Interaction with sound for participatory systems and data sonification

Liu, D.

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.*

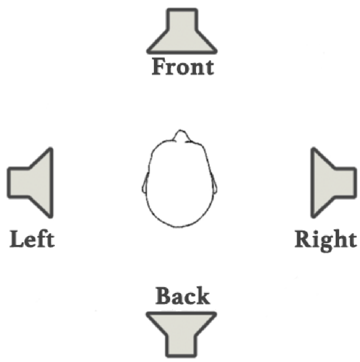


Figure 4.1: Positions of four speakers setup.

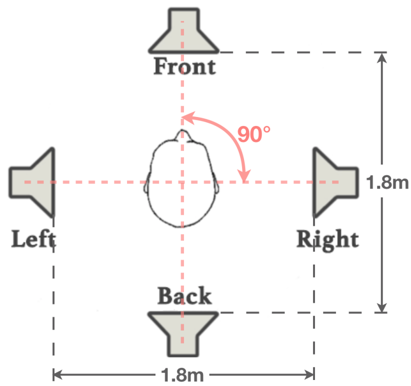


Figure 4.2: Implementation of the four speakers setup.

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.

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).

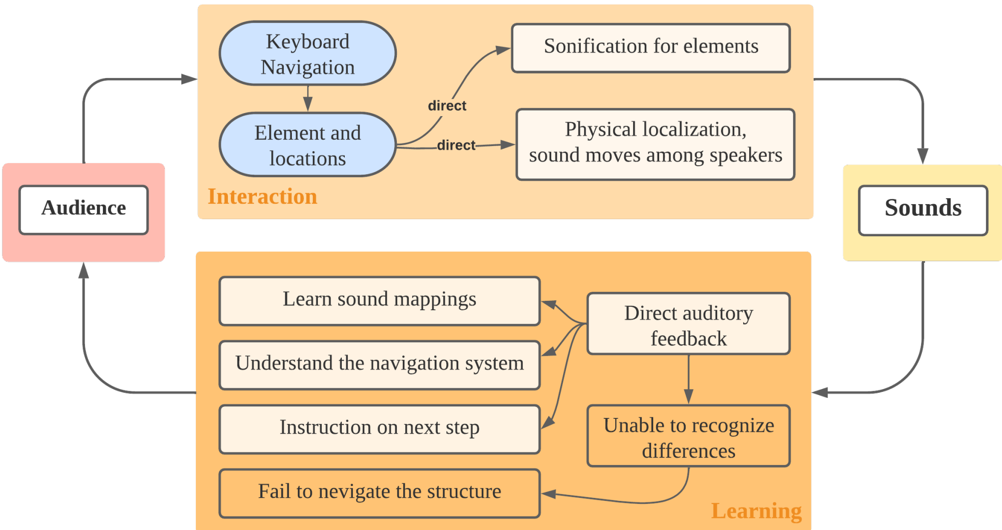


Figure 4.5: Ideal framework for interactive navigation in a molecular structure. The color code refers to the elements of the dialogue model: subject-audience, verbal-actions, adjective-sounds.

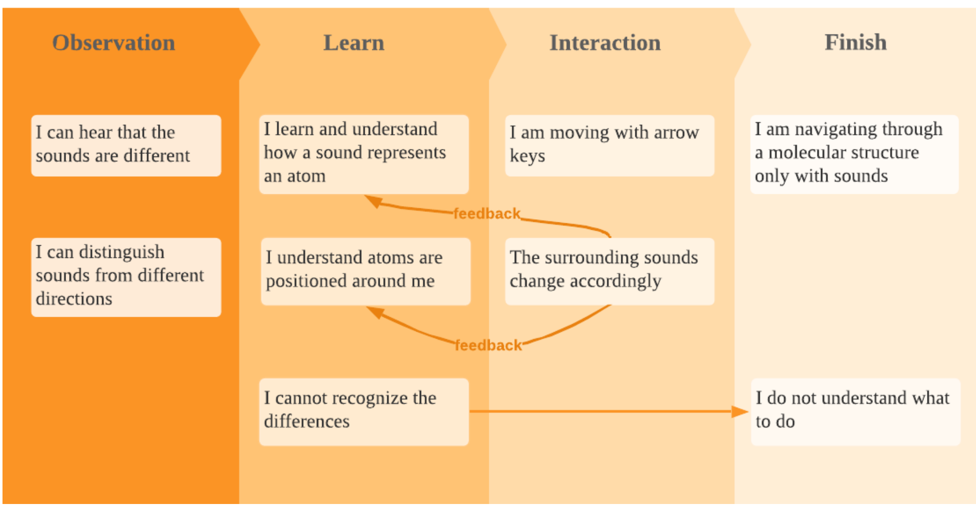


Figure 4.6: Participation journey map of interactive navigation in a molecular structure.

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:

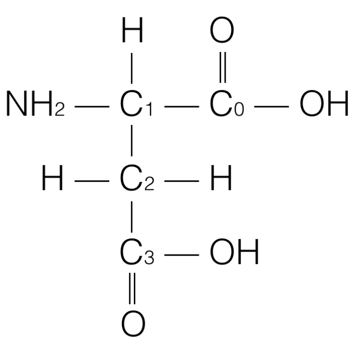


Figure 4.7: The structural formula of Aspartic acid.

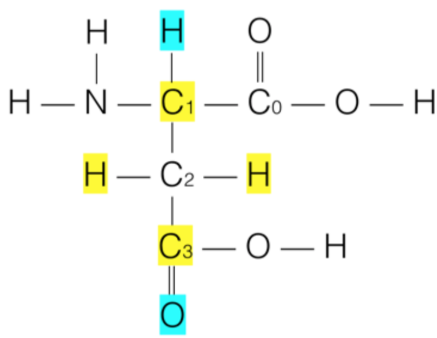


Figure 4.8: The structural formula of Aspartic acid for the larger area sonification.

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.*

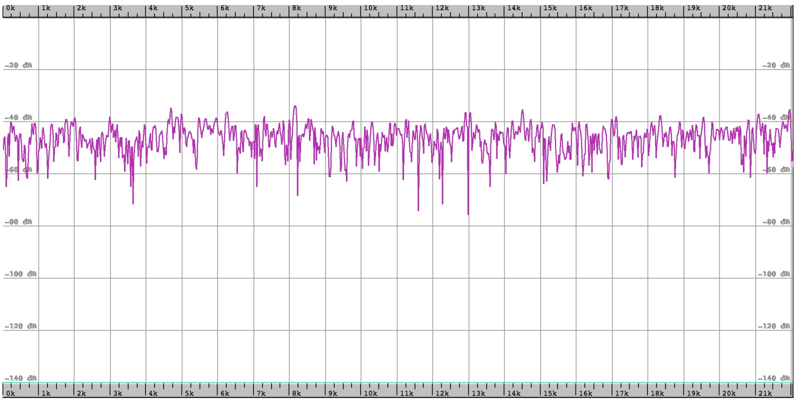
Envelope

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).

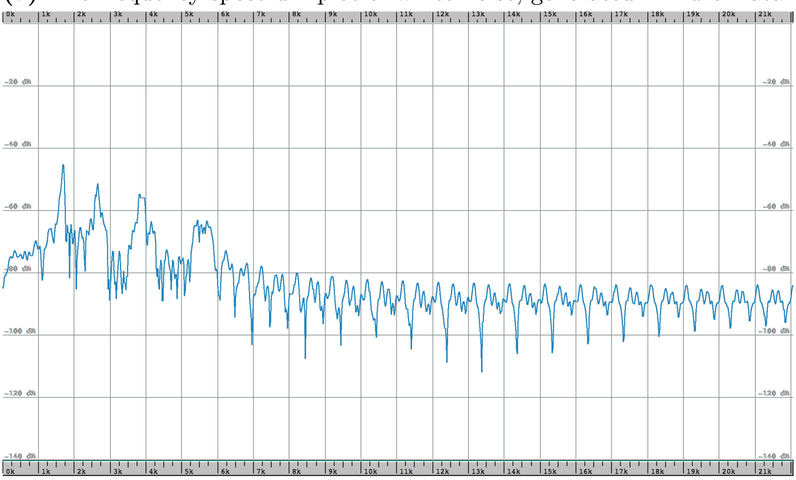
Synthesis setup I: Our initial attempt is to use different drum samples because the timbre of different parts from a drum set (e.g. bass drum, snare drum, hi-hat) can be easily distinguished and these percussion sounds are short and easy to localize. In our first prototype, hydrogen was mapped to closed 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

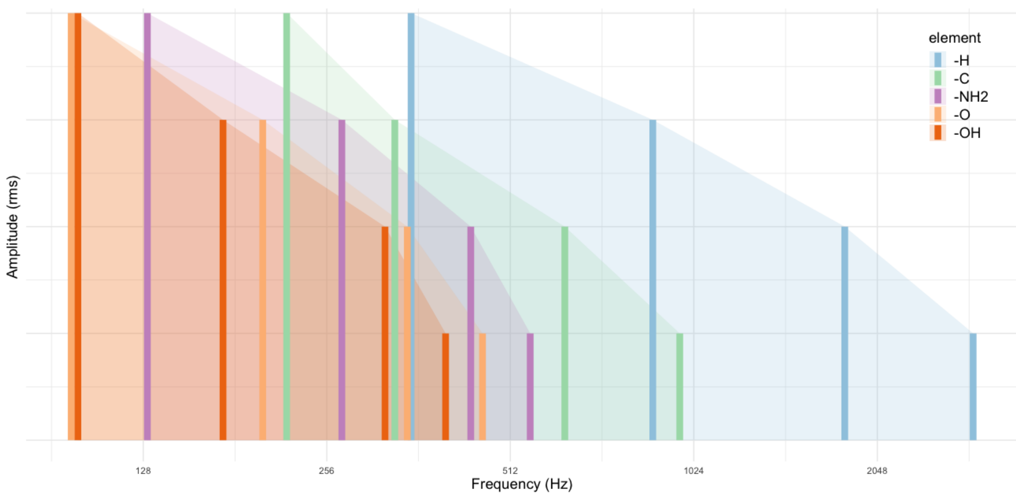


Figure 4.10: Frequency components for each element, synthesis setup 3. The shaded areas indicate regions of overlapping frequencies.

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).

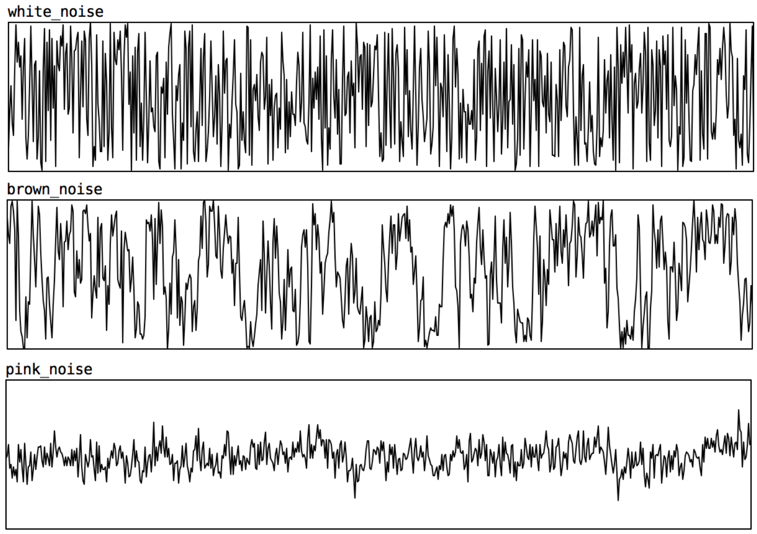


Figure 4.11: Three temporal structures of colored noise, showing amplitude changes over time (generated in Pure Data). The x-axis represents time, while the y-axis represents amplitude. The colored noise exhibits distinct patterns and fluctuations in its amplitude, providing a visual representation of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

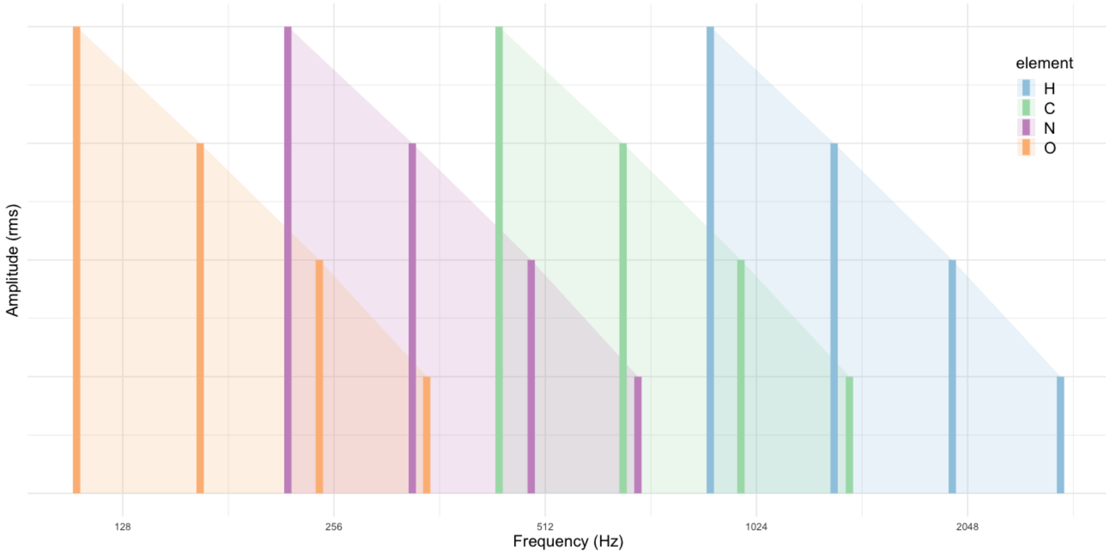


Figure 4.12: Frequency components for each element, synthesis setup 4, with octave separations. The shaded areas indicate regions of overlapping frequencies.

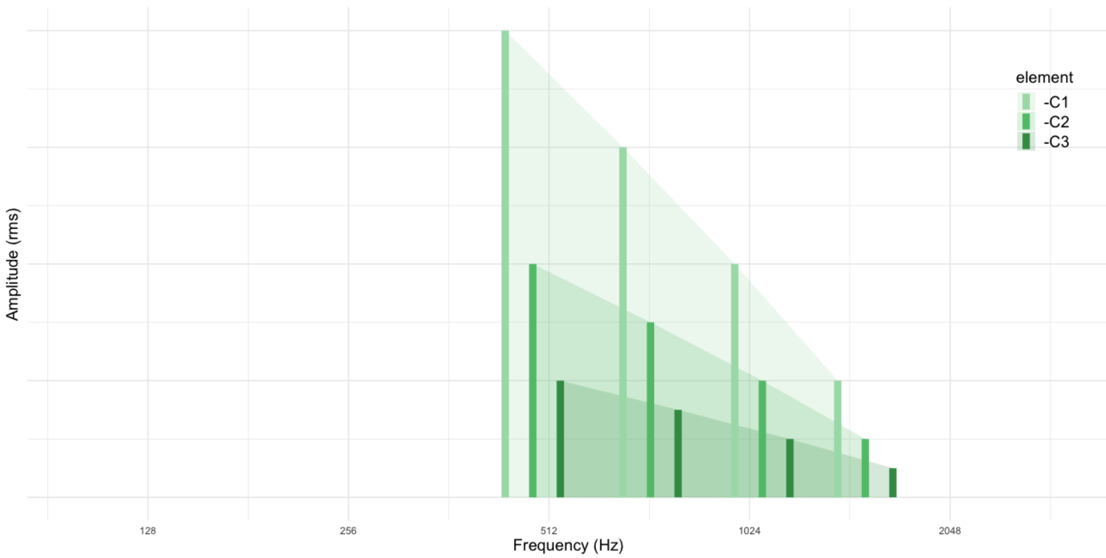


Figure 4.13: Frequency components for each carbon atoms on different layers. The shaded areas indicate regions of overlapping frequencies.

	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).