounds with respect to their distance to the center of the sphere. The center of this sphere corresponds to the listener’s position, and the time order corresponds to depth information. I will use the abbreviations SW (shock wave) rendering for the data-sonogram-inspired approach and PS (panorama scan) rendering for the approach inspired by the auditory image representation.

As a consequence, the drawback of the PS method is that sounds need depth cues. Contrarily, in the SW method items of equal distance are likely to mask each other and their position is hence difficult to distinguish. However, in this special application the two methods are conceptually not so different but rather similar. This is because the virtual listener position is chosen to be far away, which resulted in two effects: First, the wave front of the impulse is very flat when reaching the data and hence is equivalent to a plane scanning the depth axis. Second, more surface normals deviated less from pointing towards the listener. For the SW method, the shortest delay in time 0 corresponds to the biggest extension of the sculpture in the xy plane towards the listener.

5.2.2.3 Mapping Data Features to Sound Parameters

The mapping scheme which affected sound parameters beyond the temporal organization of the data-to-sound mapping is depicted in Figure 5.8.

For the sonification of the sculptures, the horizontal mapping to the stereo panorama for the PS rendering is adopted from Meijer (1992). For both renderings, PS and SW, the vertical positions of the triangles are mapped to the center frequencies of the filters. Sound design trials indicated that it is aesthetically pleasant to use a wide frequency range spanning from 800 to 8000 Hz. This range makes it necessary to use the psychoacoustic amplitude compensation as mentioned above. Additionally, instead of mapping directly to frequency f the mel m scale is used, for which the formula $f = 700(10^{m/2595} - 1)$ suggested by O’Shaughnessy (1987) is applied.

The next 3 features from the data domain *distance to listener*, *surface normal*, *triangle size* are mapped in a many-to-many mapping paradigm to three properties from the signal domain, namely *gain*, *reverb*, *attack*. This mapping was found after a long phase of experimentation aiming to create a strong variation and hence perceptual contrast in the sonification for the 36 different positions around

the sculpture. During the design process, the most challenging aspect was to balance the amount of the attack so that the sonification appealed to the connotative listening mode of the action-sound-object, in practical terms this means to negotiate between narrow band versus wide band sound grains, which need to give a sufficient attack sound and at the same time must not mask all concurrent sound grains.

The distance to the listener is mapped to gain according to the distance attenuation function described in Peters et al. (2012). As a consequence, surface triangles that are far away are proportionally attenuated. Further, distance is mapped to the mixing parameter of the reverb signal of a small room, in order to use the de-correlation of the sound signal as an additional distance cue. The distance to the listener is also mapped to the attack. As a result, surface triangles that are far away from the listener have a smoother attack and hence the signal is less broadband and the ringing of the filter becomes more noticeable. However since the gain is reduced at the same time, the overall presence of the sound appears weaker compared to proximate surface triangles.

The surface normal is also mapped to gain and attack in an analogous way to distance. Triangles that are facing the listener have a gain of 0 dB which decreases linearly to -100 dB as they turn away from the listener to an angle of $\pi/2$. The linear mapping function is chosen, because it ensures that elements of the surface that are not directly facing the listener are attenuated but still noticeable.

The size of a surface triangle is also mapped to gain, as mentioned above the whole surface is normalized to 1. We have also experimented with a root-mean-square normalization as it was introduced for VBAP systems by Pulki (1997). In this normalization however, small surface triangles were barely noticeable.

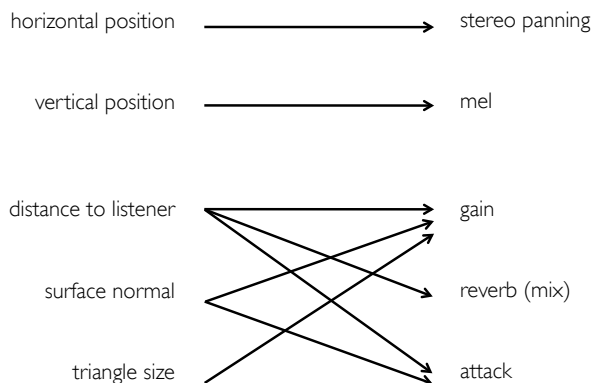


Figure 5.8: Mapping data features to signal parameters

5. MAPPING IN AUDITORY AUGMENTATIONS

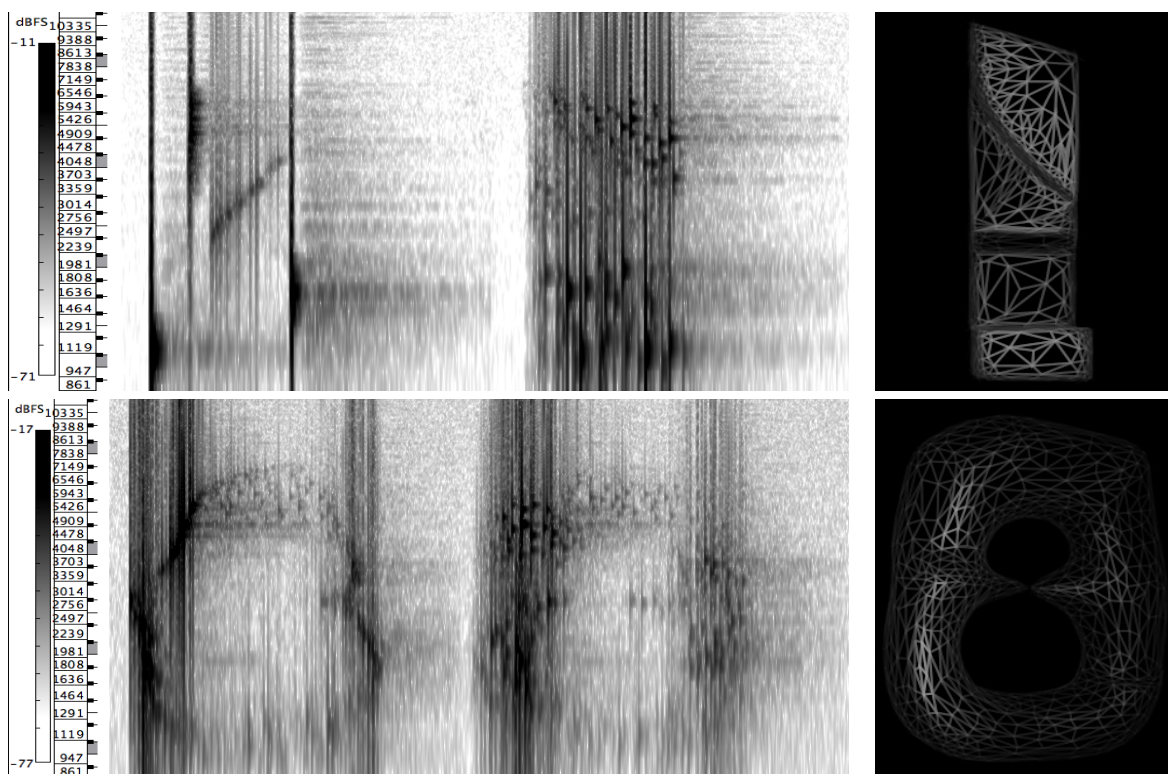


Figure 5.9: Spectrograms and mesh representation of the sculptures. At the top both depict a sculpture from Mount Royal, corresponding to replica B from Figure 5.11. At the bottom they depict a sculpture from Henry Moore corresponding to Figure D. See Section 5.2.2.4 for a detailed description.

5.2.2.4 Participatory Sound Design

The use of echolocation as an interaction paradigm and the design of the audible sculptures for the blind was an interesting challenge from the sound design point of view. The interpretation of sound in echolocation is related to listening skill. This means that the time scale on which salient sound events occur is typically short for the accustomed ear. This does not mean that untrained users don't hear these events, but rather that inexperienced listeners don't know at what time scale attention needs to be focused. Through self-observation during the sound design, the length of the sonification was continuously getting shorter and shorter the longer I worked on this project.

This is why we chose to closely involve a skilled member of our target audience during the sound design process. Through this ethnographic and participatory design approach we wanted to ensure that our target audience finds familiar sonic structures, which would allow them to appreciate the shape of the sculpture and potentially develop a mental representation of it. During the design process, we consulted Michael Ciarciello who is a blind composer, mobility trainer and computer instructor at the Mab-Mackay

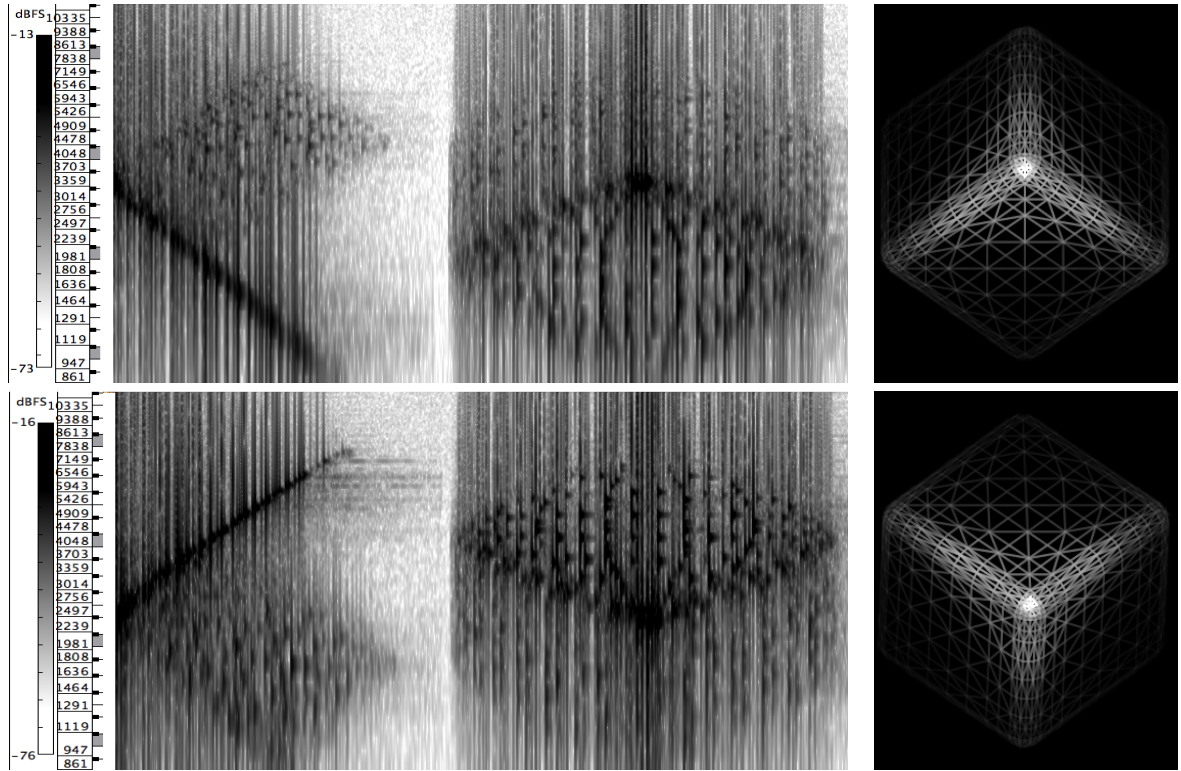


Figure 5.10: Spectrograms and mesh representation of the rotated cube in two different orientations, corresponding to replica C from Figure 5.11.. See Section 5.2.2.4 for a detailed description.

Rehabilitation Centre in Montreal. Through the collaboration with Michael the following two parameter sets (variant A and B) were identified as interesting candidates for a qualitative evaluation.

The ring time of the short impulses is set to two rather short values. For the shortest value the impulses do not have a pronounced pitch, however a distinct height on the frequency scale can be perceived. For variant A 0.05 seconds and for variant B 0.2 seconds are chosen for the ring time.

The mix factor for the dry-to-wet ratio of the reverb signal for variant A is linearly mapped between 0.0 and 0.2 from 0 distance and the most distant point of the sculpture respectively. The room size for the `FreeVerb.ar` is set to 0.2. High-frequency damping is disabled in order to have perceptually comparable effects across vertical surface elements. In variant B no reverb is used, due to the longer ring time.

5.2.2.5 Auditory features in the display

In Figure 5.9 and 5.10, the sonifications of selected orientations of the sculptures for the parameter variant A are depicted as spectrograms, with both stereo channels com-

5. MAPPING IN AUDITORY AUGMENTATIONS

bined. On the right, there is a mesh representation of the sculpture in its corresponding orientation. The two examples in Figure 5.9 show the different effects of the rendering modes, which are described in Section 5.2.2.2.

For the sculpture from Mount Royal (corresponding to replica B from Figure 5.11), the selected perspective gives two profoundly different sonifications. With respect to the distance to the listener, the SW rendering convincingly reproduces the organization of the sculpture, while the PS rendering captures the step at the bottom as well as the diagonal aspects of the top part of the sculpture.

For the sculpture by Henry Moore (corresponding to replica D) both sonifications are similarly structured in time. This is mainly due to the topological property of the sculpture, its symmetry, the orientation towards the listener, and the fact that the sculpture is scanned from left to right. Nonetheless, specificities of the surface, which break the symmetry, are noticeable and cause different sonifications in both renderings.

The cube (corresponding to replica C), which was standing on one corner, was a particularly interesting case. Due to its symmetry the SW rendering gave sonic results that made it difficult to relate it to the shape. The PS rendering however depicted the shape very literally in the spectrogram.

In all spectrograms, some properties of the mapping can be recognized. The correspondence between frequency and vertical position of a surface triangle is the most obvious. Surface triangles that are close and have a surface normal pointing towards the listener have a sharp attack and show a broadband signature across frequencies. Similarly, surface triangles which are not oriented towards the listener have a lower intensity and a smoother attack.

This equivalence between the spectrogram and the shape it represents is reminiscent to some spectrograms show in Meijer (1992). However, intensity in the spectrogram is related in our case to depth information and for variant A additional depth cues such as a reverb is added to the signal. This reverb is however difficult to identify in the spectrogram as it can be confused with the filter ring-time.

5.2.2.6 The Interactive Application for Exploration

For the exploration of the sculptures together with the auditory augmentation, we developed an application prototype that allowed us to explore models of the sculptures by touch and listen to the sonic responses of short impulses in front of a microphone.

A screen shot of the application in use is shown on top in Figure 5.11. The window of the video captured by the isight camera from a MacBook Pro is placed in the middle. The whole setup is captured through a screen-capture program.

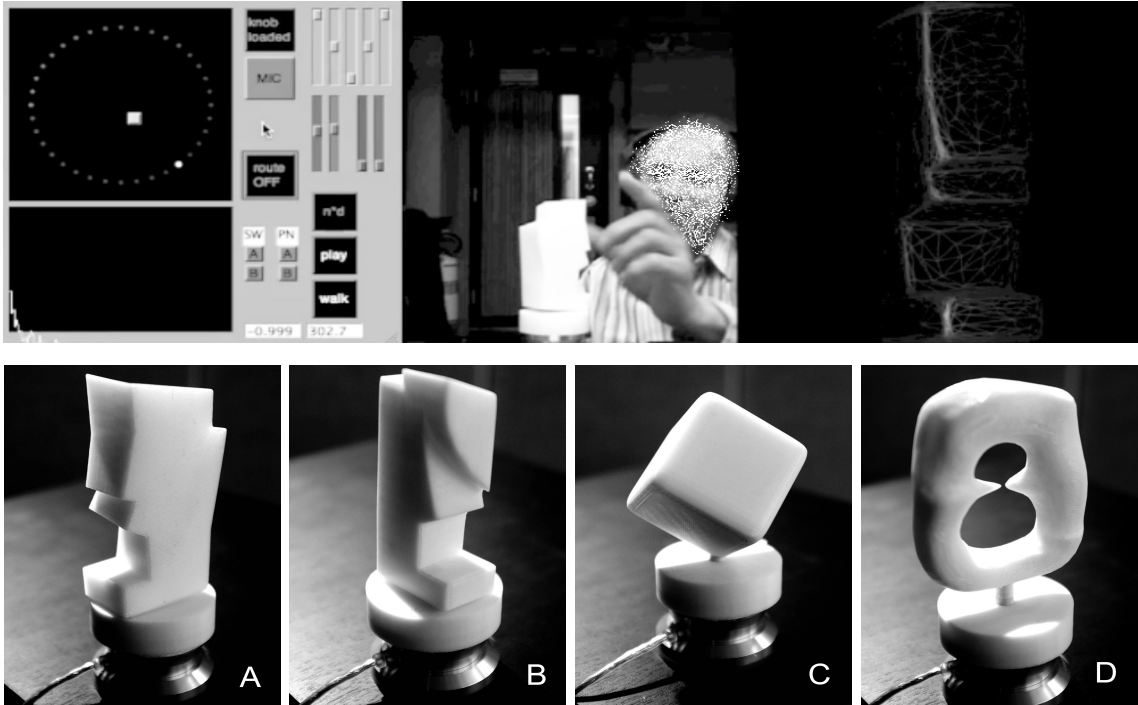


Figure 5.11: On top: Screenshot of the application for the exploration of the audible sculptures. At the bottom: Replicas of the four different sculptures. A and B are models from Mount Royal.

Replicas of the sculptures are shown at the bottom in Figure 5.11. These models, which are produced through rapid prototyping, are mounted on a Power Griffin Mate rotating knob, which gives readings for the position into which the sculpture is turned. This information is used to load the corresponding sonification into the buffers for the convolution. The real time convolution is computed on a MacBook Pro, connected to an RME fireface audio card with a DPA 4011-TL microphone used to record the impulse sounds. The convolved signals are rendered through closed headphones to the subjects. The GUI of the application shows a visual representation on the 3D mesh data which corresponds to the orientation of the replica on the Power Griffin Mate rotating knob. Buttons allow switching between variant A and B as well as the SW and PS rendering modes. The GUI provides means to bypass the convolution so that we could have an undisturbed conversation with the subjects.

5.2.3 Qualitative Evaluation

In order to get feedback in addition to the expertise from Michael Ciarciello we conducted a qualitative evaluation. Four visually impaired subjects participated (3 male,

5. MAPPING IN AUDITORY AUGMENTATIONS

one female, with 47, 40, 58, 56 years of age, two congenitally and two non-congenitally blind.) The qualitative evaluation consisted in an open exploration of the sculpture model A together with its sonifications through clicking in front of a microphone. We recorded the exploration session with video. The introduction to the setup and the exploration took about 30 minutes. We also asked the subjects to perform a small task. The whole evaluation session took about 2 hours and 30 minutes, including repeated breaks. After the exploration session with the sculpture replica A, we gave the subjects a task which they completed for the three different sculptures replicas B, C, D.

5.2.3.1 Exploration Protocol

For the qualitative evaluation we designed the following protocol. First, we gave the participants a brief introduction to the project. We used the narrative of walking around a sculpture in order to create a mindset appropriate for the task. We informed them that they would listen to the geometric properties and shapes of sculptures. We further explained that the interaction paradigm is inspired by echolocation and that the sonic response is inspired by reflecting sounds. We also mentioned that the sonic feedback to the click might deviate from their own listening experience.

The exploration session was split into two parts A and B. In A the participants explored the rendering mode SW and in B the rendering mode PN. The parts A and B were balanced across the 4 subjects. In each part, we presented to the participants the two sonification parameter variants that we had described in Section 5.2.2.4. At the end of the exploration session, we asked the subjects to tell us which parameter variant they liked for the given rendering mode. The variant which they liked best (A or B), was then used for the task described below.

For the exploration, we had implemented a setting that we called *virtual walk*. In this setting, we asked the subject not to touch the sculpture but just to listen. Before the virtual walk started, we turned the model into the starting position so that the subjects knew which orientation the first sound belonged to. We also demonstrated in which sense the sculpture would change its orientation. Then we played the 36 sonifications for a 10-degree angle increment in succession by convolving them each with a prerecorded click. The purpose of the virtual walk was to make sure that each participant heard every position around the sculpture at least once.

5.2.3.2 Orientation Identification Task

After the exploration, which was done with sculpture A, we asked the subjects to execute a task which would allow us to better assess if a mental model of the sculpture could be developed based on the sonification only. The design of this task was inspired through the work by Kim and Zatorre (2010).

This task was conducted with sculptures B, C and D. The subjects were given about 15 minutes to freely explore these replicas, then we gave them a virtual walk around them. During the task, the subjects were asked again not to touch the sculpture but to trigger the sonification of a selected position through clicks. They could trigger the sonification as often as they wanted, knowing that they would be asked to identify the orientation that corresponds with the sound heard. Next, the sound was switched off and the replica was turned into a different position. Then we asked the subject to turn the replica back into the position which they thought to correspond to the sound. For each sculpture, 3 different positions were tested through this task. We selected positions that did not have an obvious sonic signature but rather deviated from easily recognizable sounds. Like this we hoped to better test the mental model rather than memory.

5.2.3.3 Results

The results of the task are compiled in Table 5.3. The columns P1 - P4 show the result for each participant. The column *position* shows the orientations that have to be matched. The parameter setting indicates the choice that was made with respect to the sound design options A and B. Bold entries are orientations that match the position ± 20 degree. For the cube, the threefold (120°) symmetry is taken into account and orientations that match the position ± 15 degrees are bold. Dark grey entries indicate orientations which exhibited a similar profile. We chose the big margins of ± 20 and 15 degree respectively because it was not sure how the participant perceived their own orientation towards the replica. Hence they could only approximately match the heard sound with a selected orientation of the sculpture.

Due to the small sample of subjects and the qualitative nature of the evaluation by freely exploring the sculptures, we present the results case by case and do not provide in depth statistics across subjects. For the SW rendering two subjects chose A and two chose B, however participant P2 considered A but finally decided for B, shown as B_A . This gives a very small preference for A in the SW mode. For the PS rendering all subjects chose variant B, i.e. the longer ring time of the parameter set.

11 out of 36 positions were recognized in the SW rendering, and 8 out of 36 positions were recognized in the PS rendering. This poor ratio would suggest that no mental

5. MAPPING IN AUDITORY AUGMENTATIONS

| | position | P1 | P2 | P3 | P4 |
|--------------------------|----------|------------|----------------------|------------|------------|
| rendering mode SW | | | | | |
| parameter setting | | | | | |
| | | <i>B</i> | <i>B_A</i> | <i>A</i> | <i>A</i> |
| cube | 0 | 121 | 132 | 314 | 124 |
| | 180 | 121 | 80 | 263 | 282 |
| | 130 | 100 | 120 | 120 | 314 |
| MTL | 260 | 260 | 251 | 37.8 | 291 |
| | 190 | 207 | 37.8 | 69 | 302 |
| | 50 | 57 | 326 | 243.9 | 219 |
| Moore | 90 | 270 | 117 | 26.1 | 144 |
| | 330 | 302 | 17 | 77.3 | 302 |
| | 280 | 128 | 280 | 263 | 211 |
| rendering mode PS | | | | | |
| parameter setting | | | | | |
| | | <i>B</i> | <i>B</i> | <i>B</i> | <i>B</i> |
| cube | 0 | 231 | 215 | 116 | 203 |
| | 180 | 203 | 77 | 77 | 88 |
| | 130 | 148 | 295 | 96 | 33 |
| MTL | 260 | 124 | 314 | 89 | 357 |
| | 190 | 148 | 330 | 330 | 295 |
| | 50 | 57 | 77 | 254 | 65 |
| Moore | 90 | 295 | 3 | 314 | 88 |
| | 330 | 199 | 65 | 57.3 | 329 |
| | 280 | 61 | 330 | 238 | 274 |

Table 5.3: Tabular results of all participants from the orientation identification task. See 5.2.3.3 for a detailed description.

model was built. The orientations indicated in dark grey are easy to confuse with the tested one due to symmetry.

Two participants however, P1 for the sculpture MTL in the SW rendering mode and P4 for the Moore sculpture in the PS rendering mode, matched all 3 tested orientations. Participant P2 matched two positions of the cube in the SW rendering mode. Remarkably P4 deviated from the target orientation with a maximum of only 6 degrees, P1 with a maximum of 17 degrees. The probability of guessing the right orientation as being one of 9 segments in the 360 degree circle around the sculpture is about 11% in one task. The conditional not-independent probability of repeatedly guessing correctly three different orientation in succession equals approximately 0.2%.

5.2.4 Discussion

The results from the qualitative evaluation lead to the following conclusions. Since three orientations of two different sculptures were sufficiently well-matched by two different

participants, it is reasonable to assume that a mental model of 3D shapes can be built based on the sonification methods developed in this project. It is not entirely clear if the matching orientations were found on memorizing the order of the sounds from the virtual walk. Distinguishing 32 sounds, which vary fairly smoothly with the angular increment, seemed however a difficult task to remember without repetition.

Further trials need to be conducted in which we improve our setup with respect to the orientation the user assumes towards the sculpture. Another possibility is to provide additional means for the subjects to relate the orientation of the sculpture with respect to an absolute reference point.

In the SW rendering there was a slight tendency for the sound design variant A. This might be due to the fact that in this rendering, sound events often occur at the same time and it is hence more pleasant to have sounds with a shorter ring time. In the PS rendering, the sound design variant B with the longer filter ring time was preferred because of its more sequential distributions of sound in time and across the stereo panorama. No matter which variant was chosen, participants did not complain about the sounds despite the long duration of the exploratory sessions, which lasted approximately 2 hours and 30 minutes.

Subjects liked the PS rendering but performed slightly worse in it. The fact that this rendering mode is unusual from a perceptual perspective compared to the echolocation-inspired SW rendering made the subjects curious but equally made it challenging for the task.

It was for us particularly interesting to observe that all subjects explored the whole replica first and later often just touched the rotating knob to turn it. All participants only occasionally verified the orientation by touching the whole replica again. This implies that all subjects build very quickly a strong mental model of the shape and tried to match imagined orientation of the mental model with the perceived sound.

5.3 Summary and Outlook

I discussed in this chapter two parameter-mapping sonifications for auditory augmentation applications. Both applications extended the already existing work of [Bovermann et al. \(2010\)](#) by exploring possibilities of how in auditory augmentation information can be distributed in time. The organization of geometric data features as temporal structures was inspired in both cases through the MBS method of data sonograms.

In the monitoring application, the mapping of data features to time delays was controlled by user interaction in order to get a better control over the obtrusiveness

5. MAPPING IN AUDITORY AUGMENTATIONS

of the display. This was possible because the mapping to signal delays was implemented as real-time sound synthesis. For the audible sculptures, the sonification was rendered offline due to the large dataset and the augmentation was realized as real time convolution.

Complementary to the mapping in time, both applications also exploit the mapping of data features to positions in the stereo panorama. In the monitoring application, both the spatial and the temporal mapping was a one-to-many mapping. Both parameters represented the optimization progress with respect to the development of the mutation strength. In the audible sculpture application the temporal structure was coupled in a similar way to the stereo panorama in the PS rendering mode. I suspect that the subjects appreciated this rendering mode in the audible sculptures, because of this redundancy in the mapping.

In the SW rendering, the depth information of the 3D shapes corresponded more directly to the echolocation experience, and the temporal and spatial evolution of the sound was coupled through the geometry of the 3D shape. Although more natural in this sense, it was interestingly less preferred by the participants, which suggests some perceptual advantages for one-to-many mappings. However it must be noted that participants performed slightly better in the SW mode. The sonic articulation in the SW mapping was also more uniform, as louder sounds always came first and were followed by distance-attenuated sounds. It will be interesting to explore in future evaluations if the distance to the sculpture can be better estimated in the SW rendering mode compared to the PS, where distance cues were mostly represented through loudness.

The center frequency of the filters was another important sound feature exploited in both parameter-mapping approaches. In the monitoring application, the center frequencies served mostly as dynamic labels and were part of a one-to-one mapping representing the current position in the high-dimensional search space. Their function was to indicate change in the search space by altering their position in the spectrum, thereby creating a salient audible transient feature. This mapping was used in combination with the distribution in the stereo panorama and by distributing them through time delays, in order to decrease the effects of masking within one auditory augmentation compared to the version without delay, where all augmented copies of the impact sounds were rendered at once.

The center frequencies in the sculpture sonification were used as a continuous variable to represent the vertical position of the surface triangles. During the design process it was interesting to discover that only high frequency ranges gave a uniform sonic impression for the vertical information. For low frequency ranges the temporal resolution was simply not sufficient. Despite the high location in the spectral range that was

finally adopted, the spectrogram in Figure 5.9 still shows the difference in temporal resolution. The high spectral location made it necessary to use the Mel scale for a perceptually uniform mapping.

In both approaches, I attempted to create sounds that appealed to the action-sound-object of the connotative listening modes presented by Tuuri et al.. This was realized by transferring the physicality of the interaction sound to the sonification, either by realtime filtering and signal delay lines or through pre-rendering of the sonification, which was based on impact sounds and convolving it in real time with the interaction sounds. The distribution of data features in time added an echo-like effect to the action-sound-object of the connotative listening intention thereby also engaging the causal field from the denotative intention. The noticeable delays separated action and perception and thus suspended to some degree the closed feedback situation.

Mapping of elements of the σ vector to filter bandwidths in the optimization monitoring application was firstly motivated by the metaphor of the Gaussian distributed mutation. The sonic effect for the augmentation was that for open filters the augmented sound was almost identical with the interaction sound. As a result the augmentation was not really present in the beginning and listening modes proper to interaction such as *ergo audition* found a matching sonic substrate. As filters closed, the interaction sounds were augmented with more artificial sound overlays and one could experience how the listening intention shifted to pick up distinct frequencies. Similarly in the audible sculpture project, the filter bandwidth constituted the major difference for the parameter settings A and B. In 6 out of 8 cases the participants opted for the longer ring time B. Setting A also had a small reverb component to it as additional depth cue which might have also influenced the choice in favor of B. The ring time in the option B was apparently small enough and did not create too much temporal overlap.

For future auditory augmentation applications, it is interesting to note that more complex spectral contours such as in vowel sounds did not give satisfying results. The formants of one vowel occupy too many regions in the spectrum. The simultaneous use of many complex spectral contours made them mask each other. Hence there were not enough sound parameters to represent all the dimensions of the search space. Vowel transition, which led to distinct articulations for single auditory streams in Section 4.3 and which also worked well as audible labels in the data sonogram in Section 2.4.4.2 could not be used as a continuous mapping variable for the vertical position of the surface triangles. This was mostly because the spectral position of the determining formants is low compared to the finally adopted frequency mapping range. This leads to the conclusion that the more complex the data become, the smaller the spectral footprint for each sonified data point needs to be for a perceptually transparent sonification. This

5. MAPPING IN AUDITORY AUGMENTATIONS

means that mapping dimensions such as complex spectral contours as timbre or pitch as tonic mass get lost and only the height in the spectral range, gain and to some degree the attack remains as a sonic feature to which data can be mapped. It also makes the sonification shift towards higher spectral regions, so that the overall *sonic gestalt* can be articulated in a short amount of time. Datasets with many data points seem to enforce a mapping that is reminiscent of the Lombard effect (see Lombard (1911)), where in a crowded and noisy environment interlocutors involuntarily raise their voice and shift their register up in the frequency domain in order to make themselves understood.

Despite the smaller amount of mapping targets, the resulting sound of the sonification can still exhibit complex timbre developments but those are less a mapping target and rather an emergent property influenced by the mapping and the data structure. This highlights the role of listening within the PMSon design process: whilst low level mapping targets can be identified and perceptual rules such as psychoacoustic amplitude compensation and the Mel scale can be incorporated, the overall spectral evolution of the sonification needs to be assessed through reflective listening which provides feedback on how to tune the mapping of the available low level parameters.

6

From Case studies to Listening-Mode Based Guidelines for Exploratory, Multimodal and Interactive Sonifications

The integration of sonification in interactive and multimodal displays as featured in the application prototypes of the last three chapters about sonifying ancillary gestures, auditory graphs and auditory augmentations, confronts the display designer with an ever growing number of elements to consider and integrate during the design process. As outlined in the introduction of this thesis, the necessity to provide guidelines for integrating interactivity and multimodality was identified by [Hunt and Hermann \(2011, page 295\)](#) and [Frauenberger \(2009, page 167\)](#).

In this chapter, I propose integrated design guidelines for sonification in applications with multimodal and interactive elements. These guidelines are based on (A) the literature discussed in the first part of this thesis, (B) the experience made during the design processes as well as (C) the evaluations of the practical applications. The design guidelines provide a structure for listening-mode-centered reasoning about the relationship between sound, interactivity and complementing modalities between descriptive and normative purposes of sonification. The focus of these guidelines is to look at how the potential of audible phenomena to become signifiers, i.e. to refer to the data substrate, is influenced through interactivity and multimodality. Rather than rules, these guidelines take a diagrammatic form in [Figure 6.1](#) which provides an overview over the central issues in the design process.

6.1 Listening Modes and Display Purposes

In the guidelines, the following aspects will be discussed from the perspective of listening modes:

- The auditory display system, which comprises interactive and multimodal aspects (compare Hermann (2008)). Design considerations about the auditory display system need to include the aspects of repetition, interaction and complementing modalities and apply independently of the sonification method.
- Sound design will be discussed with a focus on PMSon where many sound related questions are left for the designer of the display to decide, as discussed in Section 2.3.
- The role of embodied listening modes will also be considered by integrating the experience of the vowel-based sound synthesis in the practical work of this thesis and the new taxonomy of listening modes introduced by Tuuri as discussed in Section 2.2.5.

When conceptualizing an auditory display system, I propose to identify first how the normative or descriptive purpose of the display (compare Section 2.2.7.2) can connect with listening intentions of the user. In the monitoring example in Section 5.1 for instance, the closed loop monitoring of the optimization process is an example of a normative display purpose, with the defined goal to communicate the optimization progress. The active querying of the state of the optimization in turn is an example of a descriptive display purpose, where the user more consciously turns towards the display and tries to interpret optimization states.

The term *descriptive* was inspired by an article of Grimaldi and Engel (2007) in which the authors argue for the relevance of descriptive science, which create taxonomies and systems based on the identification of forms and shapes proper to the phenomena under study. This is in essence the idea of the sound object for which Schaeffer developed a taxonomy, based on the typology and the morphology of sounds. The descriptive purpose also parallels the oscillation between description and identification in the sound object through *reduced listening* (compare Figure 2.2). Commonly highlighted advantages of sonification can be identified with the descriptive pole, which is to transmit information on a sub-symbolic level, or the notion of sound as conveying data holistically and having the potential to communicate or represent complex data

structures¹. Particularly the latter is interesting because *reduced listening* is an effort to circumvent complexity reducing mechanisms in our perceptual and cognitive apparatus. The descriptive pole can be associated with sonifications supporting exploration tasks, where sound serves a *descriptive purpose* for complex data relations. In the practical applications of this thesis, the focus in almost all cases was to build a descriptive display with varying complexity of the data, where the least complex was the auditory graph in Chapter 4 and the most complex the sonification of 3D shapes in Section 5.2.

On the other side, we find the normative display purpose where the sound is meant to inform the user about well defined events or actions. Here auditory display shows advantages for being eyes free, having a high temporal resolution, and not always requiring a directed focus and orientation. On this pole, the auditory display system has an emphasis on feedback, control and conditioning of actions in an HCI context and sound serves a *normative purpose* because the focus is less on the sound itself but rather on whether sound helps to achieve a predefined or desired goal. The explicit elements of auditory graphs in Section 4.2 and the background mode in the monitoring application in Section 5.1 are examples that come close to the normative purpose.

For the associations of listening modes with display purposes the taxonomy proposed by Tuuri and Eerola (2012) is best suited, as it is hierarchically ordered. Here the most normative purpose would be an auditory alarm, which needs to cater at first to reflexive listening modes. Alarms also illustrate the multiplicity of purposes, since they need to convey information beyond the initial reflexive reaction. A descriptive purpose as in exploratory data analysis, corresponds to reflective listening modes and here more specifically to *reduced listening*.

The diagram in Figure 6.1 is vertically organized according to the two poles, *descriptive* on the left and *normative* on the right. This diagram was conceived to support the reflection on repetition, multimodality and interactivity and their roles for the two different display purposes. In an auditory display system with a descriptive purpose, the holistic aspect of sonification, which is to potentially communicate information to the user in a subsymbolic way, is related to the practice of *reduced listening*. Since this listening mode involves higher cognitive capacities, as discussed in Section 2.2.5, it is a listening mode that supports reflection on sonic structures as a prerequisite to relate them to data structures. Here we also find a possible connection of the notion of aesthetics and the semiotics of sound in a sonification context, where aesthetics extends beyond the mere design for pleasantness. In *reduced listening*, the sound does

¹Here, the notion of complexity does not follow any particular definition but loosely tries to take into account the amount of data points, their relations and whether these data are discrete or continuous, also compare the DSM by deCampo (2007).

6. FROM CASE STUDIES TO GUIDELINES

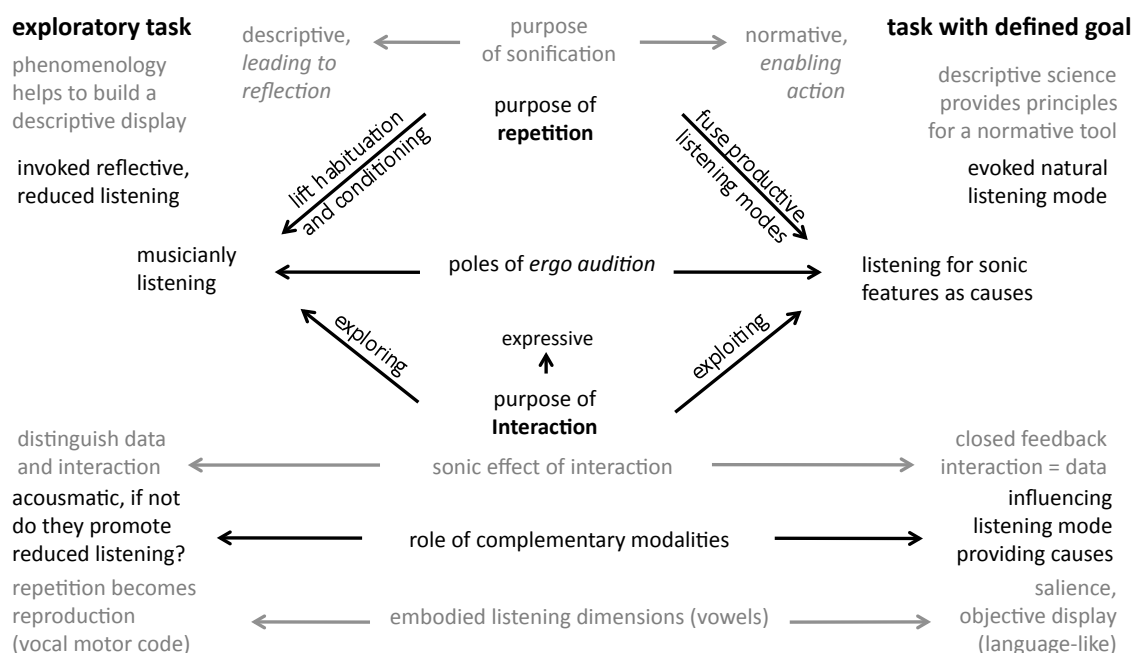


Figure 6.1: Design scheme for multimodal interactive sonifications with key aspects discussed in the sections of this chapter.

not take on at first a referential but rather an aesthetic function¹. Sound does not stand as a sign for something else but is first and foremost perceived for itself. In this sense, design understood as facilitating the perception of the sound object adopts an aesthetic dimension. This is not in contradiction with its consistent link to the data substrate, but rather a prerequisite after which less biased references to the data can be discovered.

While the detailed terminology describing the sound object has to my knowledge not been integrated in specific listening skills other than for electroacoustic composition, it can still be assumed that *reduced listening* practices are important parts of listening skill acquisitions and execution, considering that it can be understood in its weaker form as circulating through all four listening modes, see Section 2.2.1.1, or an activation of various modes, see Section 2.2.6.

Since listening intentions are an observer (listener) category, they cannot be directly influenced like sound synthesis parameters. During design cycles, and during the conceptualization of an auditory display system, questions need to be raised about the role of complementary modalities with respect to specific listening intentions as they

¹The aesthetic function is a linguistic term which I borrow here from Mukařovský (1989, page 62), his notion of the *aesthetische Einstellung*, share similarities with the *reduced listening* intention.

determine how the listener uses the sound. Tasks can also have changing demands while being executed, such as the monitoring tasks (compare Section 2.2.3 and in Section 5.1) and hence provoke or demand a change in the listening intention.

In this sense, the *evoked* potential, understood as the reaction to salient sound properties, needs to be put to use for normative aspects. For exploratory tasks, the *invoked* potential needs to be raised, which is to create auditory display systems that allow to engage premeditated listening intentions. In the diagram in Figure 6.1, the following major factors are identified and described in the sequel.

6.2 Repetition

Due to the fleeting nature of sounds as signs that evolve in time, they cannot be next to each other in the same sense as two static visual signs can coexist next to each other. If sounds are played at the same time, they can merge, split, highlight certain aspects of their differences, mask others, or even cancel each other. While even two co-occurring visual signs influence each other, their persistent nature allows reinvestigating and reevaluating in which aspects they differ or whether they are the same. For sound, reinvestigating and reevaluating depends inevitably on repetition. Or as Chion (1994, page 30) puts it for *reduced listening*: “*the descriptive inventory of a sound cannot be compiled in a single hearing. One has to listen many times over, and because of this the sound must be fixed, recorded.*”

As stated in Grond and Hermann (2012a), repetition is also an aesthetic choice favoring engagement with sound because it creates more than just one unique event which, due to its singularity, can only fascinate but cannot be understood. Paradoxically, repetition can induce both sameness and difference in sound. Sounds can differ in repetition as different listening modes can set in, reducing sound to events, objects, indices or features. In this process direct listening intentions can also become exhausted and as a consequence the formation of a more persistent sound object can emerge (compare Section 2.2.1.1). Hence in the design cycle one needs to pay attention to which option is more likely to be engaged through the context – other modalities or interaction, (compare Section 6.3) – in which the sound appears.

It is interesting to recall that the sound object was discovered partly through the repetitive experience of the closed groove (compare Chion (1983, page 13) and Section 2.2.1.1). For auditory display systems it is equally important to analyze how technology can help to unlock the effects of repetition. Repetition can be either rooted in the data structure; it can be realized through repeated playback or through interaction. For the latter, the degree of sameness depends on how the action of the user

6. FROM CASE STUDIES TO GUIDELINES

translates into sound and if the data-driven aspect of the sound remains recognizable, a distinction, which will be discussed under Section 6.3.

In Section 3.2, repetition was part of the task of repeatedly and consistently identifying events in the display. Since the display was audiovisual for most stimuli, the listening mode can be assumed to be causal. It would be interesting to evaluate if even in this case the listening intentions can be exhausted, and whether repetition would change the focus within the multimodal display from the visual to the auditory modality or vice versa. Although the length of the stimuli is likely to have exceeded the length of a sound object, that can be retained in memory, the support of the intentional unit as the *gestural-sonorous object* might allow for longer sonic sequences to be retainable in memory. It is noteworthy that in the sound-only condition as well as in the multimodal condition the selection of the identified events appeared to be more consistent, see Figure 3.5. This suggests that auditory perception is a powerful modality to store and recall time structures in repeated presentation of extended movement sequences, which are themselves not repetitive.

The effects of repetition have been encountered in the evaluation of *Singing Function* in Section 4.3. Here, sounds which were exactly the same and heard in succession were occasionally judged as being different. Interaction through a slider made it even more difficult to reproduce the listening experience. This had interesting effects for sounds that were the same, where the addition of interaction did not influence the rating of the listener. However for sounds that were in fact different, interaction made the listeners judge them as being the same. For future developments this insight might call for a discrete rather than a continuous display for the interactive mode of auditory graphs, as it was used in *Sonic Function* in Chapter 4. The hypothesis is that it is easier to compare relations in a set of well-formed but only partly ordered sound objects rather than comparing the similarity of complex sonic evolutions. Repetition in *Singing Function* occurs also on another level, namely that characteristic elements of the graph such as extrema show a similar sonic gestalt with respect to three sonic qualities: pitch contour, timbral development and change in brightness.

In the monitoring application in Chapter 5, one user gave interesting feedback, wishing for more explicit control over the display, which would essentially allow for the repeated augmentation in a selected mode. This feedback can be interpreted that auditory displays for exploration should always offer the possibility to easily repeat the sonic experience.

Although I did not test how the the sound signal after convolving it with different finger-clicks in Section 5.2 was perceived, the variation seemed not to disturb the user experience. The encounter of small differences rather encouraged repeated interaction,

which potentially pushed the listening intention towards *reduced listening* trying to identify the sonic type – the sound object – that matches all the sonic tokens, being specific instances through interaction.

With respect to repetition, I suggest paying attention to the following points in an auditory display system with a descriptive purpose:

- **Data:** Identify the potential for repetition through repetitive structure in the data. Evaluate how sonification can lead to recognizable sound objects, which at the same time provide noticeable audible contrast for variations in the data.
- **Interaction:** Identify the potential for repetition through repeated interaction. Does the interaction device allow repeating the sonic experience?¹ Carefully evaluate how the singular instance of a sonic interaction relates to all possible sonic interactions with the data set or subset and whether stable and comparable sound objects have the potential to emerge.
- **Playback:** The simplest way to repeat is to re-trigger the playback of the sonification. This can offer interesting possibilities together with recorded interaction data. This kind of repetition through playback can offer a more reflective approach towards the sound and its cause for skill learning situations beyond mere conditioning.
- **Reproduction:** Repetition can also be achieved through imitation by the listener. This applies explicitly for vowel-based sonifications, but also implicitly for all sounds that evoke forms and *gestalts* that relate to the *gestural-sonorous object*. In the latter case, the sound cannot necessarily be reproduced but the images of effort and chunking can be re-invoked by the listener.²

The first, second and the last point are partly based on the experience from the practical applications in this thesis. The third point is a conclusion based on the notion of *musicianly listening* for the context of skill learning.

¹Also compare Hermann (2008), who raises the question of repeatability in interaction.

²Strictly speaking, this does not lead to sound objects, for which the sound needs to be fixed. For the concept of vocal sketching compare Ekman and Rinott (2010).

6.3 Interaction

The importance of interaction in sonification has been summarized by Hunt and Hermann (2011) with an emphasis on the physicality of the system and the flow that is experienced by the user in closed loop applications.

As discussed in Section 2.2.4, Schaeffer conceived the idea of the sound object and listening modes tightly integrated with thoughts on sound making. I would like to propose here a complementary, listening centered approach towards interactive sonification, which is inspired by the notions of *ergo audition* and *musicianly listening*. I suggest organizing interactive sonification according to the following three categories of sonic interaction: *sonic exploitation*, *sonic exploration* and *sonic expression*. For interactive sonification as a data display, the first two are the most relevant, all three categories are schematically compiled in Figure 6.2.

Sonic exploitation is the use of specific characteristics of sound in order to support a task with a specific goal. This category corresponds to natural listening modes, where sound is used to provide information about objects and causes. In the normative sonification task, sound is exploited to provide feedback and is reduced to features that support maintaining the feedback loop. In *sonic exploitation* the listener strikes a balance by reducing sound to features that are (a) informative for the action and (b) extractable from the sound with the smallest possible effort. Although the closed-loop situation as real-time application strikes a balance between perceiving and acting in so far as the interaction data are fed back to sensory input modalities, an emphasis is put on perceiving in order to act. In the most explicit case of an auditory closed loop, the distinction of data and interaction vanishes. The monitoring example based on auditory augmentation from Section 5.1 is in this sense halfway between exploration and exploitation.

Sonic exploration corresponds to the idea of *musicianly listening*, see Chion (1983, page 39). This listening mode can be best described as playful sonic interaction with everyday objects. Here the emphasis is put on acting (conscious and not reflexive) in order to perceive and consequently to reflect on what was perceived. *Sonic exploration* is like *musicianly listening* tied to an attitude of preparedness to closely attend to potential sound features before producing the sound. The focus of the listener turns away from a specific task towards the plethora of audible features of a sound. At first, these features do not represent anything in particular but are discovered in the process of asking what is at all perceivable. Only in a subsequent step, sonic features are tied back to the physical aspects of interaction and the nature of the sounding object. For musical instruments this corresponds to the exploration of the potential of the instrument for musical expression; here the sonic features need to be tied back

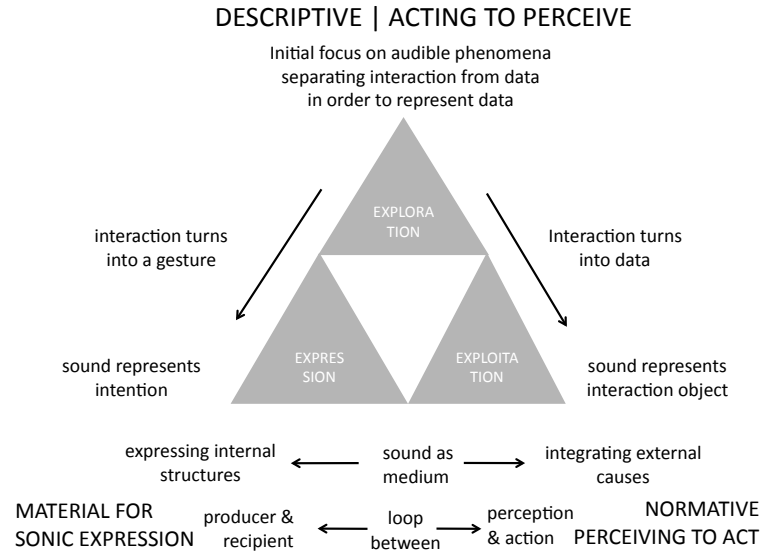


Figure 6.2: Modes of sonic interactions, for a description see text in Section 6.3

to the interaction in order to be able to reproduce the sound in a controlled way. In an auditory display system the sonic features need to be understood and perceived as indices to data properties.

Sonic expression is in my opinion a musical category, which is relevant for new instrument design (compare Section 2.3.1.3 for a PMSon-related discussion). *Sonic expression* comes after the process of *sonic exploration* if interaction and sound synthesis correlate in a way that is not only consistent but also musically interesting. Expressivity in interaction has an intentional and idiosyncratic element to it and shadows the consistent connection to data, by shifting the focus onto the sound producing gesture away from the sound producing causes, which provide in the case of sonification the link between the sound and the data. Outside the context of interactive sonification, *sonic expression* can very well play an important role in auditory display. By representing intentions, sonic expression is well suited to communicate internal states in the context of HCI, as has been explored in detail for vowel-based sounds by Tuuri (2011).

In Figure 6.2, *sonic expression* and *sonic exploitation* are both at the bottom of the triangle. In both cases interaction dominates data since the resulting sounds either communicate an intention or represent external objects and causes related to the interaction. For both cases, the option to record the interaction data for playback and repetition is worth considering (as discussed in Section 6.2) in order to move towards the descriptive corner of the triangle on the top. Despite the earlier statement that

6. FROM CASE STUDIES TO GUIDELINES

sonic expression is more of a musical category, I chose this triangular organization in the diagram to highlight that there is no simple distinction that can be made between musical sounds and auditory display. An interactive sonification with a normative purpose seems particularly incompatible with musical goals, but at the same time in both, in *sonic expression* and in *sonic exploitation*, sound takes on the role of a medium and closes loops, contrary to *sonic exploration*, where sound stands at first for itself.

In comparison with closed-loop auditory display systems, where the interaction provide the data that are turned into sound, in *sonic exploration* the listener tries to distinguish interaction and data properties (also compare Section 2.2.4). This is important for sonification because only the instrument's timbre is carrying the information about the object, which corresponds for the field of sonification to data properties. The player's timbre belongs to *sonic expression*. If it cannot be perceived as separate from the instrument, exploratory interactive sonification makes the data disappear and leaves the user with sound-making, pure and simple.

This distinction between the player's timbre and the one of the instrument can be found in the auditory augmentation in Chapter 5. The possibility to distinguish both easily is most likely due to the small but always noticeable delay between the interaction sound and its convolution with the sonification. This echo-like separation of the interaction and the sonification sound also emphasizes the purpose of action to facilitate perception. This separation of unimodal events will later also be discussed as an option for complementing modalities.

Interaction was also an integral part in the sonification of Chapter 4, where it supported the exploration of auditory graphs. As it has been discussed in the previous Section 6.2 on repetition, this kind of sustained interaction made it difficult to compare sounds that succeeded each other. The sonic effect of the interaction was very prominent and the subjects tried to find anchors such as navigating towards both extremes of the sliders, (see Figure 4.11), in order to get more control of the interface, and in consequence over the sonic output. The increased confusion of different sounds as being identical suggests that this interaction mode did not support greater attention to the sound and understanding of its global evolution. Sonic interaction contained here an aspect of *sonic expression* and hence did not induce a transparent auditory display, with respect to the underlying data.

Beyond the task of comparing auditory graphs, where interaction performed worse than playback, it is still useful to navigate towards points of interest in the display for understanding the concept of extrema or turning points in auditory graphs. However, the question remains open how sustained interaction can be connected with *sonic exploration* such that it allows for an audible distinction between interaction and data

properties. The combined mode of a continuous and a discrete sonification in *Sonic Function* in Section 4.2, which on the one hand respects the dense set of values on the x -axis, and on the other hand accentuates sound objects and makes them comparable during the interaction, offers future avenues for designing auditory graphs.

With respect to interaction, I suggest focusing on the following three points in an auditory display system with a descriptive purpose, the last two points need to be considered together with the reflections on repetition:

- **Sonic exploration:** Analyze how interaction supports sonic exploration, i.e. a turn towards all audible features of the sonification, rather than just representing the interaction itself. This point is very general and provokes thoughts about which audible aspects connect with the interaction as its cause, and if they possibly detract from an attitude of preparedness towards listening. Which aspect of the sound refers to which aspect of interaction (interaction quality, interaction progress, sound localization and proprioception, etc.), to which between the reflexive and reflective listening modes do they most likely appeal? The design goal is *acting in order to perceive*.
- **Player's and Instrument's Timbre:** Borrowing this distinction from Schaefer, this point pays attention to the distinction between the audible effects of the interaction versus the sound originating from the data. If applicable, a temporal separation between interaction and sonification can be considered in order to separate interaction and data-related sonic aspects. Under this point sonic or other modalities can be considered to introduce the distinction between both, always evaluating whether it enables reflective rather than reflexive listening intentions.
- **Order of temporal evolution:** This point focuses on how sound objects can be created so that they are retained in memory and compared. Answers to this point need to address the question of how interaction intervenes with information that is contingent with the temporal order of sounds in the sonification. If the interaction is sustained¹, like moving along a slider, a discrete sonification is worth considering. If the interaction is discrete like triggers, temporally extending sonifications can be considered.

The first point is derived from theoretical reflections on listening and sound making particularly the notion of *musicianly listening*. The second and the third point are based on the experience made in the practical applications of this thesis.

¹This example conforms with the notion of navigation in the closed loop diagram by Hermann (2008, page 6).

6.4 Complementary Modalities

Today, sonifications are often conceived within multimodal displays. For design guidelines, I propose that complementary modalities, in which an auditory information display is embedded, are assessed in terms of how they influence listening modes, and if the evoked listening modes foster the integration of the complementary information provided. In audiovisual displays for instance, the indexical listening mode has been discussed as causal listening in Section 3.1.3. In this mode audible events are perceived as emanating from visual events happening simultaneously in the visual field. This intention to attribute the sound to visual correlates can be strong (compare the notion of *synchresis* in Section 3.1.3), and as a result the perceived sound can be bent and altered to fit a fused percept (compare the ventriloquism and McGurk effect in Section 3.1.3).

From a sonification and auditory display perspective – having in its early definition an emphasis on non-speech sounds – it is interesting to note that the strong fusion and crossmodal influence of audiovisual percepts have been described in speech. Attributing the sound of the voice to a visual cause, which leads to searching for the lips from which the voice originates, presumably serves the purpose to anchor the focus and extract redundant and complementary information which supports the intelligibility of a voice by adding complementing articulatory information. In this sense the audiovisual phenomenon of speech-producing lips is complementary beyond the causal listening intention, which only attributes sound to a source and is presumably exhausted once the source is found. However for speech, it needs to be kept in mind that the percept is strongly shaped through language and is not only a sonic category.

With respect to sonification in a multimodal data display, it is interesting to ask if crossmodal percepts can be created which fuse similarly strong like voice and lip movements in speech. Such displays would constitute a complementarity beyond the causal listening intention, which only makes one modality refer to the other. The *gestural-sonorous object* seems to be a similar case. The prototypes presented in Chapter 3 developed approaches to sonify them in an audiovisual display. It remains however open how much the presence of several concurrent auditory streams and visual causes allowed to go beyond the causal listening intention.

This predominant causal listening mode in the audiovisual complex does however not mean that visual or other modalities completely impede *reduced listening*¹. Here insights from perceptual psychology on crossmodal influences can add details to the causal listening mode. As reviewed in Section 3.1.3, moving images and sound cause a stronger grouping of the percept than still images. Based on this finding it is reasonable

¹Chion (1994, page 32) even argues that the acousmatic condition might promote causal listening by prompting the listener to ask: "What is that?" (In the sense of what causes this sound).

for multimodal sonification design to consider if complementary modalities provide time-based or persistent sensory input. If sound is complemented by a time-based sensory input it is more likely, that the listening intention is causal and hence that the user needs to make a very conscious effort to attend to the sound and its details. Beyond the strict dichotomy between time-based and persistent, the intensity of the change in the other modality as well as the synchronicity can also be considered.

For the exploratory sonification of ancillary gestures in Section 3, two options can be considered to create a display that supports *reduced listening*: First, it might be worth considering to replace the dynamic visual with a movement score. Second, for relatively short time intervals, it is interesting to consider to play the dynamic visual and the sonification in succession. This separation of sensory modes is similar to the unimodal separation of interaction and sonification sounds in the echo-like convolution in Section 5.2. Further, it would be interesting to see (although difficult to test) if alternating modalities analogue to the unimodal sonic separation in Section 6.3 can provoke the emergence of the *gestural-sonorous object* as proposed by Godøy (2006). The feedback from one subject, who reported to have imagined the visual details of cubes in the audiovisual display without them suggests that such a transfer is feasible.

For the design of descriptive auditory display systems, it is hence important to provide means to select unimodal stimuli if all sensory modalities are time-based. In a combined audiovisual display with the normative purpose to segment the data, sound provides effective guidance and can help to structure the stimulus in time as shown in Section 3.2.

During the qualitative evaluation of the audible sculptures in Section 5.2, we observed that the interaction with the small sculpture replica was different from our expectation. Rather than repeatedly confirming their heard impressions through touch, the subjects compared the sound with the mental model of the sculpture which they had built at the very beginning when holding the sculpture for the first time. All participants, were closely engaged in listening to the repeatedly triggered sound. This observation suggests that the complementary haptic modality offered by the persistent object of the sculpture replica enabled a *reduced listening* intention. The sculptural object seemed to have literally provided access to the sound object, or at least to aspects of it which were relevant for an echolocation-like listening situation.

Reflecting on complementing modalities in design cycles essentially reverses the idea of affordances when sonifications are used in exploratory tasks (compare with the functional listening mode of Tuuri and Eerola (2012) in Section 2.2.5). As proposed by Gaver (1991) sound can provide hidden affordances, offering the user options on how to use an object. The sound of a moving door handle might hint at the door's weight

6. FROM CASE STUDIES TO GUIDELINES

for instance. In an exploratory task, it is important to assess how the complementary modalities provide the user with affordances on how to make use of the sound, i.e. with which listening intention one should perceive it and if a reflective intention such as *reduced listening* can be invoked.

With respect to complementary, I suggest focusing on the following points in an auditory display system with a descriptive purpose:

- **Direction of affordances:** This point focuses on the general role of sound in the multimodal display. Does the sound provide affordances on how to use the other modalities or do the other modalities influence which listening intention we exploit? For exploratory data analysis we need to ask if the way how other modalities are used promote listening modes that support a descriptive purpose.
- **Complementarity of modalities:** This point focuses on the question of whether the used modalities really complement each other by providing additional information, if they lead to a fused percept, or if they only provoke causal listening by making the sound refer to events or selected aspects in other modalities. Do amodal (embodied) percepts exist such as the *gestural-sonorous object* that match the nature of the data to be sonified and can they be addressed through the sensory modalities involved?
- **Time-based versus persistent:** If the complementing modalities are time-based and evoke the causal listening intention it is worth to consider if it is possible to provide them as a persistent display. This can for example be a visual score instead of an animated visual display.
- **Separating modalities:** In the audiovisual case for instance, it is worth considering to separate the sonification from the visual presentation of the data by alternating the presentation of both. This might help to transfer the percept of one modality to the other without the immediate use of causal listening intentions, due to synchronicity. Separation can be also considered for interaction by introducing a delay between the interaction and the sonification, this refers to the point *playback*, from the consideration about repetition, and the distinction between *the player's and the instrument's timbre* from the considerations about interaction.

The first two points are derived from theoretical reflections on listening modes related literature. The last two points are based on the experience made in the practical applications of this thesis.

6.5 Embodied Modalities

Embodied modalities constitute a hybrid design aspect: they belong to the auditory display system by incorporating aspects beyond sound. In this sense they are invoked by the listener. From a sonification design perspective we need to consider how they can be evoked through sound.

Vocal sounds appeal naturally to the embodied dimension of perception as they always refer to our own sound production capacity (as discussed in more detail in Chapter 2.4). We are therefore also 'listening' with our apparatus for speech production and not only with our ears. In this particular case the sound object is not so much a matter of a listening intention and memory but is stabilized in parallel through the vocal motor code related to human production capacities. The design goal would hence be the potential utterability of the sonification.

For the *gestural-sonorous object*, (discussed in more detail in Section 3.1.2), the design goal includes the potential gestural expression of the sound. Both design goals, utterability and or performance as a gesture are however difficult to meet if the data material is complex, as discussed in Section 2.4.4. In these cases however, vowel sounds maintain their advantage to be salient and provide at the same time various distinguishable sounds from the same category. The vocal motor code evoked through vowel sounds might also help to support the formation of a percept similar to the *gestural-sonorous object* as articulated movements.

In the vowel-based auditory graph in Section 4.3, the connection between the explorational gesture on the slider and the pitch mapping created a certain contradiction, since the left and right movement on the slider, which was mimicking progression on the x-axis, created an up and down movement in the pitch field. This contradicts the embodied experience of the larynx that moves down for low sounds and up for high sounds, which corresponds to the up and down hand movements in the singing solfège (do-re-mi). Here the embodied aspect has also something to do with an imagined vertical position of the pitch and is not only related to the vocal timbre.

Vowel sounds are nonetheless interesting with respect to specific listening intentions, for instance listening for pitch or timbre. Because of our ability to sing vowels these two dimensions can be identifiable and to some degree separable listening intentions. Pitch can be associated with the production aspects involving the vocal fold, and timbre with those of the vocal tract including tongue and palate. This partial independence of these perceptual dimensions was used in *Singing Function* in Section 4.3 and helped to add audible contrast within a function family with small parameter variations. In the sonification of ancillary gestures in Section 3.3, pitch and timbre of vowel sounds

6. FROM CASE STUDIES TO GUIDELINES

was used in combination in order to create a sonification with perceptual contrasting sonic labels.

In the data sonogram in Section 2.4.4 vowel sounds worked well as sonic labels, similar as in Section 3.3. Inspired by this experience, I tried to apply vowel synthesis during the design process of the audible sculptures. The amount of data however did not lead to any utterable sounds of reasonable length. This was the reason to experiment with other sound synthesis options, and will be discussed further in Section 6.6. As much as vowel sounds are appealing, the low frequency range in which they exhibit a distinguishable timbre limits their applicability for sonifications where many sounds need to be rendered in a short amount of time. It seems worthwhile to look into other vocally produced sounds such as consonants. They offer less flexibility to encode data since not all of them are pitched sounds, but some inhabit higher frequency ranges.

The embodied aspects of listening modes are not only restricted to the sonic material of vowel sounds. They can also be found in the action-sound-object of the connotative aspects of listening intentions (see Section 2.2.5), as it was exploited in the chapter about auditory augmentation in Section 5. Here the sounds need to appeal to a plausible interaction with objects such as an echo that is bouncing off from a surface. This is why a great amount of time was spend in the design process to find ways to make use of the attack of the sound as a mapping target, which led to the involved many-to-one mapping topology.

For an auditory display system it is advantageous to evoke embodied listening modes, as they allow one to naturally engage with the sound of the sonification. The applicability of the spectral contours of vowel sounds as sonic material to represent complex data has however shown some limitations, which will be discussed in Section 6.6. Also the action-sound-object first and foremost makes the interaction meaningful but does not necessarily help to relate to the complex temporal evolution of the sound, which is what carries information about the shape of the sculptures in Section 5.2. This demonstrates that the meaning-making potential of embodied listening modes is a point of entry for engagement but cannot serve fully the descriptive purpose of exploratory data analysis for which reflective listening intentions need to be invoked in parallel.

With respect to embodied modalities in design guidelines I suggest focusing on the following points in an auditory display system with a descriptive purpose. The last two points coincide if *listening mode hearing (3)* is subsumed in reflective listening modes:

- **Utterability, expressibility:** Create sonifications on time scales that correspond to those of the embodied correlates of vocal utterances and gestures in order to put these correlates to use for the formation of the sound object or the *gestural-sonorous object*. This design goal corresponds to the goal of reproduction as an element of repetition. Based on the findings from Hutchins (compare Section 2.4), vowel-based sonifications would allow for sonic user modeling, by imitating the user’s voice by approximating the user’s register.
- **Entry point towards more reflective listening modes:** Embodied modalities need to be understood as points of entry to engage with the sound. Listening intentions of the action-sound-object need sounds that suggest an appropriate interaction as the cause of the sound. If the listener needs to engage with sonic features beyond those referring to the sound causing action, more reflective listening intentions need to be invoked. In the case of interactive sonification this can be for instance *musicianly listening*.
- **Meaning-making potential and bias:** Embodied listening modes offer a potential to perceive sonifications as meaningful prior to ‘decoding’ them but can also create a perceptual bias. In vowel-based sonifications for instance, it can be necessary to engage *listening mode hearing (3)*, in order to identify variations of specific sonic values such as pitch or spectral changes. Otherwise vowel-based sonifications remain “*a voice and nothing more*” (compare Dolar (2006, page 3)).

The first two points are based on the experience made in the the vowel and gesture related applications. For the last point compare Grond and Hermann (2012a).

6.6 Parameter-Mapping and Sound Design

The connection between parameter-mapping and listening modes cannot be a direct one, since parameters are a category of the signal and the listening modes depend on the intentions of the user. Nonetheless certain relations between listening intentions, data properties and how they can be mapped to synthesis parameters can be traced. As far as the three mapping targets *physical* (pure tones), *psychophysical* (pitch, as tonic mass) or *perceptually coherent complexes* (timbre, as spectral contour) – see PMSon in Section 2.3 – are concerned, their accessibility as direct mapping targets depends on the amount of data and their complexity which has to be represented through sound. Interestingly, all these mapping targets relate to perceivable values that belong to *listening mode hearing (3)*. As noticed in Section 2.2.6, the abstract subjective

6. FROM CASE STUDIES TO GUIDELINES

listening mode hearing (3) seems to be subsumed in the reflective modes proposed by Tuuri and Eerola (2012), which includes *reduced listening*. In multimodal displays, these mapping targets can be related to other listening intentions. In Figure 6.1, the arrow *fusing productive listening modes* means in audiovisual displays as in Section 3.3 for instance: *hearing (3)* labeled auditory streams as causes (*listening (1)*). In this case mapping polarities need to fit all involved sensory modalities, possibly taking advantage of embodied amodal structures.

During the mapping design process across all practical applications, I made the following observations: for data that can be represented by a single pitched auditory stream with a low fundamental frequency, spectral envelopes such as formants are well supported through the narrowly spaced partials and hence result in a recognizable timbre, which remains fairly stable with moderate variations of the fundamental frequency. Also the pitch remains reasonably stable with the variation of the spectral envelope. An example of this mapping possibility is given in *Singing Function* in Section 4.3.

The low fundamental frequency, however, makes the sonification "slow" since more display-time is needed for pitch to emerge as a salient audible feature. Also the narrowly spaced partials create a large spectral footprint, thereby potentially masking concurrent sounds if more than one auditory stream is needed. The higher the fundamental frequency gets the less the distantly spaced partials are able to support the spectral envelope and the mutual influence of timbre and pitch increases. As a consequence pitch and spectral contour become less suitable as independent mapping targets for continuously varying data features. For multi-stream sonifications like those for ancillary gestures as in Section 3.3, or the data sonogram in Section 2.4.4.2, spectral envelopes together with pitch can still serve as sonic labels.

For the amount of data that had to be mapped in the audible sculpture applications involving auditory augmentation in Chapter 5, synthesis schemes with short attack sounds and a small spectral footprint had to be developed, which was discussed in Section 5.3. These attacks were still controlled in their spectral characteristics but the timbre of the sound representing a single data point did not have a complex spectral contour anymore. Nonetheless, the overall sonification exhibited a complex timbral development – particularly those of the 3D shapes in the audible sculptures. This however was an emergent sonic property rooted in the data structure, but not a direct mapping target.

Based on this observation, I conclude as a general rule that with increasing data complexity direct mapping targets vanish, starting with timbre as spectral contour, followed by the pitch of complex tones, leaving at last frequencies. The latter, however, can still be perceptually corrected through amplitude compensation for equal loudness

and the Mel scale for perceived height in the spectrum, as it was applied in Section 5.2. From a listening mode point of view, this means that embodied listening intentions that relate to spectral sound properties such as in vowel sounds have a certain limit in terms of the complexity of the data that they can represent.

The timbral development of sonifications of many data points like those of the audible sculptures in Section 5.2 increasingly results in sounds to which it becomes difficult to relate through embodied listening intentions alone, beyond the connotative intention of the action-sound-object that was exploited through interaction sounds. This means that with increasing data complexity, sonifications demand to invoke more reflective listening modes.

This observation concerning data complexity and mapping targets also illustrates the relationship between ASA and listening modes. While the first provides bottom-up principles how *auditory gestalts* can emerge, the latter suggests how we can potentially make sense of these *auditory gestalts* in connection with the purpose for which they were made in auditory display, which is to represent abstract data.

For a hierarchically-organized auditory display, which incrementally maps more data details as additional sonic features as in *Singing Function* in Section 4.3, it is desirable that each hierarchy corresponds to an identifiable aspect within *listening mode hearing (3)* such as pitch or timbre for instance. Using embodied listening intentions evoked for instance through vowels can help to identify and to some degree separate these aspects through the possibility to reproduce them even after the sonification was heard.

Mapping audible information in space needs to take into consideration that locating sounds mostly addresses the reflexive listening modes by turning where the sound comes from. This might be one reason why the scanning mode from left to right in the audible sculptures in Section 5.2 was slightly preferred, because the spatial progression of the sound was predictable and hence the listeners could focus onto the sound itself. This suggests to conceive one-to-many mappings that couple reflexive and other listening modes so that the first gets exhausted after the connection has been established by the listener. A similar coupling between the location in the stereo panorama and optimization progress indicated through the filter bandwidth and delay was used in the monitoring application in Section 5.1.

The idea of the sound object and *reduced listening* is particularly useful when the sound designer experiments with different mapping topologies. In this thesis several mapping topologies have been applied, exploiting various audible dimensions. For one-to-one mappings, the listening intention of the designer comes from the expectation to identify signal parameters as distinct features in the perceived sound. In increasingly

6. FROM CASE STUDIES TO GUIDELINES

complex mappings it is often difficult to identify if the mapping topology noticeably contributes to the sonic result. In these situations in the iterative design process one often has the desire to get a “*fresh pair of ears*”, compare Chion (1983, page 47). Since this wish will stay unfulfilled, it seems worthwhile for the sonification designer to practice *reduced listening* and possibly adopt terms from the taxonomy of the sound object in order to get a critical perspective on what we perceive during the design process.

- **Hearing (3) and other listening modes:** Identify which sonic values from *listening mode hearing (3)* correspond to sound synthesis parameters. Evaluate whether it is necessary to be able to identify the individual perceptual dimensions and how embodied listening intentions – evoked through vowel sounds for instance – can help to separate them. Pay attention to the spatialization of sounds in the display and how reflexive listening modes - when locating sounds - interfere with *reduced listening*.
- **Availability of mapping targets and ranges:** Analyze which mapping targets are available with respect to the complexity and the amount of data that need to be sonified. With increasing data amount and decreasing display time, more attention needs to be paid to principles of ASA in order to create sound objects and collections thereof with satisfying perceptual contrast (compare what I called the sonification-Lombard effect in Section 5.3). With disappearing direct mapping targets, sonic features emerge indirectly and more reflective listening intentions need to be invoked to describe and identify audible phenomena in the auditory display.
- **Preparing data for shapes and events:** Analyze the potential of data preparation (data reduction or data derivatives) in PMSon with respect to listening modes by focusing on the creation of sound objects according to the DSM or by focusing on articulations according to embodied listening modes like the *gestural-sonorous object*.
- **Listening in the design cycle:** During the design of PMSon, reflective listening modes need to be cultivated in order to critically assess the effect of the selected mapping targets and the applied mapping topologies onto the sonic result. The two reflective listening modes have here a complementary role with respect to the specific bias of the sonification designer knowing how data have been mapped. Critical listening, helps to identify if all data have been audibly mapped. *Reduced listening* helps to judge the salience of each data feature within the complete sonification.

All points are derived from the theoretical considerations and the experience gained in the practical applications of this thesis.

6.7 Concluding Remarks

The listening-based design guidelines focus on auditory display systems with a descriptive purpose, which I presented as a result of the practical and theoretical parts of this thesis. These guidelines provide a structure for reasoning during the design process of sonifications which have multimodal and or interactive elements. Listening intentions are a user category and not directly controllable like the sound signal. However they allow us to formulate initial design decisions and further they provide a framework to make informed steps in the design process. While for auditory display systems with a normative purpose the completion of a task always allows measuring the success of the designed artifact, auditory display systems with a descriptive purpose are more difficult to evaluate as they tend to lack a measurable goal. With these guidelines, design decisions can be critically reflected upon and hypotheses can be built on the phenomenology of listening, thereby aptly informing the design process for each individually conceived display.

6. FROM CASE STUDIES TO GUIDELINES

7

Conclusion

In the introduction to this thesis, the need of design guidelines for multimodal and interactive sonifications has been identified, based on a review of existing contributions to auditory display related to design such as those by Gaver (1991), Barrass (1997), deCampo (2007) and Frauenberger (2009). The focus on multimodal and interactive aspects follows the need to theoretically structure this field which has been identified and described by Hunt and Hermann (2011, page 295).

In order to focus the design guidelines on the sonic aspects of auditory display systems, I reviewed the historic and recent literature on listening modes. In this synthesis of the contributions to the field, I reviewed the 4 natural listening modes in Section 2.2.1, and the term of *reduced listening* together with the *sound object* in Section 2.2.1.1 all proposed by Schaeffer (1966). This review was based on the comments on Schaeffer's work by Chion (1983), which is available to the English readership as of 2009. I compared the terminology of this foundational work in Section 2.2.2 with the terms *musical listening* and *everyday listening* which have been proposed by Gaver (1993a) and have become the common point of reference in auditory display for the last 20 years. I included in this review in Section 2.2.5 the recent contributions by Vickers (2012) and particularly the taxonomy proposed by Tuuri and Eerola (2012), which focuses on listening intentions from an embodied cognition perspective and diversifies the connotative and denotative aspects of listening modes.

For interactive sonifications, I discussed in Section 2.2.4 the phenomenology of listening and sound making highlighting the notions of *ergo audition* and *musicianly listening*. For multimodal aspects, I reviewed the phenomenological and psychological literature in Section 3.1.3, focusing on listening modes relevant for audio vision. I also discussed in Section 3.1.2 the conceptual extension of the the sound object towards the *gestural-sonorous object* for the field of movement sonification. In order to better

7. CONCLUSION

understand the role of listening modes in auditory displays, I proposed to conceptualize displays as serving *descriptive* or *normative* purposes in Section 2.2.7.2 and refined this proposition in Section 6.1. With respect to the questions raised in the introduction, how natural listening modes and reduced listening relate to normative and descriptive purposes, I identified the *reduced listening* mode as serving descriptive purposes such as in exploratory tasks. This has been the focus for most aspects of the practical applications of this thesis. During the review of listening modes, I illustrated their relevance by concrete examples that show how they apply to sonification and auditory display.

In all practical applications the sonification methods consisted of parameter-mapping sonifications. For this method we systematically conceptualized the design cycle in Section 2.3, in the form of a diagram (see Grond and Berger (2011)). This diagram provides a basis for the systematic reasoning about the mutual influence between data, their preparation, the mapping topology and auditory factors, helping to identify how to intervene in both the data and the signal domain with the goal of translating the information into audible phenomena. I also discussed the role of listening modes in the PMSon design cycle in Section 2.3.2.

As a practical contribution to parameter-mapping sonifications, we developed in Grond et al. (2011a) a library for the synthesis of vowel-like sounds in the sound synthesis environment. This library described in Section 2.4 allows for a convenient and yet flexible synthesis and control of the spectral contour of vowel sounds. It provides a set of synthesis building-blocks to access the spectral envelopes that constitute a vowel and thereby allow control over spectral contours in timbre space. The resulting sounds provide material that appeals to embodied listening modes. This implementation has been used in sonification prototypes in Section 2.4.4 and in the practical applications in Section 3.3 and in Section 4.3.

In this thesis, we developed multimodal and interactive sonification prototypes in three different fields: two sonifications of ancillary gestures of clarinetists, developed together with a visualization of the movements; two multi-parameter-mappings for auditory graphs as a display for mathematical functions; two applications related to the field of auditory augmentation, where interaction sounds form the basis of the sonic material of the sonification.

In Chapter 3, we developed two audiovisual displays involving parameter-mapping sonifications for multivariate motion tracking data of clarinetists (compare Section 3.2 published in Grond et al. (2009) and Section 3.3). These data represent the ancillary gestures of the clarinetists and the mapping strategy was inspired by the *gestural-sonorous object*, particularly for the display developed in Section 3.3. The audiovisual display developed in Section 3.2 was qualitatively evaluated by a free annotation task of

the movements in which the sonification has shown to support the consistent segmentation of the movement over various repetitions. Subjects also preferred a visualization in which the movement elements were highlighted and where it was identifiable what aspect of the movement caused the sound. One version of the sonification included as a data preparation step a principal component analysis of the posture data. The resulting audiovisual display was least preferred by the subjects since it led to unidentifiable sound objects as discussed in Section 3.2.7. The second sonification was based on the articulation of the movements through sounds with a vowel-like spectral contour and it was studied with eye-tracking in order to find out if different mappings can influence the visual perception of movements in this display. The result was that the influence of sound on the gaze is small compared to changes in the visual presentation of the movement. However, for some moments in the movement sequences attention directing influence through the mapping polarity of the sonification was found. This chapter concluded by discussing the potential of sonification in audiovisual displays for exploratory data analysis of movements in Section 3.4. Future work with respect to the sonification of ancillary gestures in an audiovisual display needs to find more data reduction possibilities in order to provide a better matching data substrate for the *gestural-sonorous object* which corresponds to the movement visualization.

In Chapter 4, we presented two auditory graphs, one in Section 4.2 (published in Grond et al. (2010)) and one in Section 4.3 (published in Grond and Hermann (2012b)). For the integration of several derivatives in one sound stream, we presented the concept of multi-parameter-mapping, developed in order to integrate derivatives of the function in one sound stream and to support *gestalt* formation of mathematically relevant features of the curve. One goal was to make the information in the sonification more independent of its temporal evolution moving along the x-axis. The first sonification, which integrated $f(x)$ and $f'(x)$, was qualitatively evaluated by our collaborator Trixi Droßard in a pedagogical context by blind pupils. Their interaction patterns indicated that implicit sonic information was more difficult to grasp and that the localization of extrema depended on the curvature and the resulting sonic contrast. The general feedback was positive and subjects expressed their interest to use this auditory graph in other subject areas. The second sonification, which included $f(x)$, $f'(x)$ and $f''(x)$, was based on vowel synthesis. In this sonification, the function value $f(x)$ was mapped to pitch, $f'(x)$ to vowel transition and $f''(x)$ to the brightness of the vowel. I tested the auditory graph for its perceptual contrast in a discrimination task for the case of static playback and dynamic exploration with a slider. The result was that the multi-parameter-mapping paradigm increases the contrast compared to the orthodox pitch mapping approach in both the static playback and the interactive exploration of the

7. CONCLUSION

graph. Interestingly the static playback provides a stronger contrast and allows to better judge whether two sonifications are the same and thereby represent the same or different graphs, as the multi-parameter-mapping could not make the information of the auditory graph completely independent of the exploration along the x-axis. This chapter concluded by discussing the potential of multi-parameter-mappings, within the state-of-the-art in auditory graphs in Section 4.4. The possibility to focus on the temporal evolution of perceptually separable audible features which can also be individually manipulated in the signal domain has been emphasized as a particular advantage of the vowel-based approach.

In Chapter 5, we presented sonifications that belong to the recent field of auditory augmentation. The sonifications methods in this chapter are influenced by concepts from model-based sonifications and contain a parameter-mapping aspect. The first application, published in Grond et al. (2012), was developed to augment the monitoring of algorithmic processes for evolutionary optimizations. For the monitoring application three different parameter settings were developed which allowed to give access to the audible information on various levels of detail. I evaluated this application qualitatively in a small user study. Subjects preferred the setting that allowed to navigate between different presentation modes of varying information content, which indicates that it is desirable to provide auditory interfaces featuring different perspectives in order to meet different listening modes. In the second application I developed data-sonogram-inspired representations of 3D shapes in order to make sculptures perceivable for blind individuals. The interaction paradigm in this application was inspired by echolocation; Sonifications were based on convolving finger clicks in real-time with the pre-rendered sonifications of the different listener positions towards the sculpture. This application was developed in a participatory design approach and evaluated qualitatively with 4 blind subjects. Due to the small number no explicit finding can be reported but some cases have been found which suggest that this sonification allows blind subjects to develop a mental model of the sculpture. This chapter concludes by discussing the potential of auditory augmentation for complex data substrates in Section 5.3. One finding from the parameter-mapping design cycles was that for complex data substrates sounds with a small spectral footprint and higher mapping ranges in the frequency spectrum need to be used, which means that certain mapping targets disappear with increasing data complexity.

Based on the theoretical considerations and the practical applications, I proposed in Chapter 6 design guidelines for sonifications with multimodal and interactive elements. These design guidelines are conceived between two poles of auditory information display, spanning from a descriptive to a normative purpose. The design guidelines focused on

the pole of the descriptive purpose, where most of the practical applications are located. In the discussion of the design guidelines I also addressed the questions raised about aesthetics for auditory display systems with a descriptive purpose. For this display purpose the questions of aesthetics can be framed as how the overall design of the auditory display system helps to avoid the sound to be trapped as a first degree signifier. Other questions raised in the introduction such as: what is the relationship between multimodality and interaction with listening modes and display purposes? What is the potential of embodied cognition based listening modes for exploratory data sonification? How do data complexity and PMSon relate to exploratory data analysis and listening modes? have been answered in the following points of the design guidelines:

In Section 6.2, I discussed *repetition* for its role to help the promotion of reduced listening and the potential emergence of sound objects. As a source for repetition, the data structure itself, interaction, and repeated playback optionally based on the recording of the interaction has been discussed as well as the option of imitation through the listener.

In Section 6.3, I analyzed *interaction* for its potential to support descriptive display purposes within the triangle of *sonic exploration*, *sonic exploitation* and *sonic expression*. I identified sonic exploration with Schaeffer's notion of *musicianly listening* and the distinguishability of the sonic effects of the interaction versus those from the data and discussed the influence of interaction onto the temporal order of the information content within a sonification.

In Section 6.4, I discussed *complementing modalities* with respect to their influence on the direction of affordances, i.e. if they promote reduced listening or rather make the sound point towards other modalities. This point is connected to the nature of complementarity between the sonification and other modalities and how they are related to amodal percepts. As one design option I suggest to separate modalities in time as a possibility to circumvent causal listening due to synchronicity.

In Section 6.5, I discussed *embodied modalities* with respect to the design goal of utterability for vowel-based sounds, or expressibility for sounds that attempt to evoke the emergence of a *gestural-sonorous object*. I highlighted the function of embodied modalities as entry point to other listening modes, meaning that they have a meaning making potential which at the same time can be a bias, preventing engagement with more reflective listening intentions.

As the last point of the design guidelines, I discussed in Section 6.6 *Parameter-Mapping and Sound Design* with respect to how the listening mode *hearing (3)* relates to mapping targets, how the availability of mapping targets depends on data complexity, and what implications they carry for listening modes. I emphasized that data

7. CONCLUSION

preparation before the mapping needs to supports the emergence of events and shapes in the sonic result. I also addressed the role of listening modes within the PMSon cycle.

Particularly point two and three of the design guidelines try to draw the attention of the designer towards the following challenge: More input modalities do not automatically lead to better access to complex information. The complex information-bearing structures of sounds that can be perceived through the effort of reduced listening will always remain to some degree a subjective experience. Whatever method we use to make these structures more objective, be it complementing modalities and/or interaction, we need to be careful that these add-ons do not make the structure of the sound object disappear by evoking reflexive or causal listening intentions or by creating sonic interaction effects that overshadow the consistent link with the data.

In the conclusion of the practical chapters, I discussed future developments for the individual application areas of the three practical applications. For the future development of auditory information displays, the design guidelines are meant to provide a tool for systematic reasoning during the design process. While listening modes are difficult to address explicitly as a design element, these guidelines can support the systematic development of methods for the evaluation of exploratory data sonifications such as the one proposed by Vogt (2011) by grounding initial evaluation questions along the proposed central issues elaborated in this thesis. As stated in the introduction, these design guidelines address what has been commonly perceived as the creative aspect of sonification or what has been compiled as expert knowledge, for instance by Frauenberger (2009). For auditory information displays with a descriptive purpose, these guidelines help to support implicit design knowledge with theoretical reflections based on the phenomenology of listening.

Glossary

- ASA** Auditory Scene Analysis researches the basis of auditory perception in terms of the segmentation, the integration and segregation of sounds, the term was proposed by Bregman (1994), page 37
- DSM** Design-Space-Map is a design framework for auditory display that was conceived by de-Campo (2007), which provides guidelines to select from MBS, PMSon and audification the method of choice for a given data set, page 3
- ES** Evolutionary optimization seeking is a stochastic optimization heuristic inspired by Darwinian principles, page 120
- GUI** A Graphical User Interface provides the user with graphical elements to interact with a program, such as buttons sliders or menus, page 48
- HCI** Human-Computer-Interface, page 1
- MBS** Model-Based Sonifications connect data with sound through a sonification model, which is often based on physics inspired principles. (also compare Hermann (2011), 2.2.1 and 2.2.7.2), page 4
- MIDI** Musical Instrument Digital Interface is a protocol for event messages carrying information about musically relevant parameters such as, pitch, velocity, volume or others, in sonification typically used in combination with sampled sounds, page 36
- OSC** Open Sound Control is a data protocol tailored to the requirements of real time sound synthesis. This protocol has been developed by Wright et al. (2003), for further information see CNMAT (2003), page 93
- PMSon** Parameter mapping sonification involves the association of data features with sound parameters, also compare Grond and Berger (2011) and 2.3, page 4
- SC3** SuperCollider 3 is an environment and programming language for realtime sound synthesis developed by McCartney (2002) and others, page 40
- SynthDef** A Synthesis Definition in SC3 bundles unit generators in a graph function, that determines how the unit generators are interconnected. The compiled SynthDefs are the basis for synthesis nodes in the DSP chain, which are controlled by OSC messages, page 45
- TADA** Task-Data is a design framework for auditory display that was conceived by Barrass (1997) in order to design displays that are *useful for the task and true to the data*, page 3

GLOSSARY

UGen Unit generator are generators or manipulators of audio or control signals. They constitute the basic building blocks of synthesis definitions, page 44

References

- Baier, G., Hermann, T., and Stephani, U. (2007). Multi-channel Sonification of Human EEG. In Martens, W., editor, *Proceedings of the 13th International Conference on Auditory Display*, pages 491–496, Montreal, Canada. International Community for Auditory Display (ICAD), ICAD. 119
- Ballas, J. (1994). *Auditory Display: Sonification, Audification, and Auditory Interfaces*, chapter Delivery of Information Through Sound. Addison-Wesley. 10, 22, 32
- Ballora, M., Pennycook, B., Ivanov, P. C., Glass, L., and Goldberger, A. (2004). Heart Rate Sonification: A New Approach to Medical Diagnosis. *Leonardo*, 37(1):41 – 46. 119
- Barrass, S. (1997). *Auditory Information Design*. PhD thesis, Australian National University. 3, 10, 37, 171, 177
- Barrass, S. and Vickers, P. (2011). Sonification Design and Aesthetics. In Hermann, T., Hunt, A., and Neuhoff, J. G., editors, *The Sonification Handbook*, chapter 7, pages 145–171. Logos Publishing House, Berlin, Germany. 10, 31
- Bartz-Beielstein, T., Lasarczyk, C., and Press, M. (2005). Sequential Parameter Optimization. In McKay, B. et al., editors, *Proceedings of the IEEE Congress on Evolutionary Computation CEC*, page 773780. IEEE Press. 121
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective Areas in Human Auditory Cortex. *Nature*, 403(20):309 – 312. 30, 40
- Ben-Tal, O., Berger, J., Cook, B., Daniels, M., and Scavone, G. (2002). SonART: The Sonification Application Research Toolbox. In Nakatsu, R. and Kawahara, H., editors, *Proceedings of the 8th International Conference on Auditory Display (ICAD2002)*, Kyoto, Japan. Advanced Telecommunications Research Institute (ATR), Kyoto, Japan. 40
- Beyer, H. G. and P., S. H. (2002). Evolution Strategies – a Comprehensive Introduction. *Natural Computing*, 1:352. 120
- Birkhoff, G. (1933). *Aesthetic Measure*. Harvard University Press. 90
- Blattner, M., Sumikawa, D. A., and Greenberg, R. M. (1989). Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction*, 4:11–44. 22
- Bly, S. (1994). *Auditory Display: Sonification, Audification, and Auditory Interfaces*, chapter Multivariate Data Mappings, pages 447 – 470. Addison-Wesley, Santa Fe Institute Studies in the Sciences of Complexity. 33
- Bonebright, T. L. (2005). A Suggested Agenda for Auditory Graph Research. In Brazil, E., editor, *Proceedings of the 11th International Conference on Auditory Display (ICAD2005)*, pages 398–402, Limerick, Ireland. Department of Computer Science and Information Systems, University of Limerick, Department of Computer Science and Information Systems, University of Limerick. 88, 92
- Bovermann, T., Hermann, T., and Ritter, H. (2008a). AudioDB. Get in Touch With Sound. In *Proceedings of the 14th International Conference on Auditory Display*, Paris, France. inproceedings. 37

REFERENCES

- Bovermann, T., Rohrhuber, J., and Ritter, H. (2008b). Durcheinander. Understanding Clustering via Interactive Sonification. In *Proceedings of the 14th International Conference on Auditory Display*, Paris, France. inproceedings. 37
- Bovermann, T., Tünnermann, R., and Hermann, T. (2010). Auditory Augmentation. *International Journal of Ambient Computing and Intelligence (IJACI)*, 2(2):27–41. 46, 117, 118, 122, 145
- Brazil, E. and Fernström, M. (2011). Auditory Icons. In Hermann, T., Hunt, A., and Neuhoff, J. G., editors, *The Sonification Handbook*, chapter 13, pages 325–338. Logos Publishing House, Berlin, Germany. 21
- Bregman, A. S. (1994). Auditory Scene Analysis: The Perceptual Organization of Sound. *The MIT Press*. 37, 120, 177
- Cadoz, C. and Wanderley, M. M. (2000). Gesture - Music. In Wanderley, M. M. and Battier, M., editors, *Trends in Gestural Control of Music*, pages 71 – 94. Ircam – Centre Pompidou. 53
- Carlile, S. (2011). Psychoacoustics. In Hermann, T., Hunt, A., and Neuhoff, J., editors, *The Sonification Handbook*, chapter 3, pages 41–61. Logos Publishing House, Berlin, Germany. 30
- Cassidy, R. J., Berger, J., Lee, K., Maggioni, M., and Coifman, R. R. (2004). Auditory Display of Hyperspectral Colon Tissue Images Using Vocal Synthesis Models. In Barrass, S. and Vickers, P., editors, *Proceedings of the 10th International Conference on Auditory Display (ICAD2004)*, Sydney, Australia. 40
- Chion, M. (1983). *Guide To Sound Objects: Pierre Schaeffer and Musical Research, (English translation John Dack and Christine North 2009)*. Éditions Buchet/Chastel. 11, 14, 15, 16, 17, 19, 23, 24, 25, 28, 54, 153, 156, 168, 171
- Chion, M. (1994). *Audio-Vision: Sound on Screen*. New York: Columbia University Press. 11, 18, 26, 54, 153, 160
- Chion, M. (1998). *Le Son*. Editions Nathan, Paris. 11, 25, 28
- Chion, M. (2008). Les douzes oreilles / Die zwölf Ohren. In Meyer, P. M., editor, *Acoustic Turn*, pages 563–600. Wilhelm Fink Verlag, München. 55
- Chion, M. (2009). *Film, a Sound Art; translated by Claudia Gorbman*. Columbia University Press. 18, 54, 55, 73
- CNMAT, editor (2003). *Proceedings of The Open Sound Control Conference 2004*, volume 1, Berkeley, CA, USA. 177
- Darwin, C. (1859). *On the Origin of Species*. John Murray. 120
- deCampo, A. (2007). Toward a Data Sonification Design Space Map. In Scavone, G. P., editor, *Proceedings of the 13th International Conference on Auditory Display*, pages 342–347, Montreal, Canada. Schulich School of Music, McGill University, Schulich School of Music, McGill University. 3, 29, 33, 35, 58, 151, 171, 177
- Dolar, M. (2006). *A Voice and Nothing More*. MIT Press. 165
- Droßard, T. (2010). Sonifikation als Methode im Mathematikunterricht mit blinden und sehbehinderten Schülerinnen und Schülern. Hausarbeit zur Erlangung des ersten Staatsexamens im Fach Blindenpädagogik, Fakultät für Erziehungswissenschaft, Psychologie und Bewegungswissenschaft der Universität Hamburg. 97, 103
- Droßard, T., Grond, F., and Hermann, T. (2012). Interaktive Sonifikation mathematischer Funktionen als Unterrichtsmethode für blinde und sehbehinderte Schülerinnen und Schüler. *blind sehbehindert*, 132(1):42–52. 87, 99, 102
- Duden (2001). *Duden. Herkunftswörterbuch. Etymologie der deutschen Sprache. 3., völlig neu bearbeitete und erweiterte Auflage*. Dudenverlag, Mannheim/ Leipzig/ Wien/ Zürich. 28

- Edwards, A. D. N. (2011). Auditory Display in Assistive Technology. In Hermann, T., Hunt, A., and Neuhoff, J., editors, *The Sonification Handbook*, chapter 17, pages 431–453. Logos Publishing House, Berlin, Germany. 87
- Effenberg, A. O. (2005). Movement Sonification: Effects on Perception and Action. *IEEE MultiMedia*, 12 (2):56–69. 52
- Ekman, I. and Rinott, M. (2010). Using Vocal Sketching for Designing Sonic Interactions. *Designing Interactive Systems archive, Proceedings of the 8th ACM Conference on Designing Interactive Systems Aarhus, Denmark*, pages 123 – 131. 155
- El-Shimy, D., Grond, F., Olmos, A., and Cooperstock, J. (2012). Eyes-Free Environmental Awareness for Navigation. *Journal on Multimodal User Interfaces*, 5(3-4):131–141. 131, 135
- Eriksson, M. and Bresin, R. (2010). Improving Running Mechanics by Use of Interactive Sonification. In *Proceedings of ISON 2010, 3rd Interactive Sonification Workshop, KTH, Stockholm, Sweden*. 52
- Fant, G. (1960). *Acoustic Theory of Speech Production*. Mouton: The Hague. 42
- Ferrari, A., Benedetti, M. G., Pavan, E., Frigo, C., Bettinelli, D., Rabuffetti, M., Crenna, P., and Leardini, A. (2008). Quantitative Comparison of Five Current Protocols in Gait Analysis. *Gait Posture*. 57
- Fishwick, P. A., editor (2006). *Aesthetic Computing*. Leonardo Book Series, The MIT Press. 10
- Flowers, J. H. (2005). Thirteen Years of Reflection on Auditory Graphing: Promises, Pitfalls, and Potential New Directions. In *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*, Ireland, Limerick. 33, 89, 90, 91
- Flückiger, B. (2001). *Sound Design die virtuelle Klangwelt des Films*. Marburg: Schüren. 55, 56, 69, 84
- Fogel, B. D. (1966). *Artificial Intelligence through Simulated Evolution*. Wiley, New York. 120
- Frauenberger, C. (2009). *Auditory Display Design – An Investigation of a Design Pattern Approach*. PhD thesis, Queen Mary, University of London, Mile End Road, EN1 4NS, London, UK. 3, 149, 171, 176
- Gaver, W. (1989). The SonicFinder: An Interface that Uses Auditory Icons. *Human-Computer Interaction*, 4:67–94. 21
- Gaver, W. (1991). Technology Affordances. In *Proceedings of CHI'91*, ACM, New York, pages 79–84, New Orleans, Louisiana, . 2, 161, 171
- Gaver, W. W. (1986). Auditory Icons: Using Sound in Computer Interfaces. *Human-Computer Interaction*, 2(2):167 — 177. 21
- Gaver, W. W. (1993a). How Do We Hear in the World? Explorations in Ecological Acoustics. *Ecological Psychology*, 5(4):285–313. 2, 6, 11, 15, 19, 171
- Gaver, W. W. (1993b). What in the world do we hear? An ecological approach to auditory source perception. *Ecological Psychology*, 5(1):1–29. 19, 20, 21
- Godbout, A. and Boyd, J. E. (2010). Corrective Sonic Feedback for Speed Skating: A Case Study. In *Proceedings of the 16th International Conference on Auditory Display (ICAD2010)*, pages 23–30, Washington, DC. 52
- Godøy, R. I. (2006). Gestural-Sonorous Objects: Embodied Extensions of Schaeffer’s Conceptual Apparatus. *Organised Sound, Cambridge University Press.*, 11(2):149–157. 27, 40, 53, 54, 59, 60, 70, 161
- Goïna, M. and Polotti, P. (2009). Elementary Gestalts for Gesture Sonification. *NIME08*, pages 150–153. 53
- Grimaldi, D. and Engel, M. (2007). Why Descriptive Science Still Matters. *BioScience*, 57(8):646–647.

REFERENCES

150

- Grond, F. and Berger, J. (2011). Parameter Mapping Sonification. In Hermann, T., Hunt, A., and Neuhoff, J., editors, *The Sonification Handbook*, chapter 15, pages 363–397. Logos Publishing House, Berlin, Germany. 6, 11, 34, 35, 38, 88, 89, 172, 177
- Grond, F., Bovermann, T., and Hermann, T. (2011a). A SuperCollider Class for Vowel Synthesis and its Use for Sonification. In Worall, D., editor, *Proceedings of the 17th International Conference on Auditory Display (ICAD-2011)*, Budapest, Hungary. OPAKFI. 6, 40, 43, 46, 47, 48, 49, 172
- Grond, F., Droßard, T., and Hermann, T. (2010). SonicFunction, Experiments with a functionbrowser for the blind. In *Proceedings of the 16th International Conference on Auditory Display*, pages 15–21, Washington D.C. ICAD. 7, 87, 89, 91, 92, 93, 97, 99, 100, 101, 102, 104, 113, 173
- Grond, F. and Hermann, T. (2012a). Aesthetic Strategies in Sonification. *Artificial Intelligence and Society AIFS*, 27(2):213–222. 22, 32, 153, 165
- Grond, F. and Hermann, T. (2012b). Singing Function, Exploring Auditory Graphs with a Vowel Based Sonification. *Journal on Multimodal User Interfaces*, 5(3):87–95. 7, 87, 88, 104, 105, 106, 107, 108, 109, 111, 113, 173
- Grond, F., Hermann, T., Verfaillie, V., and Wanderley, M. M. (2009). Methods for Effective Sonification of Clarinetists’ Ancillary Gestures. In Kopp, S. and Wachsmuth, I., editors, *Proc. 8th Int. Gesture Workshop*, Lecture Notes in Computer Science, pages 171–181, Berlin, Heidelberg. Springer Verlag. 6, 51, 53, 57, 58, 62, 63, 64, 65, 66, 67, 68, 172
- Grond, F., Kramer, O., and Hermann, T. (2011b). Interactive Sonification Monitoring in Evolutionary Optimization. In Worrall, D., editor, *Proceedings of the 17th International Conference on Auditory Display (ICAD-2011)*, Budapest, Hungary. OPAKFI. 119, 122
- Grond, F., Kramer, O., and Hermann, T. (2012). Balancing Saliency and Unobtrusiveness in Auditory Monitoring of Evolutionary Optimization. *Journal of the Audio Engineering Society*, 60(7/8):531–539. 7, 119, 121, 122, 123, 124, 125, 126, 128, 174
- Großhauser, T. and Hermann, T. (2010). Multimodal closed-loop Human-machine Interaction. In *Proceedings of ISON 2010, 3rd Interactive Sonification Workshop*, KTH, Stockholm, Sweden. 53
- Guastavino, C. and Cheminee, P. (2003). Une approche psycholinguistique de la perception des basses fréquences : conceptualisations en langue, représentations cognitives et validité écologique. *Psychologie Française*, 48(4):91–101. 30
- Guttman, S., Gilroy, L. A., and Blake, R. (2005). Hearing What the Eyes See. *Psychological Science*, 16(3):228 – 235. 56, 84
- Harrar, L. and Stockman, T. (2007). Designing Auditory Graph Overviews: An Examination of Discrete vs. Continuous Sound and the Influence of Presentation Speed. In Scavone, G. P., editor, *Proceedings of the 13th International Conference on Auditory Display (ICAD2007)*, pages 299–305, Montreal, Canada. Schulich School of Music, McGill University, Schulich School of Music, McGill University. 88, 104
- Hermann, T. (2002). *Sonification for Exploratory Data Analysis*. PhD thesis, Bielefeld University. 35
- Hermann, T. (2008). Taxonomy and Definitions for Sonification and Auditory Display. In *Proceedings of the 14th International Conference on Auditory Display*, Paris, France. International Conference on Auditory Display. 9, 11, 150, 155, 159
- Hermann, T. (2011). Model-Based Sonification. In Hermann, T., Hunt, A., and Neuhoff, J., editors, *The Sonification Handbook*, chapter 16, pages 399–427. Logos Publishing House, Berlin, Germany. 117, 135, 136, 177
- Hermann, T., Baier, G., Stephani, U., and Ritter, H. (2006a). Vocal Sonification of Pathologic EEG Features. In Stockman, T., editor, *Proceedings of the 12th International Conference on Auditory Display*, pages 158–163, London, UK. International Community for Auditory Display (ICAD), De-

- partment of Computer Science, Queen Mary, University of London UK. 40, 41, 43, 45
- Hermann, T., Baier, G., Stephani, U., and Ritter, H. (2008). Kernel Regression Mapping for Vocal EEG Sonification. In *Proceedings of the 14th International Conference on Auditory Display*, Paris, France. International Conference on Auditory Display. inproceedings. 41
- Hermann, T., Bunte, K., and Ritter, H. (2007). Relevance-based Interactive Optimization of Sonification. In Martens, B., editor, *Proceedings of the 13th International Conference on Auditory Display*, pages 461–467, Montreal, Canada. International Community for Auditory Display (ICAD), ICAD. 37
- Hermann, T., Höner, O., and Ritter, H. (2006b). AcouMotion - An Interactive Sonification System for Acoustic Motion Control. In Gibet, S., Courty, N., and Kamp, J., editors, *Gesture in Human-Computer Interaction and Simulation: 6th International Gesture Workshop, GW 2005, Berder Island, France, May 18-20, 2005, Revised Selected Papers*, volume 3881/2006 of *Lecture Notes in Computer Science*, pages 312–323, Berlin, Heidelberg. Springer. 52
- Hermann, T. and Hunt, A. (2005). An Introduction to Interactive Sonification (Guest Editors’ Introduction). *IEEE MultiMedia*, 12(2):20–24. 91
- Hermann, T. and Ritter, H. (1999). Listen to your Data: Model-Based Sonification for Data Analysis. In Lasker, G. E., editor, *Advances in Intelligent Computing and Multimedia Systems*, pages 189–194, Baden-Baden, Germany. Int. Inst. for Advanced Studies in System research and cybernetics. 41
- Hermann, T. and Ritter, H. (2004). Sound and Meaning in Auditory Data Display. *Proceedings of the IEEE (Special Issue on Engineering and Music – Supervisory Control and Auditory Communication)*, 92(4):730–741. 25
- Hermann, T. and Ritter, H. (2005). Model-based Sonification Revisited—Authors’ Comments on Hermann and Ritter, ICAD 2002. *ACM Trans. Applied Perception*, 2(4):559–563. 4, 14
- Hermann, T., Ungerechts, B., Toussaint, H., and Grote, M. (2012). Sonification of Pressure Changes in Swimming for Analysis and Optimization. In *Proceedings of the 18th International Conference on Auditory Display*, Atlanta, GA, USA. 52
- Hetzler, S. and Tardiff, R. (2006). Two Tools for Integrating Sonification into Calculus Instruction. In *Proceedings of the twelfth International Conference on Auditory Display (ICAD2006)*, pages 281–284. 88
- Hodgkinson, T. (1986). An Interview with Pierre Schaeffer - Pioneer of Musique Concrète. In *ReR Quarterly magazine*. 13
- Höner, O., Hunt, A., Pauletto, S., Röber, N., Hermann, T., and Effenberg, A. O. (2011). Aiding Movement with Sonification in “Exercise, Play and Sport”. In Hermann, T., Hunt, A., and Neuhoff, J., editors, *The Sonification Handbook*, chapter 21, pages 525–553. Logos Publishing House, Berlin, Germany. Höner, O. (chapter ed.). 52
- Hug, D. (2009). Using a Systematic Design Process to Investigate Narrative Sound Design Strategies for Interactive Commodities. In *Proceedings of the 15th International Conference on Auditory Display ICAD2009*. 3
- Hummel, J., Hermann, T., Frauenberger, C., and Stockman, T. (2010). Interactive Sonification of German Wheel Sports Movement. In *Proceedings of ISON 2010, 3rd Interactive Sonification Workshop*, KTH, Stockholm, Sweden. 52
- Hunt, A. and Hermann, T. (2011). Interactive Sonification. In Hermann, T., Hunt, A., and Neuhoff, J., editors, *The Sonification Handbook*, chapter 11, pages 273–298. Logos Publishing House, Berlin, Germany. 2, 3, 12, 25, 149, 156, 171
- Hunt, A., Paradis, M., and Wanderley, M. M. (2002). The Importance of Parameter Mapping in Electronic Instrument Design. *Invited paper for the Journal of New Music Research, SWETS, special issue on New Interfaces for Musical Performance and Interaction*, 32(4):429–440. 36

REFERENCES

- Hunt, A. and Wanderley, M. M. (2002). Mapping Performer Parameters to Synthesis Engines. *Invited article for Organised Sound, special issue on Mapping*, 7(2):97–108. 36
- Hutchins, S. and Peretz, I. (2011). A Frog in Your Throat or in Your Ear? Searching for the Causes of Poor Singing. *Journal of Experimental Psychology: General*, Advance online publication. 40, 114
- Hutchins, S., Zarate, J. M., Zatorre, R. J., and Peretz, I. (2010). An Acoustical Study of Vocal Pitch Matching in Congenital Amusia. *J. Acoust. Soc. Am.*, 127(1):504 – 512. 40
- Janat, P. and Childs, E. (2004). Marketbuzz: Sonification of Real-time Financial Data. In Barrass, S., editor, *Proceedings of 10th Meeting of the International Conference on Auditory Display*, Sydney, Australia. International Community for Auditory Display (ICAD), International Community for Auditory Display. 119
- Jekosch, U. (2005). *Communication Acoustics*, chapter Semiotics in the context of Product-Sound Design. Springer Verlag Berlin Heidelberg. 3
- Kane, B. (2007). L’Objet Sonore Maintenant: Pierre Schaeffer, Sound Objects and the Phenomenological Reduction. *Organised Sound*, 12(1):15–24. 18
- Kepler, J. (1967). *Weltharmonik (Nachdruck)*. Max Caspar, Munich. 31
- Kim, J. and Zatorre, R. J. (2010). Can You Hear Shapes You Touch? *Experimental Brain Research*, 202(747–754). 131, 135, 143
- Kim, S.-J. (2010). A Critique on Pierre Schaeffer’s Phenomenological Approaches: Based on the Acousmatic and Reduced Listening. In *Pierre Schaeffer Conference: mediART*, Rijeka, Croatia. 18
- Kish, D. (2003). Sonic Echolocation: A Modern Review and Synthesis of the Literature. <http://www.worldaccessfortheblind.org/sites/default/files/echolocationreview.htm>. 131
- Klatt, D. (1980). Software for a Cascade/Parallel Formant Synthesizer. *Journal of the Acoustical Society of America*, 67(3):971–995. 44
- Kleiman-Weiner, M. and Berger, J. (2006). The Sound of One Arm Swinging: A Model for Multidimensional Auditory Display of Physical Motion. In Stockman, T., Valgerur Nickerson, L., Frauenberger, C., Edwards, A. D. N., and Brock, D., editors, *Proceedings of the 12th International Conference on Auditory Display (ICAD2006)*, pages 278–280, London, UK. Department of Computer Science, Queen Mary, University of London, UK. 40
- Kramer, G. (1993). *Auditory Display: Sonification, Audification, and Auditory Interfaces*. Perseus Publishing. 36
- Kramer, G. (1994a). An Introduction to Auditory Display. In Kramer, G., editor, *Auditory Display*, pages 1– 79. Addison-Wesley. 1, 9
- Kramer, G. (1994b). *Some Organizing Principles for Representing Data with Sound*. Addison-Wesley. 22
- Lewis, J. W., Talkington, W. J., Tallaksen, K. C., and Frum, C. A. (2012). Auditory Object Saliency: Human Cortical Processing of Non-biological Action Sounds and their Acoustic Signal Attributes. *Frontiers in Systems Neuroscience*, 6(1):27. 30
- Lombard, E. (1911). Le signe de l’elevation de la voix (translated by paul h. mason 2006, reviewed by m. jacques lenoir). *Annales des Maladies de l’Oreille, du Larynx, du Nez et du Pharynx*, 37:101–119. 148
- Mansur, D. L., Blattner, M. M., and Joy, K. I. (1985). Sound graphs: A numerical data analysis method for the blind. *Journal of Medical Systems*, 9(3):163–174. 88, 90, 92, 104
- McCartney, J. (2002). Rethinking the Computer Music Language: SuperCollider. *Computer Music Journal*, 26(4):61–68. 177
- McGookin, D. and Brewster, S. (2011). Earcons. In Hermann, T., Hunt, A., and Neuhoff, J., editors,

- The Sonification Handbook*, chapter 14, pages 339–361. Logos Publishing House, Berlin, Germany. 22
- McGurk, H. and MacDonald, J. (1976). Hearing Lips and Seeing Voices. *Nature*, 264:746 – 748. 55
- Meijer, P. B. L. (1992). An Experimental System for Auditory Image Representations. *IEEE Transactions on Biomedical Engineering*, 39(2):112–121. 135, 136, 140
- Meyer-Nieberg, S. and Beyer, H. G. (2007). Self Adaptation in Evolutionary Algorithms. In Lobo, F. G., Lima, C. F., and Michalewicz, Z., editors, *Parameter Setting in Evolutionary Algorithms*. Springer, Berlin. 121
- Moon, C., Lagercrantz, H., and Kuhl, P. K. (2013). Language Experienced in Utero Affects Vowel Perception After Birth: a Two-Country Study. *Acta Paediatrica*, 102(2):156–160. 40
- Mukařovský, J. (1989). *Kunst, Poetik, Semiotik, übersetzt von Erika Annuss u. Walter Annuss*. Suhrkamp Verlag, Frankfurt am Main. 152
- Nees, M. A. and Walker, B. N. (2007). Listener, Task, and Auditory Graph: Toward a Conceptual Model of Auditory Graph Comprehension. In Scavone, G. P., editor, *Proceedings of the 13th International Conference on Auditory Display (ICAD2007)*, pages 266–273, Montreal, Canada. Schulich School of Music, McGill University, Schulich School of Music, McGill University. 89, 90, 92
- Nielsen, J. (2000). Why You Only Need to Test with 5 Users. 74
- Nusseck, M. and Wanderley, M. M. (2009). Music and Motion—How Music Related Ancillary Body Movements Contribute to the Experience of Music. *Music Perception*, 26(4). 53
- O’Shaughnessy, D. (1987). *Speech Communication: Human and Machine*. Addison-Wesley. 136
- Oswald, D. (2012). Semiotik Auditiver Interfaces: Zur Geschichte von Gestaltung und Rezeption auditiver Zeichen in Computer Interfaces. In Volmar, A. and Schoon, A., editors, *Sound Studies: Kulturgeschichte der Sonifikation*. Transcript Verlag. 21
- Parise, C., Spence, C., and Ernst, M. O. (2012). When Correlation Implies Causation in Multisensory Integration. *Current Biology*, 22(1):46–49. 55
- Peters, N., Lossius, T., and Schacher, J. C. (2012). SpatDIF: Principles, Specification, and Examples. In *9th SMC*, Copenhagen, DK. 137
- Pulki, V. (1997). Virtual Sound Source Positioning Using Vector Based Amplitude Panning. *Audio Engineering Society*. 137
- Roessler, O. E. (1976). An Equation for Continuous Chaos. *Physics Letters A*, 57(5):397–398. 47
- Rohrhuber, J. (2010). \hat{S} – Introducing Sonification Variables. In *Proceedings of the Supercollider Symposium*. 35
- Rubin, B. U. (1998). Audible Information Design in the New York City Subway System: A Case Study. In *Proceedings of the International Conference on Auditory Display*. ICAD, British Computer Society. 119
- Savard, A. (2008). When Gestures are Perceived through Sounds: A Framework for Sonification of Musicians’ Ancillary Gestures. Master’s thesis, IDMIL CIRRMTC McGill University. 53, 55, 61
- Scaletti, C. (1994). *Auditory Display: Sonification, Audification, and Auditory Interfaces*, chapter Sound Synthesis Algorithms for Auditory Data Representation, pages 223 –251. Addison-Wesley, Santa Fe Institute Studies in the Sciences of Complexity. 33
- Schaeffer, P. (1966). *Traité des Objets Musicaux*. Le Seuil, Paris. 2, 6, 11, 13, 17, 18, 171
- Schaffert, N., Gehret, R., and Mattes, K. (2012). Modeling the Rowing Stroke Cycle Acoustically. *J. Audio Eng. Soc.*, 60(7/8):551–560. 52
- Schmidt, F. P. (2007). Design and Implementation of a Realtime 3D Graphics Server. Master’s thesis,

REFERENCES

- Bielefeld University, Germany. 62
- Schoon, A. (2012). Unmerkliche Eindringlinge, Versuch über akustische Kontrolle. In Volmar, A. and Schoon, A., editors, *Sound Studies: Kulturgeschichte der Sonifikation*. Transcript Verlag. 31
- Schwefel, H. (1995). *Evolution and Optimum Seeking*. Wiley Interscience Sixth Generation Computer Technology, New York. 120, 123
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27:379–423, 623–656. 10
- Shelton, R., Smith, S., Hodgson, T., and Dexter, D. (2006). Mathtrax. <http://prime.jsc.nasa.gov/MathTrax/index.html>. 90
- Smalley, D. (1997). Spectromorphology: Explaining Sound-shapes. *Organised Sound, Cambridge University Press.*, 2(2):107–126. 54
- Spence, C. and Soto-Faraco, S. (2010). Auditory Perception: Interactions with Vision. In Plack, C., editor, *Oxford Handbook of Auditory Science: Hearing*, pages 271 – 296. Oxford University Press, 1 edition. 55, 56
- Steiglitz, K. (1994). A Note on Constant-Gain Digital Resonators. *Computer Music Journal.*, 18(4):pp. 8–10. 59, 94, 124
- Stockman, T., Nickerson, L. V., and Hind, G. (2005). Auditory Graphs: A Summary of Current Experience and Towards a Research Agenda. In Brazil, E., editor, *Proceedings of the 11th International Conference on Auditory Display (ICAD2005)*, pages 420–422, Limerick, Ireland. Department of Computer Science and Information Systems, University of Limerick. 88, 105, 112, 113
- Stowell, D. and Plumbley, M. D. (2007). Adaptive whitening for improved real-time audio onset detection. In *Proceedings of the International Computer Music Conference (ICMC'07)*, Copenhagen, Denmark. 126
- Striem-Amit, E., Cohen, L., Dehaene, S., and Amedi, A. (2012). Reading with Sounds: Sensory Substitution Selectively Activates the Visual Word Form Area in the Blind. *Neuron*, 76(8):640–652. 135
- Sumbly, W. H. and Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustical Society of America*, 26(2):212–215. 55
- Truax, B. (2001). *Acoustic Communication*. Ablex Publishing. 25
- Tuuri, K. (2010). Gestural Attributions as Semantics in User Interface Sound Design. In Kopp, S. and Wachsmuth, I., editors, *Lecture Notes in Computer Science, Gesture Workshop 2009*, pages 257–268. Springer Verlag Berlin Heidelberg. 54
- Tuuri, K. (2011). *Hearing Gestures : Vocalisations as Embodied Projections of Intentionality in Designing non-speech Sounds for Communicative Functions*. PhD thesis, Jyväskylä Studies in Humanities. 25, 26, 28, 157
- Tuuri, K. and Eerola, T. (2012). Formulating a Revised Taxonomy for Modes of Listening. *Journal of New Music Research*, 41(2):137–152. 2, 6, 11, 27, 28, 29, 30, 151, 161, 166, 171
- Tuuri, K., Mustonen, M., and Pirhonen, A. (2007). Same Sound – Different Meanings: A Novel Scheme for Modes of listening. In *In Proceedings of Audio Mostly*, pages 13–18, Ilmenau, Germany. Fraunhofer Institute for Digital Media Technology IDMT. 26
- Tuuri, K., Pirhonen, A., and Hoggan, E. (2009). Some Severe Deficiencies of the Input-output HCI-paradigm and Their Influence on Practical Design. In Norros, L., Koskinen, H., Salo, L., and Savioja, P., editors, *Designing beyond the Product - Understanding Activity and User Experience in Ubiquitous Environments*, pages 363–369. Espoo, Finland: VTT Technical Research Centre of Finland. 26, 40
- Vanderveer, N. J. (1979). *Ecological Acoustics: Human Perception of Environmental Sounds*. Dissert-

REFERENCES

- tation Abstracts International*. 40/09B, 4543. PhD thesis, Cornell University, Ithaca, NY, USA. 20
- Verfaillie, V., Quek, O., and M., W. M. (2006). Sonification of Musicians' Ancillary Gestures. *Proceedings of the 12th International Conference on Auditory Display*. 53
- Vickers, P. (2005). Ars Informatica–Ars Electronica: Improving Sonification Aesthetics. In *HCI2005: Workshop on Understanding & Designing for Aesthetic Experience*. 10, 22
- Vickers, P. (2011). Chapter 18: Sonification for Process Monitoring. In Hermann, T., Hunt, A., and Neuhoff, J., editors, *The Sonification Handbook*. Logos Publishing House, Berlin, Germany. 11, 15, 24, 119, 122
- Vickers, P. (2012). Ways of Listening and Modes of Being: Electroacoustic Auditory Display. *Journal of Sonic Studies*, 2(1). 6, 11, 26, 28, 171
- Vickers, P. and Hogg, B. (2006). Sonification Abstraite/Sonification Concrète: an Aesthetic Perspective Space for Classifying Auditory Displays in the Ars Musica Domain. In *Proceedings of the 12th International Conference on Auditory Display, London, UK June 20 - 23, 2006*, pages 210 – 216. 10
- Vogt, K. (2011). A Quantitative Evaluation Approach to Sonification. In Worrall, D., editor, *Proceedings of the 17th International Conference on Auditory Display (ICAD-2011)*, Budapest, Hungary. OPAKFI. 176
- Vogt, K. and Höldrich, R. (2010). A Metaphoric Sonification Method - Towards the Acoustic Standard Model of Particle Physics. In *The 16th International Conference on Auditory Display (ICAD-2010)*. 33
- Walker, B. (2007). Consistency of Magnitude Estimations with Conceptual Data Dimensions Used for Sonification. *Applied Cognitive Psychology*, 21:579–599. 33
- Walker, B. N. and Kramer, G. (2005). Mappings and Metaphors in Auditory Displays: An Experimental Assessment. *ACM Trans. Appl. Percept.*, 2(4):407–412. 33
- Wanderley, M. M. (2002). Quantitative Analysis of Non-obvious Performer Gestures. *Gesture and Sign Language in Human-Computer Interaction*, pages 241–253. 53
- Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., and Hatch, W. (2005). The Musical Significance of Clarinetists' Ancillary Gestures: an Exploration of the Field. *Journal of New Music Research*, 34(1):97–113. 53
- Winters, R. M. and Wanderley, M. M. (2012). New Directions for Sonification of Expressive Movement in Music. In *Proceedings of the 18th International Conference on Auditory Display*, Atlanta, GA, USA. 53
- Worrall, D. (2009). *The Oxford Handbook of Computer Music*, chapter 16. An Introduction to Data Sonification. Oxford University Press. 2, 11, 33
- Worrall, D. (2010). Parameter Mapping Sonic Articulation and the Perceiving Body. In *The 16th International Conference on Auditory Display (ICAD-2010)*, Washington D.C. 33
- Worrall, D. (2011). A Method for Developing an Improved Mapping Model for Data Sonification. In *The 17th International Conference on Auditory Display (ICAD-2011)*, Budapest, Hungary. 33
- Wright, M., Freed, A., and Momeni, A. (2003). Open sound control: State of the art 2003. In *International Conference on New Interfaces for Musical Expression*, pages 153–159, Montreal. 177

REFERENCES

8

Appendix

8.1 Movement Annotation Plots

Figure 8.1, 8.2 and 8.3 compile the kernel estimated click density for all three movements, sections of these plots are shown and discussed in 3.2.6. For each of the three figures all plots are organized in the following order:

The first plot compares all stimuli with the marker velocity based sonification (A1), containing the condition with sonification only (V0) and in combination with the stick-figure visualization (V1) and the one with enhancing glyphs in the form of cubes (V2).

The second plot compares the PCA based sonifications (A2) without (V0) and with (V1) the stick-figure visualization.

The third plot compares the visualizations as stick-figure (V1) and the one with cubes (V2), both without sonifications (A0).

The fourth plot compares all stimuli with the stick-figure visualization (V1), i.e. the visualization only (A0), and together with the the marker velocity based (A1) or the PCA based (A2) sonification.

The fifth plot compares the PCA based sonification (A2) with the marker velocity based (A1) sonification, both without visualization (V0).

The sixth plot compares the visualization of the stick-figure with cubes (V2) without sonification (A0) and the marker velocity based sonification (A1).

In all plots the p value of the Kolmogorov-Smirnov test is shown on the right for the comparison of the annotation events of each stimuli pairing. The grey horizontal bar indicates the average click density $\int = 1$.

8. APPENDIX

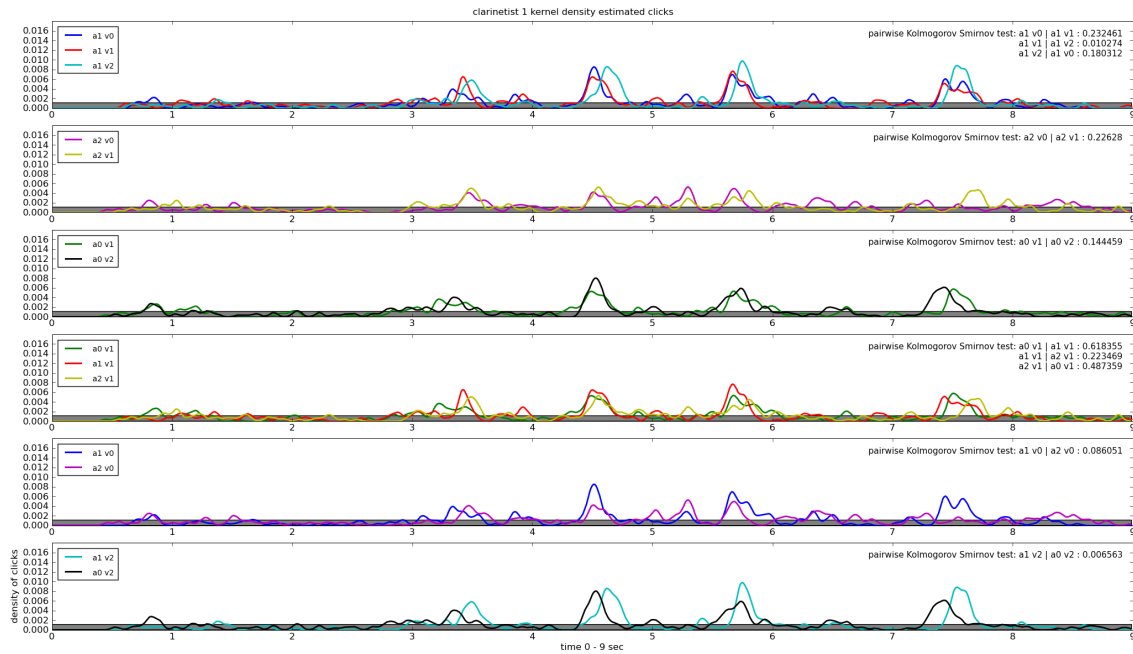


Figure 8.1: Kernel estimated click density, movement sequence 1

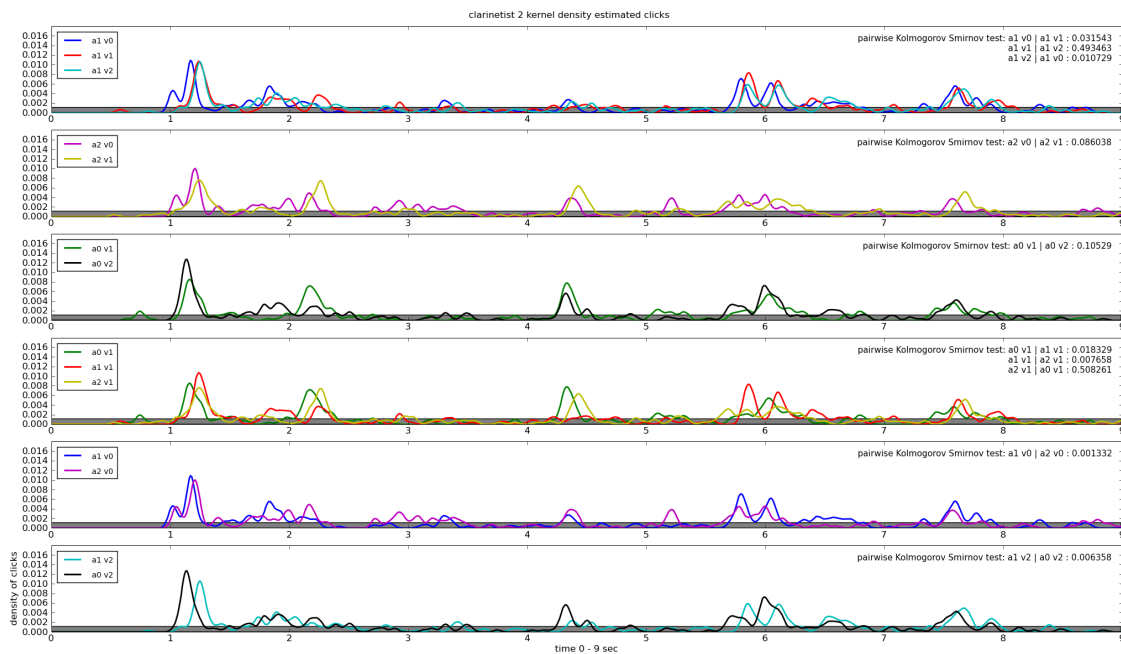


Figure 8.2: Kernel estimated click density, movement sequence 2

8.1 Movement Annotation Plots

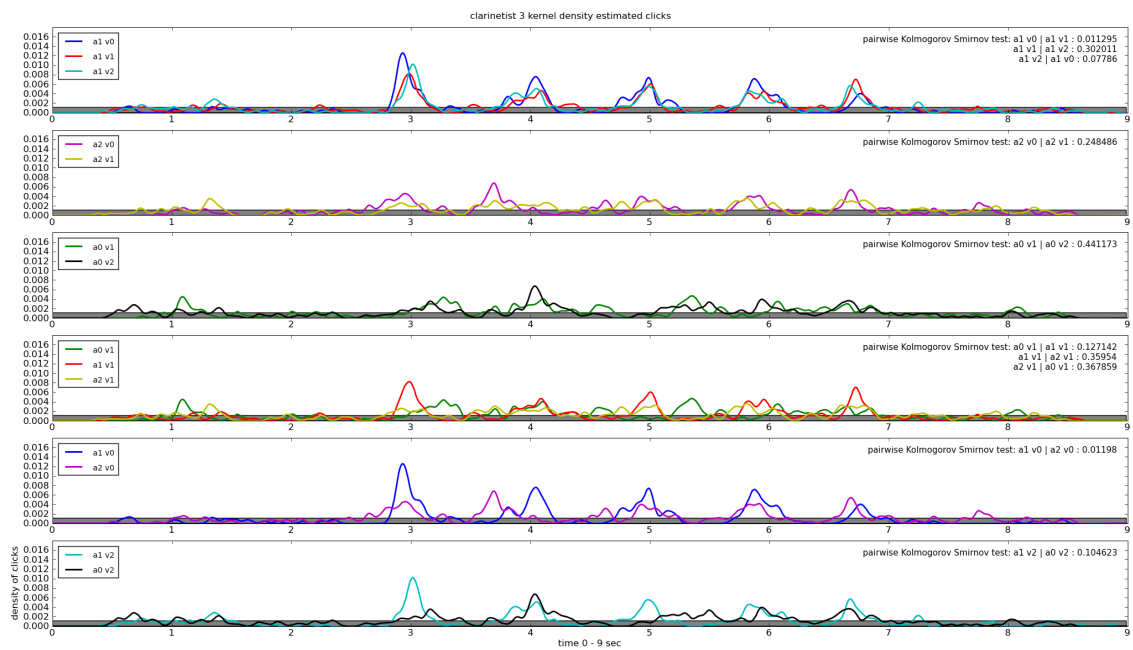


Figure 8.3: Kernel estimated click density, movement sequence 3

8.2 Gaze-density, Fixation Plots and Interactive Data Inspection

Some of the plots from Figure 8.4 are shown in Section 3.3 and discussed in Section 3.3.3. For a better comparison, the plots showing the gaze density over all four movement sequences are compiled here. Figure 8.4 shows the gaze density over the whole stimuli time of 20 seconds. Movement sequences 1 and 3 on the left, and 2 and 4 on the right, were extracted from the performance of clarinetist *A* and *B* respectively. Clarinetist *B*'s ancillary gestures included expressive back and forth movements, which can be identified as the large horizontal distribution of the upper body.

Figure 8.5 depicts the horizontal eye fixation positions on the left and the vertical eye fixation positions on the right. In each plot, the left column shows the fixations for the stick-figure (L) and the right column shows the fixations for the enhanced visualization with cubes (C). The pronounced back and forth movement of the clarinetist in sequence 2 and 4 can be identified in the left column. The vertical displacement of the eye-movement on the right shows less pronounced pattern and mostly indicates that subjects were mostly looking at the upper body.

Movement sequence 1 Figure 8.4 (top left): The top row shows the stick-figure (L) and differs from the visualization with cubes (C) in which the upper body is differentiated into two zones (head and hips) annotated as **A**. The annotation **B** shows that for this movement subjects remained for a considerable amount of time in the drift correction position for two stimuli. The most noticeable sound related difference is annotation **C** in the **Ls2** sonification, where the gaze density across subjects is focused in the feet. Further, annotation **D** shows that the attention on the knee seems more focused for **Cs2**. In the data inspection movie the moment around 7.5 seconds shows a shift of the attention towards the lower parts of the body after a movement in the knee. A similar shift is observed after 11.5 seconds and after 14 seconds, in all three cases after a knee movement.

Movement sequence 2 Figure 8.4 (top right): In this sequence a similar influence can be found for the visualization with cubes (C) annotated as the differentiation in the upper body as **A** and the attention of the gaze directed towards the knees **D** and feet **C** in **s2**. In the data inspection movie the moment between 8 and 9.5 seconds after a moving foot shows more fixations in the knee and foot area in the **Cs2** stimulus.

Movement sequence 3 Figure 8.4 (bottom left): In this sequence the differentiation in the upper body is annotated as **A**, a slightly higher density in the feet can be seen for *L* compared with the *Ls1* and **Ls2**, annotated as **C**, which is likely to be due to some fixations after a foot movement around 10 seconds.

Movement sequence 4 Figure 8.4 (bottom right): In this movement sequence, sonification **s2** shifts the focus slightly down towards the hips in the upper body annotated as **A** in the stimulus **Ls2**. For the visualization with cubes, concentration around the hips for **C** was altered through the sonifications *s1* by shifting the focus to the arms in *Cs1* and further towards the head in **Cs2** both annotated as **A**. No moments indicative for this shift could be identified in the data inspection movie. This was mostly because fixations tended to group around targets but were rarely spot on. This made it difficult to identify specific body movements, also because eye fixations always occurred always after an delay.

8. APPENDIX

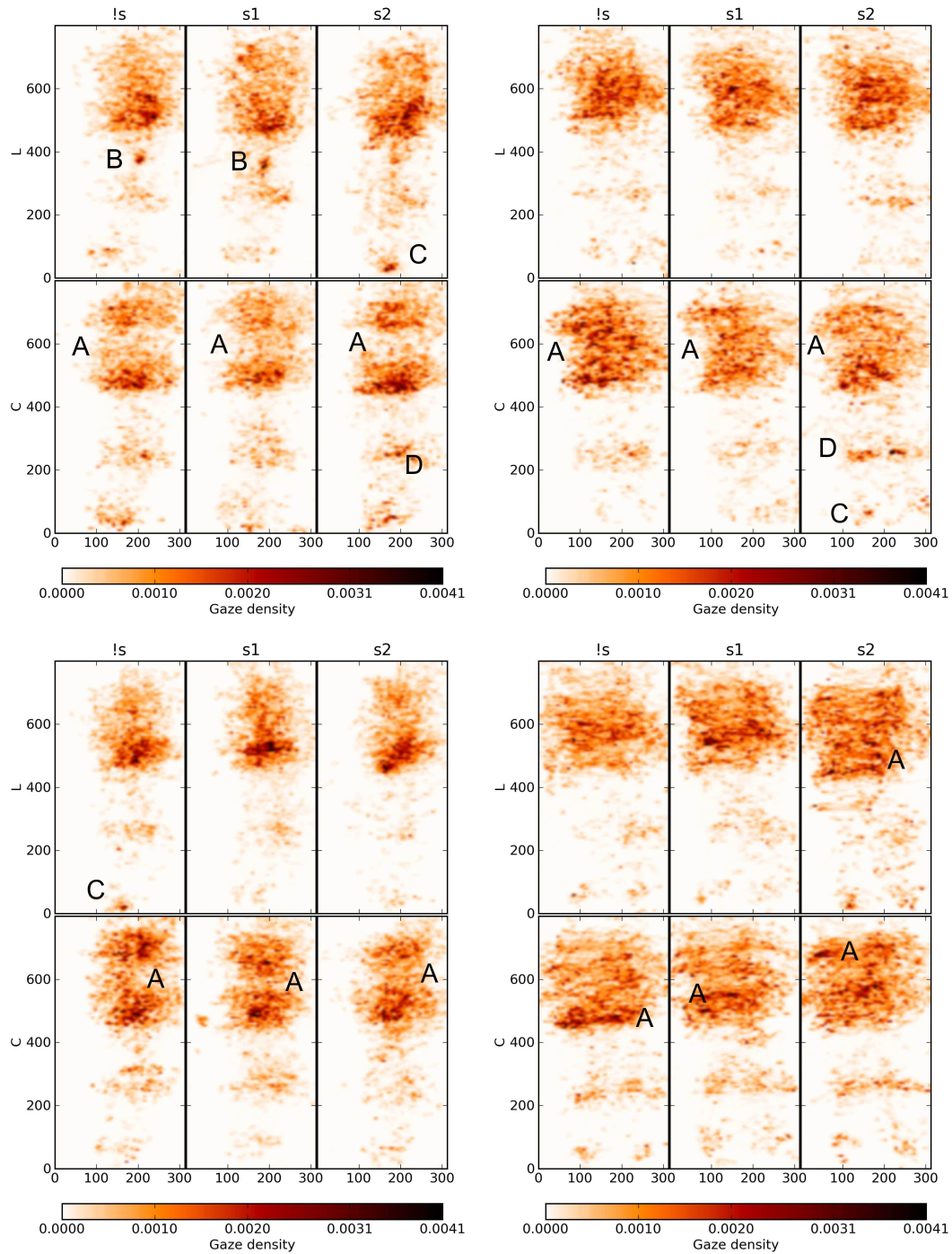


Figure 8.4: Gaze density plot for all 4 movement sequences, on top 1 left, 2 right, at the bottom 3 left, 4 right. For each movement the top row shows the gaze density for the stick-figure (L) and the bottom row shows the enhanced visualization. The axes correspond to extension of the movement, 768 pixels high and 300 pixels wide.

8.2 Gaze-density, Fixation Plots

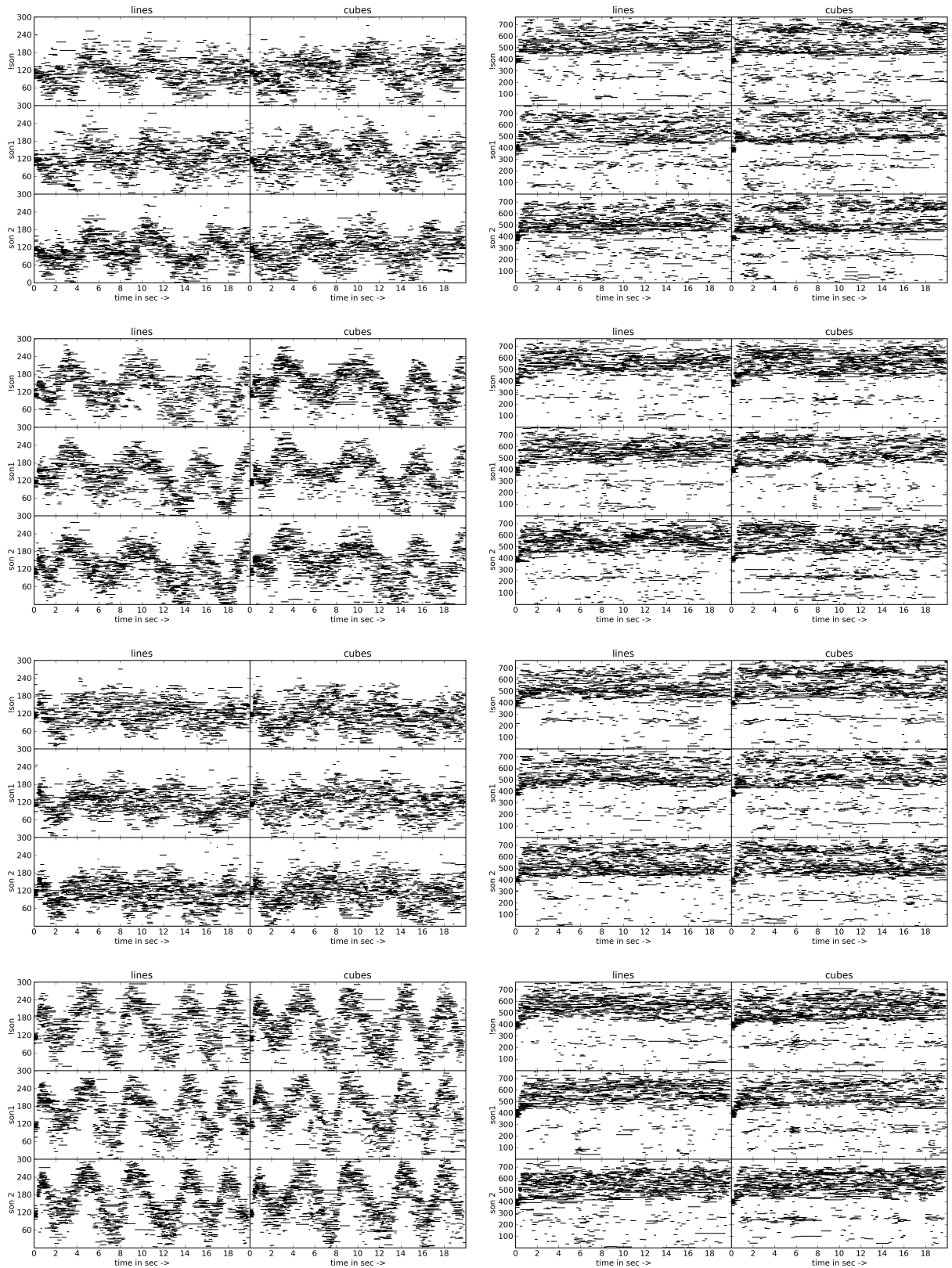


Figure 8.5: Horizontal fixations left, and vertical fixations right, from movement 1 top to 4 bottom.