Interaction with sound for participatory systems and data sonification
Liu, D.

Interaction with Sound for Participatory Systems and Data Sonification

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op dinsdag 21 november 2023
klokke 12.30 uur

door

Danyi Liu

geboren te Wuhan, China

in 1990

Promotores:
Prof.dr.ir. F.J. Verbeek
Prof.dr. A. Plaat

Promotiecomissie:
Prof.dr. M.M. Bonsangue
Prof.dr. H.C.M. Kleijn
Prof.dr. M. Fiocco
Dr. K. Fedorova
Dr. R. Saunders
Prof.dr. J.N. Kok (University of Twente, the Netherlands)

Cover design by Rentao Li (李任涛)
Layout design by Koen van der Blom and Danyi Liu
Printed by Proefschrift.nl
ISBN: 978-94-6473-270-2

The studies described in this thesis were performed at the Leiden Institute of Advanced Computer Science (LIACS), Leiden University, Leiden, The Netherlands.

The research was partially supported by the China Scholarship Council (No.201607720047).

To my mom

Contents

CHAPTER 1

General Introduction

1

1.1 Background

Sound cues are immersive. If we are blessed with a good auditory system, we use sound, i.e. auditory cues and signals to orient ourselves in the environment we live in. This thesis is about using sound in interaction with data. These data can be musical elements, abstract data, scientific data, etc. But more importantly in information systems, the data can be understood through interaction with sound elements. And it is this interaction that we will explore further. In order to get insight in this interaction, we introduce the term ***dialogue*** to analyze and define the whole interactive process (cf. definition 2.1). Dialogue is typically studied in Human Computer Interaction, in which dialogue is studied at three levels: lexical, syntactic and semantic (Dix, Finlay, Abowd, & Beale, 2003). It conveys messages of how a subject communicates with an interactive system and what the system gives back. Therefore, we have included three main elements to develop this dialogue model: *verbal*, *subject* and *adjective* (cf Figure 1.1). A verbal element involves the participation and actions taken by a subject, which initiates a dialogue at lexical level. Depending on the context and scenario, the subject can be a user, a participant, an audience member or a listener in different scenarios. In the dialogue, an adjective describes how a sound is produced by the interactive system responding to subject actions. Additionally, the interaction in this dialogue model requires the user to pick up the clues that are provided by the system through sounds. The design of sounds is essential to complement a dialogue at syntactic level so that a subject can learn the interaction. The clues are given by the total dialogue that is developed between the actions of the user and how a sound is produced through interaction, which can also assist the user to understand the system at semantic level.

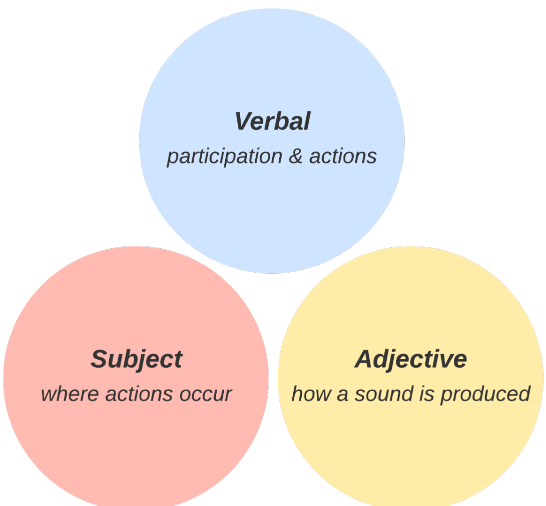


Figure 1.1: The Dialogue Model.

We will study these three elements of dialogue model to be able to propose a

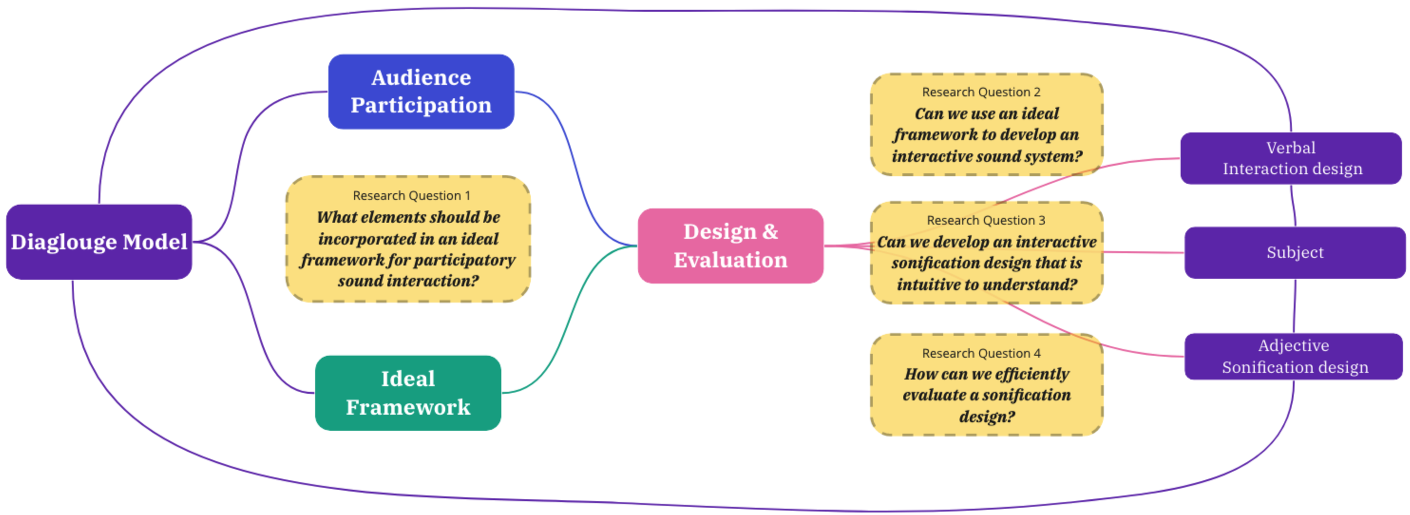


Figure 1.2: Dissertation narrative and structure. The figure presents an overview of the narrative of the dissertation, depicting the key concepts and research questions. The figure also provides a brief outline of the dissertation's structure, including the key topics, their sequence, and their interrelationship, highlighting the flow of the research inquiry.

framework that can be generally applied for sound interaction in Chapter 2 (cf. Figure 1.2). Therefore, participation of users in such interactive environment needs to be investigated first. This requires a very interdisciplinary approach. Subsequently, we intend to extrapolate our findings to an ideal but yet feasible framework. Meanwhile, we consider it necessary to study *data sonification*, in which the sounds are used to represent data in an interactive way and present users to understand data from sound. We will discuss how a sound is produced using different sonification design and how to make it learnable by a subject in Chapter 3 & 4.

1.2 Participatory Sound Interaction Models

Music interaction, as a valuable area of study, is actively used and therefore chosen to generalize concepts. In order to investigate interactive systems in the dialogue model that produce sound or relate to computer mediated music making, we start with participatory musical performance, in which audience is the main subject.

The field of Human Computer Interaction (HCI) is an important additional ingredient to this research. The universal Model-View-Controller (MVC) design pattern for interaction with computers is often used for building interactive software programs. The controller mediates input from an operator and converts it to commands for model and view; the model handles the command according to data and the rules of an application; and the view is the output that represents the data handled by the model such as charts and graphs (Gamma, Johnson, Vlissides, & Helm, 1995). Likewise, Van Troyer proposed three basic components to construct an interaction model in *participatory musical performances*: capture, effect, and performance model (Van Troyer, 2012). This is an extension to the MVC design pattern. While capture controls audience input, effect represents the outcome of audience manipulations. Next, the performance model processes and translates input to output.

In a participatory musical performance, participants are involved in the sound interaction with the intention of producing sound or music (cf. definition 2.2). From the works of Bayliss *et al.* (Bayliss, Lock, Sheridan, & Campus, 2004) and other researchers, we considered it essential to structure interaction phases in different sub-models like the audience model, the environment model, and the output mapping model. In this way, sub-models can be used to generate different interaction models. In another work, Bilda *et al.* proposed various interaction modes and phases indicated from the model of engagement (Bilda, Edmonds, & Candy, 2008), based on their observations and analysis of audience's intentions and expectations during their experience with interactive artworks. Hence, in Chapter 2, we design a participation journey map (cf. definition 2.4) to visualize the process that a subject experiences through uncovering the moments of observation, learning and interaction. This map presents a holistic view aiming to identify and analyze the audience participation form for the discussion of the interactions models. From our analysis in section 2.4 we derive a number of performance models that capture the interactions in different manners. As indicated, we study these participatory musical performances to get more insights in dialogues (cf. definition 2.1) for interaction with sound. Subsequently we will propose an ideal framework for participatory sound interaction.

1.3 Data Sonification: Interaction & Design

In the context of sonification and auditory display, sound has been used to represent complex data, enhance visualizations, as well as support the understanding of items in an educational context. Considering the MVC design pattern, sound is the view and sonification design is the model handling how data is transformed into sounds. Several approaches are distinguished from each other such as the use of earcons, auditory icons, parameter mapping sonification (PMSon) and model-based sonification (MBS) (Hermann, Hunt, & Neuhoff, 2011).

From a review on research in the area of participatory musical performance, we found that an interactive design would help a participant to understand and learn a sonification design and simultaneously the meaning of the sounds. For example, in SoundBounce, participants were able to throw and bounce a virtual ball to each other with smartphones (Dahl & Wang, 2010). The movement of the virtual ball was sonified with frequency modulation synthesis. The melodic pitch got higher, and the sound became louder as the ball rose. Additionally, the sound crossfaded from thrower to receiver. The interaction with the virtual ball is simulated with the affordances found in the real physical world by representing the changes in sounds. This example shows the possibility of an interactive slightly musical performance using a sonification approach (PMSon), which was succeeded in an intuitive interaction between the participants and the sound.

In order to study sound interaction in an audience focusing on a sound dialogue, we designed an implemented Bǎi/摆 as a research object aiming to achieve aforementioned intuitive interaction using data sonification. Bǎi/摆 is an interactive sound installation that uses a pendulum speaker as the interface to interact. The audience is free to move the pendulum speaker. The physical movement of the pendulum is sonified in a way that the control parameters for the sound synthesis are mapped (PMSon). The noticeable sound generated directly from the pendulum speaker can help the audience understand how their actions are being used for the sound and create an intuitive interaction form. Meanwhile, there are six surrounding speakers reacting the pendulum as independent entities (MBS).The sounds generated from the surrounding speakers employ both direct and indirect sonification approaches, which result in a harder understanding of the sound design. This may lead further exploration about the installation so that the audience can navigate through different types of sound composition and reach a continuous interaction. This navigation is a form of interaction to understand the sound and behaviour of the installation. The interaction is bidirectional between the audience and the system. The development process of this installation is a case study that involves both interaction design and sonification design.

In a more abstract case, we use our findings in a data sonification design where a mapping between sound and data needs to be understood. To that end, we use four sounds to represent four chemical elements (H, C, N, O). In order to understand the effectiveness of the mapping and how adjectives can influence a dialogue at semantic and syntactic level, we have designed different sonification designs and implemented an interactive sonification system which the participant navigates through the network of carbons in amino acids structures. In this study, we were interested in multiple concurrent sound sources. We assume such interactive navigation form would help the participants to learn the meaning of the sounds and understand a certain specified area of a molecular structure. Accordingly, participants can recognize and localize the surrounding chemical elements only with auditory signals. Thus navigation is an interactive method for the participants to perceive data and understand sound.

1.4 Evaluation of Data Sonification

From case studies, we learn how participants interact with a system using data sonification designs and we visualize the process with participation journey map. Subsequently, we design experiments to evaluate the system.

In the field of HCI, the System Usability Scale (SUS) has been commonly used to measure usability for interactive systems and applications. However, it is rather general and might not be applied to the field of auditory display or sonification, because of the individual differences in item interpretation. However, a usability framework can possibly be applied when evaluating the efficiency and the effectiveness of a sonification design, depending on the goals that are intended to be achieved in context of use. In previous evaluations of sonification applications, users were given various tasks during a series of experiments. Ibrahim *et al.* reviewed ten types of tasks that were used for measuring usability properties such as effectiveness, efficiency and satisfaction (Ibrahim, Yassin, Sura, & Andrias, 2011). For the experiment, such task design provides possibilities to obtain insights in factors that may influence the sonification design. We have conducted two cycles of experiments to evaluate, and further develop, our sonification design. The first experiment used a pretest-posttest design including training part to evaluate how easy the four element sounds can be identified and recognised. In the second experiment, two conditions of sound were tested using a within-subject design to investigate how many sounds can be maximally recognized and localized. In this way, we have been able to evaluate the learnability, immediacy and other aspects of this sonification design.

1.5 Research questions

Given the discussion presented in the introduction, we here formulate the research questions for this thesis:

RQ1 What elements should be incorporated in an ideal framework for participatory sound interaction?

It is essential to design an ideal framework for participatory sound interaction in the research presented in this thesis. The ideal framework can be generally applied in following research topics. We will answer this research questions in Chapter 2. Formulating the interaction models can be done through the overview of previous related artworks and research. By identifying the roles of the participants and performance models, it becomes possible to discuss related aspects such as: contribution of audience, interaction design, feedback of a system, sound production, etc. Auditory results and proper feedback, visual or otherwise, are possible to assist the audience with learning and understanding the interactive system and performance. Additionally, a responsive system can maintain the interaction between the audience and the system.

RQ2 Can we use an ideal framework to develop an interactive sound system?

Based on the ideal framework we propose in Chapter 2, we develop a sound installation Bǎi/摆 as a case study to answer this research question (cf. Chapter 3). While the audience is interacting with a pendulum speaker, the motion data collected from the speaker are sonified as a feedback responding to the audience. The diverse data and sonification design result in a stochastic system, which takes time to understand. Accordingly, the audience may involve into the interaction to learn, which makes the interaction responsive and ongoing. In order to introduce how an engaging and continuous interaction design is achieved, we will analyze the installation from three aspects: physical interface design, sonification design and experience of the audience.

We will also answer this research questions in Chapter 4, by presenting a deterministic interactive system designed for a single participant. Comparing to the stochastic system designed in Chapter 3, this system is developed in a way that is easier to learn and understand. We use atoms that are sonified in an environment surrounding the participant. In this way a participant can navigate a molecular structure through sounds. The ideal framework is applied to the system design, in which participants are expected to learn the sound mappings with the feedback from interacting with the navigation system. The participation journey map proposed in Chapter 2 will be used to analyze the experience.

RQ3 Can we develop an interactive sonification design that is intuitive to understand?

It is important that a sonification design is easy to understand and learn in an intuitive way (cf. definition 3.3). The development of Bǎi/摆 shows the sonification design of the movement of the pendulum as well as the interaction between the pendulum and surrounding speakers. It is addressed that the audience is able to explore the system or the interactive form without too many instructions (see Chapter 3). The sonification design for the molecular structures aims to achieve a faster and intuitive recognition and localization of the different concurrent sounds, which does not require too much practice and reduces cognitive load (see Chapter 4).

RQ4 How can we efficiently evaluate a sonification design?

We will answer this question in Chapter 5 & 6. Two cycles of experiments are designed for evaluating the sonification design, in order to get insight in factors that may affect individual performance in identifying and localizing concurrent sound sources.

Validation 1 includes a pretest-posttest design with a training phase, aiming at the learnability and effectiveness of the sonification design. The comparison between the results from the pretest and the posttest enabled us to look at different aspects such as element type, directions, durations, etc.

Validation 2 uses a within-subject design focusing at the identification and localization of multiple concurrent sound sources. The development of the two experiments and the analysis of the results may give clues to answer the question.

1.6 Structure of this Thesis

The structure of this thesis is as follows, Chapter 2 investigates several interaction models derived from existing participatory musical performances that are using mobile devices, *i.e.* cellphones. Several potential directions are indicated for the development of an engaging and ongoing interactive dialogue, aiming at answering the RQ1.

Chapter 3 presents the design and development of an interactive sound installation, which relates to RQ2 & RQ3.

Chapter 4 describes an interactive form of sonification in which the participant is able to navigate through a molecular structures of amino acids over the network of carbon atoms. This chapter proposes to use irregular but easy to recognize sounds for the representation of multiple concurrent sound sources.

The sound design is evaluated in Chapter 5 and 6, where two experimental approaches are involved. These two chapters pertain to RQ4.

Results obtained in other chapters are discussed in Chapter 7, answering the research questions and giving relevant perspectives for future studies.

This thesis contains examples of multimedia material, including sounds and videos. Sound samples and video fragments are referred to with a **QRCodes**. The internal camera on a mobile device such as smartphone or tablet can be used to scan the QRcodes, which will then lead to the material.

1.7 Contribution of this Thesis

- **Liu, D.**, & van der Heide, E. (2017). Interaction models for real-time participatory musical performance using mobile devices. In *Proceedings of the 2017 International Computer Music Conference, ICMC 2017, Shanghai, China, October 16-20, 2017*. Michigan Publishing. `http://hdl.handle.net/2027/spo.bbp2372.2017.051`

- **Liu, D.**, Kroese, J., & van der Heide, E. (2018). The development of bǎi: An oscillating sound installation. In *Interactivity, Game Creation, Design, Learning, and Innovation*, (pp. 69–79). Springer. `https://doi.org/10.1007/978-3-030-06134-0_8`

- **Liu, D.**, & van der Heide, E. Interactive auditory navigation in molecular structures of amino acids: A case study using multiple concurrent sound sources representing nearby atoms. (2019) In *Proceedings of the 25th International Conference on Auditory Display, ICAD 2019*, (pp. 140–156), Newcastle, UK. `https://doi.org/10.21785/icad2019.049`

- **Liu, D.**, & van der Heide, E. Evaluating the spatial sonification of the molecular structures of amino acids using multiple concurrently sounding sources. (2021) In Proceedings of the*26th International Conference on Auditory Display, ICAD 2021*, to appear.`https://doi.org/10.21785/icad2021.013`

1.8 Other publications

- Dekker, L., Peeperkorn, M., **Liu, D.**, & Verbeek, F. J. (2019) An Alternative Interaction Paradigm for DNA-Sequence Data. In *Proceedings of the International Conferences Interfaces and Human Computer Interaction 2019 Game and Entertainment Technologies 2019 and Computer Graphics, Visualization, Computer Vision and Image Processing 2019*, (pp. 343-347). IADIS Press. `https://doi.org/10.33965/ihci2019_201906C045`

- Papagiannis, A., **Liu, D.**, Zammit, A., Gultyaev, A. P., & Verbeek, F. J. (2019) Interacting with RNA Secondary Structure through Sonification. In *Proceedings of the International Conferences Interfaces and Human Computer Interaction 2019 Game and Entertainment Technologies 2019 and Computer Graphics, Visualization, Computer Vision and Image Processing 2019*, (pp. 43-49). IADIS Press. `https://doi.org/10.33965/ihci2019_201906L006`

CHAPTER 2

Interaction Models with Sounds for Real-time Participation

This chapter is partially based on the following publication:

Liu, D., & van der Heide, E. (2017). Interaction models for real-time participatory musical performance using mobile devices. In *Proceedings of the 2017 International Computer Music Conference, ICMC 2017, Shanghai, China, October 16-20, 2017*. Michigan Publishing.

2.1 Introduction

Sound interaction, as an interdisciplinary art form, it is relevant to the study of the arts, audience, behavior, techniques, interactions and so forth. Considering interaction with sound, participatory performance that involves music is an interesting area to gather examples on how such interaction can evolve. We consider the interaction consists of a system that communicates with participants, where a dialogue is possibly developed (cf. definition 2.1). If more people are involved this should be considered computer mediated interaction. In general, we start to consider systems in which the interaction is co-located and synchronous. In addition, remote but synchronous systems are considered. In terms of interaction this states whether or not the participants are in the same place as the system or not. For the further understanding we first introduce three major concepts that are important to the development of interaction models.

As indicated, crucial to the interaction is the dialogue:

Dialogue

Definition 2.1 *Dialogue has an interactive component, which is developed between the actions of the subject (i.e. audience) and the reactions of a system. It possibly also provides a means of learning.*

A dialogue can be applied in different context. As a paradigm we first investigate participation in musical performance defined as:

Participatory musical performance

Definition 2.2 *It is an immersive form of sound performance that directly invites the audience to be a part of, or influence the performance in real time.*

Participants are essential to dialogue and performances.

Participants

Definition 2.3 *In participatory musical performances, participants are the people who take part in a performance and partly contribute to the performance result.*

Besides the traditional western concert performance practice, the participation of audience in performances is becoming an emerging field. The current ubiquity of *mobile devices*, mobile phones or smartphones, makes it possible for large numbers of participants to interact with sounds in real time. The participants in such musical performance could be either just audience members or the audience in combination of professional performers. All considered audience takes different roles from performers, to composers, to editors, etc., varying from work to work. While some systems provide a relatively simple and passive participation form, other systems can develop a dialogue with the audience and achieve a complex participation form. Referring back to the ***dialogue model*** (cf. Figure 1.1), the verbal element enables audience input and actions, while the subject represents the entity or role that initiates the dialogue. Furthermore, the adjective provides descriptive information on how the system responds to the subject's actions, particularly in terms of sound production.

In this chapter, we first review documentation and publications of a series of real-time participatory musical performances, from which we deduce the relevant aspects as well as a journey map of participation. Then we discuss several forms of audience participation by mainly considering in three aspects: active / passive, direct / indirect, with / without limitations. From our analysis we derive a number of performance models that capture the interactions in different manners. Meanwhile, we assess usability issues for the interaction design of performances and follow the explanation of unidirectional and responsive interaction based on the approaches how sounds were designed and produced. Lastly, we propose possible directions of efficient sound interaction design for further research.

We provide video examples of some related work through QRcodes, which can be scanned with a mobile device for viewing. The QRcodes are numbered and given in the page margin.

2.2 Participants

In some existing participatory musical performances, participants could be divided into two groups, novices and masters (Lee & Freeman, 2013; Freeman et al., 2013). Miletto *et al.* considered a *novice* a music beginner, a person who lacks musical knowledge or who is just learning the rudiments of music (Miletto et al., 2011). We have excluded works that were designed for professional performers, as these required substantial practice or extensive prior musical knowledge. Thus the works we review in this chapter were not aimed at mastered musicians, but attempted to engage the audience in the performances. Furthermore for definition 2.3 we characterise participant as 1) audience and performers or 2) audience only. Performers may play a leading or a collaborative role in a performance.

2.2.1 Audience & Performers

In traditional western concert performances, there has been a clear separation between audience and performers. While the audience is watching and listening at the auditorium, performers are performing arts onstage in front of the audience. The development of mobile devices and wireless technology has opened doors for a completely different approach to engage the audience, creating the possibility for the audience to interact in a performance and become a part of the creation of the performance. In this respect, one of the pioneer works is *Dialtones*. This is premiered in 2001 (Levin, 2001). The audience was asked to register their phone numbers at web terminals and specific ringtones were installed on their phones. During the performance, the phones were dialed by the performers via a computer program that allowed 60 phones to ring simultaneously. In this way, the ringtones could be orchestrated as a musical performance (cf. QRcode 2.1). The audience's mobile phones became the performance medium, although their participation was *passive* (cf. definition 2.5) in the performance itself.

QRcode 2.1

Since then, performances have been developed to invite the audience to participate in a more *active way* (cf. definition 2.6). *massMobile* (Freeman et al., 2015; Weitzner et al., 2012) is an audience participation framework developed by Freeman *et al.* in 2012. It provides a possibility for the audience to shape an onstage live performance. It was used to develop an application for the live performance *Saxophone Etudes* (Freeman, 2012). The audience could vote for various musical factors through the application, including tempo, dynamics, note duration, articulations and measures of the music, which were displayed to the saxophonist for solo improvisation in real time. In this case, a dialogue can be initiated at lexical level via the interaction between the audience and the performer.

Both works combine audience participation with performers and construct different collaborations between the audience and the performers. While the performers directed the performance in *Dialtones*, the audience actually influenced the way the performance developed in *Saxophone Etudes*.

2.2.2 Audience Only

In some participatory performances, there are no professional performers and the audience is instructed to create or join a performance individually or form a group. In *Tactical Sound Garden (2006)*, participants can choose a sound from a pre-designed sound library and plant it at a certain location by adding a GPS location (Shepard, 2006). Furthermore, they can modify the sounds planted by others and modify their volume and repetition time. All the sounds are mixed at the server-side and streamed to mobile devices that are used by the participants. As soon as someone walks into the garden, the sounds planted around him would be played in real time. When one plants or modifies a sound based on the mixture of former sounds in the garden, communication among the audience members emerges through the sounds. Each audience member contributes sounds independently, and one audience member can be seen as the contributor in others' view.

Swarmed is yet another case of audience only participation (Hindle, 2013). It applies a captive-wifi-portal allowing participants to generate audio via a webpage used on multiple mobile devices simultaneously. There are several pre-defined instruments on the webpage for the participants to choose from, which produced synthesised sounds based on gestures. The audience are able to hear the sounds they are playing on their phones independently, as well as the combination of the sounds produced by the other audience from onstage speakers.

SoundBounce shows the possibility for multiple audience members to be able to perform and interact with a sound in a group (Dahl & Wang, 2010). In *SoundBounce* a sound is regarded as a ball. Pre-defined gestures allow the players to throw and bounce the virtual ball to each other according to compass data using their mobile devices (cf. QRcode 2.2). The movement of the ball (or sound) is sonified through FM synthesis. Although the premiere of *SoundBounce* was initially performed by a group of professionals, we still included it in our audience category. *SoundBounce* utilises a physical metaphor to develop the mobile interaction and is designed in an intuitive way for a group of participants to perform and interact with each other. We see this as an equally meaningful approach in an audience only context.

QRcode 2.2

All aforementioned works are examples of audience only participation performances and the auditory results are non-deterministic. *Tactical Sound Garden* is not a performance with a beginning and an end and can therefore be seen as an installation format. *Tactical Sound Garden* and *Swarmed* have a similar approach in the sense that the participants make their own contributions and are not 'forced' to interact with each other. Interaction between the participants is an implicit requirement of *SoundBounce* since the ball has to be thrown and caught. In all cases the participants can take a pause and just listen to what is going on.

2.3 Participation Journey Map

If we consider features of an interactive sound system, we take as a starting point that the feature is available to one or more participants. The system can be fully self-operational or can have interference with a performer. In the latter case the conditions, i.e. rules, of the system can change during the interaction. We investigate these systems in relation to the interaction that is provided. The interference of the performer/operator is, for now, of less interest.

In this section, we focus on the audience rather than professional performers. Participation journey map is a well-known tool to investigate participation across various states and how individuals transition between these states(Mast, de Vries, Broekens, & Verbeek, 2021). We analyze different forms of audience participation based on a user journey map (see Figure 2.1, cf. definition 2.4), which consists of three main stages: observation, learning, and interaction. Some participation forms are passive and may include little interaction, while other forms may in-

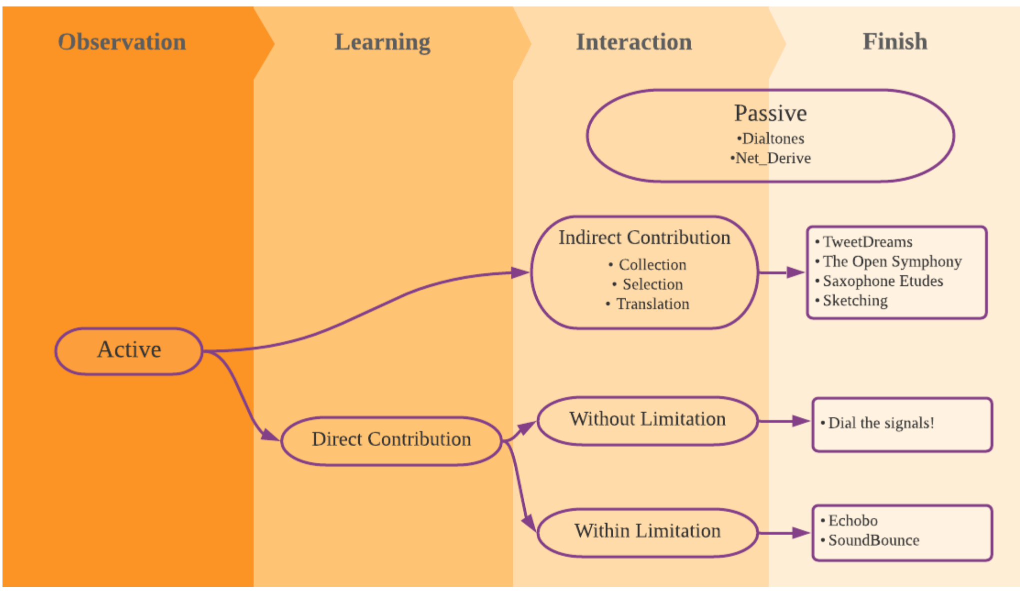


Figure 2.1: A journey map of audience participation.

volve a more complex progression that requires the audience to learn for effective interaction.

Participation Journey Map

Definition 2.4 *A tool to visualize the process that a participant experiences through when interacting with a system.*

2.3.1 Passive Participation

We first discuss the passive form of participation and define it as follows.

Passive participation

Definition 2.5 *Passive participation is a form in which the audience does not influence the result of a performance, but are nevertheless aware of participation.*

In *Dialtones* (Levin, 2001), the audience did not need to take any action but only brought their phones along. This innovative way of performing sounds tried to get participants involved in the performance, still the passive role might make participants feel surprised more than engaged.

In *Net_Dérive (2006)*, the participants were given a broader palette of actions to follow. There were several paths for participants to choose and follow around a specific location, i.e. a gallery (Tanaka, 2006). While the participants were walking, ambient sounds were automatically recorded by the participants' phones at regular intervals. These recordings were mixed at server-side and played in the gallery. Although the participants could choose different routes around the gallery, the process of recording was automatic.

Compared to *Dialtones*, in *Net_Dérive* the participants were activated more. But they did not have the opportunity to fully decide the recorded material. Both works require the audience to be present and participate but the audience has little or no influence on the sounding result itself. Thus in journey map, passive participation skips the stages of observation and learning (see Figure 2.1). Additionally, the way of interaction is quite limited.

2.3.2 Active Participation

The interaction of an active participant with a system requires from the participant to observe and pick up the clues that are provided by the system in order to understand the system (see Figure 2.1). The clues are given by the total dialogue (cf. definition 2.1) that is provided by the system. Visibility is an essential quality of allowing participants to easily recognise the clues (cf. definition 2.7).

Active participation

Definition 2.6 *Active participation signifies that the audience makes choices through interaction to influence and construct a performance via interaction.*

Visibility

Definition 2.7 *Visibility is a degree to measure whether the reaction from a system (feedback) is noticeable and understandable for the audience (Dix et al., 2003).*

In this section we will make a distinction between two forms of active participation: direct and indirect contribution.

Direct Contribution

Direct Contribution

Definition 2.8 *Direction contribution indicates the condition in which the audience is directly involved in the production of sound. The auditory results can be a clear feedback for the audience to perceive and help them learn the system (cf Figure 2.1).*

An example of direct contribution is that actions of participants are directly utilised as the input of sound generation or to trigger audible events. The audience is likely to become aware of how everything functions from aforementioned conditions, which brings possibility of learning to the audience (see Figure 2.1). In *Dial the signals!* (Ligna & Röhm, 2003), a matrix of mobiles phones was exhibited as an installation and the numbers of the mobile phones were passed to the audience to dial. Every tone the phones played was broadcasted by several radio stations and as a live stream on the internet. The audience had full control of deciding which phone to dial and their involvement was direct. Furthermore each of the phones corresponded to a specific sound, which provides a clear visibility of the auditory feedback for the audience to track. The audience is able to learn the mapping between the phone numbers and tones accordingly (see Figure 2.1). Meanwhile, the audience gets complete freedom and there are no specific choices made regarding the development over time. It is impossible for the designers of this work to predict which phone would be dialed first or in which order a series of phones would be dialed.

In some other works, specific interaction rules are used to limit the direct contribution from the audience. In *Echobo* (Lee & Freeman, 2013), an eight-key keyboard was displayed on the individual mobile phone as an instrument. The audience was instructed to play the keys of the chord selected by the composer or a master musician (cf. QRcode 2.3). The sequence of playing keys was, however, not fixed. The keys were marked with a black arrow and an electronic piano sound was generated directly from the phone as the harmonic backdrop of the performance. The harmonic structure was controlled by providing limited keys for the audience to play. Accordingly, the audience collaborated with the performer in harmony, while having partial freedom to play the instrument.

QRcode 2.3

In *SoundBounce* (Dahl & Wang, 2010), the audience performed with certain mapping rules pre-defined by composers and developers. The movement of the virtual ball was sonified through frequency modulation synthesis, which can be categorised in parameter mapping sonification approach (PMSon, cf. definition 4.3). The melodic pitch got higher and the sound became louder as the ball rose. Additionally, the sound cross-faded from the thrower's phone to the receiver's phone. *SoundBounce* is an example showing how motion data has been sonified in an intuitive way for participants to interact with. Limitations were set up to control the randomness and unpredictability in *Echobo* and *SoundBounce*, to improve the structure and visibility of auditory results.

Indirect Contribution

A participatory performance using interpretation or manipulation of the audience's contribution might have a more indirect nature. Therefore, we define indirect contribution as follows:

Indirect Contribution

Definition 2.9 *The audience's active participation is passed through a selective and translation process, whereas the audience members themselves do not control or produce sound directly or instantaneously. Indirect contribution reduces the visibility of feedback.*

So if the contribution is indirect, it is harder for the audience to observe or track their contribution due to the lack of feedback. While being indirect, the audience plays a crucial role in the performance. For example in *TweetDreams*, the audience was asked to tweet during the performance (Dahl et al., 2011).

Tweets with a certain hashtag were able to be searched and collected. The specific hashtag and the rate of appearance of each tweet were determined by the performer, furthermore the tweets were used as an input for an algorithmic composition (cf. QRcode 2.4). The combination was indirect in this example, thus it was harder or impossible to learn how the input data (tweets) had been transformed into sounds (see Figure 2.1).

QRcode 2.4

The Open Symphony is another example of indirect contribution. Four professional musicians improvised in accordance to a score resulting from the audience's votes (Zhang et al., 2016). The audience could vote different playing modes for different players through a user interface, including single notes, melodies, free improvisation, silence, etc. The result was displayed as graphic notations on a big screen from left to right. With a limited amount of features to vote, the audience contributed compositional resources to the performers to improvise.

2.4 Performance Model and Sound Production

The interactive experience is constructed through a dialogue between the audience and the system. Such dialogue is possibly developed in a performance (cf. definition 2.1). By participating in an interactive sound based system, an audience member can control or influence the auditory outcome of the performance stepwise (cf. definition 2.10). Therefore, we first introduce the concepts of *performance* and *performance model*. In regard to the sound production in each performance, we analyze it with respect to the concepts of *deterministic* and *stochastic*.

Performance

Definition 2.10 *A performance consists of the join of actions of the audience and the reactions of an interactive system, which involves how a sound is produced through interaction.*

Performance model

Definition 2.11 *A performance model describes the connections among audience members and between the audience and the performance system.*

Deterministic System

Definition 2.12 *There is a fixed set of mapping rules between input data and output sounds.*

Stochastic System

Definition 2.13 *The output sounds are generated from the input data via a model with randomness and mutations.*

Schraffenberger and van der Heide considered mutual influence between audience and interactive artwork as an important underlying principle of interaction (Schraffenberger & van der Heide, 2015). Some performances might have an evolving interactive dialogue, whereas the interaction model could be static in other performances. Accordingly, we will categorise and discuss two models in this section, the inherent performance model (cf. Figure 2.2) and the responsive performance model (cf. Figure 2.3). In order to align with the dialogue model proposed in Chapter 1 (see Figure 1.1), same color coding is used in Figure 2.2 & 2.3 to analyze the elements from a dialogue in both performance model.

2.4.1 Inherent Performance Model

In *Dial the signals!* (Ligna & Röhm, 2003), participants were free to dial the mobile phones exhibited in the installation. The corresponding phone responded immediately with a specific ringtone. Although there is possibilities for participants to learn the mapping between ringtones and phone numbers and choreograph the combination of ringtones to create a larger composition. There is no varying interactive dialogue taking place between the actions of the audience and the reactions of the system. The interaction was unidirectional and ended when a phone was rang (see Figure 2.2a). In similar fashion, in *Moori* (Kim, 2011), the audience responded to a list of guided questions sent by the performer to form a story. The text messages then were spoken with text to speech software. Despite the questions that were designed like a narrative script to keep the audience answering the questions, the interaction ended when one answer was spoken (see Figure 2.2b). Both aforementioned works are examples for a deterministic system (cf. definition 2.12), where there are simple fixed relations between the actions of the audience and the sounds being played. The performance model is linear and unidirectional. Therefore, there was no real opportunity for an interactive dialogue.

In *TweetDreams*, tweets were used as input for an algorithmic composition (Dahl et al., 2011). The algorithm brings a stochastic system creating a more diverse auditory outcome (cf. definition 2.13). A dedicated computer program analyzed the selected tweets and checked whether one was new as a root or was

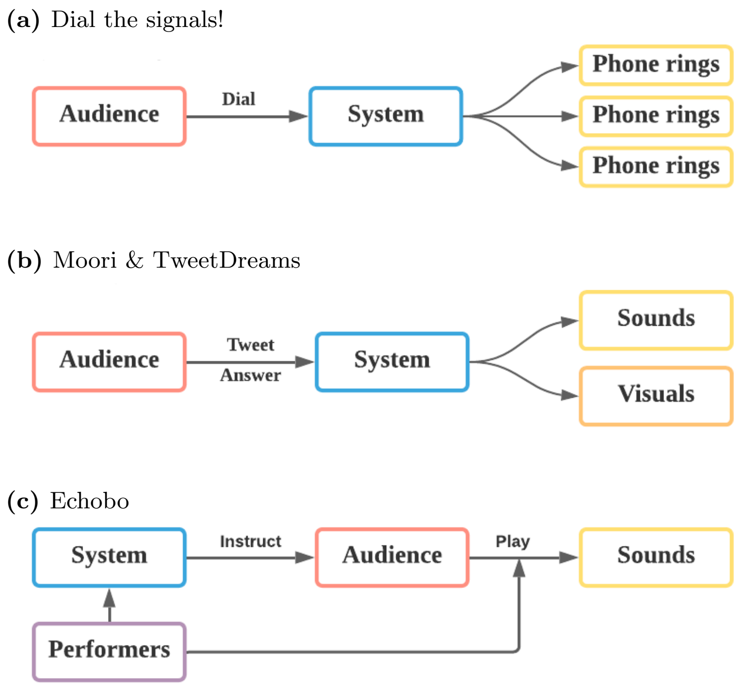


Figure 2.2: Diagram of Three examples of Inherent Performance Model.

related to an existing root. When a tweet starts a new root, a new melody will be chosen from pre-composed melodies, otherwise the melody will be mutated from its parent melody. This is another example how model based data sonification (MBS, cf. definition 4.4) has been used in real-time audience participation performance. Although the visualisation of tweets gave the audience feedback as clues to track and locate their own contribution, the musical transformation may not be clear enough for a true interactive dialogue.

In *Echobo* (Lee & Freeman, 2013), there was one musician controlling the chord progression, and an acoustic instrument player performing melodic material. The audience can play only with the harmony defined by the musician. Still, the audience was free to decide the order of keys they played. The performer would probably affect the decision of the audience about the order. The unidirectional interaction started with the chord selection and ended when the audience finished playing the notes from the chord without any further dialogue.

In order to characterise this kind of performance model, we introduce the concept of ***inherent performance model***. From the diagram (cf. Figure 2.2), it can be deduced that the interaction in the inherent performance model is often unidirectional.

Inherent performance model

Definition 2.14 *The inherent performance model is a form of interaction that does not provide an actual interactive dialogue between the audience and the interactive system. The audience is not able to fully understand the system due to the lack of perceivable feedback.*

2.4.2 Responsive Performance Model

On opposed to the inherent performance model, we introduce the concept of ***responsive performance model***.

Responsive performance model

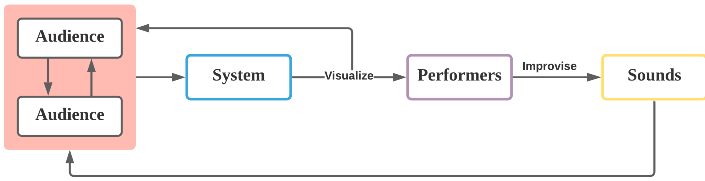
Definition 2.15 *The responsive performance model provides a constant interaction between the audience and the performance, in which the audience is motivated by various forms of feedback from the system.*

As an example considered, the improvisation of the performers could give a clear feedback to the audience and create a continuous interactive dialogue (cf. definition 2.1). It may also make the audience feel that they were engaged and influencing the performance. Because the audience might be curious about, or anticipate on how their contribution is being translated or performed.

In *Sketching* (Freeman et al., 2013), the audience could draw different shapes through a web page developed within *massMobile* (Weitzner et al., 2012). Each shape was assigned to an instrument. Other features such as colour, size, opacity, border were mapped to different musical parameters. Performers from a jazz band would pick and play a jazz standard and improvise based on the drawings from the audience, or entirely improvise with the drawings. The audience might draw a new shape or adjust the features according to the previous performance, and try to develop the performance further. This is a typical form of a responsive performance model.

In *The Open Symphony* (Zhang et al., 2016), the graphic notation presented the individual and collective feedback for both the audience and the performers. The decision of a single member of the audience may be influenced by the other

(a) The Open Symphony



(b) Tactical Sound Garden

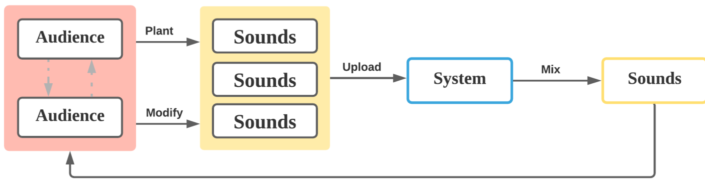


Figure 2.3: Diagram of Two Examples of Responsive Performance Model.

audience members as well as by the performers (see Figure 2.3a). Differently, in *Saxophone Etudes* (Freeman, 2012), although the points selected by the other audience members were presented on mobile of each audience member, the audience did not know the results of voting for each musical factors, such as speed, dynamics, etc.

In *Tactical Sound Garden* (Shepard, 2006), the participant was not only able to plant a sound, but also modify the sounds planted by others. Influenced by the mixture of the sounds in the garden, one participant may come up with various ideas about planting and modifying sounds. Such interaction remained among the participants and the sound mixture. In addition, it extended the experience of the audience beyond the time that the audience is interacting themselves. Therefore, a responsive performance model not only provides better visibility to stimulate constant interaction with the system, but also provides the clear interaction in a group of participants.

2.5 Discussion

In contrast with visual feedback, the auditory feedback might be less noticeable and understandable when the contribution of the audience is indirect. This depends on whether the audience is capable of associating the musical outcome with their actions or contributions. Sound design or the adjective of a dialogue can possibly improve visibility and affordance (see Figure 1.1). Of the performances that we have reviewed in this chapter visual feedback is used to create insight in the performance and make the interaction more engaging. On the other hand, the visual feedback might also distract the audience from the actual auditory result.

Our interest focuses on the interaction with sounds and therefore initially we concentrate more on the sound and music than the visuals. In this context, we prefer a form of interaction in which the contribution of the audience and the auditory results can be clearly perceived, which possibly provide feedback for the audience to understand the interaction (cf. Figure 2.4).

The inherent performance model (cf. definition 2.14) has the characteristic that lacks an evolving interactive dialogue, therefore we argue that it would be harder to keep the audience participating in a performance with low visibility. The responsive performance model (cf. definition 2.15), however, could lead to a more dynamic and ongoing form of interaction.

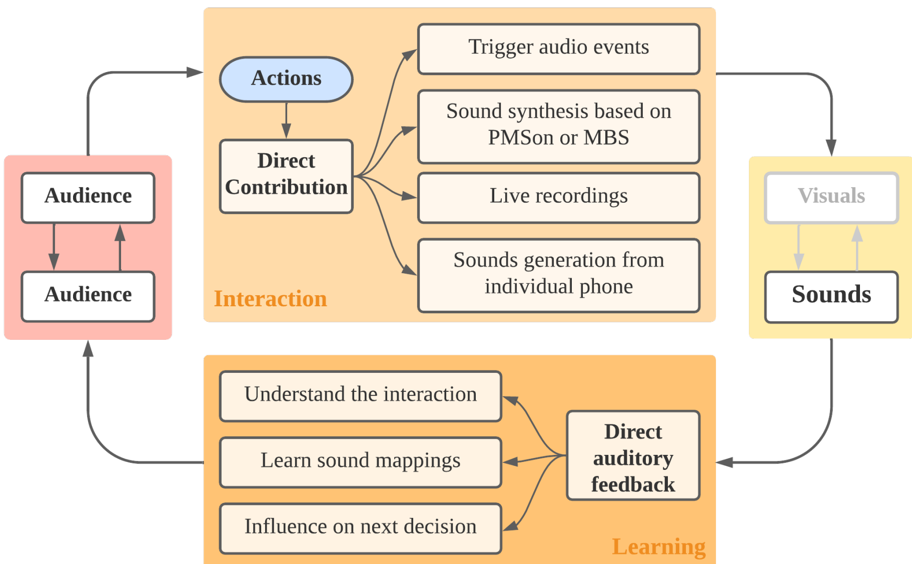


Figure 2.4: Proposed Framework for Future Research. A responsive performance model combining with direct contribution and direct auditory feedback, aiming to achieve a constant loop between interaction and learning. The color code refers to the elements of the dialogue model: subject-audience, verbal-actions, adjective-sounds.

Although efforts have been made to maintain the audience participating longer by elicitation, for instance, questioning (Kim, 2011) or continuous instruction (Lee & Freeman, 2013). The feeling of participating could be broken at any time during the performance since the audience does not get enough hints of participation from the result. Furthermore, a static form of interaction could be in the way of developing a true dialogue between the audience and the performance system. In most of the discussed works in section 2.4.2 the participation form has an indirect nature.

It is worth a try to combine the approaches used for direct contribution in the context of a responsive performance model (cf. Figure 2.4). The proposed framework attempts an explicit description of a dialogue between the audience and the system. It includes two main components from the participation journey map (cf. Figure 2.1), interaction and learning. In this case, sound will play an important role to help the audience to learn the interaction rules and understand the interactive system. While learning can motivate the audience to better interact with the system, the auditory feedback from the interaction can boost the learning process.

Additionally, a stochastic system can bring along an element of surprise (cf definition 3.2). In that case the dialogue might develop in an unpredicted way as the feedback has a stochastic element. However, it may result in low visibility of the feedback. This is also seen in case if the system is operated through a performer that changes the state of the system and thereby changes the expectations that were acquired by the participants. In this case, a clear feedback of sound design is required to achieve obvious affordance and improve the visibility.

2.6 Conclusion

The mobile device is a widely available medium for the audience to participate in a performance in real time, but it is not necessarily used. We have chosen not to focus on the device or technology itself but rather on the possible forms of interaction. Participatory musical performance is a great start point for our journey of exploring and learning about sound interaction.

In the course of our review, we have presented the descriptions of *participants* (cf. definition 2.3). The audience turns out to be the participants when they take part in an interactive system. Performers could be included to lead or collaborate with the audience in a performance but are not a requirement. The journey map has been an essential tool for us to visualize the participation experience from observation, learning to interaction (see Figure 2.1). It is also the foundation for the proposed framework (cf. Figure 2.4).

We used the audience as an example to investigate the behaviors of subjects in a dialogue. We have categorised audience participation forms from their behaviors via the participation journey map (cf. Figure 2.1). There is little interaction existing in passive participation. Differently, active participation supports a wider range of interactions. In some works, certain rules were set up to limit the possible results and create more control over the performance. Still, participants can be aware that their decisions are forming and influencing the performance if their auditory contribution is direct. Indirect contribution employs a system to collect choices of participants, and translate those materials into compositional resources to create a performance. Participants experience less clear direct feedback of interaction from this form.

We have distinguished two performance models. We have analyzed the interactions among audience members, and between the audience and the performance. Subsequently, we have labelled existing works with inherent or responsive performance models. Aspects such as interaction, feedback, randomness, control have been discussed. We are particularly interested in exploring the experience of a responsive performance in combination with direct contribution in order to create a more engaging and interactive dialogue experience (cf. Figure 2.4). The system's responses can vary based on the actions of the audience and the context of the interaction. By considering the three elements of the dialogue model, we can create more engaging and tailored dialogue systems that enhance the audience's experience. Accordingly, we need to investigate interaction design (verbal) and sound design (adjective) further. While interaction design can influence a dialogue at semantic level, including how the participants comprehend the system. Various sound design approaches need to be examined what works for better auditory feedback and how to realize a dialogue at syntactical level for learning.

CHAPTER 3

The Development of an Interactive Installation using Sonification: Bǎi/摆

This chapter is partially based on the following publication:

Liu, D., Kroese, J., & van der Heide, E. (2018). The development of bǎi: An oscillating sound installation. In *Interactivity, Game Creation, Design, Learning, and Innovation*, (pp. 69–79). Springer.

3.1 Introduction

In Chapter 2, we have concluded an ideal sound interaction framework as a responsive dialogue between the audience and the performance (cf. definition 2.10). While the audience participation and interaction starts the dialogue at lexical level, the semantic level of this dialogue involves auditory feedback and influences the audience's comprehension. A sound design complements this dialogue at the syntactic level by assisting the audience to learn and understand the system. We found such dialogue could be developed through a loop between the progress of learning and interaction (cf. Figure 2.4). This may achieve the possibility of a more continuous interactive system we proposed in section 2.5 for audience participation. We intend to apply the ideal framework to the interaction design of a sound installation, to examine whether an installation can constantly engage the audience through the interaction (cf. Figure 3.2).

As a starting point of such installation we would develop a musical instrument for audience participation. Hereby we first define a musical instrument as:

Musical instrument

Definition 3.1 *A musical instrument is a device which can be controlled to produce musical sounds.*

Since the invention of the loudspeaker, investigations, composers and artists have explored various ways of using speakers ranging from multi-channel speaker setups and hemispherical speaker designs to speaker sculptures and wearable speaker-based instruments. While speakers are often used in static positions, Gordon Monahan's *Speaker Swinging*, first performed in 1982, did apply a moving speaker as a *musical instrument* in a live performance. Three performers each swing a loudspeaker in circles with a sine or square wave as source signal (Monahan, 1982). The resulting sound is subject to the Doppler effect and the acoustic properties of the space. In 1968, Steve Reich pioneered the pendulum principle in his *Pendulum Music* (Reich, 1968). The performance involves phasing feedback tones resulting from suspended microphones swinging above the speakers (cf. QRcode 3.1). *Spatial Sounds (100dB at 100km/h)* by Marnix de Nijs and Edwin van der Heide (2000, 2001) is an interactive installation using a moving speaker. The installation interprets the position and movement of a visitor and reacts to it both with its movements and the real-time generated

QRcode 3.1

sound. In return, the visitors react to the installation and go through different experiences and emotions (van der Heide, 2010).

What these works have in common is that they exploit the physical properties of a moving sound source ,or microphone, in their design. The development and possibilities with loudspeakers are interesting and they provide a good basis for interaction and dialogue. However, more physical interaction can be added to the design. Therefore we will further investigate this in an interactive sound installation. We translate some of the ideas of Reich to this setup and investigate a moving speaker as the basis of a pendulum (cf.Figure 3.1). We decided to comply the moving sound source with our framework and use a speaker setup to achieve a responsive dialogue (cf. Figure 3.2). Additionally, we intend to investigate a motion data sonification design to determine whether the audience can understand the relation between the sound and the movement.

This chapter describes and reflects on both the technical and artistic decisions that were made during the design and development of an interactive sound installation, Bǎi. It covers the design goals and a short reflection upon what we have achieved so far.

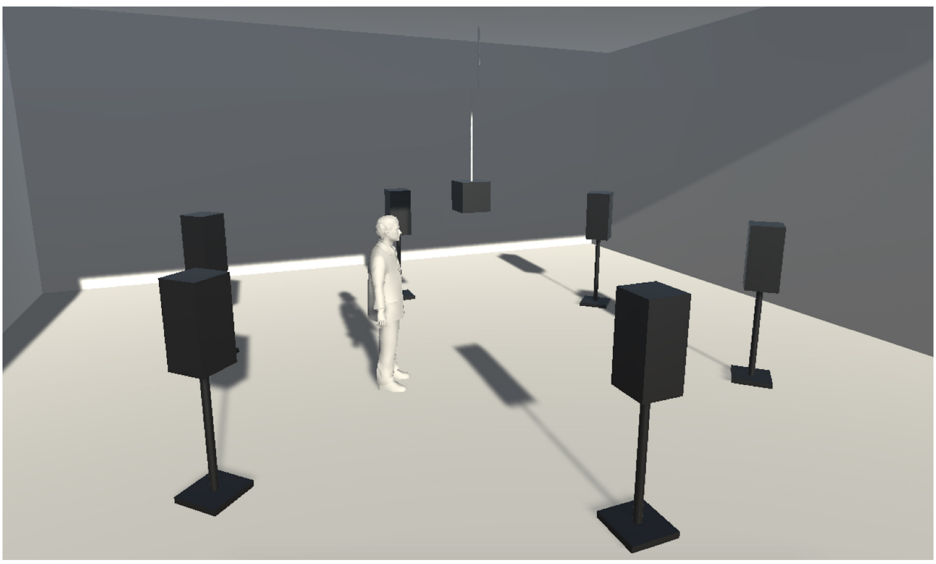


Figure 3.1: 3D model of the space setup for Bǎi/摆: an Interactive Sound Installation.

3.2 Interaction Design for Participation

In our approach we interpret the term interaction as a dialogue between the audience and the installation (cf. definition 2.1). We aim at developing the interaction and behaviour of the installation as surprising and intuitive.

Surprise

Definition 3.2 *In a surprising dialogue the two parties communicate and react to each other while neither of the two parties is fully predictable, nor has full control over the situation.*

Intuitive

Definition 3.3 *In an intuitive installation, the interactive form is easily understandable, so that the audience is able to explore the system without specific instructions.*

The interactive behaviour is not static but develops in order to realise an interesting ongoing dialogue. We have set a number of goals to help us achieve this:

1) use analogies between the physical input and sonic output of the system,
2) give the audience the experience of interacting with a system that reacts to their input but also surprises them with its own unpredictable behaviour,
3) make the audience aware that their actions impact the way the system behaves, without being able to fully control it,
4) make the audience perceive the system as 'beautiful', but also 'upset' or 'dangerous' through the changes of its behaviour (see section 3.3).

At the core of the interaction design is to experiment the ideal framework proposed in Chapter 2 in an interactive sound installation. We started with a moving speaker in the form of a pendulum as an interface for the audience to actively interact with (cf. definition 2.6). The pendulum speaker is suspended from the ceiling, surrounded a 6-speaker setup standing on the floor (see Figure 3.1).

Through pushing, pulling and rotating the pendulum speaker, the audience can set it into different oscillating motions (see Figure 3.2). We choose for the installation not to have a fixed form of interaction but rather have it alternative between different rules, and therefore different modes of behaviour (cf. in section 3.2.4). Furthermore, we have chosen a stochastic system for sound production in which the surrounding speakers react to the pendulum as it approaches them resulting in a dynamically changing sound environment. At first, it may

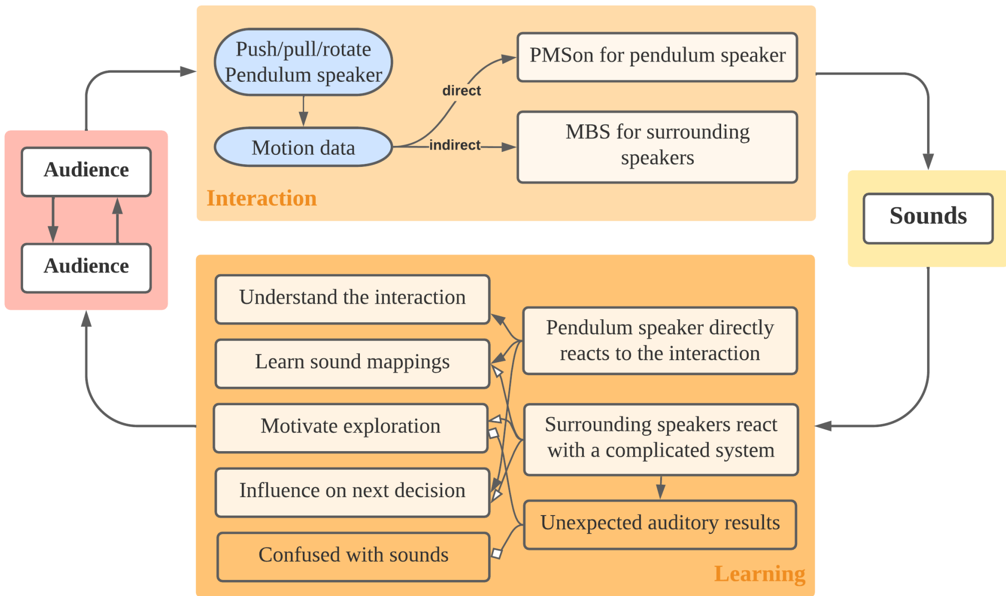


Figure 3.2: Ideal Framework for Băi. The color code refers to the elements of the dialogue model: subject-audience, verbal-actions, adjective-sounds.

seem that the environment reacts to the motions in a predictable manner. However, the self-movement of the pendulum influences the behaviour of the system, even when the audience does not directly interact with it; this brings unforeseen results. Combined with the fact that physical labour is needed to restrain the pendulum, this leads to a tense dialogue between the participant and object, struggling for control.

3.2.1 Technical Requirements

First of all, the position of the pendulum needs to be known to the system. An HTC Vive base station is mounted on the wall inside the room, and emits infrared signals. An HTC Vive tracker is placed on top of the pendulum speaker, in order to continuously collect the absolute position and orientation data of the speaker in the room. The data is transmitted to a computer running a patch in Pure Data (Pd), a real-time graphical programming environment for audio and graphical processing [1]. In Figure 3.3, we show how we have programmed

[1] Pure Data, `https://puredata.info`

3

the sensor interpretation, the rules for the interactive behaviour and the real-time sound synthesis for the surrounding speakers in Pure Data (version 0.50). Furthermore, Pure Data is controlling a software synthesizer in Ableton Live (version 9) for the sound generation of the pendulum speaker.

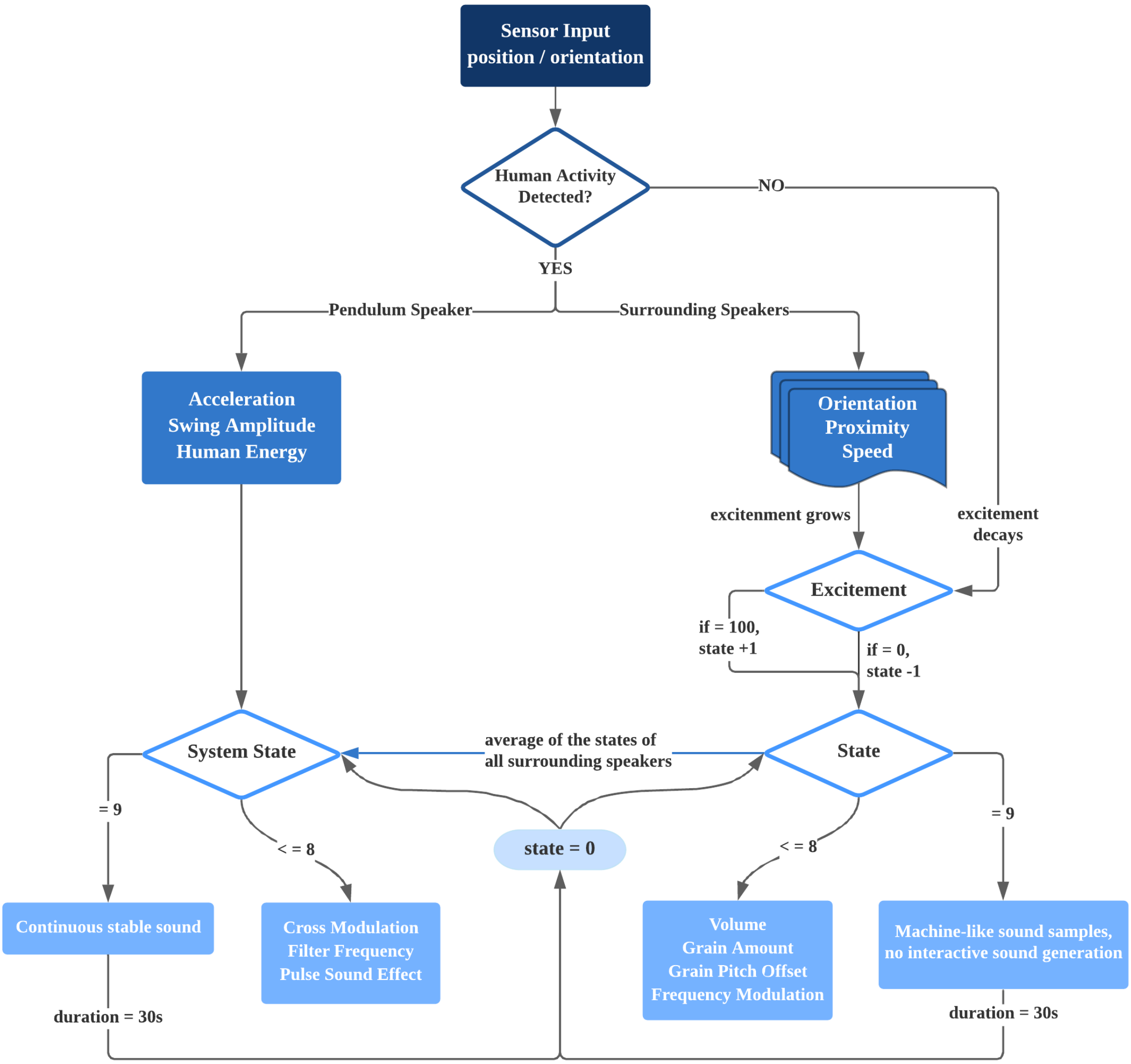


Figure 3.3: Process workflow of software development.

3.2.2 The Pendulum Speaker

From initial and preliminary experiment, we have observed that the pendulum itself has a strong and clear form of, what we would call, natural behaviour. It is a clear inherent physical behaviour, an oscillating motion slowly decaying because of the friction with the air. This makes interacting with the pendulum speaker, and thereby the installation, not so much a process of having full control over the system, but rather a process of using and directing the behaviour of the pendulum. The audience can interact with this behaviour by accelerating, holding and rotating the speaker. It can swing in linear or an ellipsoid orbit. After moving the speaker, it will continue to oscillate corresponding to the new energy applied to it. We decided to try to distinguish the natural motion of the pendulum from the audience interacting with it. In order to do this, we have developed an algorithm that learns the period, phase and amplitude of the swinging behaviour, we then analyze and compare the current phase and position of the pendulum with the predicted natural movement. This way human interruptions of the natural movement can be immediately detected and its energy can be quantified by calculating the amount of deviation. The detected human energy put in to the installation is used to influence the sonic and interactive behaviour of both the pendulum speaker and the surrounding speakers (see Figure 3.3). After interacting with the speaker, the algorithm stores the new swing movement and interprets it as the new natural movement. We believe that this direct form of interaction, realised in this way, gives the audience a feeling that the pendulum is alive and able to respond to the audience's actions. We intend this to result in a playful and physically intensive interactive endeavour.

In order for the pendulum speaker to produce swings that would not move too fast or tilt too high, the length, weight and mounting point of the cable that holds the pendulum are important design parameters. We established a minimum cable length of 3 meters. The mounting point of the cable is placed 0.5 meters above the pendulum's centre of weight, to keep the speaker relatively stable.

3.2.3 The Space

For the first presentation of the installation we chose to use six surrounding speakers placed in a circle around the pendulum (see Figure 3.1). Each of the surrounding speakers is functioning as a separate entity that individually reacts to the movement of the pendulum. It is designed in a way that a surrounding speaker reacts when the pendulum moves towards it. Since the pendulum can swing 360 degrees in the horizontal plane, the arc length or the distance between two neighboring speakers cannot be too wide so that the audience can clearly recognise the interaction between the pendulum and the surrounding speakers. Accordingly, the minimal amount of surrounding speakers is six since 60 degrees is an optimal angle for a small room setup. If the exhibited space becomes larger, it is recommended to add more surrounding speakers.

The installation is not meant to only interact with a single audience member. The swinging movement in space makes it possible for multiple audience members to interact with the installation in the same session. In that case, the audience does not only interact with the speaker but also interacts with each other through the installation. Furthermore, the audience can play different roles and alternate between engaging with the installation or just observing the progression (See Figure 3.5).

3.2.4 Software Development

One of our goals for the experience of the installation was to give the audience the feeling that they are interacting with a system that has a form of autonomous behaviour. The installation was designed to noticeably react to the audience, but also have a certain amount of unpredictability in how it will react. Furthermore, in order to motivate the audience to interact with the installation for longer periods of time, we chose to let the behaviour evolve as a result of the amount of energy that the audience puts into the installation.

Excitement and State

In our system, each of the surrounding speakers forms a separate entity that produces its own characteristic sound. The character of the sound is determined by calculating two main features for each speaker: ‘excitement’ and ‘state’.

$$Speaker\#\,[E, S] = [0 - 100, 0 - 10]$$

These features were implemented in order to achieve an evolving form of interactive behaviour. The state determines both the character of the sounds that are produced and how the speaker reacts to the movement of the pendulum.

The level of excitement ranges from 0 to 100. It is a parameter to describe how 'excited' a speaker is within its current state. It is continuously updated by an algorithm that uses:

1) audience interaction — how much energy has been put on the pendulum speaker, which is measured by the acceleration differentiated from natural movement;

2) the proximity of the pendulum to the speaker;

3) the duration the pendulum is within a certain proximity of the speaker;

4) the speed of the pendulum;

5) the level of excitement of its neighbouring speakers.

The excitement level is calculated for each speaker separately. When a speaker's level of excitement reaches 100, it shifts to the next state. When its level of excitement decreases to 0, it falls back to the previous state. There are 10 states in total, starting from 0. Each state has its own mapping strategies for sound production (cf. section 3.3). They are designed in such way so that the audience perceives a clear change in sound and interactive behaviour and gets challenged by the new interactive behaviour.

We have added some additional rules to the state changes. A speaker can only switch to a new state when the state difference between the speaker and its neighbouring speakers is less than "3". Otherwise, it will not change state and influence the neighbouring speaker's level of excitement instead and wait for it to get to a higher state. Due to these rules, the system as a whole evolves as a result of the individual speakers' behaviour.

The pendulum speaker also has its state. It is determined by, what we refer to as the system state. The system state is the average of the states of all surrounding speakers. An important exception to these general rules is that when the system state equals to "8", all of the speakers' states will shift to "9 - chaos". This state lasts 30 seconds as a clear stage and builds up to a point where there is no return, because the feedback mechanisms in the system drive it into a state of uncontrollable chaos where all speakers stop being influenced by the audience (cf. QRcode 3.2). We will discuss this special state in detail in Section 3.3.

QRcode 3.2

Increase and Decrease of Excitement

The level of excitement increases while the pendulum is close to a surrounding speaker and it slowly decays while the pendulum is further away from it, this is modelled with parameters of growth rate and decay rate. The growth rate of the excitement varies with the amount of audience interaction. More interaction results in a higher growth rate. When the pendulum is following its natural movement the growth rate will start to decrease and the speakers start 'cooling down'. Each surrounding speaker has its own growth rate related to the audience interaction. We expect that the implementation of growth rate would add a responsive nature to the system. We have decided to make the decay rate increase once a surrounding speaker reaches state 4, which means its level of excitement will decline faster and its state will easier fall back to its previous state. Thereby it becomes harder for a surrounding speaker to reach a state higher than 4, especially when there is no continuous human activity detected by the pendulum. We assume this rule would help to keep the audience interacting with the installation. To avoid that the states would alternate too fast, a minimum time that a state lasts has been defined.

3.3 Sonification Design for Motion Data

The participation experience is largely depending on the sound design, i.e. adjective element, as this comprises the syntactic level of the dialogue. We do not want the pendulum speaker to only act as an interface for triggering sounds in the surrounding speakers but intend to create a responsive performance model (cf. definition 2.15) for its interactive and expressive behaviour. In other studies, for example, Livingstone and Miranda introduced the term 'responsive sound environment' as *"a system that regenerates a soundscape dynamically by mapping 'known' gestures to influence diffusion and spatialization of sound objects created from evolving data"* (Livingstone & Miranda, 2004). In order to achieve the responsiveness, the surrounding speakers not only react to the movement of the pendulum speaker but the pendulum speaker also expresses its own movement in its sound and clearly reacts to people touching and moving the speaker.

To that end, we have chosen to use two different sound synthesis techniques to make a clear sonic distinction between the pendulum and the surrounding speakers. The pendulum generates machine-like (i.e., low to mid frequencies) sounds, whereas higher frequency sounds are generated from the surrounding speakers. PMSon (cf. definition 4.3) and MBS (cf. definition 4.4) are used to interpret and translate the input data into sound. These mappings make the installation react both directly and indirectly to the interaction with the audience. This gives the audience a sense of control, but at the same time makes the sounds resulting from the interaction somewhat unpredictable. Meanwhile, the sound results of the installation can develop from calm and peaceful to chaotic and aggressive. This was chosen to make the audience perceive the installation as beautiful and calming when handled with care, but also dangerous and distressing when handled aggressively.

3.3.1 Sound from the Pendulum Speaker (PMSon)

We use U-he Diva, a virtual analogue synthesizer in Ableton Live 9, to generate the sound for the pendulum speaker [2]. The control parameters are calculated in Pure Data and sent to Ableton Live via MIDI. The machine-like sound is produced by two oscillators passing through a voltage-controlled filter (VCF). Using cross-modulation, a sawtooth oscillator and a sine wave oscillator modulate each other's frequencies. The VCF is a low-pass filter, that filters the sounds of both oscillators.

As we have discussed in Chapter 2, *SoundBounce* (Dahl & Wang, 2010) successfully used a physical metaphor and PMSon design to realize an intuitive interaction with sound. Similarly, we use a metaphor of a mechanical machine and parameter mapping approach (PMSon) for the sound design of the pendulum speaker, in order to provide an alternative of direct contribution (cf. definition 2.8) that is intuitive to recognize and understand. The control parameters for the sound synthesis are derived from the pendulum's own physical behaviour. To that end, the amount of human activity is mapped to the amount of the frequency modulation and the filter frequency (cf. Figure 3.4). The audience can 'power on' this machine by putting energy into the pendulum. The pendulum's position is used to play a single midi note that gets triggered every time the pendulum travels a specific distance in space. Furthermore, the amplitude of the pendulum swing controls the velocity of each midi note. When the pendulum

[2] Diva Homepage, `https://u-he.com/products/diva`

swings, it generates pulsed sound effects. The linear acceleration of the pendulum is mapped to cross-modulation between the oscillators, to make it sound like a machine engine that is operating and accelerating. Meanwhile, the rotation rate or the angular speed of the speaker is mapped to the pitch of the synthesised sound. The faster the pendulum rotates, the higher the sound. In this way, the amount of human activity is used to create direct auditory feedback when the audience interacts with the pendulum speaker. The more energy the audience is trying to put into the pendulum, the more active and powerful the machine will be, and the more dynamic the sound will be. The parameters gradually decline again when no one touches it. The state of the pendulum speaker is used to make it sound more aggressive. When the pendulum speaker reaches state 9, it stops triggering midi notes but generates a continuous and stable sound. The cross-modulation and low-pass filter are removed and the pitch goes much lower. The machine turns out to be 'over-excited', and cannot be controlled or influenced by the audience any more.

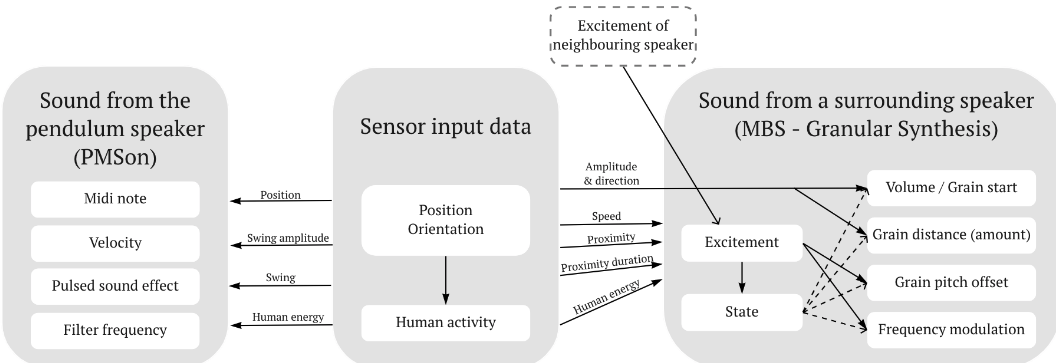


Figure 3.4: Mappings between the input data and the sound parameters.

3.3.2 Sounds from the Surrounding Speakers (MBS)

In contrast to the synthetically generated sounds from the pendulum speaker, the surrounding speakers produce a more natural sound. The sounds are generated using a granular synthesizer built in Pd. Each of the speakers have their own individual synthesizer that uses the same sound sample but with a pre-edited different pitch. The original sample is a recorded hit of a bell (cf. QRcode 3.3).

QRcode 3.3

When the pendulum hangs exactly at its equilibrium point in the centre of the space, the surrounding speakers will not generate any sound at state 0. But when the pendulum moves towards one of the surrounding speakers the sample of that speaker is played in full length. It sounds like the audience is using the pendulum speaker to hit the surrounding speakers, and 'awake' them.

As the interaction progresses, the granular synthesizer is used as a polyphonic sample playback engine in a later state.The distance between the pendulum and each of the surrounding speakers is mapped to the grain distance which sets the rate at which the grains are triggered and results in overlapping grains with a variable density. Currently, up to 100 overlapping grains can be generated resulting in dynamic and rich sonic textures. We have chosen this approach because we believe that this behaviour makes it intuitive for the audience to perceive what kind of effect the pendulum speaker has on each of the surrounding speakers. Next to that, the distance value is also mapped to the start point of each grain player. There is a clear and loud hit at the beginning of the sample. From observations we found that using the start point was an optimal parameter as opposed to the use of volume control since it applies the natural decay of the sound. The closer the pendulum moves towards a surrounding speaker, the louder sound it produces. This behaviour can be easily understood by the audience and it is intended to help them to understand the behaviour of the implemented excitement.

We have applied a model-based sonification (MBS) approach to implement different mapping strategies for the different states, in order to create distinct sonic characteristics for each state. Initially we implemented 5 states. As the state of a speaker increased, the sound transforms from stable harmonic tones into abstract and unrecognizable synthetic noise. However, this approach resulted in large and sudden changes between the states. Therefore, we decided to implement more states to be able to transform the original sample in a more gradual way. This makes it easier for the audience to perceive changes of the system while navigating through the various states. In order to create more complex dynamics in the playback of the grains, we have decided to add frequency modulation for states above 3 as well as to randomise the start point and pitch within a specific range of each grain. Consequently, the grains create a more complex, and use a wider sonic range. The original sampled sound gets dispersed because the hits are intensified and blurred as the speaker reaches higher states. The sound becomes more and more chaotic as the speaker gets excited. In state 9, all of the surrounding speakers play the full length of the original sample with frequency modulation and repeat at a random interval. The sounds become more machine-like compared to the sound in state 0, and assimilate into the pendulum's synthesised sound. After reaching the highest state the system 'cools down' and needs a little rest before it starts responding again starting in state 0.

3.4 Observations and Discussion

During a three-day exhibition at the NIME 2018 conference (see Figure 3.5), some observations of the audience interacting with the installation were made. We also had informal conversations about the work with some of the visitors. Although we did not use a strictly formal method for reviewing, our observations gave us some preliminary indications of how the audience reacts to, and interacts with the work. We use the same participation journey map (cf. Figure 2.1) to visualize audience experience at four stages (see Figure 3.6). We noticed that, at first, many visitors were mostly observing the installation instead of interacting with it. Some mentioned that 'they were not sure if they were allowed to touch the work'. After interacting with the installation, most of the visitors that we observed independently discovered the different forms of movement that the pendulum speaker reacts to, without the need for specific instructions. This seems to indicate that the basic form of interaction is intuitive and PMSon design may provide a more direct feedback for the audience to understand the interaction.

Most of the visitors also seemed to quickly notice that moving the pendulum towards a surrounding speaker resulted in this speaker reacting by playing a sound (see Figure 3.5b). Some visitors specifically mentioned that the interaction reminded them of handling a bell or wind chimes. Thus, it seems that the direct and noticeable sound results, as reference of direction contribution (cf. definition 2.8), can help the audience understand the interaction and navigate through different types of sound composition.

However, it seemed to not be clear in all to the audience that the sounds of the installation were able to develop from sounding calm to sounding aggressive and that the effects of the audience's interaction with the sounds would then also change. For some visitors, this was due to them handling the pendulum speaker so gently that the installation would always sound calm and not aggressive. Other

(a) Interaction 1: Gentle movement

(b) Interaction 2: Push hard

Figure 3.5: The audience interacting with Băi at NIME 2018 (Faces have been intentionally blurred to protect the privacy and anonymity of the individuals depicted).

(c) Experience for multi participants

(d) A contemporary dancer experimented with the speaker

Figure 3.5: The audience interacting with Băi at NIME 2018 (Faces have been intentionally blurred to protect the privacy and anonymity of the individuals depicted).

visitors did put enough energy in the installation to make it sound aggressive, but seemed to not be fully aware of how their actions altered the sounds. Although we also noticed that with the current setup, visitors needed some explanation before being able to experience the full dynamics and concept of the installation (see Figure 3.6). The design of the pendulum speaker might lack some obvious perceivable affordances. Moreover, a clearer distinction of sound design between different 'states' of the installation might help visitors to discover and understand the installation more easily. The MBS approach can achieve a more complex and indirect sound result (cf. definition 2.9). Lastly, our observations indicated that visitors had quite varying sensations while experiencing the installation. Some visitors avoided close proximity to the pendulum, but to others purposefully stood right under the pendulum to 'get a rush of it swinging right over their head' (see Figure 3.5c). Some visitors experienced the installation while laying on the floor and reported that it was a calming experience to them. This indicates that the audience was able to perceive both sensations of beauty and danger, which we aimed to convey with the installation.

A meta level of participation happened unexpectedly during the exhibition. There were two contemporary dancers who experimented choreography with the pendulum speaker (see Figure 3.5d). They gained understanding the rules of the

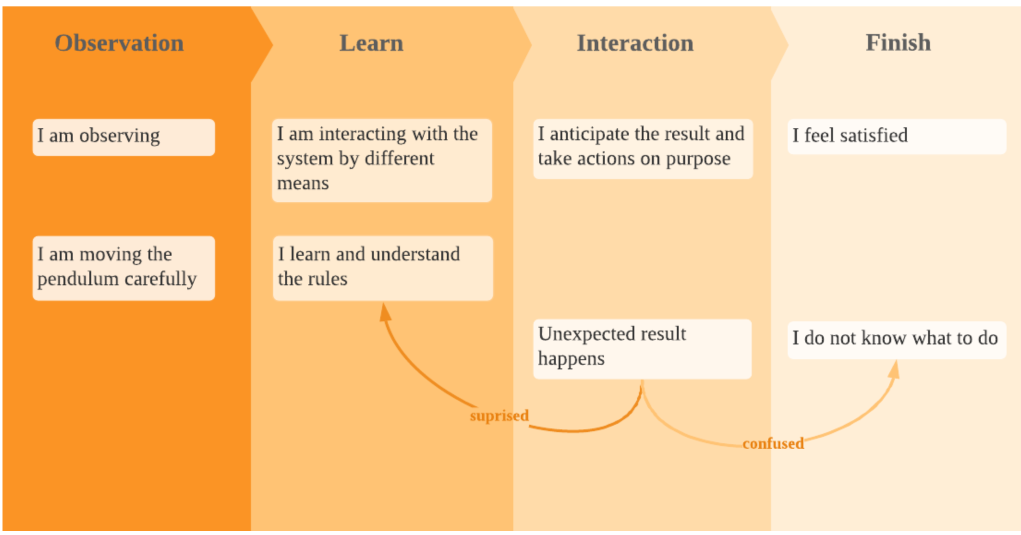


Figure 3.6: A participation journey map of Bǎi.

QRcode 3.4

changes of the sounds quick. Therefore, they moved the pendulum speaker in an intended artistic way and adapted their dance according to the changes in the sounds (cf. QRcode 3.4).

We believe that a good interactive installation should explain itself to the audience and establish a dialogue with the audience. In other words: it should steer the audience in such a way that it reveals its behaviour to the audience. Our initial observations indicate that there is still some room for improvement, but the original concept can be transferred.

3.5 Conclusion

We have developed Băi as a case study for the ideal framework proposed in Chapter 2. A pendulum speaker has been used as an expressive control interface by sensing its position, speed and rotation. Besides performing its own natural movement, the speaker gives both physical and auditory feedback to the sensed input. These aspects serve as a solid foundation for developing a dialogue between the audience and the installation. In this manner, we have achieved a dialogue model at semantic level, where the audience is able to pick up the clues that are provided by the system through sounds.

We have mapped out audience experience in a participation journey map to address how feedback assists the audience in learning and understanding the rules of the interaction (cf. Figure 3.6). We found PMSon approach is easier for the audience to recognise how the actions have a direct impact on the sound of the pendulum speaker (cf. Figure 3.2), which has contributed to the goal of achieving a good visibility of the auditory feedback.

Furthermore, we have developed the installation in a responsive way to challenge the audience when they are adapting their behaviour to the interface and interacting with sounds. While the audience are controlling and interacting with a system that noticeably reacts to them, the system also has its own behaviour and thereby a certain amount of unpredictability. Additionally, the pendulum speaker interacts with the surrounding speakers. While the movement of the pendulum speaker triggers sounds in the surrounding speakers it also influences the excitement and state of them. In return, the state of the pendulum gets affected by their state. We have constructed a dynamic relationship where the states of the surrounding speakers shift up and down, depending on the intensity and duration of the audience's input. On one hand, unexpected auditory results can surprise the audience and achieve continuous dialogue. On the other hand, they might be confusing and bring the interaction to an end (cf. Figure 3.6). In short, MBS approach establishes a stochastic sound system that is able to be developed constantly according to the interaction. While it can achieve a more dynamic and continuous interaction, unexpected auditory results may cause confusion for the audience and hence it takes more time and efforts to get understood (cf. Figure 3.2).

There are several options for the audience to engage in the installation (cf. Figure 3.5). One might just stand alone and observe the installation or walk around and move the pendulum speaker. It is possible for others to join simultaneously and either observe or join the interaction.

While our installation is based on a complex system, we have shown that the responsive interaction method (cf. definition 2.15) is easily understood, especially with the sound results that can be recognized and acquired. Meanwhile, the installation provides different levels of engagement depending on the interaction chosen by the audience. The initial observations of the prototype have given us some indications, but for a good review of the interaction of the audience with the installation, a thorough study would be a good addition to take this project to a next level.

CHAPTER 4

Interactive Auditory Navigation in Molecular Structures: A Case Study Using Multiple Concurrent Sound Sources Representing Atoms

This chapter is based on the following publication:

Liu, D., & van der Heide, E. Interactive auditory navigation in molecular structures of amino acids: A case study using multiple concurrent sound sources representing nearby atoms. In *Proceedings of the 25th International Conference on Auditory Display, ICAD 2019*, (pp. 140–156), Newcastle, UK.

Liu, D., van der Heide, E., & Verbeek, F. J. Design and evaluation for the sonification of molecular structures using multiple concurrently sounding sources. (Publication in preparation)

4.1 Introduction

In Chapter 3, we studied a stochastic interactive system, i.e. Băi, using dynamic data. The installation was relatively hard to completely and instantaneously comprehend. Băi achieves a stochastic interactive system and Chapter 3 gives a detailed example of how it has been developed, however, the interactive sonification design is too complicated and the dialogue (cf. definition 2.1) itself is hard to be evaluated in a good manner. As a consequence we decided to take one step back and investigate the dialogue in a more realistic situation which can be evaluated accordingly. In this chapter, we will introduce a deterministic interactive sonification system designed for a single participant. We investigate the representation of data through the use of sounds, enabling the perception and comprehension of information through auditory cues. We have selected data for sonification that are relatively simple, have a spatial structure and that can be mentalized. To that end we have chosen to work with organic molecules, and more specific amino acids. Particular sounds are designed to represent the type and position of the atoms as they surround us. It is accordingly important that the sonification is easy to learn and understand in an intuitive way (cf. definition 3.3, RQ3). The research described in this chapter is to see how sonification can work in an interactive learning environment, and to find the right design for such sonification that can then be evaluated further.

In the context of auditory display and sonification, there are several common approaches such as earcons, auditory icons, parameter mapping sonification (PMSon) and model-based sonification (MBS) (Hermann et al., 2011). All of these approaches are rooted in the functioning of the human auditory system, which derives three auditory dimensions that are commonly used in auditory display: loudness, pitch and timbre (Neuhoff, 2011). With these primary features, humans are able to separate and identify different sound sources, each with their own characteristics. Additionally, Carlie showed that human auditory system is sensitive to differences in the duration of a sound longer than 10ms, generally considered the smallest detectable change increases with the duration of the sounds (Carlile, 2011). This brought us to the idea that duration could therefore also be used as a parameter for sonification.

Earcons

Definition 4.1 *Short structured auditory messages that can be used to effectively communicate information in a human-computer interface* (Hermann et al., 2011, pp. 358).

Auditory Icons

Definition 4.2 *Sounds are likely to be familiar to users from their everyday life. They represent objects and events in applications* (Hermann et al., 2011, pp. 326).

Parameter Mapping Sonification (PMSon)

Definition 4.3 *PMSon involves the association of information with auditory parameters for the purpose of data display* (Hermann et al., 2011, pp. 363).

Model-Based Sonification (MBS)

Definition 4.4 *A general framework or paradigm for how to define, design and implement specific, task-oriented sonification techniques* (Hermann et al., 2011, pp. 403).

While auditory icons (cf definition 4.2) are meant to represent events directly, earcons (cf definition 4.1) are synthesized sounds which require a learning process to relate the indirect sound to a specific meaning. When a continuous data stream is involved, it is probably more effective to use PMSon (cf definition 4.3) with predetermined relations between the chosen auditory features and the information the data contains. Differently, MBS (cf definition 4.4) often uses a dynamic model that can include interaction, and utilizes sound to help to analyze a specific data task. We have found in previous chapters that PMSon could provide a direct auditory feedback which works well in the context of audience participation. It was also intuitive to learn and understand.

Due to the fact that molecular structures possess a spatial organization, sonification is considered to be a potentially effective approach for representing them in three-dimensional space, rather than relying solely on a visual representation. For example, the spatial arrangement of atoms in a molecule can be represented as specific sounds based on their location in a structure. Moreover, incorporating sonification of a specific area surrounding an atom and enabling navigation through the structure can help manage the complexity of multiple sound objects occurring simultaneously. We intend to empower listeners to mentally perceive and comprehend the arrangements of atoms in a spacial context, and facilitate a cognitive understanding of the molecular structures through auditory cues.

In our daily lives we are used to navigate through sound environments consisting of multiple sources that not only indicate their positions but also communicate information to us. In laboratory environments, listeners are often presented with rather simple auditory stimuli and listening tasks in order to learn more about our spatial perception. Many studies investigated the localization of diverse sound stimuli in the form of single sound sources positioned at various azimuths and elevations (Stevens & Newman, 1936; Hartmann, 1983; Lokki et al., 2000; Letowski & Letowski, 2012). Relatively few studies, however, focused on our ability to localize two or more concurrent sound resources (Divenyi & Oliver, 1989; Brungart et al., 2005). In this chapter, we illustrate and discuss the approach we have taken to develop an interactive sonification system using multiple sound sources that are spatialized in the horizontal plane around the listener. We propose using a simple four-speaker setup in which the positions of the speakers correspond to the directions of the sound sources (see Figure 4.1). As a starting point we are using amino acid molecules. We investigate how we can sonify the structural formula of amino acids. These molecules are relatively easy and are well-known by life-sciences researchers. From our experience, in the future, we aim to extend our work to other structures, such as RNA, including folding and amino acids sequences.

Our ability to perceive the direction of a sound and estimate the origin of a sound is referred to as sound localization. This can work through a process known as binaural hearing. In horizontal plane, our localization relies on a combination of multiple acoustic cues: a) interaural time/phase differences (ITD/IPD), b) interaural intensity differences (IID) and c) the spectral shape (Popper, Fay, & Popper, 2005). An enormous amount of research has been conducted on spatial hearing and the ability of a human to localize sound, both using headphones, as well as in free-field setups with loudspeakers. Stevens and Newman conducted experiments in the open air, already in 1936. Sounds were produced by a speaker which could be moved noiselessly in a circular orbit in the horizontal plane. They concluded that noise was localized more easily than any of the pure tones (Stevens & Newman, 1936). Later, Hartmann tested and compared the performance of localizing continuous pure sine tones, broadband noise and complex signals with loudspeakers in a room. The result indicated that azimuth judgement became more precise when the spectral density, i.e. the frequency content, of the sound became richer and more complex (Hartmann, 1983). In 2000, Lokki *et al.* did an auditory navigation experiment in which the subjects were asked to move in a virtual space with arrow keys of a keyboard and find a point-shaped sound source with a random-position (Lokki et al., 2000). For this study, the headphone was used as the sound reproduction equipment. They tested three different factors: a) audio stimuli with different spectra including pink noise, artificial flute sound and recorded anechoic guitar sound, b) different panning methods for the positioning of the sound, and c) different acoustical conditions: direct sound, combined with early reflections, combined with reverb. The results proved that noise is the easiest stimulus to localize, and reverberation complicates the navigation. Letowski *et al.* pointed out that sound sources producing impulse sounds (e.g. firearms) are easier to be localized than sources emitting continuous or slowly rising long tones in closed spaces (rooms) (Letowski & Letowski, 2012). These studies have investigated different aspects that may affect the localization accuracy of single sound sources. On the other hand, Brungart et al. conducted an experiment in which 14 different continuous, but independent, noise sources were turned on in a sequence within a geodesic sphere consisting of 277 speakers (Brungart et al., 2005). Each time when a new source was added, the listener was asked to localize it. They found that localization accuracy was modestly better for the sounds with rapid onsets than 1-second ramp onsets. Additionally, accuracy declined as the number of sources increased but was still higher than expected on the basis of chance when all 14 sources were on.

In our study, we let the sounds represent the type and position of the atoms, i.e. the spatial organization of the molecule, as they surround us. In this way we explore a molecule in which data can be perceived and comprehended through the representation of sounds. It is accordingly important that the sonification is easy to learn and understand in an intuitive way (cf. definition 3.3, RQ3). The research described in this chapter is to find the right design for such sonification that can then be tested further.

In order to be able to localize and identify the multiple surrounding atoms as fast as possible, our considerations and choices for the sound design were influenced by the features mentioned above. We will explain our design choices in detail in Section 4.3. Binaural recording examples of different sonification designs are presented through QRcodes, which can be scanned with a phone to listen to. For optimal experience, it is recommended to use a stereo device, such as headphones, to fully perceive the immersive binaural recording effect.

4.2 Interaction Design

The visual field of the human eye has a limited arc while sounds is perceived omnidirectional. Sounds could therefore reveal the existence of something in space that is otherwise difficult to be observed. Although We are very much attracted to the three-dimensional structures of proteins, especially the folded parts where amino acids interact with each other. The initial focus is on simpler molecules i.e. the family of amino-acids. This progression allows for a step-by-step exploration of molecular structures, starting with foundational elements before moving on to more complex entities.

The aim of our research is to sonify multiple surrounding objects simultaneously in the horizontal plane, and to test whether they can be perceived, localized and identified by means of interactive navigation. We started with exploring the structural formulas of different amino acids in two dimensional schematics. Unlike written chemical formulas, the structural formulas provide a geometric representation of the molecular structure. To simplify the localization task, our first step has been to transform the formulas into flat graphical ones with identical bond angles of either 90 or 180 degrees, and identical bond lengths (see Figure 4.7). We are aware that this is an extreme simplification of the actual structure but it simplifies the sound spatialization in such a way that the speakers always correspond to the actual directions of the sound sources. It relates more to how a molecule is drawn on paper than to its spatial three-dimensional shape.

4.2.1 Speaker Setup

Different from the common quadraphonic speaker setup, we place the four speakers around us from the front, left, back and right (see Figure 4.1). We have specifically chosen to make the speaker positions correspond to the location (or direction) of the intended sonified atoms.

It is not necessary to create a *phantom sound source* (cf. definition 4.5) in between the speakers and thereby we avoid potential negative effects of spatialization techniques. Such negative effects were found in previous research when we compared the localization performance for spatialized sound sources with both quadraphonic and octophonic speaker setups (Liu, 2016). It was concluded that the sound virtually positioned between two speakers is difficult to be perceived. During the experiment, some participants felt that sounds from 'middle' were sometimes missing (Liu, 2016).

Phantom sound source

Definition 4.5 *A sound source is perceived or localized as a point between two speakers.*

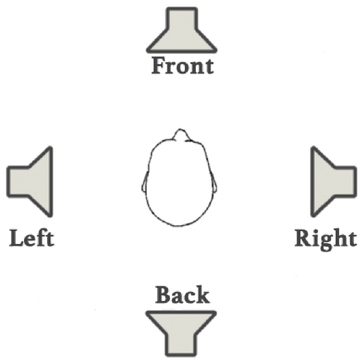


Figure 4.1: Positions of four speakers setup.

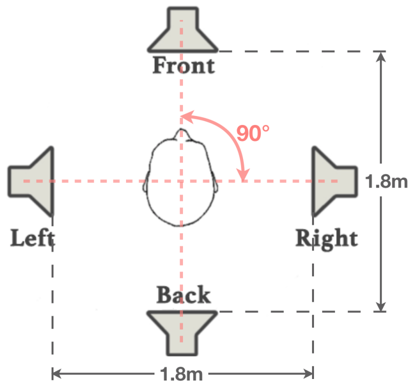


Figure 4.2: Implementation of the four speakers setup.

4

In this design, we sonify the atoms that are connected to a certain carbon atom. This atom is then virtually positioned at the center of the speakers and it will not be audible. Thus it is possible to navigate through over the network of carbon atoms. The navigation method will be described in the following section. The detailed implementation of the speaker setup can be viewed in Figures 4.2 and 4.3. All four speakers are Apart SDQ5P[1] speakers. This speaker setup was the starting point for the design and has been used for the experiments in Chapter 5 & 6.

4.2.2 Interactive Navigation of Structural Formulas

In the past decades, structural biology developed into dealing with the molecular structure of biological macromolecules, like proteins, made up of amino acids or DNA/RNA built from nucleic acids. Atoms are organized in a complex ordered 3D manner and thus form a macromolecule. Grond *et al.* developed SUMO, an

[1]APART SDQ5P is a stereo loudspeaker set. The active speaker is equipped with a stereo 2 x 30 watt amplifier (Link to technical specifications).

Figure 4.3: Speaker setup used in developing the sound design and doing experiments.

open source software environment to sonify chemical structure data contained in PDB files[2]. They implemented acoustic signatures for each amino acid, where different amino acids had different sounds, and parameterized earcons (cf. definition 4.1) were used to distinguish pairwise distances and conformation differences of amino acids (Grand & Dall Antonia, 2008). SUMO shows how sonification can be complementary to a visual display of macromolecules. Two years later, Grond *et al.* combined visualization, sonification and interaction in their application to represent the possible secondary structures of an RNA sequence. The application was designed to turn RNA structures into auditory timbre gestalts according to the shape classes they belong to, on the different abstraction levels (Grond et al., 2010). Thereby, it became possible for the users to quickly compare structures based on their sonic representation. Additionally, the users were able to learn the meaning of the sound by selecting the visual pieces and playing back the corresponding sound. Compared with sonifying the structures as a whole part in (Grand & Dall Antonia, 2008), such interactions provide an interesting and effective way for the users to discern the meaning of the sounds and thus perceive the chemical structure of molecules.

In previous studies, we have used sound to enhance the existing structural visualization of static data. Is it conceivable for the listeners to follow the structures when the visuals are removed? What kind of method could help the listeners to learn the meaning of the sounds when there are multiple concurrent sounds? Previously we investigated navigation in a virtual environment. Direct environment was divided into an 8-connected grid encompassing the avatar (cf. Figure 4.4) that was solely represented by sound using the arrow keys on the keyboard. (Liu, 2016). The participants were able to navigate in an audio-based maze.Sound samples of bird and water were used to indicate obstacles that were not allowed to pass. Most of participants did manage to localize surrounding sound sources and thereby could find a way

Figure 4.4: An example of surroundings in the audio maze.

[2]PDB is a standardized file format saving macromolecular structure data, which contains the positions in x/y/z of all atoms belonging to the corresponded molecule and other relevant information.

4

out of the maze. The sound sources around the avatar changed smoothly as a feedback of movement. Additionally, such forms of interactive navigation could assist the participant to perceive the representation of surroundings with just sounds (Liu, 2016).

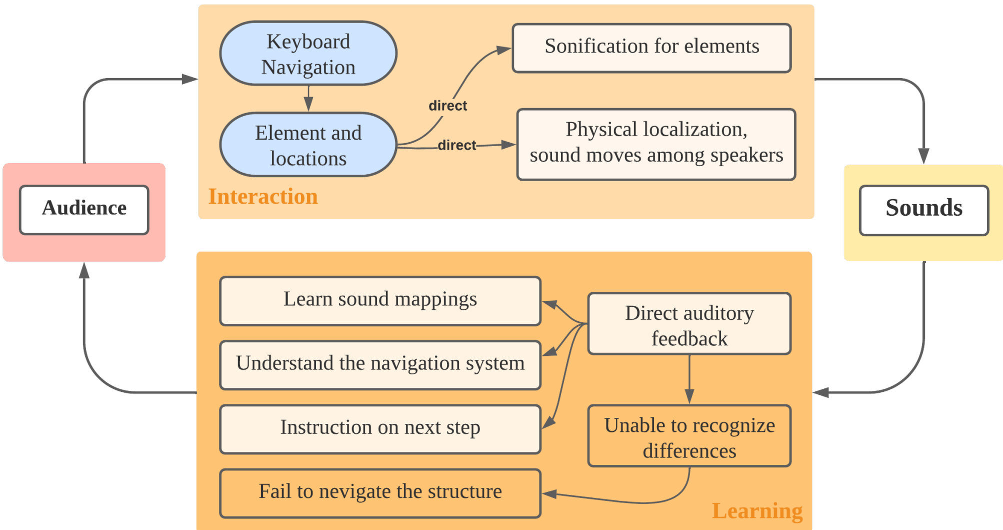


Figure 4.5: Ideal framework for interactive navigation in a molecular structure. The color code refers to the elements of the dialogue model: subject-audience, verbal-actions, adjective-sounds.

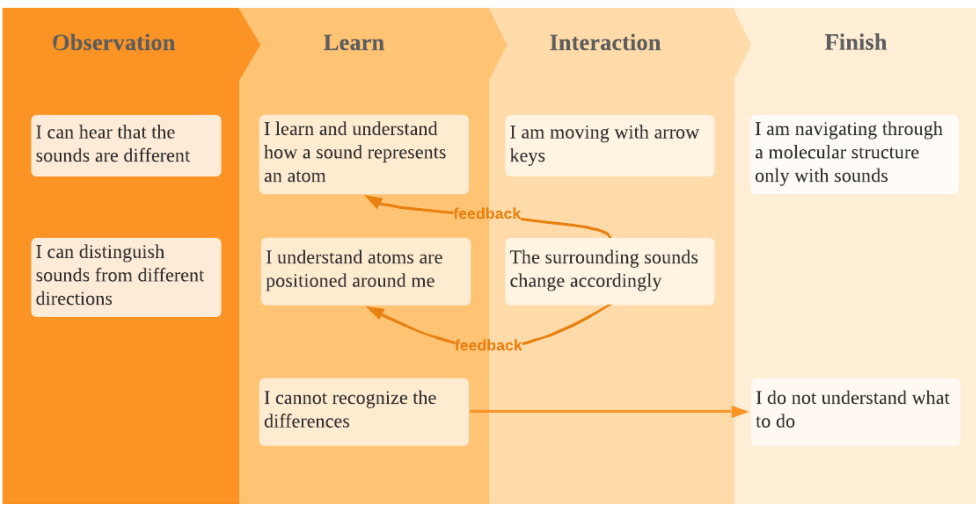


Figure 4.6: Participation journey map of interactive navigation in a molecular structure.

In our sound design, we would like to only use sound to represent the structural formulas of amino acids. We take into consideration that a carbon backbone is an essential part of all amino acids, therefore we would enable the listeners to navigate the structures by moving over the carbon atoms. The navigation task provides opportunities for the listeners to explore the structure step by step. At the same time it allows the listeners to focus on a part of the molecular structure (see Figure 4.5).

In previous chapters, we proposed a four-stage participation journey map to observe and analyze user behaviour (cf. Figure 3.6). Here we utilize the same map to conceptualize how a listener would navigate within an interactive sound environment (see Figure 4.6). By analyzing the journey through the four stages, we are able to identify potential problems and make adjustments to improve the overall effectiveness of the sonification design. Our assumption is that such sound interaction would help the participants to learn the meaning of the sounds and thereby understand the molecular structures.

Navigation Rules

The design aims to prevent listeners from getting lost while navigating through the structures. This requires the listeners to establish a mental model of a molecular structure and obtain understanding of how to navigate through a structure and what actions can be taken in a structure. Such ideation process relies heavily on an intuitive and informative interaction design. Providing feedback can be effective in enabling the mental mapping between the real and virtual world (Alkemade, Verbeek, & Lukosch, 2017).

Our starting point for a design has been the 20 regular amino acids. The carbon backbone of these amino acids consists of a central carbon atom bonded to four groups: an amino group (-NH_2), a carboxyl group (-COOH), a hydrogen atom, and a variable side chain (R-group) that differentiates one amino acid from each other. Therefore, the common elements are carbon (C), hydrogen (H), oxygen (O), nitrogen (N), while other elements like sulphur (S) and selenium (Se) are found in the R groups of specific amino acids. The carbon chain attached to the central carbon atom is referred to as C_1 (see Figure 4.7), which is next to the C_0 from carboxyl group. Starting from the central carbon, there are several carbon atoms connected and forming the skeleton structure. Therefore, we chose for a navigation method where the participant is able to explore the structure by moving from one carbon atom to its neighboring carbon atom(s). The starting point of navigation is C_0. In this case, the participant cannot move to the right, but only to the left where C_1 is located (see Figure 4.7). In our design, a feedback sound will be produced in the form of an alarm sound indicating an illegal move - an attempt to move into a direction that is not a carbon atom. From previous research, we learned that providing feedback sound can assist participants in forming a mental model to remember positions and orientations. This, in turn, proves helpful for navigation in a virtual auditory environment (Liu, 2016).

Concurrent sound sources implementation

There are two approaches to sonify the atoms that are connected to the current carbon position. In our first phase of development, the -NH_2 and -OH groups are exceptions to this rule and are considered as independent groups and sonified as such. In this phase, only the four atoms/groups connected directly to the current carbon position, will be sonified. For example, with reference to Figure 4.7, the listener arrives to the position of C_0, only -OH, =O and C_1 will be audible. As they audibly observe and navigate the molecular structure, they are able to form a mental representation or understanding of the structure. In support of this mental modelling, we introduce the principle of sound layers:

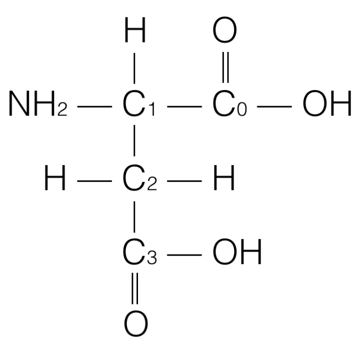


Figure 4.7: The structural formula of Aspartic acid.

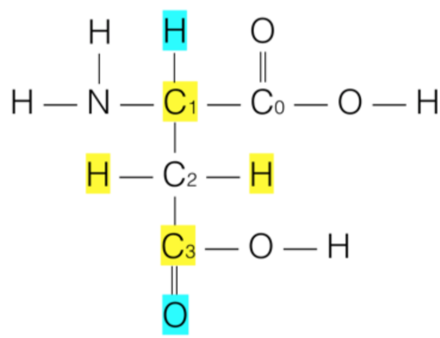


Figure 4.8: The structural formula of Aspartic acid for the larger area sonification.

Layer

Definition 4.6 *A molecule is considered to consist of layers of atoms. First layer atoms (groups) are the ones that directly connected the current carbon position. Second layer atoms are the ones behind the directly connected atoms.*

In a next phase we decided to sonify two layers of atoms. The groups will be decomposed into single atoms (see Figure 4.8). Accordingly, N connected to C_1 and H connected to -O are audible (see Figure 4.8). For example, if the listener moves to C_2, the yellow highlighted atoms are on the first layer and the blue highlighted atoms are on the second layer. Thus, up to eight atoms will be audible at the same time. This initial setup will allow us to assess larger structures in the future.

4.3 Sonification Design

This section describes our sound design choices and how we developed from a preliminary design to a final design. The structure is as follows: first we further review sonification approaches that have specially been used in the fields of chemistry and biology. Next we propose possible sound synthesis setups with expert evaluation in section 4.3.2. Then we discuss various sonification approaches, i.e. sound composition, from the perspective of motivation and pros & cons in section 4.3.1.

In recent decades information sonification in the fields of chemistry and biology has been focusing mostly focusing on DNA sequences and macromolecular structures. Many different designs have been made to sonify and represent objects (e.g. amino acids, proteins, nucleotides) as well as events. For example, a) single note is mapped directly to string data derived from a DNA sequence (Munakata & Hayashi, 1984; Temple, 2017), b) short musical phrases are formed by the Morse code of the amino acids, nucleotides and nucleotide pairs (Shi, Cai, & Chan, 2007; Temple, 2017), c) parameterized earcons help the users to distinguish similar but different structures such as amino acids. Different parameters in a sound synthesizer can be mapped to the different features of an object or event (Grand & Dall Antonia, 2008; Grond et al., 2010; Tek et al., 2012), and d) pre-recorded samples are used as auditory icons to represent events extracted from simulation progress (Rau et al., 2015). In these studies, sonification was often utilized to enhance the visual display of complicated structures. However, it remains unclear whether the listeners are able to recognize and comprehend the sounds without the visual input.

For our approach it is essential that the interacting participants can both identify and localize the atoms purely by means of sound. This brings us to the question how the atoms should sound? There are no metaphorical approaches for atoms that are already familiar to us in daily life and therefor auditory icons are not applicable in our context. Therefore we considered earcons as a way to establish a mapping strategy between the atoms and their sonic representation. Earcons (cf. definition 4.1) can be defined as short, structured musical messages, where different musical properties of sound are associated with different parameters of the data being communicated (Hermann et al., 2011). The relations between the earcons and the atoms are supposed to be understood and acquired by the listeners. The goal of our sound design is to be able to easily recognize and distinguish the different sounds from each other, even if they sound simultaneously.

We will introduce three different synthesis setups and discuss how the design evolves from one to the other.

4.3.1 Sound Synthesis Techniques

We need a specific sonification enabling listeners to build a mental model, thereby understanding the proprieties of a structure. Therefore, we have experimented with different designs regarding how to sonify the different atoms and how to deal with time, considering the rhythmical structures. The aim of our sonification is to represent as many surrounding atoms as possible; this means as many concurrent sounds as possible. In this manner the observer/listener will be able to localize and identify the atoms in as little time as possible.

We use the Pure Data[3] sound programming environment (version 0.50) for both the interactive navigation and the real-time sound synthesis. The clone function in Pure Data enables us to modify different parameters of each synthesizer independently and send the signal/sound to the assigned speaker. We have benefited from previous experience with the function used in Băi (cf. section 3.3.2).

In order to discuss sound synthesis as clear as possible, we first define the

[3]Pure Data, `https://puredata.info`

relevant concepts:

Amplitude

Definition 4.7 *An amplitude represents the loudness of a sound wave.*

Bandpass filter

Definition 4.8 *A bandpass filter attenuates the frequencies above and below a certain passband. The **center frequency** represents the midpoint between the lower and upper cutoff frequency. The **bandwidth** is the difference between the upper and lower cutoff frequencies.*
*The **Q factor** is defined as* **Q = center frequency / bandwidth** (Cipriani & Giri, 2010, pp. 304-308).

Components

Definition 4.9 *A sound waveform can be calculated as the sum of frequency components. Frequency component has independent amplitude and frequency* (cf.Figure 4.9b).

Duration

Definition 4.10 *Duration is the length of time that a signal and thus a sound lasts.*

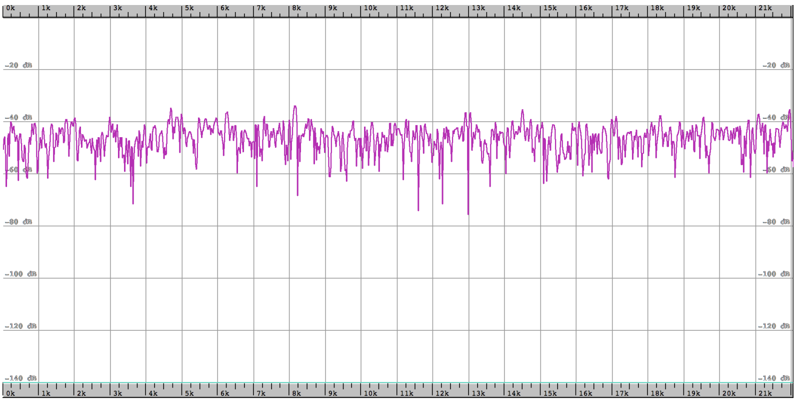
Envelope

Definition 4.11 *An envelope represents macro-level changes in amplitude over time, presented as curves and/or straight line segments that connect the positive peaks found in the sound wave* (Cipriani & Giri, 2010, pp. 24).

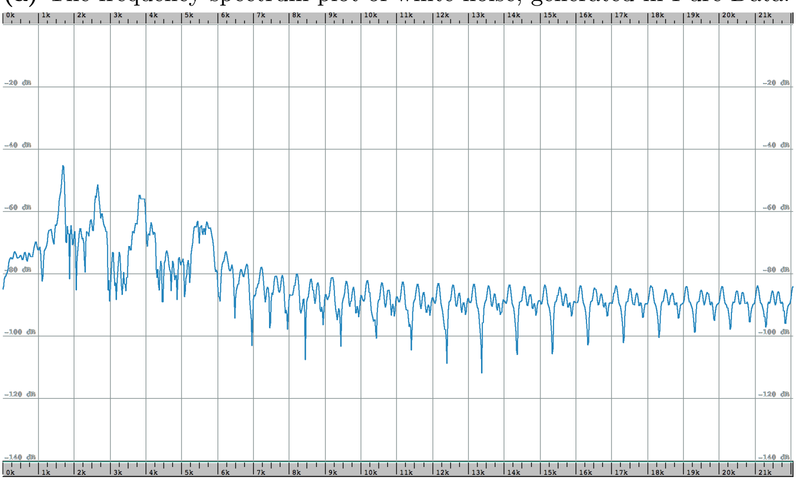
Synthesis setup I: Our initial attempt is to use different drum samples because the timbre of different parts from a drum set (e.g. bass drum, snare drum, hi-hat) can be easily distinguished and these percussion sounds are short and easy to localize. In our first prototype, hydrogen was mapped to closed hi-hat sounds every 400ms, carbon produced snare drum sounds every 1.6s, oxygen and groups generated bass drum sounds every 3.2s. The drum samples, however, might be distracting since the listeners can recognize them and may have problems to relate them with chemical elements. So, in conclusion we rejected this setup as the sounds were not abstract enough.

Synthesis setup II: To ensure a higher level of abstraction and avoid any concrete associations, we decided to explore the use of filtered white noise. By applying different amplitude envelopes, we aim to achieve a more abstract sound.

As we have to characterize different sounds for each element, the center frequency of the bandpass filter is inversely proportional to the atomic mass. The lighter atomic mass an element is, the higher filter frequency. This means that the sound that represents hydrogen has the highest frequency setting and the oxygen sound has a lower filter frequency than the carbon sound. The amplitude envelope



(a) The frequency spectrum plot of white noise, generated in Pure Data.



(b) The frequency spectrum plot of filtered white noise which consists of four frequency components, using four bandpass filters, generated in Pure Data.

Figure 4.9: In the frequency spectrum plot, the horizontal axis represents frequency (Hz), and the vertical axis represents the amplitude of the signal (dB) at each frequency. It can be observed how a certain frequency band can be extracted by a bandpass filter (1.7kHz, 2.6kHz, 3.8kHz and 5.5kHz).

enables different durations and loudness developments for each of the elements. The oxygen sound is the longest because its mass is the heaviest. While the single atoms have a clear and sharp start, the groups have a longer attack time. For example, the frequencies of a single oxygen atom and the -OH group are the same, but -OH has a slower attack time and longer duration at the sustain level. The filtered noise sounds are more abstract than the drum samples. In this design we use pitch as the main feature because the changes are easily perceivable and distinguishable.

Hartman examined a tone with a fundamental frequency of 200Hz and 11 harmonics up to 5800Hz and concluded that the mixing of components within a single critical band plays a significant role in the ability to localize the sound (Hartmann, 1983). We intend to achieve a similar improvement in the ability to localize a sound by using the four frequency components for each of the sounds that we designed.

Synthesis setup III: In order to obtain a richer spectrum in each sound representation of an atom, we added three more bandpass filters to extract four distinct frequency components (see Figure 4.9b). As shown in Figure 4.10, the frequency components made up for hydrogen are much higher, which are 352Hz, 877Hz, 1811Hz, 2941.1Hz. As a group, -OH relates to oxygen and the frequency components of -OH are slightly lower than oxygen. Both of them start with

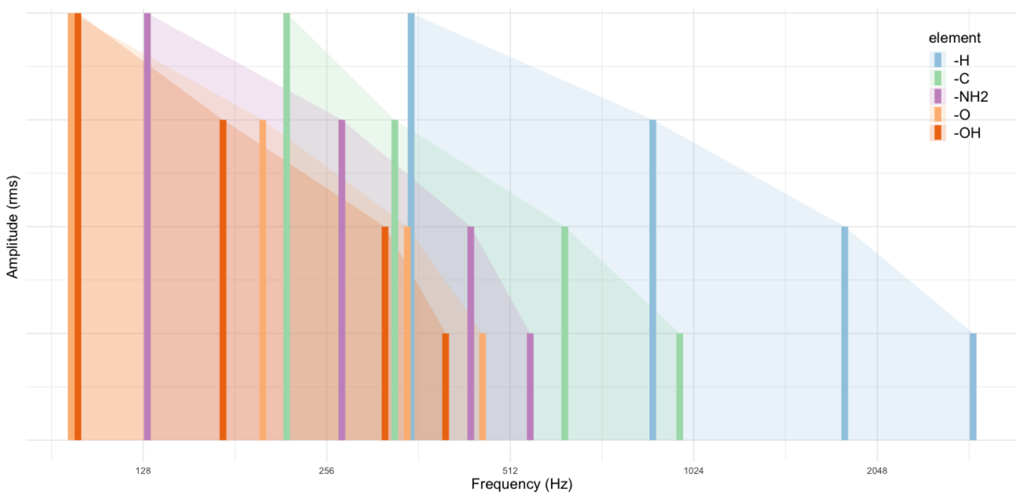


Figure 4.10: Frequency components for each element, synthesis setup 3. The shaded areas indicate regions of overlapping frequencies.

100Hz, then oxygen develops with 201Hz, 350Hz, 461.1Hz and -OH includes 173Hz, 331Hz, 401Hz.

The main problem of this sonification design is that it is hard to separate the sounds from each other when two or more of the same elements are played together. The similar frequency components produced from identical atoms may cause frequency masking (cf. definition 4.12). Moreover, if they are positioned in a row (meaning in the same direction), merging (cf. definition 4.13) may happen. Suppose that the threshold for a sound A is found to be 40 dB SPL. A second sound B is then presented and the threshold of A is measured again. Sound A has a threshold of 52dB when determined in the presence of sound B. The increase in threshold indicates that sound A becomes less audible or more difficult to detect in the presence of sound B. This phenomenon demonstrates how the presence of one sound can impact the perception of another sound. We will discuss this problem and propose *synthesis setup 4* in section 4.3.2.

Frequency masking

Definition 4.12 *When two or more sounds share similar frequency ranges, they can interfere with each other, making it challenging for the listener to distinguish specific sounds. Masking or frequency masking occurs when the threshold of one sound increases in the presence of another sound (Gelfand, 2016).*

Merging

Definition 4.13 *It is a phenomenon where two or more sound sources combine or blend together perceptually, creating the perception of a single unified sound. Merging can happen when sounds have similar spectral characteristics, temporal patterns, or spatial locations.*

4.3.2 Sound Composition

Multiple concurrent sound sources can create a complex and challenging sound environment to be perceived by the listener. When multiple sound sources are present, it might be more difficult to focus on a single sound or distinguish between several different sounds. Brungart *et al.* used a sequential localization process to examine localization accuracy in 360 degrees in a complicated sound environment. Each time, the listeners were asked to localize one newly activated sound source, but the previous played sources would remain. The sound sources were physically localized with 277 independently-addressable speakers which formed a geodesic sphere. Furthermore, each source was separated by 45 degrees from all the other sources. Brungart *et al.* pointed out that this method could avoid that sources originated from same direction, as well as help to reduce proximity-dependent effects of the individual masking (cf definition 4.12) on the target (Brungart et al., 2005).

Our design does involve multiple sound sources played in parallel and thus concurrent. The various frequency components contribute to be able to segregate one object from the others. Nevertheless, in our design there are only four speakers representing four directions, sound sources could be positioned in a row and produced from one same speaker. There are other possible methods to solve the merging problem when sources are concurrent and even played on one speaker.

In this section, we will introduce an evolutionary approach of how we learn from previous design and make changes accordingly. The approach involves creating a population of sonification designs with small variations, and then iteratively refining and improving them. We will investigate the use of the synthesis setups proposed in Section 4.3.1, in combination with different patterns. In order to evaluate and compare the performance of each design, we have defined four criteria depending on the nature of the design problems and goals. All the designs were rated on a 5-point Likert scale (cf. Table 4.1 & Appendix A, Expert review checklist):

1. Learnability, whether a design is easy to understand and learn.
2. Immediacy, whether a design can achieve fast recognition, without too much working memory loaded.
3. Segregation, whether a design can solve problems of merging and overlapping. Listeners are able to segregate different sound sources from a relatively complex auditory scene.
4. Localization, whether a design can assist localization task.

We will present six sonification designs, each accompanied by its underlying motivation, design description, and a discussion of their respective advantages and disadvantages. The evaluation results of these designs are summarized in Table 4.1. All of these designs started with the implementation of atoms on the first layer (cf. definition 4.6). Furthermore we have extended some of the designs and sonified the atoms on the second layer (cf. definition 4.6).

Rhythmical Pattern

In the field of sonification and auditory display, one can choose between melodic or rhythmical patterns. Most research has focused on melodic patterns. There is little relevant research on rhythmical patterns. Rhythmical patterns could be regarded as a sound character to enhance and help the listeners to distinguish and localize multiple sound sources played simultaneously.

Motivation: We would like to investigate whether the sequenced nature is able to help the listeners to distinguish the different elements.

Design I: We divided 4 speakers as 4 beats in a bar, and play a counter-clockwise sequence (front - left - back - right) with a fixed tempo. This way the sounds can be played sequentially [4]. We implemented the *synthesis setup 3* in this design, envelope and duration differences are combined with the bandpass filter groups.

QRcode 4.1

Pros & Cons: This design is a way to solve the problem of the overlapping sounds. However, it takes 2.4 seconds to finish a bar which might be a bit long for the listener to recognize and remember the sounds. It is still possible after several times of repetition but we would like to accelerate the process to achieve an even faster and intuitive recognition of the different sounds in a (near) simultaneous way.

Design II: Besides the envelope and duration differences, we assigned different repetition speeds to different elements. However, the position always determines the beat where the sound starts to play[5]. For example, when the listener positions on C_1 (see Figure 4.7), the hydrogen sound repeats at 600 bpm and synchronous to the first beat of the bar. The sound that represents -NH_2 repeats at 45 bpm is synchronous to the second beat in the bar. The carbon sounds repeat at 80 bpm synchronous to both the third and the forth beat.

QRcode 4.2

Pros & Cons: When all four speakers start to play sounds together, it is clear and direct for the listeners to notice the similarities and dissimilarities among them. One of the disadvantages of this design is that each element has an independent and distinct speed that can affect listeners to perceive different

[4] A binaural recording example of navigating in the structural formula of Aspartic acid with rhythmical pattern, Design 1 (cf QRcode 4.1, scan to listen).

[5] A binaural recording example of navigating in the structural formula of Aspartic acid with rhythmical pattern, Design 2 (cf QRcode 4.2, scan to listen).

tempos at the same time. In addition, the resulting pattern can be chaotic and annoying when there are various elements sonified together.

Bouncing Pattern

Imagine a ball is lifted at a certain height and then released, when it hits a surface it will create a sound, lose some potential energy and bounce into the air again, but lower than the original height. It keeps bouncing until its potential energy is zero and it stops.

Motivation: We intend to give each sound a more constant and independent character. Loops of a bouncing pattern could create a more characteristic pattern for the listeners to identify. They can possibly be compared when concurrently played.

QRcode 4.3

Design III: We consider the atoms as balls, falling from different heights and having various bouncing patterns. The height relates to atomic mass. Like hydrogen falls at a lower height and produces shorter bounces. Each element has a different bouncing speed and duration. A decay envelope is used to control the decrease in bounce period[6].

Pros & Cons: The bouncing pattern is easy to understand and the impact sound at the starting point of each loop is always clear. However, it might be complicated and confusing at some point compared with the previous designs of rhythmic patterns, whereas further bounces quickly speed up and become rather intensive. Another potential problem is that when there are atoms of a same element that generate sounds, the bouncing pattern is also the same. Such bouncing sounds could be mixed up together and challenging for a listener to separate one from the other, even though they are coming from different speakers. Furthermore, this design will sound rather confusing when a larger area of the structure is sonified.

Irregularly Triggered Bandpass Filter Banks (ITBFB)

The bouncing patterns moved us away from regular patterns and brought us to the idea of a granular structure sound, which may create a more abstract sound texture.

[6]A binaural recording example of navigating in the structural formula of Aspartic acid with the bouncing pattern, Design 3 (cf QRcode 4.3, scan to listen).

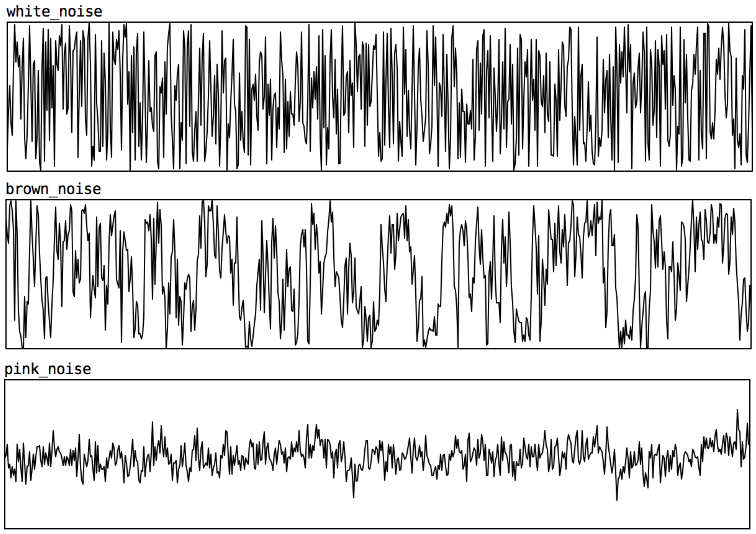


Figure 4.11: Three temporal structures of colored noise, showing amplitude changes over time (generated in Pure Data). The x-axis represents time, while the y-axis represents amplitude. The colored noise exhibits distinct patterns and fluctuations in its amplitude, providing a visual representation of its temporal characteristics

Colored Noise

Definition 4.14 *The colored noise is determined by the power spectrum of noise signal.*
White Noise has the same energy at all frequencies.
Pink Noise has a spectrum that energy decreases as frequencies get higher (3dB per octave).
Brown Noise has a spectrum that energy drops as frequencies get higher (6dB per octave).

Motivation: We aim to create a more continuous but irregular pattern in order to avoid merging (cf definition 4.13) problem that we had in Design 3.

Design IV: We used colored noise (cf. definition 4.14) in combination with a comparator with a variable threshold as a way to generate random impulses with random amplitudes for each of the elements separately. The amplitude (cf. definition 4.7) changes vary a lot from white, pink and brown noise (see Figure 4.11). By choosing between different types of noise varying the threshold we can generate different impulse patterns with different desired densities. According to previous choices, we give the lighter elements an intensive but (light) pattern and the heavier elements and groups a more extensive pattern with a larger range of amplitude changes.[7] Due to the irregular signal impulses, all the sounds have their own non-repetitive structures. This means when two or more identical atoms are represented, they still possess individual irregularities in their structures. We use the impulse patterns as input signals for banks with four bandpass filters that we used before (*sound synthesis III*). Now, even when there are multiple sound sources generated together, the differences will still be recognizable.

QRcode 4.4

Pros & Cons: The irregular structure is experienced as a kind of granular-like texture. This makes it easy to recognize the sounds and the listeners are not required to remember the rhythmical patterns and compare them with each other. Now we can play the different sounds concurrently and they can all be identified simultaneously.

Motivation: We are curious to know if we can sonify even more atoms in parallel by expanding the sonification of the second layer around the carbon atom. Instead of sonifying -OH and -NH_2 as groups, we represent each individual atom on separate layers (see Figure 4.8). Designs V and VI will provide a detailed explanation of how we achieve this expanded sonification.

Design V: In order to enhance the sensation of distance of atoms in the second layer, reverb is probed and employed. The amplitude of the direct sound of the atoms from the second layer is one third of the ones from the first layer while the amount of reverb is the same. When the listener moves to on C_1, C_2 and C_3 are then sonified seperately (see Figure 4.8). The distance determines the loudness so the sound of C_2 is louder than C_3. Moreover, the Q factor (cf. definition 4.8) of the bandpass filter of C_3 is slightly higher than C_2, which results that C_3 has more resonance and becomes less sharp and intensive[8].

QRcode 4.5

Pros & Cons: This design is likely to solve the problem that the more intensive sound mask a less intensive sound. In *synthesis setup 3*, some frequencies were too low or too close to each other, which may have resulted in a negative effect on separation and localization when two layers of sound sources are sonified simultaneously.

[7] A binaural recording example of navigating in the structural formula of Aspartic acid with ITBPFB, Design 4 (cf QRcode 4.4, scan to listen).

[8] A binaural recording example of navigating in the structural formula of Aspartic acid with ITBPFB, Design 5 (cf QRcode 4.5, scan to listen)

Motivation: We aim to enhance the segregation and localization when sonifying two layers of sounds simultaneously. This is intended to make it easier for listeners to differentiate and identify sounds from each layer and perceive their spatial differences.

Design VI: We have retained the irregular structure as it effectively provides immediacy to the sound recognition. Moreover, it simplified the process of remembering specific patterns, allowing listeners to intuitively perceive the differences. However, adjustments were made to the synthesis setup in order to achieve a more distinct and perceptible character for each sound. In this regard, we introduce the *synthesis setup 4* as part of our next iteration.

QRcode 4.6

Synthesis setup IV: We used a fixed interval size between the atoms and expanded the range of filter frequencies used. This adjustment resulted in larger frequency differences between the sounds, making them more distinguishable and aiding in their separation and localization. For example oxygen is increased to 110Hz, nitrogen starts with 220Hz, carbon has 440Hz and hydrogen gets 880Hz. While oxygen and nitrogen remain with a less dense pattern, the resonance of the bandpass filters for these two elements is higher than for hydrogen and carbon[9]. To ensure clear differentiation between identical elements positioned in the same direction, we have given the elements in the second layer a slightly higher pitch. The difference is carefully calibrated to be small enough that it is clearly identified as the same atom but large enough to be able separate the sounds from each other and avoid merging. There is a fixed ratio between two neighboring atoms. For example, if there are three carbon atoms positioned in a row at the same direction, the closest carbon C_1 is made up of 440Hz, 661Hz, 973Hz and 1389Hz and louder than other carbon atoms. The second carbon consists of 484Hz, 727.1Hz, 1072Hz and 1528Hz and the third carbons frequency components also have a 10% increase (see Figure 4.13). However, it remains to be determined through future research what the maximum number of layers is that the listeners can segregate.

Pros & Cons: From expert review, the pitch differences are clear and easy to be recognized in general. Combined with other features, density and reverb, it can help the listeners to separate and localize sound sources from same directions but different layers. However, it is still unknown what the maximum amount of objects is that the listeners can segregate. Additionally, auditory masking should

[9] A binaural recording example of navigating in the structural formula of Aspartic acid with ITBPFB, Design 6 (cf QRcode 4.6, scan to listen).

be considered when there are two or more layers of sound sources are positioned around. As this sonification design seems feasible we arranged an evaluation method to further investigate. This will be addressed in Chapter 6.

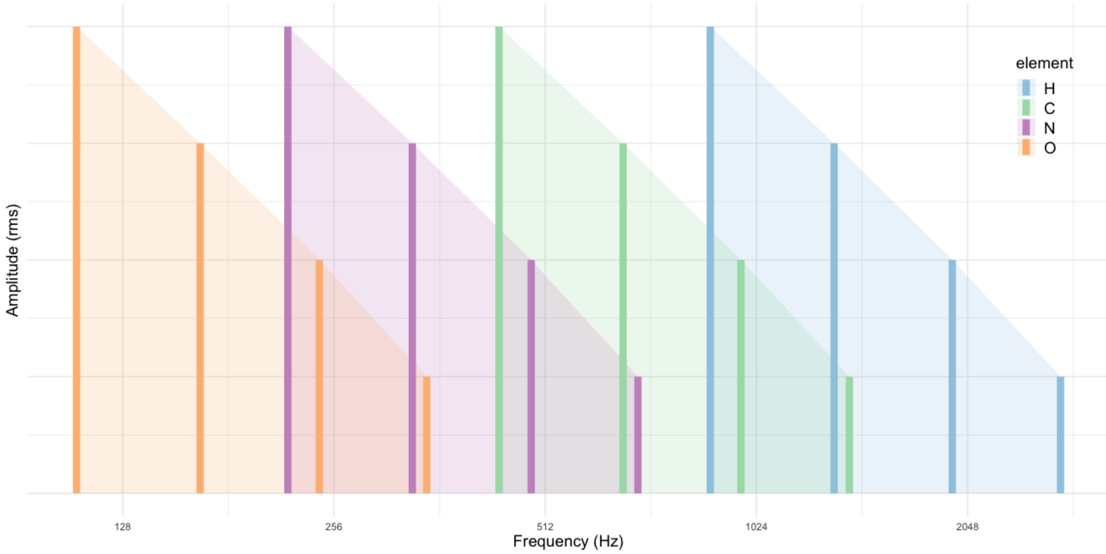


Figure 4.12: Frequency components for each element, synthesis setup 4, with octave separations. The shaded areas indicate regions of overlapping frequencies.

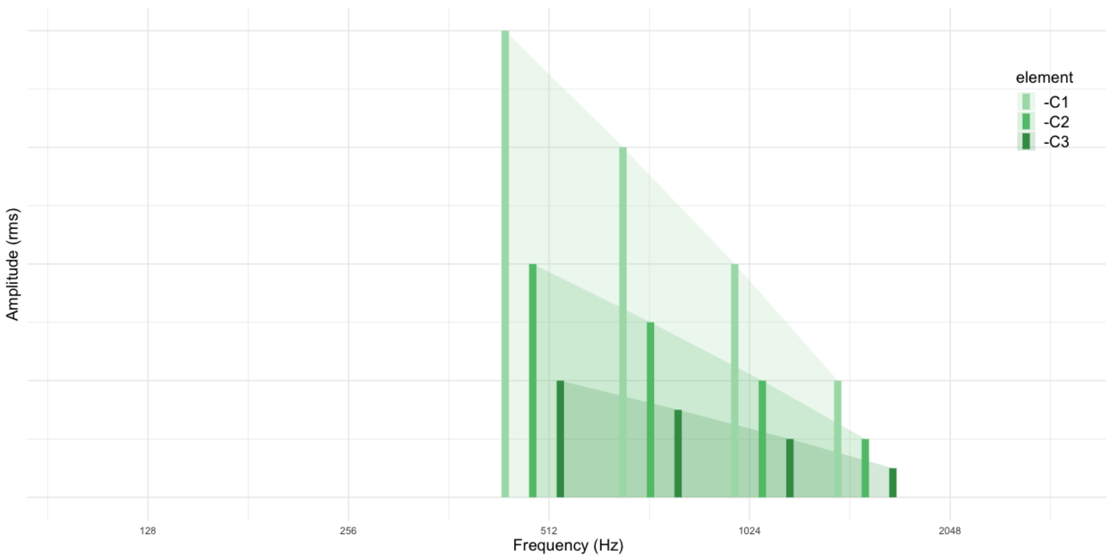


Figure 4.13: Frequency components for each carbon atoms on different layers. The shaded areas indicate regions of overlapping frequencies.

	Learnability	Immediacy	Segregation	Localization
Design I	2.5	1	3.5	3.5
Design II	1.5	2.5	2	3.5
Design III	3	3	2	3
Design IV	4	4	3.5	4
Design V	4	4	3	3.5
Design VI	4	4.5	3.5	4

Table 4.1: All pros & cons analysis from the designs 1-6 are assembled in this Table. It presents the ratings on a 5-point Likert scale, where a score of 5 indicates the highest rating. The ratings were provided by two experts who evaluated the designs (cf. Appendix A, Expert review checklist).

4.4 Conclusion and Discussion

In this chapter, we have discussed several designs to implement a spatial and interactive sonification for chemical structures, as a test object we probed amino acids. We have uncovered that the design and production of sound is a critical element of the dialogue model (cf. Figure 1.1), as it has a significant impact on the listener's understanding of the system (cf. Figure 3.6 & 4.6). The way in which sound is created and designed can greatly influence the listener's perception of the dialogue, i,e, system's responses to their actions. Therefore, a careful consideration of sound design is essential to ensure that the listener can comprehend and better engage with the interactive system.

We started with the concept of earcons (cf. definition 4.1) in order to achieve the immediacy of sound recognition and localization. Unlike conventional earcons, such as time-based melodies or other sequentially played sound samples, in our study we focus on concurrent sounds. We first used fixed sound samples for the rhythmical patterns and then changed to real-time synthesized sound using banks of bandpass filters. While the repeating rhythmical patterns and bouncing patterns may have a shallow learning curve, the irregular impulses allow for a faster and simultaneous recognition of the atoms without a separation period. In our final design (cf. *Design VI*), we combine frequency and irregular density as two main features for the sonification, to help the listeners to identify multiple simultaneous sound sources. By doing this we have expanded our design that started with earcons toward parameter mapping sonification (cf. definition 4.3).

By using the evolutionary design approach, we have optimized previous designs and explored new design possibilities. Our process started by formulating the motivations for the different designs. We have then evaluated the resulting sounds of each design with pros & cons and objectively rate from four criteria of learnability, immediacy, segregation and localization. The results are summarized in Table 4.1; here the higher score for Design VI is clear. This allowed us to refine the sonification designs that would effectively communicate the intended message to the listeners.

Our next step would be to play an even larger area of concurrently sounding atoms. We already established that making light variations in frequency, density and loudness may (partially) solve the merging problem of multiple identical atoms coming from the same direction. The sound changes are regarded as auditory feedback from the interactive navigation, which may influence the localization accuracy and improve the segregation. In addition, it would be possible to realize a richer spectrum while avoiding auditory masking (cf definition 4.12).

All of the sonification designs mentioned above require the listeners to learn from the interaction. Our design is such that through the interaction, listeners may begin to recognize the rules of the mapping, i.e. how a certain sound corresponds to a particular atom. Based on the interactive navigation within the structure, this learning process allows listeners to develop a mental model of the structure that is presented (see Figure 4.6). Overall, we postulate that by actively interacting with the system and learning from the sounds, listeners can build up a relatively comprehensive understanding of the structure. Further experimental investigations (see in Chapter 5 & 6) are considered to evaluate the sound design choices and the assumptions that have been derived from pros & cons in Section 4.3.2. Meanwhile, we will consider to include active head movement in our research (see section 5.2.3 & 6.2.3), which has proven to reduce front/back confusion and improve localization in elevation (Thurlow & Runge, 1967; Kato et al., 2003).

CHAPTER 5

Evaluating the Sonification of Molecular Structures Using Multiple Concurrent Sound Sources: Validation I

This chapter is based on the following publication:

Liu, D., & van der Heide, E. Interactive auditory navigation in molecular structures of amino acids: A case study using multiple concurrent sound sources representing nearby atoms. In *Proceedings of the 25th International Conference on Auditory Display, ICAD 2019*, (pp. 140–156), Newcastle, UK.

Liu, D., & van der Heide, E. Evaluating the spatial sonification of the molecular structures of amino acids using multiple concurrently sounding sources. In *Proceedings of the 26th International Conference on Auditory Display, ICAD 2021*, to appear.

Liu, D., van der Heide, E., & Verbeek, F. J. Design and evaluation for the sonification of molecular structures using multiple concurrently sounding sources. (Publication in preparation)

5.1 Introduction

In Chapter 4 we presented our sound design for molecule sonification explaining how specific sounds are assigned to different types of atoms as well as different spatial locations within the molecule. The pitch of the four sounds corresponds to the atomic weight of each element, the lighter the element, the higher the assigned pitch (see in Figure 5.1).

Unlike time-based melodies or other sequentially played sounds, our design focuses on concurrent sound sources. A combination of elements is played simultaneously and each sound originates from its own specific location, i.e. speaker position. In order to make the elements to be easily and quickly recognizable, every sound has its own irregular amplitude pattern whereby the density depends on the atom type (cf. Design VI). In summary, when two or more identical atoms are being played on the same speaker, they share the same pitch but each atom

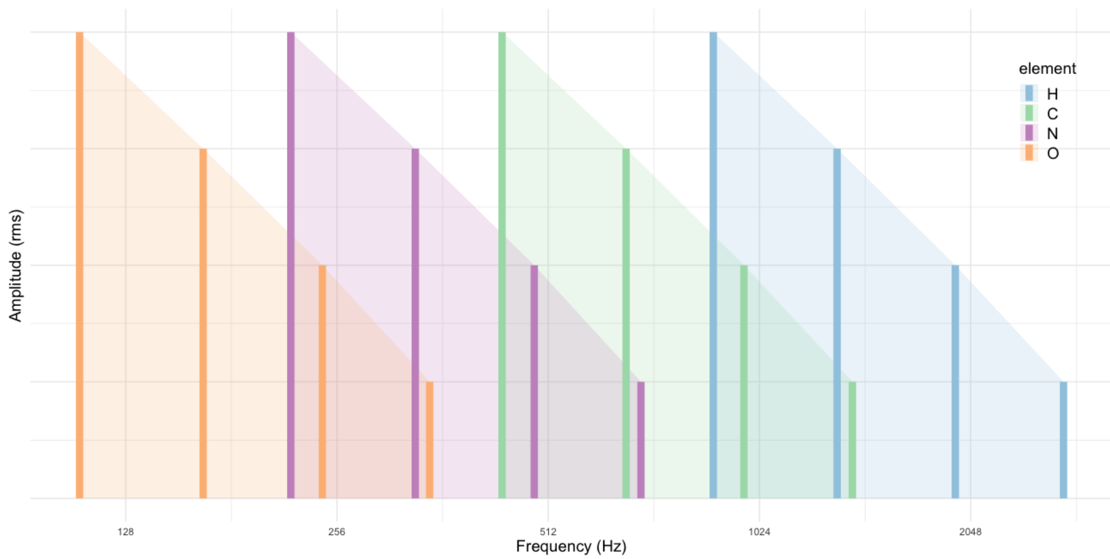


Figure 5.1: Frequency components for different elements. The graph shows the frequency components of four certain sounds representing four elements, highlighted using different colors. The shaded areas indicate regions of overlapping frequencies. The intervals between the components are identical for the each of the four sounds. The ratios are 1 : 1.5 : 2.2 : 3.2. The first (lowest) partial of oxygen has a frequency of 110 Hz, the first partial of nitrogen is 220 Hz, the first partial of carbon is 440 Hz and the first partial of hydrogen is 880 Hz.

has its own irregular, and thereby asynchronous, amplitude pattern. In this manner, we aim to avoid a merged perception (cf. definition 4.13) of two or more identical atoms. We have chosen to give the lighter elements a more dense, and thereby faster, pattern and the heavier elements a less dense, and thereby slower, pattern. From expert review, the assumption is that it is intuitive to associate a faster pattern with a lighter atom.

This chapter is dedicated to the validation of our sound design. We focus on Design VI. The validation itself follows an experimental design that is processed. We will mark the validation in this chapter as Validation 1; this is to discriminate it from the validation in the next chapter (cf. Chapter 6, Validation 2). We here explain the design and implementation. The starting point for the validation is to test two assumptions. These are:

Assumption A *Participants are able to learn and comprehend the sonification design and perform better with practice.*

Assumption B *Our sonification design can achieve immediacy of sound recognition and localization.*

In previous studies, sonification applications have been evaluated; participants were given various tasks during a series of experiments. Ibrahim *et al.* reviewed ten kinds of tasks that were used for measuring usability properties such as effectiveness, efficiency and satisfaction (Ibrahim et al., 2011). One of the ten tasks they described is an identification task, which can be used to investigate the ability of sounds to be uniquely perceived and recognized. In this task different objects or events have to be correctly identified by the subjects with their associated sounds. Accordingly, we decided to involve such identification task in our experiments to investigate the listeners' performance of identification with different combinations of sound sources and matching them with corresponding elements and positions.

Bruce and Walker used a pretest-posttest design to measure outcomes before and after implementing five experimental training conditions and evaluating their impact on sonified graph identification. Participants were randomly assigned one of them. The training conditions were with or without feedback, such as the disclosure of the correct response, guidance of a visual prompt or an interactive presentation with both voice-over and visual explanation. The study showed that practice with feedback may be more effective compared to other scenarios

(Walker & Nees, 2005).

For Validation 1, we divided the experiment into three stages: 1) a pretest, 2) a practicing session with feedback of correct answers, and 3) a posttest similar to the pretest. This way we can evaluate the learnability of our sonification system by comparing the results of the pretest and posttest. According to the conclusions of Bruce and Walker, we would assume that participants would be able to learn and comprehend our approach and perform better with practice and feedback. Calculation of the effect size (cf. definition 5.1) is therefore also necessary, in order to measure the amount of gain when comparing pretest and posttest results (York, 2016).

Effect Size

Definition 5.1 *The effect size is the amount of gain measured in terms of standard deviations if you are comparing pretest and posttest scores.* (York, 2016, pp. 80).

5.2 Experiment Design

The experiment to further evaluate our sonification design as discussed in this chapter is referred to as Validation 1. A pretest-posttest design has been used to investigate the extent to which the participants can learn and remember the mappings between the sounds and elements. The difference in performance between pretest and posttest can provide an indication of the learning rate, i.e. how quickly or effectively an individual learn the sonification over time.

Only a first-layer of sounds is presented in this experiment, consequently, up to four sounds are positioned around the participant at the same time. An important aspect of Validation 1 is to investigate the immediacy - the time it takes to recognize the sonified elements that are surrounding the listener. The irregular structure is experienced as a kind of granular-like texture. This approach does not require participants to remember a concrete sound or a specific rhythmical pattern and compare with each other (cf. section 4.3.2).

5.2.1 Materials

The options available for positioning elements in each direction amount to five, which include four elements (i.e. H, C, N, O) and the option of a zero element. Considering the possibility of choosing one of the five options for each direction, this would result in a total of up to $625(5^4)$ possible combinations. However, due to practical limitations, it was not feasible to implement all of these combinations during the experiment. Instead, we looked through possible combinations of directly connected carbon atoms among the structures of the 20 natural amino acids. From these, we selected 14 specific molecular structures (see Figure 5.2) that were used in the experiment.

QRcode 5.1

QRcode 5.2

QRcode 5.3

QRcode 5.4

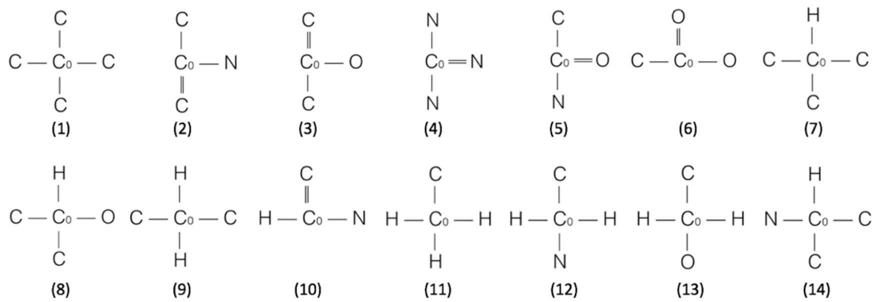


Figure 5.2: 14 structural formulas for Validation 1.

Our design is based on the irregular impulses generated by differently colored noise in combination with a comparator with a variable threshold (cf. Design VI), which results in random impulses. In order to avoid the auditory differences of generating a same element in real-time during the experiment, we decided to use pre-recorded samples of all possible combinations[1]. For the experiment it is important that each participant is exposed to the same sonification of a structure.

We have chosen not to ask the participants to finish the questions as soon as possible, in order to avoid causing anxiety. Therefore, we chose to use two different playback durations: i.e. four seconds and eight seconds. This enables us to compare the listeners' performance between the two different durations.

5.2.2 Software and Hardware

All sounds were synthesized with Pure Data using the clone function. The application was developed on a PC with 32GB RAM and internal High Definition Audio Device, supporting 6-channel output. Sound samples were recorded and then played back in Pure Data (version 0.50).

[1]Recording example of hydrogen, carbon, nitrogen and oxygen (cf QRcode 5.1 to 5.4, scan to listen).

The GUI for the experiment (cf. Figure 5.3) was programmed in Processing (version 3.5.3)[2]. For the statistical analysis we used R (RStudio version 1.2.5)[3] and Microsoft Excel (version 16.38).

5.2.3 Experimental Procedures

The experiment consisted of four phases (cf. Appendix B, instructions).

In phase 1, the participants were introduced to the four different sounds representing four elements H, C, N, O. They were informed that the perceived frequency irregular pattern had been mapped to the atomic weight of each element. They were instructed to press the keys for H, C, N, O on the keyboard to playback the corresponding sounds. Once they felt they were able to recognize the sounds, they proceeded to the pretest. The participants were told that sounds would come from four directions, with up to one sound source on each direction. Additionally, they were allowed to change their head orientation during the experiment.

Phase 2 encompasses the pretest; in this phase, a total of 28 (2×14) recordings were played to the participants. Half of these sound samples had a duration of 4 seconds and the other half had a duration of 8 seconds. The order in which these samples were played back was randomized for each participant. During a

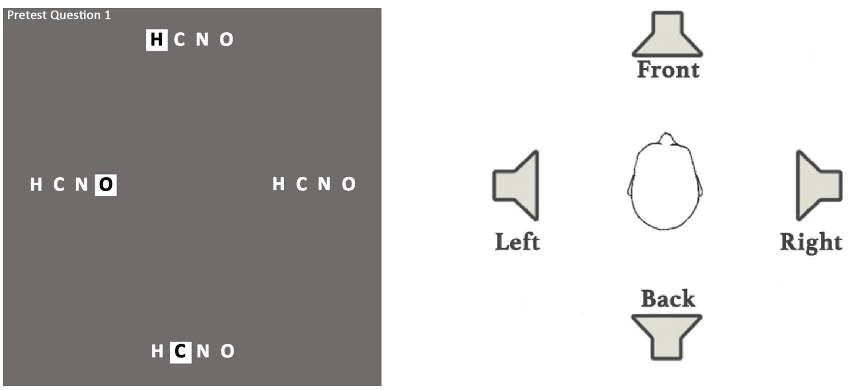


Figure 5.3: A screenshot of the user interface for the participant to indicate the sounds they heard during the experiment of Validation 1. In the user interface, the up and down directions correspond to the front and back speakers, while the left and right directions correspond to the left and right speakers.

[2]https://processing.org/
[3]https://www.rstudio.com/

structure was played the participants were asked to indicate, in a simple screen-based interface, for each direction which element they heard (H, C, N, O or none) originating from that position (see Figure 5.3). In the User Interface, the elements were displayed in the order of molecular weight, from lightest to heaviest (H, C, N, O). Participants selected the corresponding atoms using a mouse and pressed ENTER to proceed to the next structure. If they did not hear any sound from a particular speaker or were unable to identify the sound, they could leave it blank, which would be automatically marked as nothing ("-").

Phase 3 was a training session. 18 training examples were prepared in this part and participants would get feedback upon providing their answers (see Appendix B, training session). The questions were designed in a way to lead the participants to learn and get familiar to the sounds. At the beginning, one element sound was given as a reference so that the participants could compare and recognize different sounds, from two sounds to four sounds. The localization task was added later. In the last six questions, participants were given how many atoms they could hear and were asked to point out their directions and name each atom.

In phase 4, participants took the posttest after completing the training part. The posttest included the same 28 recordings. The order of playback for the recordings was randomized for each participant, ensuring a unique sequence for each individual.

After the posttest, the participants were individually interviewed about their experience and strategy when doing the tests. For example, 1) were the sounds from four directions (equally) clearly heard? 2) how did they identify the element, according to the pitch, the density or both?

The aim of the experiment is to gather and analyze appropriate evidence to either accept or reject Null Hypothesis as stated below:

H0 *There is no significant difference in performance between the pretest and posttest measures.*

H0 *There is no significant difference in performance between the 4-second and 8-second recordings.*

5.3 Experimental Results

In total, 27 participants participated in the experiment; 17 male and 10 female participants. 93% of them were from the age of 20 to 30 years old and 3% were from the age of 31 to 50 years old.

Correctness Rate

We mentioned before that for each of the 14 presented structures in the pretest and the posttest, we recorded the answers given for each of the 4 directions. To calculate a **correctness score** per presented structure, we utilize the following scoring system: each correctly identified element in a given direction contributes 0.25 points. Consequently, the total correctness score per question can range from 0 to 1, where 0 represents all atoms identified incorrectly and 1 represents all atoms identified correctly.

The *correctness rate* is determined by summing up the total correctness score and dividing it by the total score, then multiplying by 100%.

$$Correctness\,rate = (\frac{Correctness\,score}{N * p}) * 100\%$$

where:

N = the number of structures evaluated, i.e. N = 28

p = the correctness score per question, i.e. p = 1

Effect Size

As mentioned previously, effect size is a valuable step in measuring the amount of gain when comparing pretest and posttest results (cf. definition 5.1). We used Cohen's d to analyze the data obtained from the experiment.

Cohen's d

Definition 5.2 Cohen's d is a standardized measure of the effect size, representing the difference between the pretest and posttest results, in terms of standard deviation units (York, 2016).

The resulting value of Cohen's d is commonly used to interpret the magnitude of the difference between the pretest and posttest results. According to the benchmarks proposed by Cohen, an effect size of .2 is small, an effect size of .5 is medium, and an effect size of .8 is large (York, 2016, pp. 80). However, these generic benchmarks might not be applicable to all areas of study, as different domains might exhibit smaller effect sizes (Valentine & Cooper, 2003). Consequently, we decide to incorporate Cohen's U_3 as an additional measurement to compare the differences between pretest and posttest results.

Cohen's U_3

Definition 5.3 Cohen's U_3 describes the percentage of scores in the lower-meaned group that are exceeded by the average score in the higher-meaned group (Valentine & Cooper, 2003, pp. 5).

Figure 5.4 displays the results of both the pretest and posttest, comparing the performance of participants before and after the training part. We have used a paired t-test on the correctness rate of all the participants. The p-value is 1.351e-10, which is below the significance level 0.05. Therefore, we reject null hypothesis *there is no significant difference in performance between the pretest and posttest.*

The mean of the correctness rate for the pretest is 57.2% and for the posttest is 75.9% (cf. Table 5.1). Cohen's d = 1.607 suggests that the difference in gain between the pretest and posttest scores is substantial. U_3 indicates that 95% of the posttest results are above the mean of the pretest results (cf definition 5.3). This suggests that the training has resulted in a noticeable improvement in performance. This can also be clearly observed in Figure 5.4), where the participants' performance in the posttest was a lot better. These findings support that the training session had a significant effect on the learning rate of participants. Mean-

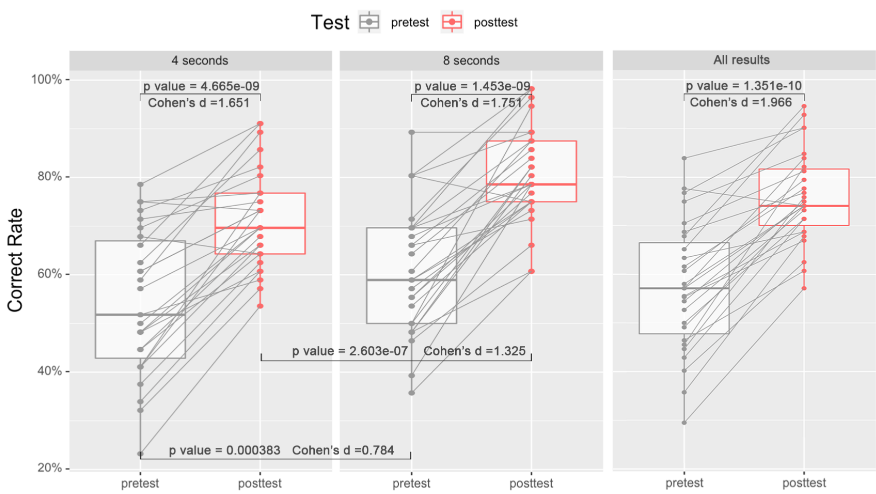


Figure 5.4: A visual comparison of correctness rate between pretest and posttest for different recording durations (4 seconds, 8 seconds and all). Each data point displays the correctness rate of a participant, while the lines connecting them illustrate the individual changes in correctness rate between the pretest and posttest.

5

while, the individual training part was around 5 to 7 minutes, consequently it was concluded that people were able to learn this sonification design in a relatively short period. This indicates that participants are able to learn and comprehend the sonification design and perform better with practice.

Additionally, we reviewed the correctness rate of the 4-second and 8-second recordings separately (see Figure 5.4). The p-value shows a significant difference between the results of the 4-second and 8-second recordings in both pretest (0.000383) and posttest (2.603e-07). Therefore, this is in favour of a rejection of null hypothesis *there is no significant difference in performance between the 4-second and 8-second recordings.*

Table 5.1 displays the effect size between two different durations, indicating a smaller effect size in the pretest (d = 0.784, U_3 = 68%) compared to the posttest (d = 1.325, U_3 = 81%). This seems to imply that the duration difference may have a slightly greater influence on the performance of the participants in in identification and localization in the posttest. We also compared different recording durations between the pretest and posttest individually. The effect size is similarly large for both the results of the 4-second recordings (mean diff = 17.4%, d = 1.651, and U_3 = 91%), and the results of the 8-second recordings (mean diff = 19.9%, d = 1.751, and U_3 = 95%). We conclude therefore that there is a statistically significant change of the correctness rate after the practice for both durations of recordings.

Comparison between	p-value	Cohen's d	Cohen's U_3
Pretest & Posttest	1.351e-10	1.966	95%
4-second & 8-second (Pretest)	0.000383	0.784	68%
4-second & 8-second (Posttest)	2.603e-07	1.325	81%
Pretest & Posttest (4-second)	4.665e-09	1.651	91%
Pretest & Posttest (8-second)	1.453e-09	1.751	95%

Table 5.1: The table presents the results of the pretest and posttest, including two playback conditions (4 seconds and 8 second). The p-values indicate the statistical significance of the differences between pretest and posttest results. The Cohen's d provides an estimate of the magnitude of the observed changes and U_3 represents the percentage of results in the lower-meaned group that are exceeded by the average results in the higher-meaned group.

5.3.1 Elements

Table 5.2 shows that there is a significant difference in the correctness rate for all the elements between the pretest and posttest. The fifth column '-' represents the zero element, i.e. the situation where no sound was played or identified. From the results in Table 5.2 it can be observed that the correctness rate of nitrogen is relatively low as well as the p-value of the difference between two tests. The correctness rate for hydrogen (70.7%) and oxygen (63.9%) were already higher in the pretest.

	H	C	N	O	-
p-value	0.0001207	2.449e-08	0.002381	9.564e-05	0.01457
Mean - pretest	70.7%±24.1%	47.2%±19.7%	43.3%±25%	63.9%±12.4%	81.2%±19.1%
Mean - posttest	89.2%±12.7%	70.9%±13.6%	59.5%±25.2%	77.8%±10.3%	88.9%±12%

Table 5.2: The table presents the differences in correctness rate (%) for different elements (H, C, N, O, -) between pretest and posttest conditions, including p-values and measures of mean with standard deviation.

In the stacked bar chart (see Figure 5.5), the x-axis represents the elements that were played including zero element ('-'), while the y-axis shows the result of the element as identified by the participants. '-' represents the situations where no atom/sound was heard. In the pretest, hydrogen was wrongly identified as carbon (16.8%) while carbon was wrongly identified as both nitrogen (22.7%) and hydrogen (16.5%). Nitrogen was often mistaken for oxygen (24.8%) and carbon (19.4%). Oxygen was mostly misidentified as nitrogen (14.5%) or nothing (11.4%). In the posttest, hydrogen's correctness rate reaches to 89.2% and it was mainly mistaken for carbon (7.5%). The correctness rate of carbon increased from 47.2% to 70.9%, it was still misidentified as nitrogen (16%) but less mistaken for hydrogen (7%). Oxygen's correctness rate increased to 77.8% and misidentification rate as nitrogen decreased to 10.5%. The correctness rate of nitrogen was improved to 59.5% but still below average (75.9%). Nitrogen was often wrongly identified as oxygen (22%) and carbon (16%). Inferences of possible explanations will be discussed in the section 5.4.

Figure 5.6 presents more detailed information about the influence of playback duration on element identification. In general, participants performed better with 8-second recordings. It can be observed that there were more times of nothing

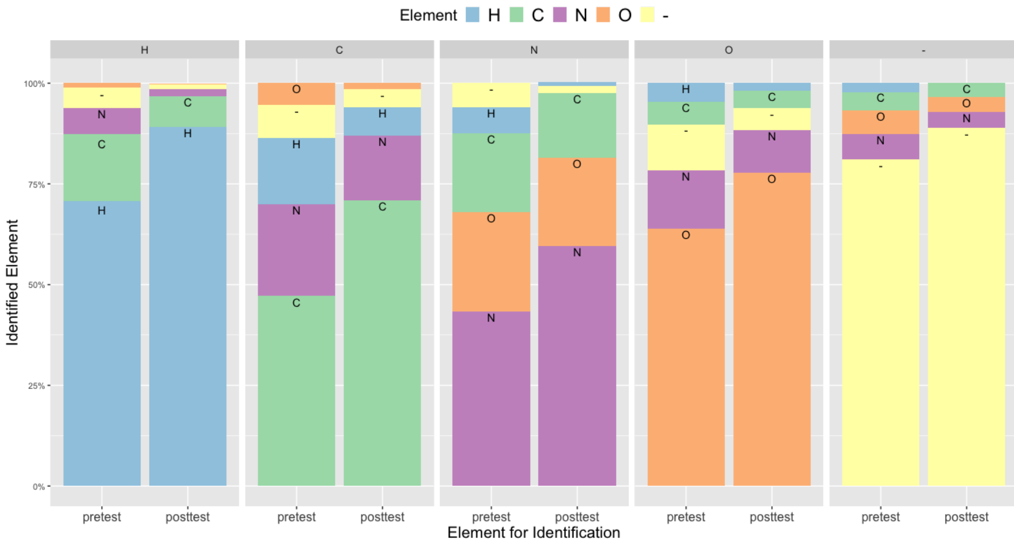


Figure 5.5: Stack diagram depicting the percentages of correct identifications and misidentifications of elements during the experiment, ordered from low to high atomic mass.

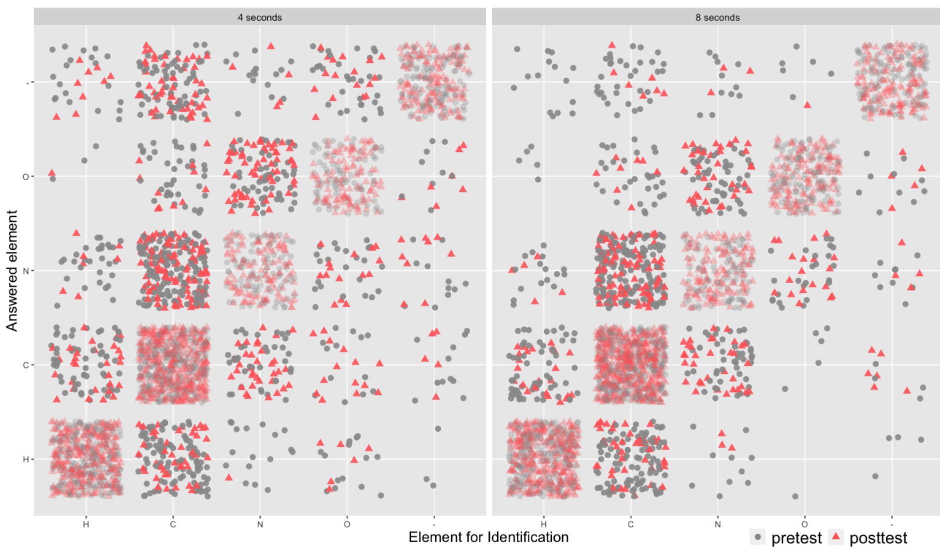


Figure 5.6: Distribution plot illustrating the accuracy of participants' identification of target elements with both 4 seconds and 8 seconds recordings. The x-axis represents the elements to be identified (H, C, N, O, -), and the y-axis represents the elements that participants answered.

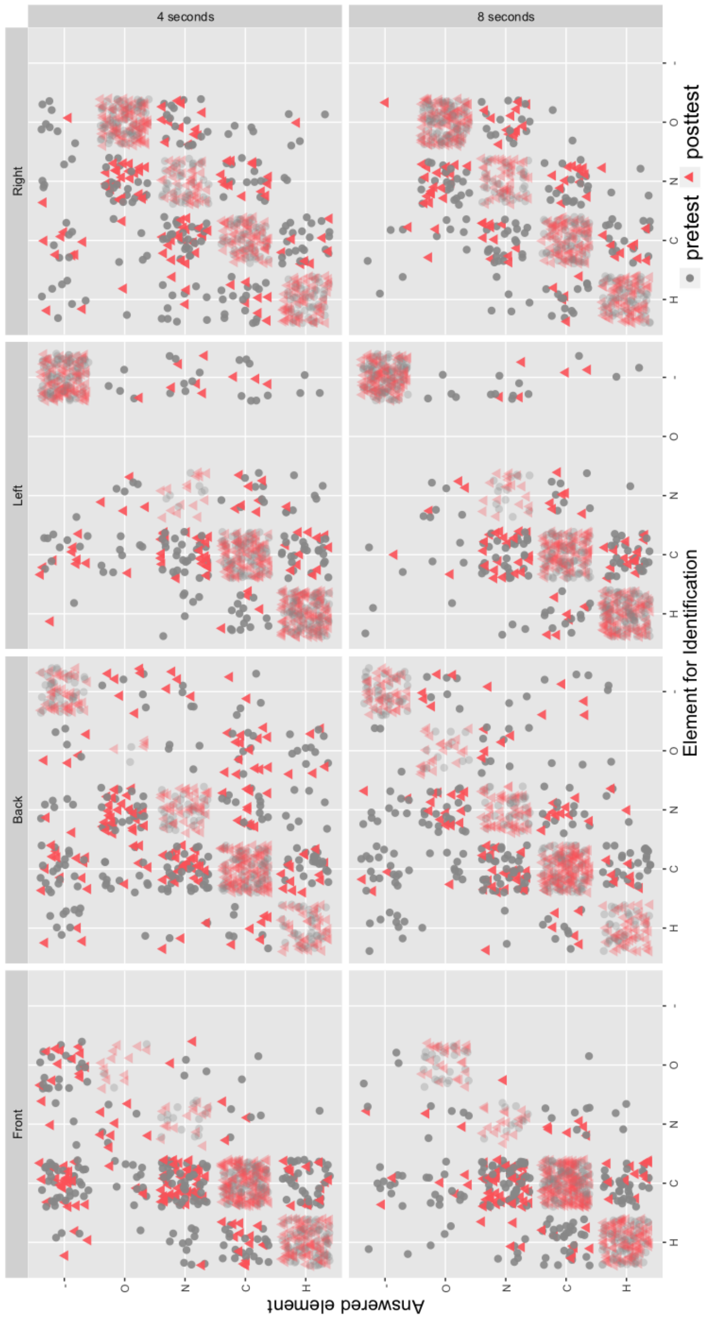


Figure 5.7: Distribution plot depicting participants' responses for identifying elements positioned from four directions (front, back, left and right). The x-axis represents the elements to be identified (H, C, N, O, -), and the y-axis represents the elements that participants answered. Please turn the page 90 degrees counterclockwise for optimal viewing.

heard or nothing identified in 4-second recordings, especially in the pretest. It could be that because of the short duration the participants may not have had sufficient time to localize and recognize the sounds from the four directions. After the training phase, the correct identification of hydrogen and oxygen is higher than the other two elements, even for the 4-second duration. As for carbon and nitrogen, participants made less mistakes with 8-second duration.

5.3.2 Directions

The influence of sound direction is shown in Figure 5.7. We can observe that the front direction in the 4-second recordings was often wrongly identified as nothing, and in most cases for carbon or oxygen, while the back or rear direction in both 4- and 8-second recordings was sometimes wrongly identified as no atom. Some participants commented that they may have paid less attention to the sound from back speaker or only notice it at a later time in the pretest. Carbon was mistaken for nitrogen and hydrogen from all directions. While nitrogen was mistaken a lot for oxygen from back speaker, and for carbon and oxygen at right speaker. In general, wrongly identified positions of the posttest were less compared to the ones of the pretest and participants performed better with 8-second duration recordings from both front and back speakers. In the pretest, the performance of front and back speaker were worse.

5.3.3 Structures

From the analysis, as visualized in Figure 5.8, we concluded that the error rate of most structures in the pretest is lower than in the posttest. Additionally, participants performed better with 8-second recordings, especially after practice. It can be observed that the identification between 4-second and 8-second differs a lot in structures 5, 6, 13, 14. There are three atoms in structure 2, 3, 4, 5, 6, 10 (cf. Figure 5.2). But the error rate of structure 4 and 10 are higher even in the posttest (see Figure 5.8). This implies that the identification of nitrogen alone might be hard. There is no overall indication that it would be easier to identify structures containing three atoms than structures containing four atoms. If we look through the structure 1, 7, 9 11, there is a transformation from four carbon atoms to the combination of carbon and hydrogen atoms (cf. Figure 5.2). Together with Figure 5.7, we found that it would be easier and

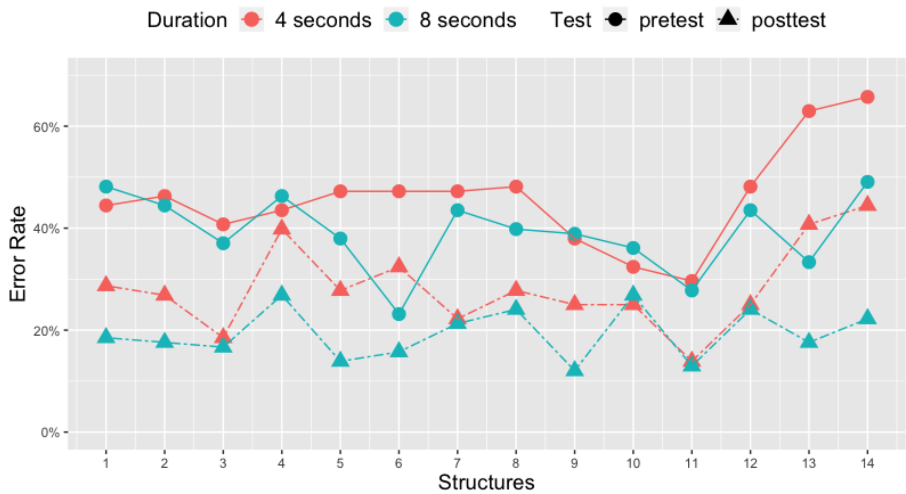


Figure 5.8: The plot illustrates the error rate (%) for each of the 14 structures (cf. Figure 5.2) in Validation 1. The lines connecting the data points within each group serve as a visual grouping mechanism, indicating that the data points belong together based on the shared characteristics of color (representing recording duration, with 4 seconds and 8 seconds) and shape (representing test type, with circles for pretest and triangles for posttest). It is important to note that the lines do not carry any specific meaning or interpretation within the context of this figure.

faster for the participant to identify and separate one carbon atom from the other three hydrogen atoms in structure 11, which has lowest error rate in both tests. Moreover, it would be easier and faster to recognize one hydrogen atom and three carbon after the practice (structure 7). This may suggest that the sound of hydrogen is easier to learn and remember than the sound of carbon.

5.4 Conclusion and Discussion

From our study we have concluded a statistically significant difference in the performance between the pretest and posttest assessments, with a p-value of 1.351e-10. Specifically, we found a relatively high learning rate, as demonstrated by a substantial increase in correctness rate from the pretest (M = 57.2%, σ = 13.4%) to the posttest (M = 75.9%, σ =9.5%), with a relative large effect size (Cohen's d = 1.966, U_3 = 95%). From the results we conclude that our sonification design is learnable and people are able to learn it relatively quickly.

Based on the results of the posttest, which showed an average correctness rate of 80.7% for the 8-second recordings, it is clear that with sufficient exposure to the sounds, participants can quickly and accurately identify and locate the first layer of sounds; up to four simultaneously playing sources. Thus, it can be concluded that Design VI can achieve immediacy of sound recognition and localization.

The results show that the sounds with highest and lowest pitch, i.e. hydrogen and oxygen, are easier and faster to be identified in both the pretest and posttest. Without the highest or lowest sound(s) as a reference, it becomes harder to identify carbon and nitrogen alone which have pitches in the range of the highest and lowest pitches. It might be confusing for the participants to identify whether the sound is from one of the middle two pitches or the lowest/highest one. Or when there were several concurrently sounding sources, it becomes harder to distinguish the ones whose pitches are in between.This suggests that the range of pitches used in the sonification design should be carefully considered and iteratively tested in order to optimize the ease and accuracy of identification.

During the individual interviews, we found that most of the participants had identified the elements according to their pitch differences. The irregular patterns, where each type of element has its own density, can be an important feature for separating concurrently sounding sources. We assumed that using density would help to avoid the merged perception of two or more identical atoms. However, density might not be a perceptually salient feature to be the most easily perceived and remembered by the participants. There were a number of participants (3) mentioning that they were unable to perceive the pitch differences and found the density differences more distinct. They described density differences as the ‘speed’ of each sound, with certain sounds perceived as ‘faster’ (more dense) and others as ‘slower’ (less dense). Nevertheless, most participants would not use it as main feature to identify the sounds of the elements, especially when they have to combine it with the pitch differences to identify sounds. This suggests that we should have considered the individual factors, for example, people with perfect pitch or background in music training might be more sensitive to differences between sounds and able to identify the sounds quicker.

Besides that there are other factors which may influence the element identification if we manually assess the identification results of each participant. Participants were able to find the relation among two or three sounds from either the frequency or the pattern differences in the pretest already. Common mistakes were made, such as H-C-N combinations were mistaken for C-N-O in structure 10, 12, 14. For example, the error rate of structure 11 is less than 12, where the nitrogen atom at back direction is the only difference between two structures. In the structure 11, hydrogen atoms were identified and localized correctly but carbon was mistaken for nitrogen. In structure 12, the identification of hydrogen was correct most of times, while carbon was sometimes still mistaken for nitrogen and nitrogen was sometimes mistaken for oxygen. The existence of other sound sources might help participants compare and thereby identify the sounds. In structure 1 (cf. Figure 5.2), 16 participants mistook carbon for hydrogen or nitrogen in the pretest. Most of them wrongly identified the structure as either four nitrogen atoms or four hydrogen atoms. After practice, 5 participants made mistakes either in 4-second or 8-second recording of structure 1. There was an exemption who identified it correctly in the pretest but mistook it for nitrogen in the posttest. This may confirm our previous conclusion that identification of single element (C or N) is harder due to the lack of reference. 15 participants mistook structure 4 (cf Figure 5.2) for three carbon or oxygen atoms in the pretest. Left direction was identified correctly at most of times, while some people mistook it for oxygen or nitrogen. 19 participants mistook carbon for nitrogen or hydrogen in structure 6 (cf. Figure 5.2). Carbon might be confused here since there are only two sounds being played. Another observation was that only 3 participants identified the front oxygen correctly with 4-second recording while most of time it was marked as nothing.

It seems that the duration of the sound exposure has a significant impact on the participants' ability to identify and localize the sounds, The results suggest that four-second duration might be too short for the participants to recognize all the sounds correctly particularly when there were multiple elements present.The data from structures 9 and 12 showed that most participants were able to identify hydrogen correctly within both the 4 and 8-second durations, but mistakes still occurred in both exposure durations. It appears that longer exposure durations may be necessary for more accurate identification and localization of the sounds.

5.5 Further Analysis and Future Development

We used fourteen structures that are more regularly found in amino acids, partially from backbones. However, the total time of each element that appeared in each of the directions from the structures varies a lot (see Table 5.3). For example, the oxygen atom was never positioned on the left and both oxygen and nitrogen appeared relatively fewer times than hydrogen and carbon. We could derive specific structures from those basic 14 structures if we combined the error rate results, For example, a different element can be added from the left in structure 4, then it can be demonstrated whether a referred element sound could improve the identification of nitrogen. Structure 12 and 14 are similar and both have high error rate. If we rotate structure 14 counterclockwise by 90 degrees, the only difference would be the carbon on the right. Then we could compare whether the different amount of carbon and hydrogen atoms would influence the difficulty level of identification.

In order to make carbon and nitrogen obviously identifiable, especially when there is no other element to refer, we consider to add changes to the sound design. For example, more distinct pitch or density difference could be applied. Alternatively, timbre differences can be applied. Like the heavier element can have a more sustained gloomy or dark sound.

Next, we intend to sonify more than one surrounding layer of atoms simultaneously in an extended version, by simulating the reverb of a surrounding space and change the loudness of the direct sound depending on the distance of the atom in relation to the current position (cf. Design VI). We will include this feature in Validation 2 for further evaluation (see section 6.1).

Direction	H	C	N	O	-
Front	8	16	2	2	0
Back	4	12	6	2	4
Left	8	10	2	0	8
Right	6	8	6	8	0
Total	26	46	16	12	12

Table 5.3: Distributions of each element on different directions, from the structures used in Validation 1.

CHAPTER 6

Evaluating the Sonification of Molecular Structures Using Multiple Concurrent Sound Sources: Validation II

This chapter is based on the following publication:

Liu, D., & van der Heide, E. Interactive auditory navigation in molecular structures of amino acids: A case study using multiple concurrent sound sources representing nearby atoms. In *Proceedings of the 25th International Conference on Auditory Display, ICAD 2019*, (pp. 140–156), Newcastle, UK.

Liu, D., & van der Heide, E. Evaluating the spatial sonification of the molecular structures of amino acids using multiple concurrent sound sources. In *Proceedings of the 26th International Conference on Auditory Display, ICAD 2021*, to appear.

Liu, D., van der Heide, E., & Verbeek, F. J. Design and evaluation for the sonification of molecular structures using multiple concurrently sounding sources. (Publication in preparation)

6.1 Introduction

In the experiment described in Chapter 5, we only considered sonification of the first layer of atoms in the molecule to investigate factors that may affect individual performance in identifying and localizing concurrent sound sources. As a further elaboration on experimental evaluation of our sonificaiton design, we would like to take it a step further by incorporating additional sound sources. This will involve adding the second layer of atoms from the molecule. In the experiment described in this Chapter, referred to as Validation 2, our objective is to investigate the maximum number of atoms (i.e., sounds) that listeners are capable of recognizing and localizing at a time.

In order to create the suggestion of distance we simulated the reverb of a surrounding space and change the loudness of the direct sound depending on the distance of the atom in relation to the current listening position[1] (cf. Design V). Additionally, based on the results obtained from Validation 1, we have considered several potential improvements in our sonification design from three aspects. Therefore, we refer to these improvements as Design VII. The aspects include:

QRcode 6.1

1) Pitch: We have raised the pitch for hydrogen and carbon sounds by one octave (see Figure 6.1), so that there is a now two-octave interval between the carbon and nitrogen atoms. We hope this modification contributes to correctly identifying the elements and avoiding the confusion that we have seen in Validation 1.

2) Timbre: In addition to the increased pitch interval we have added some changes in timbre. We have increased the differences between the sounds, which is accomplished by fine-tuning the q-factors of the bandpass filters for the individual partials of the individual sounds (cf. Figure 4.9).

3) Density: With respect to the density feature, we used the same settings for all the elements except for oxygen. The irregular repetitive pattern has been increased a bit in density so that there will not be too long a period between two consecutive impulses of the sound and thereby resulting in a bit more continuity. Although the introduction of reverb allows us to create a different sensation of distance for the elements in the first layer and the elements in the second layer, the reverb also blurs the sounds for a short period and therefore it becomes a little more difficult to distinguish the sounds from each other especially when many

[1] Recording example featuring sounds of four elements on different layers (cf QRcode 6.1, scan to listen).

objects are present. We therefore decided to give the sounds a bit a sharper attack by not only using the generated irregular impulses to excite the bandpass filters but to also mix them with the output and thereby make them directly audible. This more impulse-based attack makes it easier to detect and localize the individual sounds.

The aim of Validation 2 is to assess the ability of the listeners to identify and localize two layers of sounds surrounding the listening position. Through this experiment we want to evaluate to which extent our sonification design enables the participants to distinguish the positions of the layers from each other. We have two assumptions regarding participants' performance with two layers of sounds:

Assumption C *It would be more challenging to identify and localize the second layer of sounds compared to the first layer.*

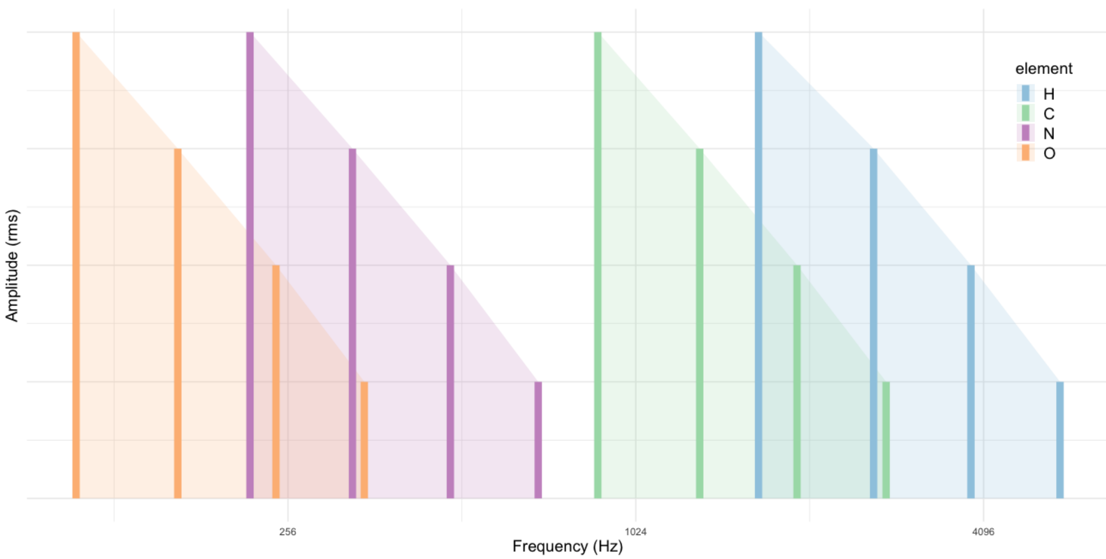


Figure 6.1: Frequency components for different elements. The x-axis represents frequency in Hz, and the y-axis represents amplitude in rms. The graph shows the frequency components of four certain sounds representing four elements, highlighted using different colors. This representation emphasizes the relative amplitudes of the components in the ratios of 4:3:2:1. The first (lowest) partial of oxygen has a frequency of 110 Hz, the first partial of nitrogen is 220 Hz, the first partial of carbon is 880 Hz and the first partial of hydrogen is 1760 Hz. The filled quadrilaterals indicate the frequency domain of a certain atom. The overlap of the ranges is clear, yet all atoms have a distinct pattern.

6

Assumption D *Participants would be able to separate the two sound sources on different layers originating from the same direction.*

6.2 Experiment Design

Our objective is to examine the performance of participants when exposed to two layers of sounds. To achieve this, we have designed two different conditions for the experiment (see Figure 6.2). In condition 1, the sounds from the first layer are played initially, and after 10 seconds are the sounds from the second layer joined. In condition 2, all sound sources are played simultaneously for 20 seconds, with each direction potentially containing up to two layers of sounds.

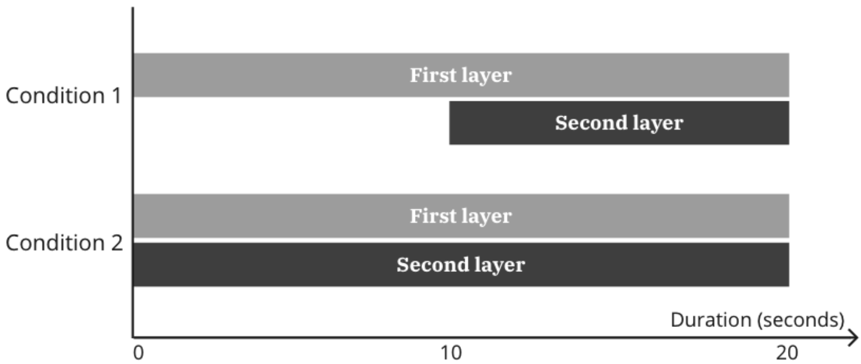


Figure 6.2: Visualization of the sequential presentations of sound sources in two conditions. The x-axis represents duration in seconds. The graph illustrates distinct timing patterns in the experimental setup.

To enable a direct comparison between the conditions within the same participants, we have chosen a within-subject design for the experiment.This has the advantage that the overall level of performance of the individual subject can be assessed in a good manner (Lane et al., 2017). For example, some subjects may more skilled in localizing sound sources or recognizing pitch differences, disregarding the condition they are in. By comparing the performance of a subject in one condition to the performance of the same subject in the other condition, individual differences could be better controlled. Furthermore, to reduce the influence that practice may cause a better performance for the second presented condition, the order of the two conditions was counterbalanced. Ideally, half of the subjects start with condition 1, and the other half of the subjects start with condition 2.

6.2.1 Materials

From the 14 structures used in the previous experiment, we specifically chose the structures 1,2,6,7,8,14 (see in Figure 5.2), because we have measured a lower error rate in the posttest test. We extended these structures by adding the second layer atoms based on combinations that are found among in amino acids. This resulted in 8 structures that were used for Validation 2[2].

QRcode 6.2

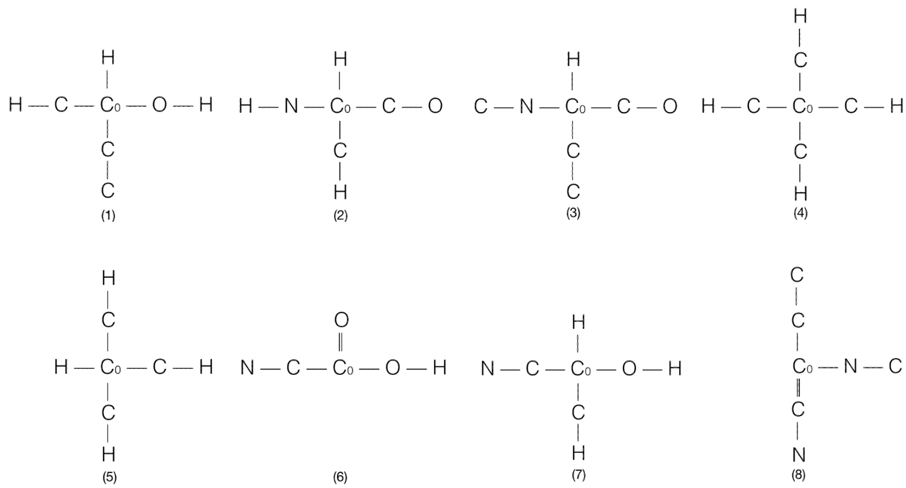


Figure 6.3: 8 structural formulas for Validation 2 (2 layers). Structure 1 is an extension of structure 8 from Validation I. Structure 3 is derived from structure 14, Structure 4 is derived from structure 1, Structure 6 is derived from structure 6, Structure 8 is derived from structure 2, Structure 9 is derived from structure 7.

6.2.2 Software and Hardware

The application was developed on a Macbook Pro with 16GB RAM with a LEAGY sound card[3]. All the sounds were generated in Pure Data (version 0.50) in real time.

The GUI (cf. Figure 6.4) for the users to indicate the sounds they heard was programmed in Processing (version 3.5.3)[4]. For the statistic analysis we used R (RStudio version 1.2.5)[5] and Microsoft Excel (version 16.38).

[2]Recording examples of structure 1 and 2 (cf QRcode 6.2, scan to listen).
[3]An external audio device supporting 6-channel output (Link to the product.)
[4]https://processing.org/
[5]https://www.rstudio.com/

6.2.3 Experimental Procedures

The experiment for Validation 2 consisted of four phases (see Appendix B, instructions).

In phase 1, an introduction to the four sounds was given to the participants identical to Validation 1. After they felt they were able to recognize the sounds, they would start phase 2.

Phase 2 was a training session including 16 questions (see Appendix C, training session). The questions were designed to guide the participants to get familiar with the concept of layers as well as multiple concurrently sounding sources step by step. At beginning, they were asked to identify either the element type or the layer number. Harder questions for localization and identification of multiple objects were given follow up later[6]. During the training session, the participants were informed that sounds would come from the four surrounding speakers and there would be up to two sound sources on each speaker simultaneously.

QRcode 6.3

In phase 3 and 4, the participants had two different conditions[7] of sound tests (cf. Figure 6.2). Participants were told that a maximum of 8 sound sources will be positioned around and each direction will contain up to two layers of

QRcode 6.4

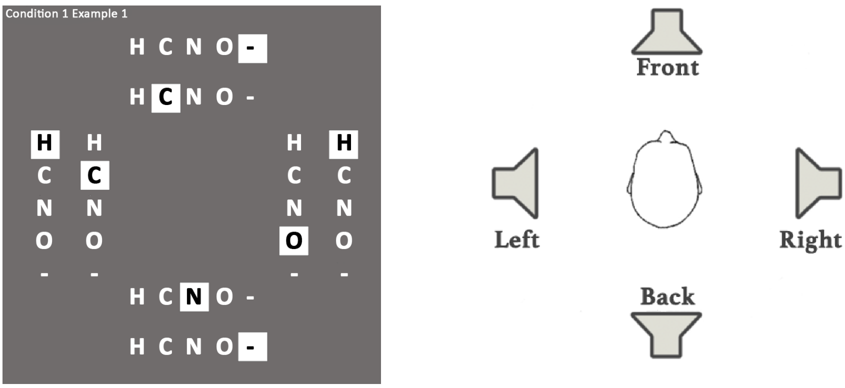


Figure 6.4: A screenshot of the user interface for the participant to indicate the sounds they heard during the experiment of Validation 2 (2 layers). In the user interface, the up and down directions correspond to the front and back speakers, while the left and right directions correspond to the left and right speakers. The inner circle options correspond to the first layer sounds, while the outer circl options correspond to the second layer sounds.

[6]Recording example of sound sources added around one by one (cf QRcode 6.3, scan to listen).

[7]Footage of Validation 2, included two conditions (cf QRcode 6.4, scan to watch).

sounds. Participants were randomly assigned to start with one of the conditions. In condition 1, 8 sets of sounds were played in a randomized order. Participant were instructed that, for each set of sounds, the first layer will be played at first and the second layer will be added after 10 seconds. In condition 2, same 8 sets of sounds will be played in a randomized order. Participants were instructed that, for each set of sound, all sound sources will be played simultaneously for 20 seconds. During the time that a structure was played the participants were able to choose in an interface which elements they heard originating from each direction and layer (i.e., H, C, N, O or -, in Figure 6.4).

Participants were told to choose '-' if they were sure no sound was played from a certain position, otherwise they had to choose a corresponding element that was most close to what they heard. In both conditions, at the onset of a session participants were given three examples to get familiar with the interface as well as the way the sounds were played. During the whole experiment, participants were allowed to change the head orientation.

The aim of the experiment is to gather and analyze appropriate evidence to either accept or reject Null Hypothesis below:

H0 *There is no significant difference in performance between Condition 1 and Condition 2.*

6.3 Experimental Results

The experiment was performed with a total of 35 participants, 19 female and 16 male participants. 97% of them were in the age group 20-30 years and 3% were in the age of 31-50 years (cf. section 5.2). None of the participants have participated the experiment for Validation 1. While each of the 8 structures had a playback time of 20 seconds, the total duration for each condition was approximately 5 minutes, including the time participants spent answering in the user interface. The experiment results had a balanced distribution, with 18 participants starting with condition 1 and 17 participants starting with condition 2.

Correctness Rate

We recorded the answers given for each of the 4 directions in both conditions. To calculate a **correctness score** per presented structure, we utilize the similar scoring system as in Chapter 5: each correctly identified element in a given direction and layer contributes 0.25 points. Consequently, the total correctness score per question can range from 0 to 2, where 0 represents all atoms identified incorrectly and 2 represents all atoms identified correctly.

The *correctness rate* is determined by summing up the total correctness score and dividing it by the total score, then multiplying by 100%.

$$Correctness\ rate = (\frac{Correctness\ score}{N * p}) * 100\%$$

where:

N = the number of structures evaluated, i.e. N = 8

p = the correctness score per question, i.e. p = 2

From Figure 6.5 we can observe that, participants performed better in condition 1. To assess the significance of differences between two conditions, a paired t-test is applied on the correctness rate of all the participants. The p-value is 1.051e-05, which is far below the significant level 0.05. This indicates rejection of Null Hypothesis *there is no significant difference in performance between the two conditions.*

Since first layer sounds were played separately in condition 1, the average correctness rate for first layer sounds identification in condition 1 is 79.2%, and

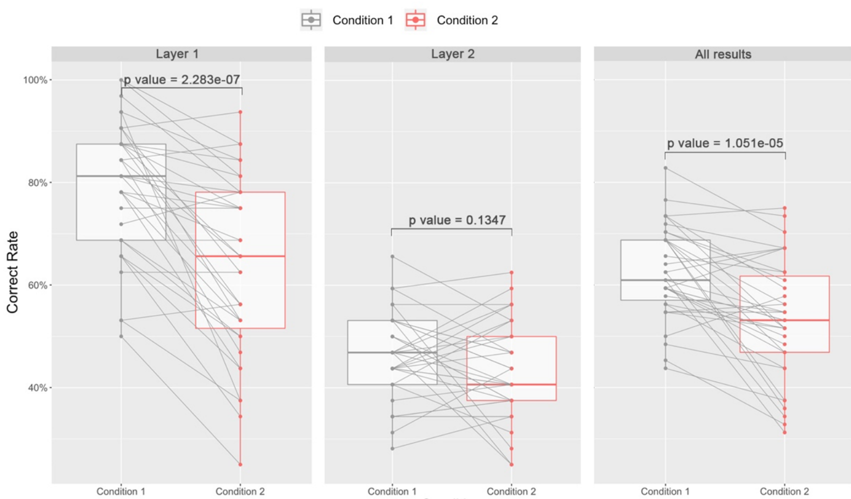


Figure 6.5: A visual comparison of correctness rate between condition 1 and condition 2 for different layers of sounds. The data points displayed as circles to illustrate the individual changes of the average correctness rate.

63.6% for condition 2. The p-value calculated for comparing the performance on layer 1 between the two conditions is 2.283e-07, indicating a statistically significant difference. From the results, however, there does not appear to be a significant difference between the second layer sounds when comparing the two conditions (p-value = 0.1347). The average correctness rate for identifying second layer sounds n in condition 1 is 46%, and in condition 2 it is 43.2%.

6.3.1 Elements

When referring to 'elements', we are indicating the abstract representation of the sounds. On the other hand, when mentioning 'atoms', we are referring to the individual objects within a chemical structure.

From the data in Table 6.1, it can be seen that the participants performed better for the sounds positioned on the first layer than second layer in both conditions. In condition 1, the correctness rate for all the identified elements positioned on first layer are all above 72%, especially the correctness rate of hydrogen and oxygen reached 82%. There was less of a chance to misidentify nitrogen with oxygen or confuse carbon with nitrogen. In condition 2, all the sounds were played in parallel. Participants can identify the first layer sounds relatively well and the overall correctness rate for all elements on the first layer is above 55%.

	H	C	N	O	-
Condition1-Layer1	83.4%	77.0%	72.4%	82.9%	91.4%
Condition2-Layer1	73.1%	56.7%	68.6%	67.1%	88.6%
Condition1-Layer2	35.1%	22.9%	47.6%	62.9%	74.6%
Condition2-Layer2	31.6%	21.1%	41.9%	65.7%	72.1%

Table 6.1: The table presents the results of correct identifications (%) of four elements and zero element ('-'), in different conditions.

However, it turned out that participants had similar performance for the second layer sounds with the ones in condition 1. It seemed to be more difficult for the participants to identify and localize the sounds from the second layer for both conditions, the average correctness rate for second layer sounds is around 44.6% when we combine the results for both conditions. The correctness rate of

6

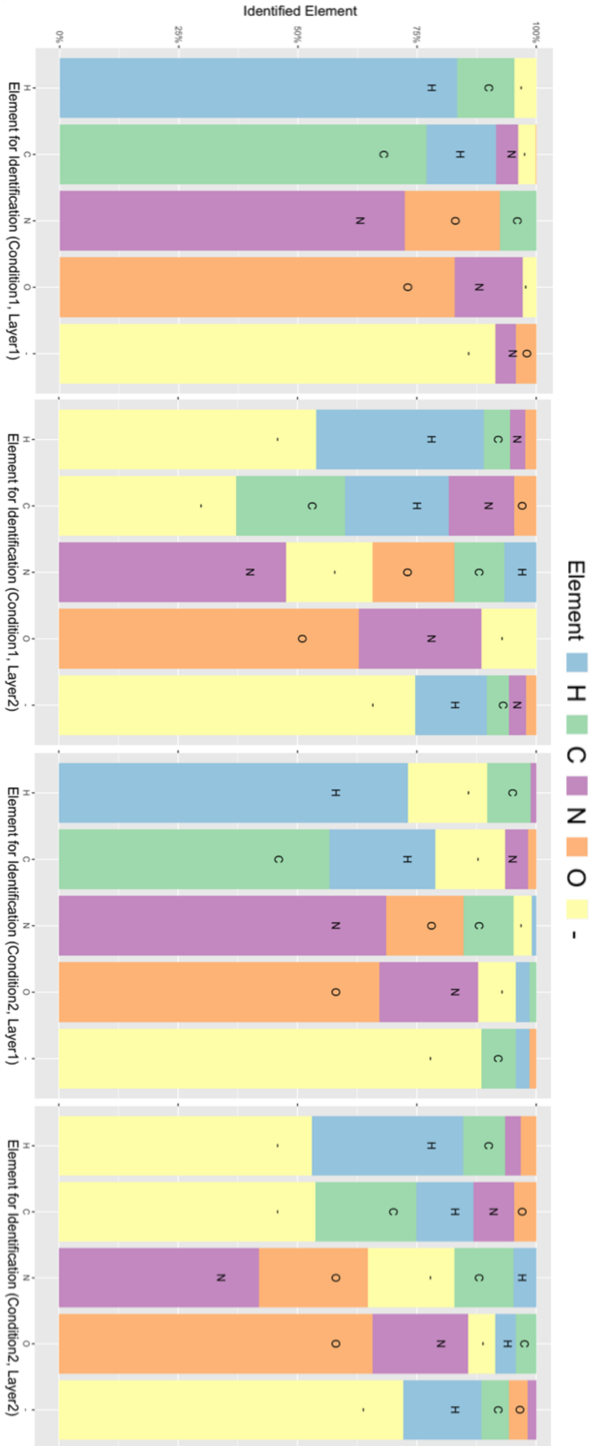


Figure 6.6: Stacked bar plot of element identifications. The y-axis displays the percentage of identifications, with the bars stacked in descending order from highest to lowest percentage. The different colored segments within each bar represent the proportions of correct and misidentifications of different elements.

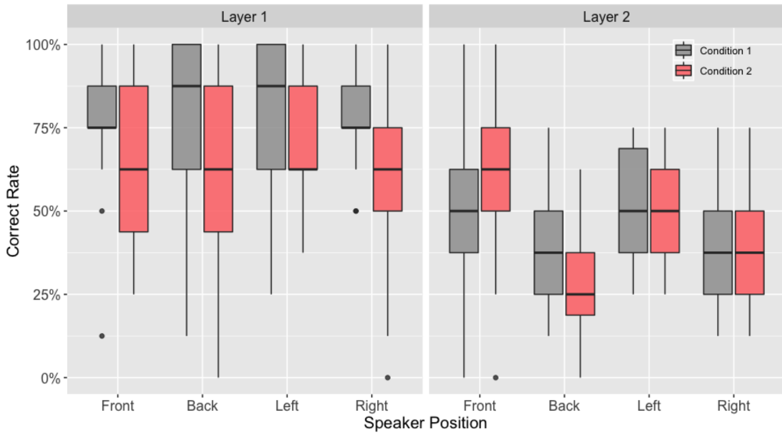


Figure 6.7: Boxplots with whiskers representing the distribution of correctness rates (%) for four directions (front, back, left and right) in both conditions. The boxplots provide an overview of the variations in correctness rates across conditions, directions, and two layers of the sounds. The whiskers indicate the extent of the data beyond the box, showing the range of values excluding outliers. Outliers are represented by individual points outside the whiskers.

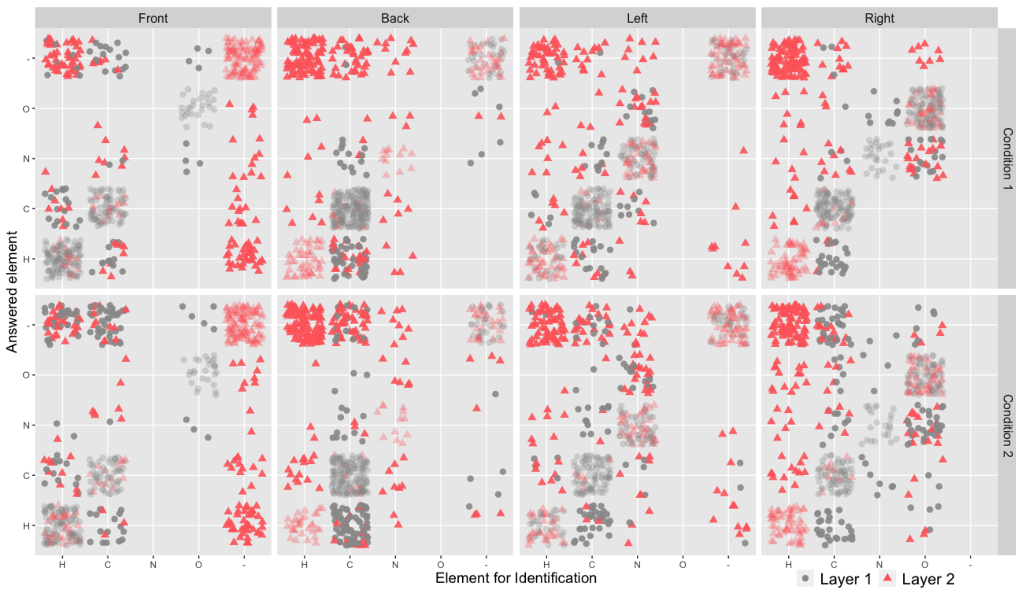


Figure 6.8: Distribution plot illustrating the accuracy of participants' identification of target elements in both conditions, from four directions (front, back, left and right). The x-axis represents the elements to be identified (H, C, N, O, -), and the y-axis represents the elements that participants answered. Shape and color are used to denote two layers of the sounds (first layer as grey circles and second layer as red triangles).

both hydrogen(35.1%, 31.6%) and carbon(22.9%, 21.1%) are low. More than half of hydrogen atoms were marked as no sound heard in condition 1 or mistaken as on the first layer in condition 2 (see Figure 6.6).

6.3.2 Directions

From Figure 6.7, we can observe that in general the participants performed better for the front and left speakers. The average correctness rate for sound sources positioned on the first layer from left (80.7%) and right (80%) speakers are high in condition 1. Participants perform worse with the second layer sounds from the back speaker so average (28.3%) goes down for back sounds in condition 2.

The performance for different directions is influenced by both the elements presented and possible differences in our abilities to localize and distinguish the sounds from each other. The hydrogen sound from the front speaker was confusing for participants to localize which layer it was on. Both the hydrogen and carbon on the second layer from the back speaker were difficult to distinguish. The first layer carbon from the back speaker was mostly misidentified as hydrogen in both conditions (see in Figure 6.8). It is assumed that distinguishing between front and back directions is more challenging compared to the left-right distinction. This might due to the shape and placement of our ears, which would allow for better localization and differentiation of sounds in the left-right distinction. The left-right distinction is primarily determined by the differences in sound arrival time and amplitude between the two ears. While the front-back distinction is more complex and relies on additional cues such as head orientation, and reflections from the surrounding environment.

6.3.3 Observations from Training

During the training session, participants were asked to identify and locate all sound sources in four structures containing six or seven sound sources playing simultaneously. (see Appendix C, training session). The average correctness rate and order of identification for each sound source have been recorded and are shown in Figure 6.9. The results showed that most participants correctly identified at least 5 sound sources, whereas some participants were able to identify 6 or 7 sources. Additionally, the left sounds were generally identified more quickly. In structure Q13, more than half of the participants could identify the nitrogen sounds from all directions, but the left and right sounds seemed to be easier to identify. In structure Q14, 29% of participants were able to identify the carbon positioned on the second layer from left, rest of the participants were unable to identify it even after a hint was given. In structure Q15, the oxygen positioned on the first layer from left was identified the fastest. On the other hand, the oxygen positioned on the second layer from front was more difficult to hear, resulting in it being the last one to be identified in order. In structure Q16, second layer hydrogen from right was identified last in order. Additionally, only 11% of participants were able to identify the second layer carbon from the back, with a hint was given.

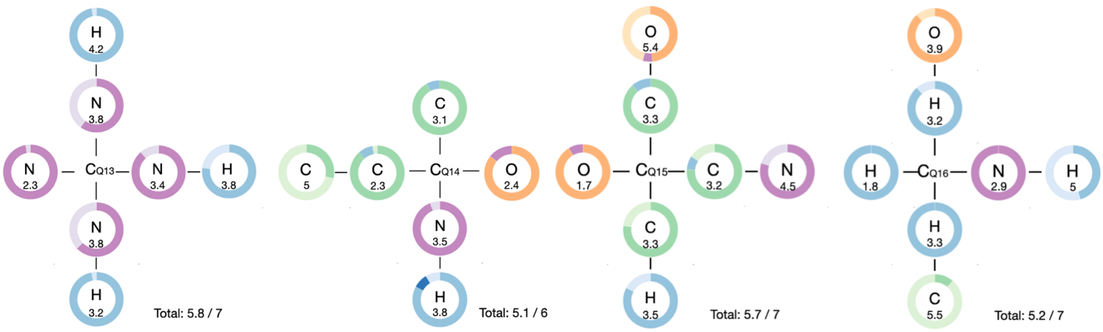


Figure 6.9: Visual representation of the correctness rate identifications for questions 13 to 16 during the training phase (see Appendix C, training session). The size of each colored segment represents the proportion of correct identifications for each element, with larger segments indicating higher correctness rates. The numbers below each element represent the average order of when an element or sound source is identified correctly in the structure, with smaller numbers indicating earlier or faster identification of the sound source from other sources. The numbers in the bottom right corner of each question represent the average number of correctly identified atoms in each structure.

6.4 Conclusion and Discussion

The aim of Validation 2 is to investigate the maximum number of atom sounds that participants are capable of recognizing and localizing using our sonification design in a spatialized environment of concurrently sounding sources. It was unexpected that there was no significant difference between the two conditions for the second layer sound identification. Some participants mentioned that although in condition 1 they did not have to identify the layers themselves, the 10-second duration they had for identifying the second layer sounds might be too short, which could indeed have negatively influenced their performance.

The correctness rate of second layer hydrogen and carbon is fairly low. It seems that higher pitches with the more dense patterns may be difficult to localize. This could be due to the reverb used. In contrast, the reverb settings that were employed may work well for the lower frequency sounds such as nitrogen and oxygen, which can still be perceived and identified on the second layer.

In condition 2, the first layer carbons were often misidentified as hydrogen atoms, and second layer hydrogen atoms were frequently not heard. Combining the results rendered in Figure 6.9) with the observation from each participant's detailed raw result, we conclude that this typically occurred when there was a hydrogen atom on the second layer, such as in the C-H combination. In this case, the hydrogen atom created the illusion of being on the first layer, resulting in its failure to be identified on the second layer. Separating hydrogen and carbon sounds when they are coming from the same direction seems to be difficult. Similarly, this occurs when a first layer hydrogen from the front is combined with a first layer carbon from the back (structure 1, 2, 3 in Figure 6.4); in this case, only the hydrogen sound is identified as the first layer sound. Based on these results and the participants' individual feedback during the training session, we think that auditory masking may occur:
1) when there are identical elements positioned around, the first layer one is might be able to mask the second layer one, even if they are not coming from the same direction.
2) left and right sounds might mask or make it more difficult to identify the front and back sounds.

Due to the occurrence of auditory masking, it still remains uncertain to draw a conclusion for Assumption D. Further research on masking effects necessary to gain a better understanding on its impact on the results. While it might be challenging to completely eliminate masking, there may be adjustments and modifications we can make to the sound design to mitigate its effects.

Although carbon and nitrogen were confusing for the participants to identify in Validation 1. The changes made in Validation 2, including the increased pitch interval between the nitrogen and carbon sounds and the added more articulated attack, appear to have improved the performance of element identification for this experiment. In Validation 1, the average correctness rate in the posttest (8-second) for carbon is 71.6% and for nitrogen was 63.4%. In Validation 2, the average correctness rate in condition 1 (layer 1) for carbon was 77% and nitrogen was 72.4%. In addition, the rate of mixing up carbon and nitrogen atoms was relatively lower in Validation 2 (see Figure 6.6). Although the participants and the test materials differed between two experiments, the results suggest that the identification of each sound became more intuitive for participants without the need for other sounds as reference.

The results of Validation 2 demonstrate that as the number of presented sound sources increases beyond four, it becomes more challenging to identify the second layer of sounds (cf. Assumption C). Nevertheless, our experiment revealed that it is still possible to differentiate between 6 or 7 sound sources within the given time frame. However, a few participants have mentioned that the time frame was somewhat limited.

With our setup and experiments we have developed sonification systems to present concurrently sounding sources in a spatial configuration and used a systematic approach to evaluate its qualities and limitations. The sounds we have designed for the mappings to chemical element can be applied to other objects such as sequences of nucleotides or RNA/DNA coding fragments.

6.5 Limitations and Future Development

In both Validation 1 and 2, we have used a restricted set of chemical structures that are based on the chemical structure of amino acids. As a result of the limited materials we selected, certain elements or combinations of elements were only present in certain positions. Oxygen, for example, never appeared on the back and nitrogen appeared only a few times from the front. We suggest that future research focuses on the possible masking effects between different sounds, both regarding sounds that share a speaker and sounds that are separated spatially using different speakers. The four-speaker setup raised a challenge when representing multiple sound sources, particularly with distance differences. In the future, it would be interesting for us to explore sound source separation and localization using different sound systems, such as an arrangement with 8 speakers to accommodate two layers of sound sources. This expanded setup could provide additional insights into the participants' ability to distinguish and locate the sound sources accurately.

An inevitable fact was that the interior setup of the experiment was not optimal. There were variations in the acoustic conditions for the different directions (left, right, front, and back). The presence of windows on the left, a brick wall on the right, and a wall of monitors in the front created differences in sound reflections and consistency. As a suggestion for future experiments, we recommend conducting the study in a more controlled acoustic environment that minimizes reflections and ensures a more consistent sound field across all directions. This would help to eliminate potential confounding factors and provide a more controlled testing environment for evaluating sound source separation and localization.

Overall, both validations were part of an exploratory research study aiming at testing the design concept and examining the variables, i.e. pitch, density and direction, that may potentially affect the identification and localization performance. As an exploratory study, the focus was on investigating and understanding the relationships between these variables rather than formulating a specific regression model based on the experiment results. While the participants' backgrounds were not explicitly considered in this study, future research could explore the potential impact of participants' musical background and training on the identification and localization performance. This would require a much larger and diverse participant group to ensure that the differences can be rendered significant. Nonetheless, the present research serves as a foundation for further study and offers valuable insights into the potential variables influencing the identification and localization performance. In future studies, we intend to dive deeper into certain variables, such as pitch, within a larger sample size and in an acoustically controlled studio. This approach would allow for more detailed observations and analysis, potentially leading to the formulation of regression models or uncovering more specific relationships between variables.

CHAPTER 7

Conclusions and Discussion

The research presented in this thesis is driven by personal curiosity as main motivation. The aim is to enhance the understanding of the chosen topics (i.e., interaction models, data sonification: interaction & design, evaluation of data sonification). Additionally, the research seeks to contribute new findings and perspectives to these areas of study.

We presented a narrative in Chapter 1, which revolves around the connections between sound and data, based on the three elements of the dialogue model (cf. Figure 1.2). In this chapter, we build upon this narrative and complement it by integrating the main results from previous chapters and aligning them with the corresponding research questions (see Figure 7.1).

It begins by exploring the current state of audience participation in scenarios where sound is controlled by or mapped to other forms of data. Through our investigation, we uncovered an essential relationship between learning and interaction, which led us to propose an ideal framework as the foundation for our subsequent research (see Chapter 2). Based on the insights gained from earlier findings, we proceeded to implement and experiment with various approaches to interactive sonification design (see Chapter 3 & 4). Our analysis and evaluation encompassed both interaction design and sonification design, considering the three elements of the dialogue model (see Chapter 3 & 5 & 6). Ultimately in this chapter, we delve into the discussion of potential directions for future research and development.

7.1 Conclusions

In **Chapter 1**, we introduced the topics covered in this thesis and established connections between them. One of the primary objectives of this thesis is to explore the interactive experience with sound and data. We utilized the dialogue model to define the interactive process and divided it into three components: subject, verbal and adjective (see Figure 1.1). ***While the subject represents the entity or role that initiates the dialogue, the verbal element enables subject input and actions. The adjective provides descriptive information on how the system responds to the subject's actions, particularly in terms of sound production.*** Through an expert analysis of the participants and the usage of sound interaction, we proposed an ideal interaction framework, which was then employed in the interaction and design of data sonification. In

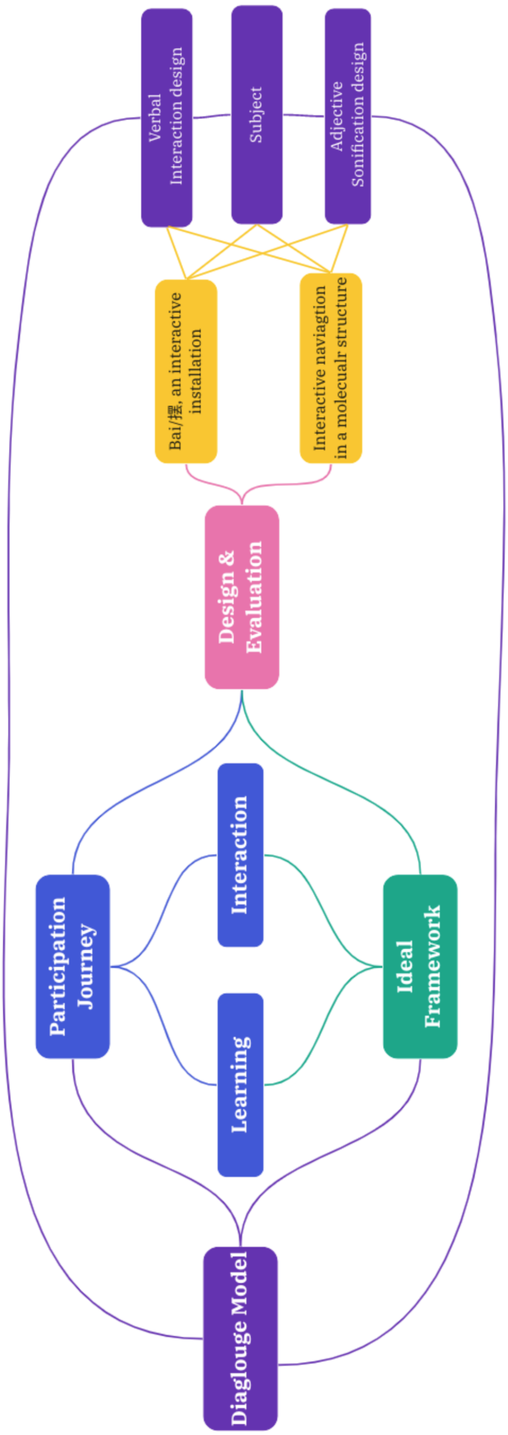


Figure 7.1: Dissertation narrative and structure. The figure presents a condensed outline that integrates the main findings from previous chapters and aligns them with the corresponding research questions.

the latter portion of our work, we explained the design and implementation of user evaluation techniques to evaluate the sonification designs.

7.1.1 Framework for Participatory Sound Interaction

In **Chapter 2**, we reviewed a series of real-time participatory musical performances and analyzed the dialogues between the audience and the systems in such interaction, in order to answer the first research question:

> *RQ1 What elements should be incorporated in an ideal framework for participatory sound interaction?*

We used the audience as an example to investigate the behaviors of subjects in a dialogue. We formulated and analyzed several interaction models mainly from audience participation forms and performance system types. We revealed that an immersive and ongoing interactive environment can be developed into an ideal framework. Therefore, audience, as the main subject of a dialogue, is able to initiate a dialogue at the lexical level by actively interact with the system. Such environment includes *a direct contribution of the audience as well as a responsive interaction*, which could provide the audience with direct auditory feedback on how they are engaged and influencing the performance. Accordingly, the audience is assisted to understand the dialogue at the semantic level with direct auditory feedback from interaction. Sound design, as the adjective of a dialogue, is the key element to complement the dialogue at the syntactic level. Additionally, we formulated a participation journey map to analyze and highlight two key components of the ideal framework: learning and interaction (see Figure 2.1). In conclusion, we consider this ideal framework as ***a responsive dialogue between the audience and the performance system, aiming to achieve a constant loop between interaction and learning*** (see Figure 2.4).

7.1.2 Interactive Sound System from Framework

We have utilized the ideal framework to develop both stochastic and deterministic interactive systems. In **Chapter 3**, we specifically focused on the development of a stochastic interactive system, the sound installation Băi, in response to the second question:

RQ2 Can we use an ideal framework to develop an interactive sound system?

On the basis of what we have discussed and concluded in Chapter 2, we intended to apply the ideal framework in Băi and examine whether an installation can constantly engage the audience through the interaction. A "pendulum" speaker has been used as an interactive interface for audience participation. The installation requires direct contribution of the audience and sound generation from the pendulum speaker gives the a clear environment to the audience to tack and understand the interaction. Together with a setup of six surrounding speakers, the installation offers a dynamic and spatially responsive sound environment for the audience to explore. Moreover, exercising too much control over the pendulum causes the installation to quickly spiral into chaotic and unpredictable behaviour. The unexpected movements of the pendulum speaker and sound results from the surrounding speakers may challenge the audience and keep the interaction continuous. This, in combination with the fact that some physical labor is needed to restrain the pendulum, leads to a tense dialogue between the audience and object, struggling for control. Meanwhile, the swinging movement in space makes it possible for multiple audience members to interact with the installation at the same time. The members in the audience may have influence on each other and form a more diverse interaction accordingly. Thus, ***an intense and ongoing interactive system is created by constructing a dynamic and responsive relationship between the audience and the installation using sound*** (cf. Figure 3.2).

Furthermore, the development of this stochastic sonification system contributes to answering the third research question:

RQ3 Can we develop an interactive sonification design that is intuitive to understand?

The term intuitive is defined as without the need for instructions (cf. definition 3.3). The algorithms are used to translate the motion of the pendulum into different modes of behaviour of the sound environment, in both direct and indirect ways. At first, it may seem that the real-time synthesized sounds react to the motion in a predictable manner, which could be easily understood by the audience as the first step. Especially the noticeable sound generated from the pendulum speaker, as well as the bell sound from the surrounding speakers. The implementation of excitement added variations to the interaction and sonification. The movements resulting from the interaction cause the sounds in the environment to change between different states of relative stability and chaos. However, the distinct audible characteristics for each state are not hard for the audience to discover. In conclusion, ***a distinguishable auditory feedback that directly reacts could assist the audience to comprehend the mappings between sound and data, accordingly explore the interactive sonification design in an intuitive way***.

7.1.3 Sonification Design from Framework

In **Chapter 4**, we addressed an interactive form of sonification with a deterministic character. To that end we considered molecular structures with a carbon backbone as a vehicle to investigate data sonification. With this approach we have further explored second and third research questions:

> *RQ2 Can we use an ideal framework to develop an interactive sound system?*

In order to provide listeners with direct auditory feedback, we used navigation as the primary mode of interaction, allowing them to perceive and track the changes in sound (cf. Figure 4.5). By sonifying a specific area surrounding an atom, we ensured that listeners could focus on a limited number of simultaneous sound objects, usually no more than four. Through an analysis of the potential flow of the listener's experience throughout participation journey map (cf. Figure 4.6), we concluded that ***interactive navigation is an effective approach that offers listeners the opportunity to explore the structure step by step, enabling them to learn and understand how the sound changes in response to the movements***. *This interactive form allows for an engaging and immersive experience where listeners can actively navigate the structure and discover the sound mappings.*

> *RQ3 Can we develop an interactive sonification design that is intuitive to understand?*

The term intuitive is illustrated as easy to learn and fast to recognize (cf. definition 3.3). Sonification design or sound design, therefore, plays a crucial role for the participants to understand. Essentially, each chemical element in the molecule has its own characteristic sound mapping to help and enhance the listeners to identify and localize in a molecular structure. We conclude that ***using a metaphor can help the listener to acquire the meaning of sounds fast as well as decrease confusion***, like building up a connection between atomic mass (light/heavy) and pitch (high/low). We proposed pitch as the main feature because the changes are easily perceivable and distinguishable. According to the atomic mass differences, the lighter element was mapped to a higher pitch while the heavier element was mapped to a lower pitch. We presented a population of designs starting with earcons and arriving at a model-based sonification. In that regard we discussed the possibility to achieve the immediacy of sound identification and localization. Finally, we ended with a pattern design of irregular impulses, which can assist the segregation of multiple concurrent sound sources.

7.1.4 Evaluation of Sonification Design

In **Chapter 5** and **Chapter 6**, our main focus was on evaluating the sonification design of concurrent sound sources, as described in Chapter 4, with the following question:

RQ4 How can we efficiently evaluate a sonification design?

We conducted two cycles of validations, where the evaluation methods were designed based on usability testing principles and aligned with the goals and hypotheses we had regarding the sonification design.

In Validation 1, the pretest-posttest design was sufficient to evaluate the learnability and effectiveness of the sonification design. Only the first layer (cf. definition 4.6) of sound was sonified because we aimed at whether the participants could learn the mappings between the sounds and elements and identify them. We used two different durations of recordings to evaluate the fast recognition of the sonified elements. Modifications on sounds have been implemented based on the results from Validation 1, for the purpose of improving the listeners' performance of identifying elements.

In Validation 2, we aimed to investigate how well the listeners are able to identify and localize two different layers of sound. We prepared two conditions of sound tests and used a within-subject design for the evaluation. In condition 1, only the first layer sounds were played and after 10 seconds the second layer sounds were added to them, In condition 2, the sounds of the two layers were played simultaneously from the beginning. The results collected from two conditions tests are valid to evaluate the factors that may influence the individual performance of the sound identification and localization, besides how many sounds listeners are able to hear and localize maximally. In conclusion, ***both validations effectively served the purpose of testing the design concept and exploring the variables, i.e. pitch, density and direction, that could potentially impact the identification and localization performances****. The results were analyzed using statistical methods, which have contributed to the advancement of evaluation and provide a solid basis for future investigation in the field of data sonification.*

7.1.5 Navigation through Sonification

In this thesis we have discussed about interaction with sound, within the context of audience participation and data sonification. A responsive system can provide a bidirectional interaction for participation. Moreover the feedback from the system can assist the subject to better understand the interaction.

Sound can be a medium to convey information of data. In contrast to visual perception, sound perception offers a broader spectrum of elements for interpretation, including pitch, volume, duration, rhythm, and more (Malikova, Adzhiev, Fryazinov, & Pasko, 2020). Interactive navigation is a good way helping to understand such relation between sound and data. Therefore, considering navigation with respect to the sound interactions that we have studied seems appropriate.

Navigation, as a concept in interaction, follows the principles of attention; i.e. the classical sequence of orienting, searching, filtering and expecting. For the learning of a sound-based interaction, first the orienting, searching and filtering are important. We can see this for the molecular sonification where the sounds provide the clues for the positions of the atoms and the navigation over the molecular structure feeds the attention sequence. To enable navigation as a feature, we first need to evaluate whether a sonification design is understandable and learnable. Different experiments were conducted based on the specific features of the sounds we aimed to investigate. If the learning substantiates, the expecting is fulfilled and the sonification is internalized

Navigation is an essential aspect of understanding an information space. We have started from a simple information space, represented by a molecule, which had a limited number of data elements and thus limited navigation possibilities. Nevertheless, without the navigation, the structure of the data cannot be understood. With respect to this simple information space, we demonstrated that sound interaction can support navigation. In order for this to be efficient, an individual sound associated with a data element should be recognizable, assisting in the localization of the data element. These two sound features address the attention assets of orienting, searching, and filtering. The expecting asset comes with the understanding, which relies on following the reasoning from the sensitivity tuple. This reasoning allows for answering questions regarding the information space, i.e., “Where can I go from here?” and “How do I get there?” The reduced information space of a molecule serves as a good example to study this.

With the sound design that has been developed for the molecule information space, the understanding is deterministic. This should mean that repetitive learning will decrease the error rate.

If the information space has dynamics; the reasoning is more difficult. Such can be seen in Băi, which is stochastic in nature. From reasoning over navigation in Băi, some prediction can be accomplished and the dynamics of the sound mappings with respect to behavior of the system can be revealed. The interesting aspect of the system is that it will never be the same, only exhibiting similar behavior. This as opposed to the molecule information space, where the state space is limited and is completely defined from the beginning of the sonification. In the stochastic system, the sensitivity tuple is more difficult to comprehend as the dynamic operation of the system requires continuous adaptation and reflection by the subject.

Together, these two distinct information spaces, with their completely different state spaces, provide interesting information for sound design in interaction and sonification. They demonstrate two manners of interactive navigation. While the navigation in Băi reveals the comprehension of interactive behavior and sound mappings. The navigation in a molecular structure focuses on the perception and understanding of an information space through abstract sounds.

In both case, navigation is an important feature for understanding and comprehending the information space – in one case from movement to an abstract sound space (Băi) and in the other case from abstract sounds to an ideation of data (Molecule). To this end, we have put effort in the evaluation of the sonification design so as to assess if it is understandable and learnable. We have investigated the features of the sounds with different experiments accordingly. As long as the sound can achieve the aim of recognition and localization, the sonification design can be applied to an extended use. Our work has pioneered the use of sound in interactive virtual environments and paves the way for other sonifcation design with more complex information spaces, deterministic or not.

7.1.6 Dialogue for Sonification

The navigation process is guided by the dialogue system, which plays a crucial role in facilitating learning and interaction and providing feedback. We have acknowledged the importance of the dialogue and we develop the dialogue from three components. We have distinguished the verbal, the subject and the adjective part. The adjective part is responsible for the sounds produced by the system, serving as auditory feedback. These sounds offered in the dialogue enable the participant to infer over the interactions and reflect on the responses from the system. These three components together form the complete sonification experience.

The interaction and navigation have physical parts, in the dialogue this is the verbal part. Subjects are able to generate actions based on the verbal part, which typically offer affordances, like a pendulum (Bǎi) or a Graphical User Interface (Molecule). Through navigation and interaction with these affordances, participants gain understanding and receive feedback that allows them to make inferences about future states or next steps. So, in conjunction with the navigation, the dialogue contributes to making the system learnable and understandable. The design of the feedback, i.e. the adjective part, is crucial to make the system work. For this matter the sound element is pivotal in the system.

7.2 Future work

In Chapter 1-6 we have obtained key findings and presented conclusions regarding the chosen topics (i.e., interaction models, data sonification, interaction & design, evaluation of data sonification). Building upon these insights, we would like to extend the application of our current findings to broader interactive participation context including data and sound, like audience participation in museums or student participation in classrooms. Moreover, data analysis can be supported by the sound representation in laboratory settings. In this section, we will discuss about the possibility of future research.

Interactive Experience

We have explored and discussed a lot about interactive experience in various contexts. We are curious to know how the *ideal framework* can be applied in the context of exhibition, museum, classroom and lab, where includes audience participation. Therefore new ways of potential interaction involving visual and physical perception can be created for the visitors or participants. For example, it is worth exploring and evaluating the interactive experience using extended reality (XR) techniques, as these have the potential to enhance exhibits and installations, transforming museums into vibrant and dynamic spaces. During COVID-19 many museums were kept closed, while some of them started to organise online events and provide virtual tours as compensation. Such digital interactive experience without physically presence is an asset derived from this special period. Similar situations happened in the context of education. The rise of online learning brings the challenge of how to improve the efficient study experience of students with a more interactive and engaged virtual environment. It is a great opportunity to explore the possibility of interactive experience over distance or virtually, as well as to study whether a *responsive interactive system* can be developed in such context.

Virtual Auditory Environment

Nowadays, virtual reality explores possibilities to enhance human perception and extend their immersive experience not only with visual feedback, but also with audio, haptic and other sensory information. Mazuryk and Gervautz summarized several benefits that auditory information can offer from previous virtual reality research, amongst them are spatial orientation cues, perception ability of information that is outside of visual display, possibility of parallel perception of many information streams (Mazuryk & Gervautz, 1996). Novo pointed out that an Auditory Virtual Environment (AVE) is like a Real Auditory Environment, composed of sound sources, a medium and a receiver. It aims to create situations in which human perceive the auditory events that correspond to a virtual environment(Novo, 2005). In our study, the evaluation of the sonification approach not only provides valuable insights in the context of auditory display contributes to the broader field of AVE. Although the results we got from the experiments still require further development and testing, *the key findings regarding concurrent sounding sources, including localization, segregation, and identification*, are related to the research fields of both auditory display and AVE. Factors such as masking effects, immediate sound recognition and localization, and the exploration of maximum concurrent sources might play significant roles in these domains. Moreover, there is a potential development in the field of extended virtual experience such as auditory navigation, either using headphones or an ambisonic system. We believe that our research paves the way for future investigation and advancements in these fields.

APPENDIX A

Supplementary materials for Chapter 4

Expert Review Checklist

Checklist

	Poor	Fair	Average	Good	Excellent
Learnablity					
I can recognize the similarities and dissimilarities among the sounds without introduction.					
After a brief introduction I can understand it and repeat it.					
Immediacy					
I am able to recognize one sound fast. (within 2 seconds)					
I do not have to remember each sound intentionally.					
Segregation					
I can easily separate one sound from other sounds when played in parallel.					
The sound is clearly heard when there are four same sounds from four directions.					
Localization					
The sound is easy to be localized from all directions.					
The sound is easy to be localized on all layers.					

APPENDIX B

Supplementary materials for Chapter 5

Instructions for Experiment I

Experiment Instructions

Thank you for agreeing to participate in our sonification study. The experiment will take you approximately 25-30 minutes. There are four speakers around you, which will play sounds during the experiment. You are free to change your head orientation, but please do not move the chair. The experiment consists of four phases:

Phase 1 Introduction (~ 2 minutes)

We designed four different sounds representing four chemical elements H, C, N, O. You can press the keys of h, c, n, o on the keyboard to playback the corresponding sounds, and press ESC to stop playing.

Phase 2 Pre-test (~ 7 minutes)

You will hear 28 sound samples, the duration of each sample will be 4 or 8 seconds. The sounds may come from four different directions (speakers) around you. Each direction will contain up to one sound source. You can use mouse (left click) to choose corresponding elements from each direction as you hear. You can change your head orientation during the test. You are allowed to leave uncertain part blank. You can press ENTER to start, each time when you finish answering please press ENTER again to go to the next question.

Phase3 Practice with feedback (~ 5 minutes)

You may have already found the differences among sounds. In this phase, you will hear several sound samples and be asked several questions as practice to get familiar with the sound design. Feedback will be provided on your answers.

Phase 4 Post-test (~ 7 minutes)

You will hear 28 sound samples, the duration of each sample will be 4 or 8 seconds. The sounds may come from four different directions (speakers) around you. Each direction will contain up to one sound source. You can use mouse (left click) to choose corresponding elements from each direction as you hear. You can change your head orientation during the test. You are allowed to leave uncertain part blank. You can press ENTER to start, each time when you finish answering please press ENTER again to go to the next question.

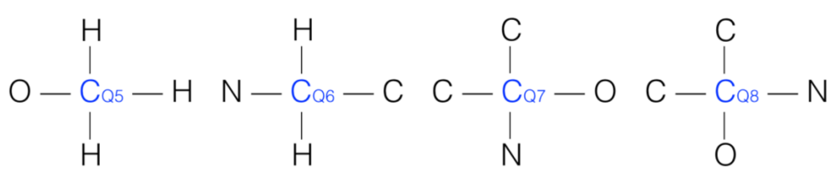
Training Session in Experiment I

No.____ Date__________

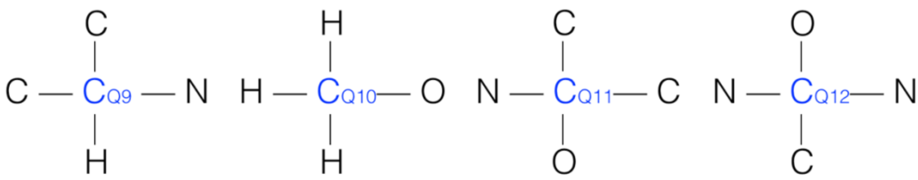
Practice with feedback

1) There are two sound sources around you. It is C from left, and what do you hear from right?
2) There are two sound sources around you. It is O from left, and what do you hear from right?
3) There are two sound sources around you. It is H from left, and what do you hear from right?
4) There are two sound sources around you. It is N from left, and what do you hear from right?

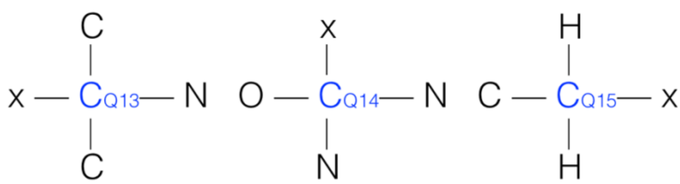
5) There are four sound sources around you. How many nitrogen atoms can you hear? Please point them out (direction).
6) There are three sound sources around you. How many hydrogen atoms can you hear? Please point them out (direction).
7) There are four sound sources around you. How many carbon atoms can you hear? Please point them out (direction).
8) There are four sound sources around you. How many oxygen atoms can you hear? Please point them out (direction).



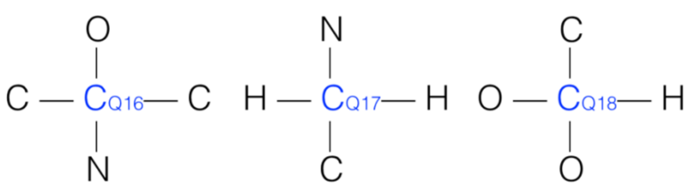
9) There is a carbon atom in front of you, what are the other three atoms around you? Please point out their directions and name them.
10) There is a hydrogen atom from your left, what are the other three atoms around you? Please point out their directions and name them.
11) There is an oxygen atom from your back, what are the other three atoms around you? Please point out their directions and name them.
12) There is a nitrogen atom from your right, what are the other three atoms around you? Please point out their directions and name them.



13) You will hear three sound sources, please point their directions and name each atom.
14) You will hear three sound sources, please point their directions and name each atom.
15) You will hear three sound sources, please point their directions and name each atom.



16) You will hear four sound sources, please point their directions and name each atom.
17) You will hear four sound sources, please point their directions and name each atom.
18) You will hear four sound sources, please point their directions and name each atom.



APPENDIX C

Supplementary materials for Chapter 6

Instructions for Experiment II

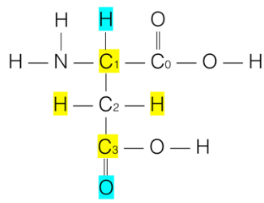
Experiment Instructions

Thank you for agreeing to participate in our sonification study. There are four speakers around you, which will play sounds during the experiment. You are free to change your head orientation, but please do not move the chair.

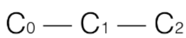
Phase 1 Introduction (~ 3 minutes)

Element: We designed four different sounds representing four chemical elements H, C, N, O. *We used pitch and density as two main features for the sound design in accordance with the weight differences of the four elements.*

Layer: If you are standing on C_2 (please see the figure right beside):
a) First layer: only sonifying the atoms directly connected to the current carbon position (marked as yellow ones).
b) Second layer: also sonifying atoms behind the directly connected atoms (marked as light blue ones).



Example: Now you are standing on C_0 (see the figure below), you will hear three sounds adding one by one from your right. C_1 is on the first layer, C_2 is on the second layer and H is on the third layer. *On one hand, the distance determines the loudness and the sound of C_1 is the loudest. On the other hand, C_2 has slightly higher pitch and more resonance, which becomes less sharp and intensive. Reverb is employed to enhance the sensation of distance of atoms in the second layer.*

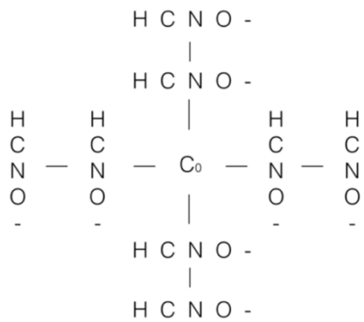


Phase 2 Training (~10 minutes)

You may have already found the differences among sounds. In this phase, you will hear several sound samples and be asked several questions as practice to get familiar with the sound design. Feedback will be provided on your answers.

Phase 3 Test - Condition 1 (~ 5 minutes)

You will hear 8 set of sounds. In each set there will be maximally 8 sound positioned around you. The first layer of sounds will be played first. After 10 seconds, the second layer of sounds will be played. Each layer will contain up to 4 sounds from four directions. Please use mouse (left click) to choose corresponding elements from each direction and layer as you hear. You can change your head orientation during the test. Please choose “-” if there is no sound heard.



Interface: You will hear three examples before the test, to get familiar with the interface.

Phase 4 Test - Condition 2 (~ 5 minutes)

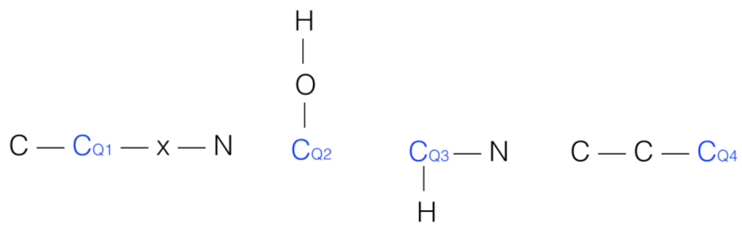
You will hear 8 sound samples, the duration of each sample will be 20 seconds. The sounds may come from four different directions (speakers) around you. Each direction will contain up to two layers of sound sources. You can use mouse (left click) to choose corresponding elements from each direction as you hear. You can change your head orientation during the test. Please choose “-” if there is no sound heard.

Training Session in Experiment II

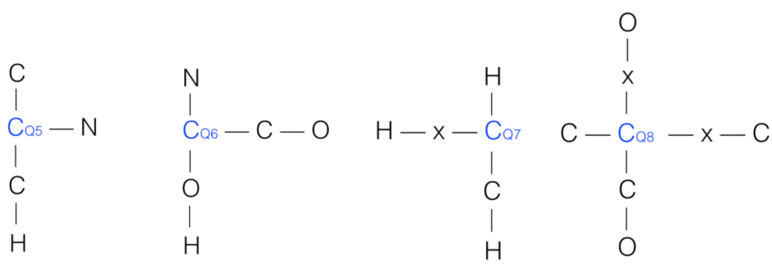
No.____ Date__________

Training

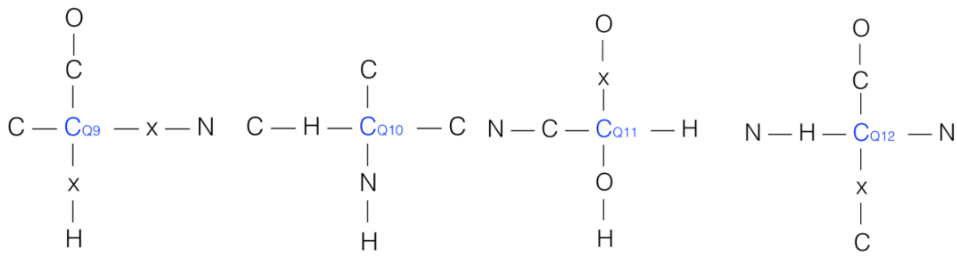
1) There is a nitrogen atom on the second layer. Now a new sound is added, which layer is this sound object on?
2) There is an oxygen atom on the first layer. Now a new sound is added, which layer is this sound object on?
3) There is a hydrogen atom on the second layer. Now a new sound is added, which layer is this sound object on?
4) There is a carbon atom on the first layer. Now a new sound is added, which layer is this sound object on?



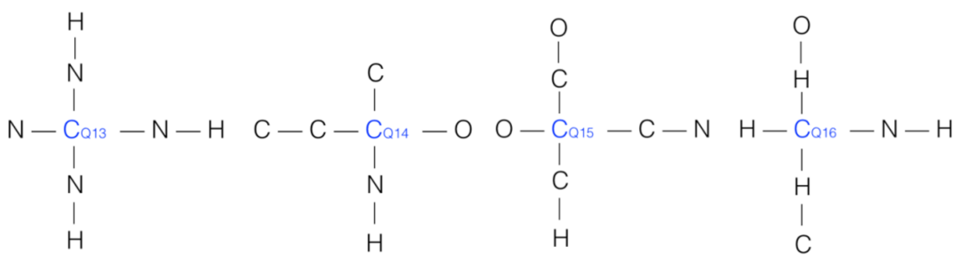
5) There are several sounds around you. How many nitrogen atoms are positioned on the first layer? Please point them out (direction).
6) There are several sounds around you. How many oxygen atoms are positioned on the first layer? Please point them out (direction).
7) There are several sounds around you. How many hydrogen atoms are positioned on the second layer? Please point them out (direction).
8) There are several sounds around you. How many carbon atoms are positioned on the second layer? Please point them out (direction).



9) There is a carbon atom in front of you, four atoms will be added around you one by one? Please point out their directions, layers and element name.
10) There is a hydrogen atom from your left, four atoms will be added around you one by one? Please point out their directions, layers and element name.
11) There is an oxygen atom from your back, five atoms will be added around you one by one Please point out their directions, layers and element name.
12) There is a nitrogen atom from your right, six atoms will be added around you one by one? Please point out their directions, layers and element name.



13) You will hear seven sound sources, please point out their directions, layers and name each atom.
14) You will hear six sound sources, please point out their directions, layers and name each atom.
15) You will hear seven sound sources, please point out their directions, layers and name each atom.
16) You will hear seven sound sources, please point out their directions, layers and name each atom.



References

Alkemade, R., Verbeek, F. J., & Lukosch, S. G. (2017). On the efficiency of a vr hand gesture-based interface for 3d object manipulations in conceptual design. *International Journal of Human–Computer Interaction*, *33*(11), 882–901.

Bayliss, A., Lock, S., Sheridan, J. G., & Campus, B. H. (2004). Augmenting expectation in playful arena performances with ubiquitous intimate technologies. *Pixel Raiders*, *2*, 6–8.

Bilda, Z., Edmonds, E., & Candy, L. (2008). Designing for creative engagement. *Design Studies*, *29*(6), 525–540.

Brungart, D. S., Simpson, B. D., & Kordik, A. J. (2005). Localization in the presence of multiple simultaneous sounds. *Acta Acustica united with Acustica*, *91*(3), 471–479.

Carlile, S. (2011). Psychoacoustics. *The sonification handbook*, 41–61.

Cipriani, A., & Giri, M. (2010). *Electronic music and sound design* (Vol. 1). Contemponet.

Dahl, L., Herrera, J., & Wilkerson, C. (2011). Tweetdreams: Making music with the audience and the world using real-time twitter data. In *Proceedings of the international conference on new interfaces for musical expression* (p. 272-275). Oslo.

Dahl, L., & Wang, G. (2010). Sound bounce: Physical metaphors in designing mobile music performance. In *Proceedings of the international conference on new interfaces for musical expression* (p. 178-181). Sydney.

Divenyi, P. L., & Oliver, S. K. (1989). Resolution of steady-state sounds in simulated auditory space. *The Journal of the Acoustical Society of America*, *85*(5), 2042–2052.

Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2003). *Human-computer interaction.* Pearson Education.

Freeman, J. (2012). *Saxophone etudes: for solo saxophone (any voice) with audience* *participation via mobile phones.* Retrieved from `http://distributedmusic.gatech.edu/jason/music/saxophone-etudes-2011-solo/materials/saxophone-etudes-score.pdf` (Accessed: Feb. 27, 2021)

Freeman, J., Chen, Y., Shen, W., Weitzner, N., & Xie, S. (2013). *Sketching: for 5-7 improvising musicians with audience participation via mobile phone.* Retrieved from `http://distributedmusic.gatech.edu/jason/music/sketching-2013-for-improvis/materials/sketching-score.pdf` (Accessed: Feb. 26, 2021)

Freeman, J., Xie, S., Tsuchiya, T., Shen, W., Chen, Y., & Weitzner, N. (2015). Using massmobile, a flexible, scalable, rapid prototyping audience participation framework, in large-scale live musical performances. *Digital Creativity*, *26*(3-4), 228–244.

Gamma, E., Johnson, R., Vlissides, J., & Helm, R. (1995). *Design patterns: elements of reusable object-oriented software* (Vol. 49) (No. 120). Addison-Wesley.

Gelfand, S. A. (2016). *Hearing: An introduction to psychological and physiological acoustics.* CRC Press.

Grand, F., & Dall Antonia, F. (2008). Sumo. a sonification utility for molecules..

Grond, F., Janssen, S., Schirmer, S., & Hermann, T. (2010). Browsing rna structures by interactive sonification. In *Proceedings of the 3rd interactive sonification workshop.*

Hartmann, W. M. (1983). Localization of sound in rooms. *The Journal of the Acoustical Society of America*, *74*(5), 1380–1391.

Hermann, T., Hunt, A., & Neuhoff, J. G. (2011). *The sonification handbook.* Logos Verlag Berlin, Germany.

Hindle, A. (2013). Swarmed: Captive portals, mobile devices, and audience participation in multi-user music performance. In *Proceedings of the international conference on new interfaces for musical expression* (p. 174-179). Daejeon.

Ibrahim, A. A. A., Yassin, F. M., Sura, S., & Andrias, R. M. (2011). Overview of design issues and evaluation of sonification applications. In *2011 international conference on user science and engineering (i-user)* (pp. 77–82).

Kato, M., Uematsu, H., Kashino, M., & Hirahara, T. (2003). The effect of head motion on the accuracy of sound localization. *Acoustical science and technology*, *24*(5), 315–317.

Kim, H. (2011). Moori: interactive audience participatory audio-visual performance. In *Proceedings of the 8th acm conference on creativity and cognition* (p. 437-438). Atlanta.

Lane, D. M., Scott, D., Hebl, M., Guerra, R., Osherson, D., & Zimmer, H. (2017).

Introduction to statistics. Citeseer.

Lee, S., & Freeman, J. (2013). echobo : Audience participation using the mobile music instrument. In *Proceedings of the international conference on new interfaces for musical expression* (p. 450–455). Daejeon.

Letowski, T. R., & Letowski, S. T. (2012). *Auditory spatial perception: Auditory localization* (Tech. Rep.). Army Research Lab Aberdeen Proving Ground MD.

Levin, G. (2001). *Dialtones: A telesymphony.* Retrieved from `https://www.fondation-langlois.org/html/e/page.php?NumPage=229` (Accessed: Mar. 03, 2021)

Ligna, & Röhm, J. (2003). *Dial the signals! radio concert for 144 mobile phones.* Retrieved from `http://ligna.blogspot.nl/2009/07/dial-signals-radio-concert-for-144.html` (Accessed: Mar. 03, 2021)

Liu, D. (2016). *Navigation in an audio maze game with quadraphonic and octophonic speaker set-ups* (Master Thesis, Leiden University).

Livingstone, D., & Miranda, E. R. (2004). Composition for ubiquitous responsive sound environments. In *Icmc.*

Lokki, T., Grohn, M., Savioja, L., & Takala, T. (2000). A case study of auditory navigation in virtual acoustic environments..

Malikova, E., Adzhiev, V., Fryazinov, O., & Pasko, A. A. (2020). Visual-auditory volume rendering of dynamic quantum chemistry molecular fields.

Mast, D., de Vries, S. I., Broekens, J., & Verbeek, F. J. (2021). The participant journey map: Understanding the design of interactive augmented play spaces. *Frontiers in computer science*, 45.

Mazuryk, T., & Gervautz, M. (1996). Virtual reality-history, applications, technology and future.

Miletto, E. M., Pimenta, M. S., Bouchet, F., Sansonnet, J.-P., & Keller, D. (2011). Principles for music creation by novices in networked music environments. *Journal of New Music Research*, *40*(3), 205–216.

Monahan, G. (1982). *Speaker swinging.* Retrieved from `http://www.gordonmonahan.com/pages/speaker_swinging.html` (Accessed: May. 17, 2018)

Munakata, N., & Hayashi, K. (1984). Basically musical. *Nature*, *310*, 96.

Neuhoff, J. G. (2011). Perception, cognition and action in auditory displays. *The sonification handbook*, 63–85.

Novo, P. (2005). Auditory virtual environments. In *Communication acoustics* (pp. 277–297). Springer.

Popper, A. N., Fay, R. R., & Popper, A. N. (2005). Sound source localization. , 272–316.

Rau, B., Frieß, F., Krone, M., Muller, C., & Ertl, T. (2015). Enhancing visualization of molecular simulations using sonification. In *2015 ieee 1st international workshop on virtual and augmented reality for molecular science (varms@ ieeevr)* (pp. 25–30).

Reich, S. (1968). *Pendulum music.* Retrieved from `http://www.ubu.com/aspen/aspen8/leadPendulum.html` (Accessed: May. 17, 2018)

Schraffenberger, H., & van der Heide, E. (2015). Audience-artwork interaction. *International Journal of Arts and Technology*, *8*(2), 91–114.

Shepard, M. (2006). Tactical sound garden [tsg] toolkit. In *3rd international workshop on mobile music technology.* Brighton.

Shi, X., Cai, Y., & Chan, C. (2007). Electronic music for bio-molecules using short music phrases. *Leonardo*, *40*(2), 137–141.

Stevens, S. S., & Newman, E. B. (1936). The localization of actual sources of sound. *The American journal of psychology*.

Tanaka, A. (2006). *Net_dérive - locative music for 3 mobile telephones and gps.* Retrieved from `http://www.ataut.net/site/Net-Derive` (Accessed: Jan. 3, 2017)

Tek, A., Chavent, M., Baaden, M., Delalande, O., Bourdot, P., & Ferey, N. (2012). Advances in human-protein interaction-interactive and immersive molecular simulations. In *Protein-protein interactions-computational and experimental tools.* IntechOpen.

Temple, M. D. (2017). An auditory display tool for dna sequence analysis. *BMC bioinformatics*, *18*(1), 221.

Thurlow, W. R., & Runge, P. S. (1967). Effect of induced head movements on localization of direction of sounds. *The Journal of the Acoustical Society of America*, *42*(2), 480–488.

Valentine, J. C., & Cooper, H. (2003). Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes. *Washington, DC: What Works Clearinghouse*, 1–7.

van der Heide, E. (2010). Spatial sounds (100db at 100km/h) in the context of human robot personal relationships. In *International conference on human-robot personal relationship* (pp. 27–33).

Van Troyer, A. (2012). *Hyperaudience: designing performance systems for audience inclusion* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Walker, B. N., & Nees, M. A. (2005). Brief training for performance of a point estimation sonification task. In *Proceedings of the 25th international conference on auditory display. (icad2005).* Limerick.

Weitzner, N., Freeman, J., Garrett, S., & Chen, Y. (2012). massmobile - an audience participation framework. In *Proceedings of the international conference on new interfaces for musical expression* (p. 21-23). Michigan.

York, R. O. (2016). *Statistics for human service evaluation.* SAGE Publications.

Zhang, L., Wu, Y., & Barthet, M. (2016). A web application for audience participation in live music performance: The open symphony use case. In *Proceedings of the international conference on new interfaces for musical expression* (p. 170-175). Brisbane.

Samenvatting

Deze thesis gaat over het gebruik van geluid in interactie in de context van auditieve participatieve systemen en data-sonificatie. Hierbij richt ons onderzoek zich op interactieve omgevingen waarin deelnemers informatie over gegevens waarnemen, dus horen, via geluidselementen. Om het interactieve proces te definiëren, maken we gebruik van het dialoogmodel, waarbij we het opsplitsen in drie componenten: het zogenaamde subject gedeelte, het verbale gedeelte en het adjective gedeelte. Met dit model wordt een beter begrip verkregen met betrekking tot de onderwerpen die in deze scriptie worden onderzocht, i.e. modellen voor interactie, data-sonificatie, ontwerp voor interactie, en de evaluatie van data-sonificatie. Daarenboven resulteert het onderzoek in nieuwe bevindingen en perspectieven in relatie tot deze onderwerpen.

De eerste stappen betreffen een verkenning van gedrag van deelnemers in een auditief en interactief systeem met betrekking tot geluid. Het publiek, de deelnemers, in een auditieve interactie dienen hierbij als voorbeeld om het gedrag van deelnemers (subjects) met betrekking tot de dialoog te onderzoeken. We hebben naar een aantal real-time participatieve muziekuitvoeringen gekeken en daarvan de dialogen tussen het publiek en het systeem geanalyseerd. In het merendeel van de onderzochte werken wordt geluid omgezet naar andere vormen van data. Onze analyse maakt duidelijk dat een immersieve en doorlopende interactieve omgeving kan worden ontwikkeld dat we benoemen als het ideale ontwerp kader (zie Hoofdstuk 2). Het gebruik van ons ontwerp kader zal ons helpen om het interactieve geluidsontwerp intuïtief te maken voor het publiek om te begrijpen. We hebben onze initiële bevindingen toegepast op twee case-studies:

In de eerste case-study passen we ons ideale ontwerp kader toe op een interactieve geluidsinstallatie - gepresenteerd als Bai (zie Hoofdstuk 3), teneinde om beter te begrijpen hoe geluid als een modus voor interactie en navigatie kan worden gebruikt. In deze installatie verandert het geluid volgens de intensiteit en duur van door de deel-

nemers gecontroleerde invoer, dit wordt gedaan door meting van de beweging van een pendule waaraan een speaker is bevestigd. Op deze manier wordt geluid gebruikt om een dynamische en responsieve relatie tussen de deelnemers en de installatie te creëren, hetgeen resulteert in een continu interactief systeem. De veranderingen in geluid dat wordt gecreëerd dienen daarbij als terugkoppeling (feedback) waardoor deelnemers kunnen redeneren over toekomstige toestanden of vervolgstappen bedenken. Daarom speelt het ontwerp van de terugkoppeling, dit is het adjectieve deel van de dialoog, een cruciale rol bij het mogelijk maken dat de deelnemers het sonificatie-ontwerp op een intuïtieve manier begrijpen en ermee kunnen navigeren. We bespreken observaties van deelnemers aan deze installatie en relateren dit aan onze benadering van geluidsontwerp.

In een tweede case-study bestuderen we een sonificatie-ontwerp van moleculaire structuren door middel van verschillende benaderingen voor een interactieve auditieve navigatie in deze moleculen (zie Hoofdstuk 4). We hebben een metafoor bedacht waarmee de verbinding tussen atomaire massa en toonhoogte kan worden gelegd. Hiermee kan een observant, het subject, de betekenis van geluiden begrijpen en daarbij de koppelingen, de metafoor, aanleren. Een evolutionair ontwerpproces van de geluidssynthese en compositie heeft een goed inzicht gegeven in hoe een geluidsontwerp stap voor stap kan worden ontwikkeld uit een molecuulstructuur, waarbij expertbeoordelingen leidend zijn voor de beoordeling van het succes.

Om de benadering voor sonificatie van moleculaire structuren te evalueren, hebben we twee verschillende validaties ontworpen en uitgevoerd zodat we de hypotheses die we willen onderzoeken kunnen toetsen. De testen worden uitgevoerd in een laboratoriumomgeving met testpopulatie van redelijke grootte. Er is gekozen voor een experimentele pre-post studie opzet om de leerbaarheid van de benadering voor sonificatie te kunnen beoordelen. Een ontwerp (within-subject) wordt gebruikt om de prestaties tussen twee condities te vergelijken. Dit stelt ons in staat specifieke kenmerken van de sonificatie te onderzoeken. Na een eerste reeks experimenten zijn aanpassingen gedaan op basis van de bevindingen. Dit formatieve onderzoek is essentieel voor ontwerpen en testen, waardoor een continue verbetering van een sonificatie-ontwerp mogelijk wordt. Het is belangrijk op te merken dat ons onderzoek een verkennend karakter heeft en is gericht op de evaluatie van een onconventionele weergave van gegevens, de benadering door sonificatie. Als gevolg hiervan moeten evaluatiemethoden, gebaseerd op gebruikerstesten, specifiek worden ontworpen om aan te sluiten bij het uiteindelijke doel van onze benadering.

Deze studies hebben aangetoond dat dialoog en interactiviteit kunnen worden ingezet om de kloof tussen complexe gegevens en menselijk begrip te overbruggen door gebruik te maken van geluid. Het onderzoek heeft het fundament gelegd voor meer en uitgebreider onderzoek met grotere populaties. De kennis opgedaan in ons onderzoek zal onderzoekers aanmoedigen om de mogelijkheden en innovaties op dit gebied verder te exploreren.

Summary

This thesis deals with the use of sound in interactions in the context of participatory systems and data sonification. We investigate an interactive environment where participants perceive information of the data through sound elements. To define the interactive process, we employ the dialogue model, breaking it down into three components: subject, verbal, and adjective. This supports the purpose of having a better understanding of the topics addressed in this thesis: i.e., interaction models, data sonification, interaction & design, and evaluation of data sonification. Moreover, it contributes new findings and perspectives to these topics.

We started by exploring behavior with respect to sound. To that end, we use the audience in a sound interaction as an example to investigate the behavior of the subjects (persons) in a dialogue. We reviewed a series of real-time participatory musical performances and analyzed the dialogues between the audience and the systems. In the majority of the related work, sound is controlled by or mapped to other forms of data. Our analysis reveals that an immersive and ongoing interactive environment can be developed within an ideal framework (see Chapter 2). Using our framework will assist us in making the interactive sound design intuitive for the audience to grasp and understand. We have applied our initial findings to two case studies:

In the first case study, we apply the ideal framework to an interactive sound installation - presented as Bǎi (see Chapter 3), to further understand how to use sound as a tool for interaction and navigation. In this installation, sound changes according to the intensity and duration of the control input from the participants, i.e., the measurement of the motion data of a pendulum speaker. In this way, sound is used to construct a dynamic and responsive relationship between the participants and the installation, which achieves an ongoing interactive system. Additionally, noticeable sound changes serve as feedback, allowing participants to make inferences about future states or next steps. Therefore, the design of the feedback, i.e. the adjective part

of the dialogue, plays a crucial role in enabling the participants to understand and navigate the sonification design in an intuitive way. We elaborate on the observations of participants of the installation and relate this to our approach to sound design.

In a second case study, we study the sonification design of molecular structures through various approaches to interactive auditory navigation (see Chapter 4). We establish a metaphor for the connection between atomic mass and pitch, which could assist the participants to understand the meaning of sounds and learn the mappings. An evolutionary design process of the sound synthesis and composition provided a good insight into how a sound is developed from data step by step based on the expert reviews.

Finally, we design and conduct two different validations to evaluate the sonification approach to molecular structures, depending on the hypotheses we aim to examine. Tests are accomplished with reasonably sized test populations in a lab setting. A pretest-posttest experimental setup is chosen to assess the learnability of the sonification approach. A within-subject design is used to compare the performance between two conditions, allowing us to investigate specific features. After the first round of experiments, adjustments have been made based on the findings from the previous experiment. This formative research is essential for design and testing, facilitating the continuous improvement of a sonification design. It is important to note that this is exploratory research aimed at evaluating an unconventional representation of data (i.e., sonification approach). As a result, evaluation methods, based on usability testing, have to be specifically designed to align with the goals of the approach.

These exploratory studies have shown that dialogue and interactivity can be employed to bridge the gap between complex data and human understanding mediated by sound. They have provided the foundation for more extensive investigations with larger populations. The lessons learned here will encourage fellow researchers to continue to push the boundaries of knowledge and innovation in this field.

中文摘要

这篇论文探讨了声音（Sound）在参与性系统（Participatory systems）及数据声音化（Data sonification）背景下的交互应用。交互环境通常是用户通过声音元素感知数据信息。为了定义这个交互过程，我们采用了对话模型，将其分为三个组成部分：主体（Subject)、言辞（Verbal）和形容词（Adjective）。这样不仅有助于更好地理解本文中所选择的主题（即交互模型，数据声音化：交互与设计，数据声音化的方法评估），而且还为它们添加了新的成果和研究视角。

我们从声音相关的行为开始探索。为此，我们以观众在声音交互中的行为为例，研究了对话中个体的行为。我们回顾了一系列实时的参与性音乐表演，并分析了观众与系统之间的对话。在大多数相关作品中，声音受到其他形式数据的控制或映射。我们的分析表明，可以在一个理想框架内开发出沉浸式和可持续存在的交互环境（参见第二章）。这个框架的应用将有助于观众更容易地理解和掌握交互声音设计。我们已将这些研究结果应用于两个案例研究:

在第一个案例研究中，我们将理想框架应用于一个交互式声音装置"摆"（详见第三章），以进一步了解如何将声音用作互动和浏览（Navigation)的工具。在这个装置中，声音会根据观众对于悬吊式音箱的控制强度和持续时间而变化，即根据摆动数据的测量结果。通过这种方式，声音在观众和装置之间构建了一个动态的、响应式的关系，从而实现了一个可持续交互的系统。此外，较为明显的声音变化，作为直接的反馈，可以帮助观众推断未来状态或是下一步的操作。因此，反馈的设计，即对话的形容词部分，能帮助参与者直观的理解数据声音化的设计方法，在多变的声音环境中遨游。我们也详细阐述了对于参与者的观察结果，并将其与声音设计所联系起来。

在第二个案例研究中，我们提出了多种不同的交互式听觉导览方案，以研究分子结构的声音化设计（详见第四章）。通过将原子质量轻重与音调高低类比，以帮助听者直观的理解声音的含义，并学习其中的映射关系。声音设计的迭代过程，为我们提供了深刻的见解。根据专家评审（Expert review）的建议，进行一步一步的修改，以得到最终的声音设计。

最后，根据需要检验的假设，本文提出了两种实验设计，以评估分子结构声音化

的方法。这些测试均在实验室环境中完成，采用了规模相对适中的测试人群。前后测实验设计用于评估数据声音化设计的易学性。而被试内实验设计则用于比较两种条件下的表现，以调查特定特征。基于第一轮的实验结果，我们对数据声音化的方案设进行了调整。这种形成性研究（Formative research）对于设计和测试的迭代至关重要，有助于不断改进声音化设计。值得注意的是，这是一项探索性的研究，旨在评估非传统的数据表达方式（即数据声音化方法）。因此，基于可用性测试的评估方法，必须通过专门设计以保持与实验目标一致。

本论文中所呈现的这些探索性研究，表明了对话与互动可用来弥合复杂数据与人类理解之间的鸿沟，特别是通过声音媒介。这些研究为更广泛的、包含更大样本的研究奠定了基础。这里所学到的经验将鼓励其他研究人员继续推动这一领域的知识和创新。

Acknowledgements

This book not only marks the end of my Ph.D. journey, but also contains all the memories behind the works and time I shared with my fellows, friends and family. It could never have started without the financial support by the China Scholarship Council (CSC). In the beginning it was not easy to make significant steps. And COVID-19 confronted me being in a lockdown in Wuhan that severely limited me. Nevertheless, I am still very happy at this moment to deliver my thesis with my sincere gratitude towards everyone who has helped me.

First of all, I would like to express my deepest appreciation to my promoter, Prof.dr.ir. Fons Verbeek. Thank you for encouraging me all the time since I started studying in Leiden. Your patience and openness brought me a great environment to conduct my research and complete this thesis. I am grateful for your guidance in crafting a great narrative storytelling, which has contributed to making this thesis easily accessible to a wider audience. Moreover, I would like to express my sincere gratitude to Edwin van der Heide. His efforts and guidance provided me with new insights that supported my research at the outset to take off. These ideas naturally developed and concluded in the various research topics. And I am grateful that Prof.dr. Aske Plaat was part on my supervision team.

Thank you, Yi Qin, my first supervisor, who brought me to the field of electronic music and interactive art. I always feel lucky to have met you at the Conservatory. Your support and enthusiasm made my time there determined and enjoyable.

Thank you, Jelger Kroese, for your playful collaboration in the Bǎi. It was so great to work with you so that we have finally made it to exhibit the installation at NIME.

A big thank-you to my parents, who always inspire and support me to explore everything. 感谢我亲爱的妈妈，无条件的支持我，爱我，让我一直拥有一条退路，

可以无所畏惧的去尝试我想做的事情。感谢我亲爱的爸爸，默默爱我，守护我，教导我做一个独立坚强的人。感谢我可爱的外婆，一直疼我护我，希望你身体健康，长命百岁。

Thanks all the people I met, worked and played with in the past years when I was in Leiden. I really treasure the time we spent together, best wishes to all of you. A special thanks to Taotao, who has designed the cover of this book. I am convinced that your talent will soon get noticed by more people. 感谢叶馥榕，罗交，张伟韬，马国栋，宋宇超，谢谢你们让我的独居生活多了许多色彩和风味，预祝一切顺利。谢谢我在Santa Clara的家人苏琰，曾嫚，总是开导我鼓励我，小虾苏聊还有机会更新。谢谢我在上海的家人，最棒的调音师李璐，这些年我们一直互励互勉，期待尽快见面。

Thank you all the people who participated my experiments, especially during the time of COVID-19 regulations.

Finally, I would like to thank Kelvin Kien Man Ko. Throughout all the time we have spent together, thank you for embracing my world, dealing with my tantrums, sharing in my frustrations, and providing emotional support. Thank you for starting next great journey with me <3.

Curriculum Vitae

Danyi Liu was born on the 19th October 1990 in Wuhan, China. After graduating from Music Technology and Arts at Shanghai Conservatory of Music and Applied Psychology at East of China Normal University, Danyi did an internship at the Institute of Acoustics and Speech Communication at TU Dresden. This marked the outset where she started to integrate her knowledge of sound design with psychology. Subsequently, Danyi pursued her Master's Degree in Media Technology at Leiden University, where she developed her interest in interactive interface design and user experience research.

Because of her interest in the interactions with diverse media types, she pursued her doctoral research in Human-Computer Interaction. Throughout her Ph.D. journey, she extensively explored and delved into interactive experiences and human-centred methodologies across various contexts. Reflecting on her decade-long journey, she has progressively transitioned towards interdisciplinary studies. Now, she continues her career as a human experience researcher and independent sound designer.