Interaction with sound for participatory systems and data sonification

Liu, D.

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

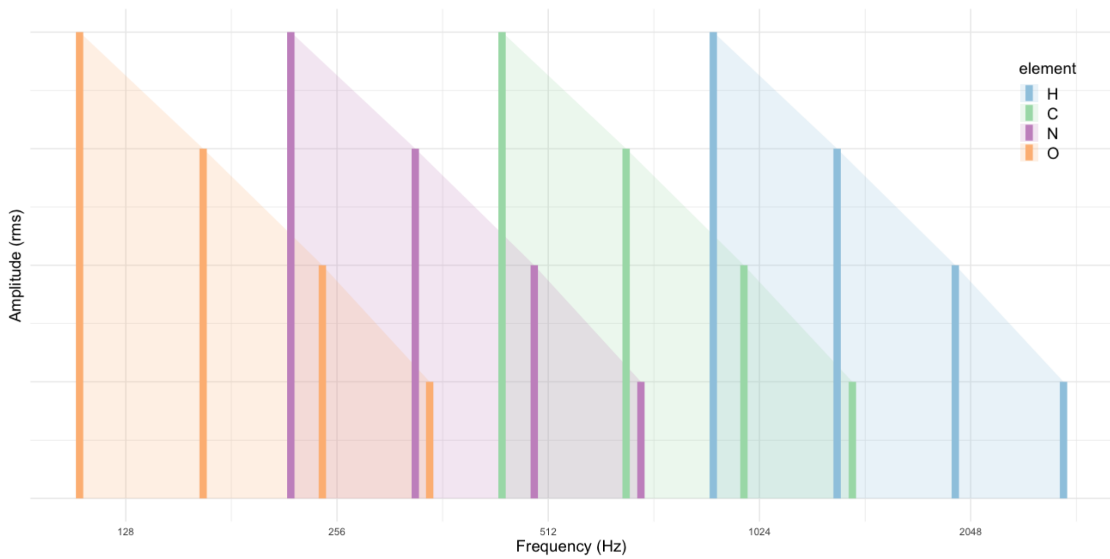

Figure 5.1: Frequency components for different elements. The graph shows the frequency components of four certain sounds representing four elements, highlighted using different colors. The shaded areas indicate regions of overlapping frequencies. The intervals between the components are identical for 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

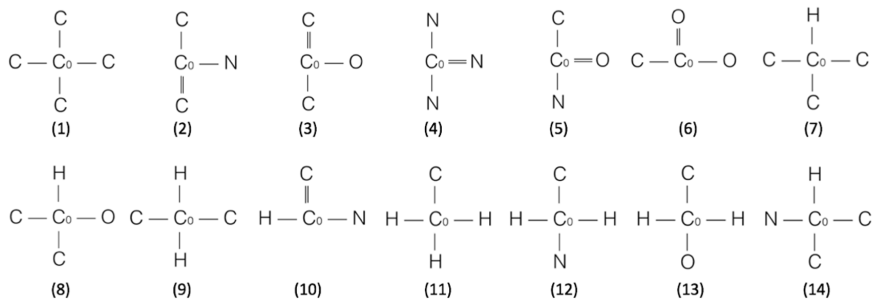

Figure 5.2: 14 structural formulas for Validation 1.

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

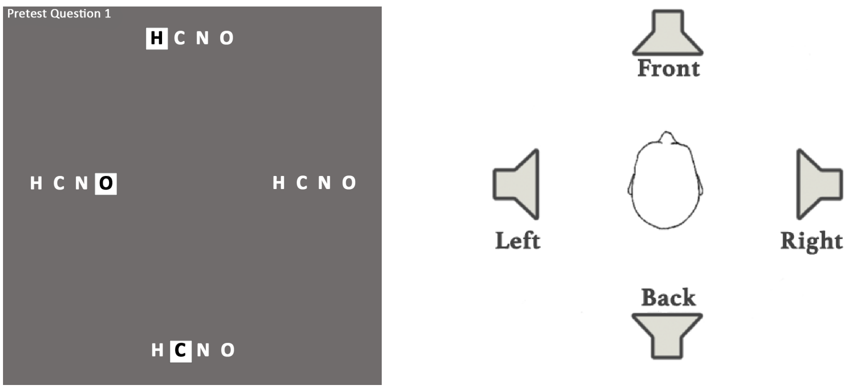

Figure 5.3: A screenshot of the user interface for the participant to indicate the sounds they heard during the experiment of Validation 1. In the user interface, the up and down directions correspond to the front and back speakers, while the left and right directions correspond to the left and right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

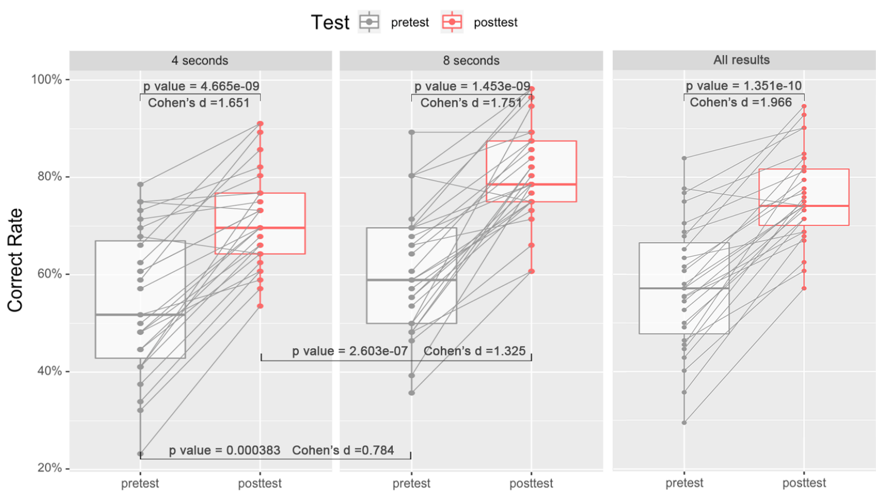

Figure 5.4: A visual comparison of correctness rate between pretest and posttest for different recording durations (4 seconds, 8 seconds and all). Each data point displays the correctness rate of a participant, while the lines connecting them illustrate the individual changes in correctness rate between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

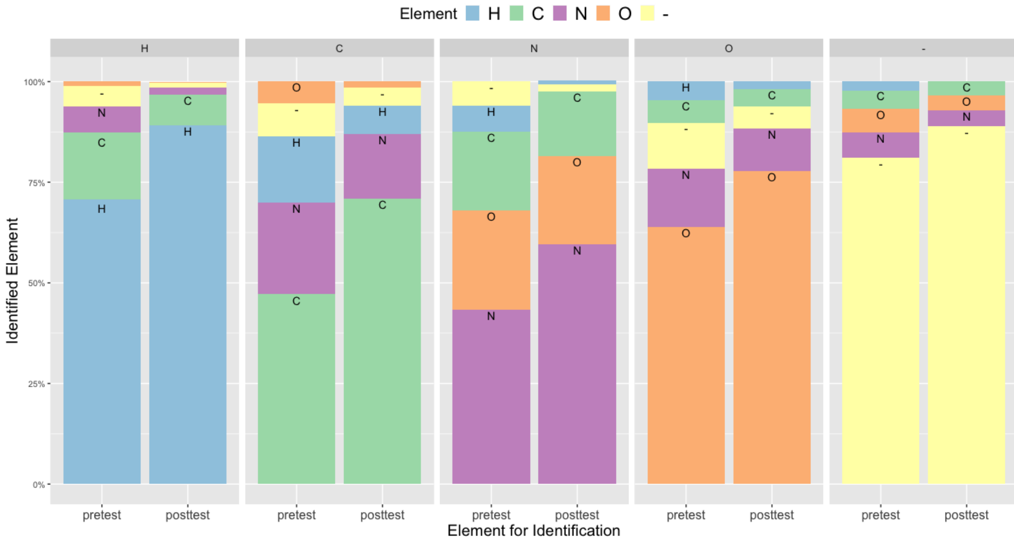

Figure 5.5: Stack diagram depicting the percentages of correct identifications and misidentifications of elements during the experiment, ordered from low to high atomic mass.

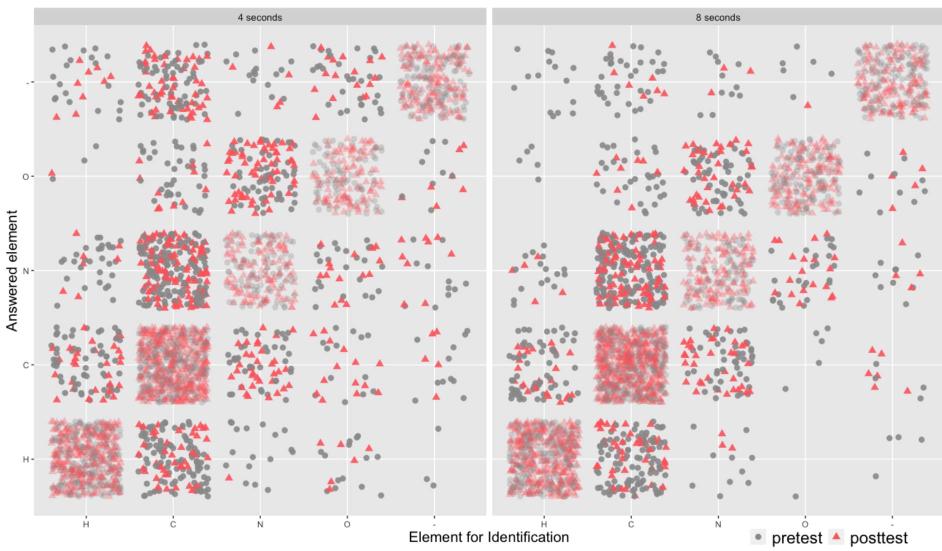

Figure 5.6: Distribution plot illustrating the accuracy of participants' identification of target elements with both 4 seconds and 8 seconds recordings. The x-axis represents the elements to be identified (H, C, N, O, -), and the y-axis represents the elements that participants answered.

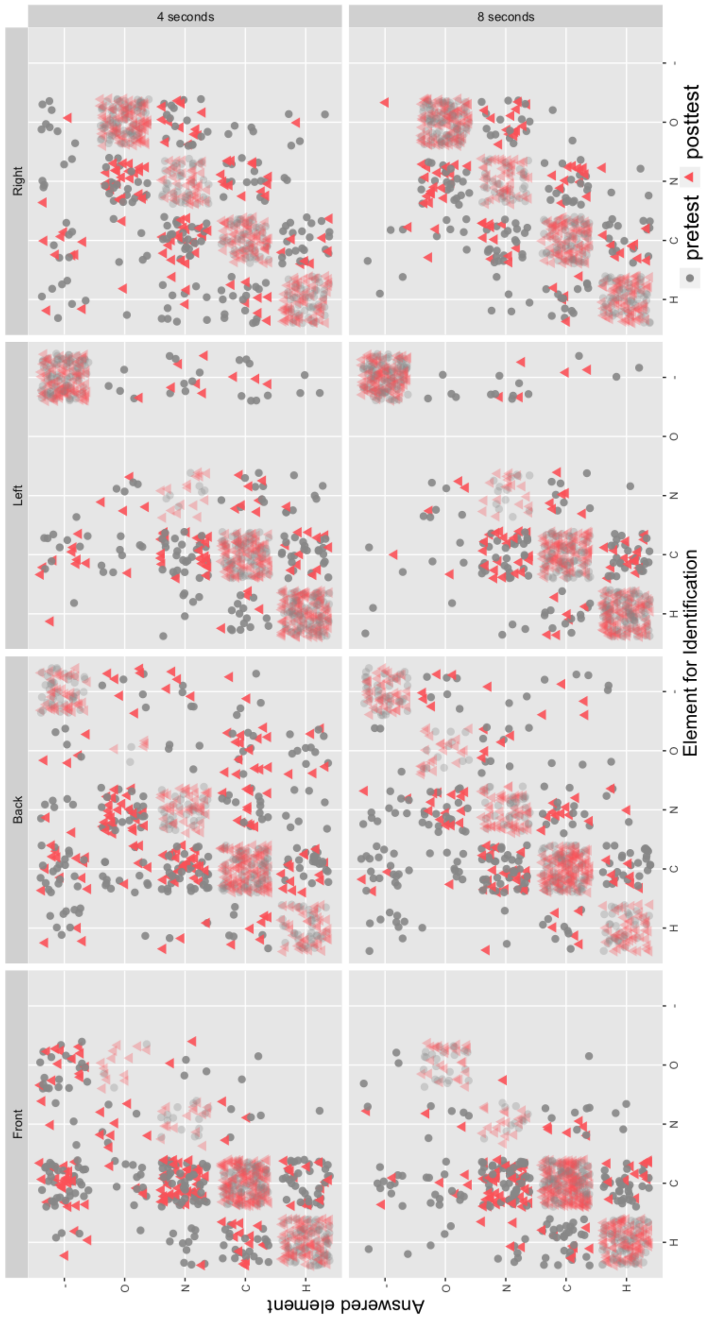

Figure 5.7: Distribution plot depicting participants' responses for identifying elements positioned from four directions (front, back, left and right). The x-axis represents the elements to be identified (H, C, N, O, -), and the y-axis represents the elements that participants answered. Please turn the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

5

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

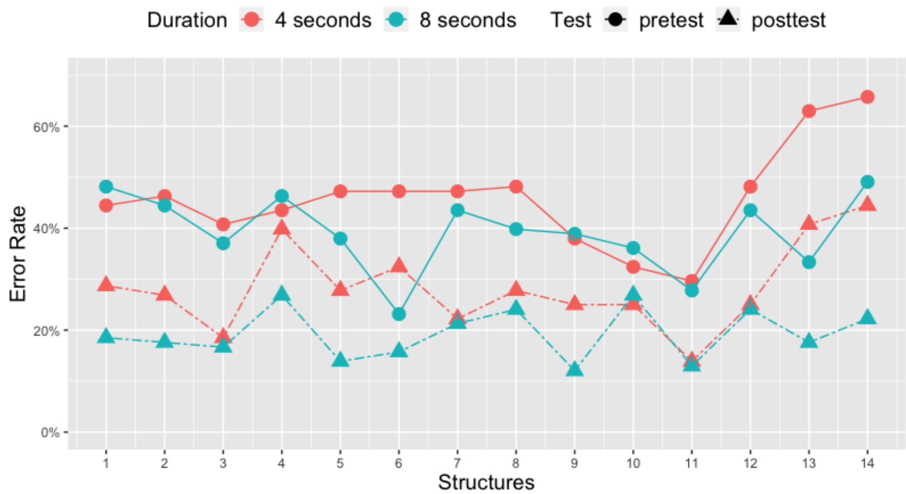

Figure 5.8: The plot illustrates the error rate (%) for each of the 14 structures (cf. Figure 5.2) in Validation 1. The lines connecting the data points within each group serve as a visual grouping mechanism, indicating that the data points belong together based on the shared characteristics of color (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.