Interaction with sound for participatory systems and data sonification

Liu, D.

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.*

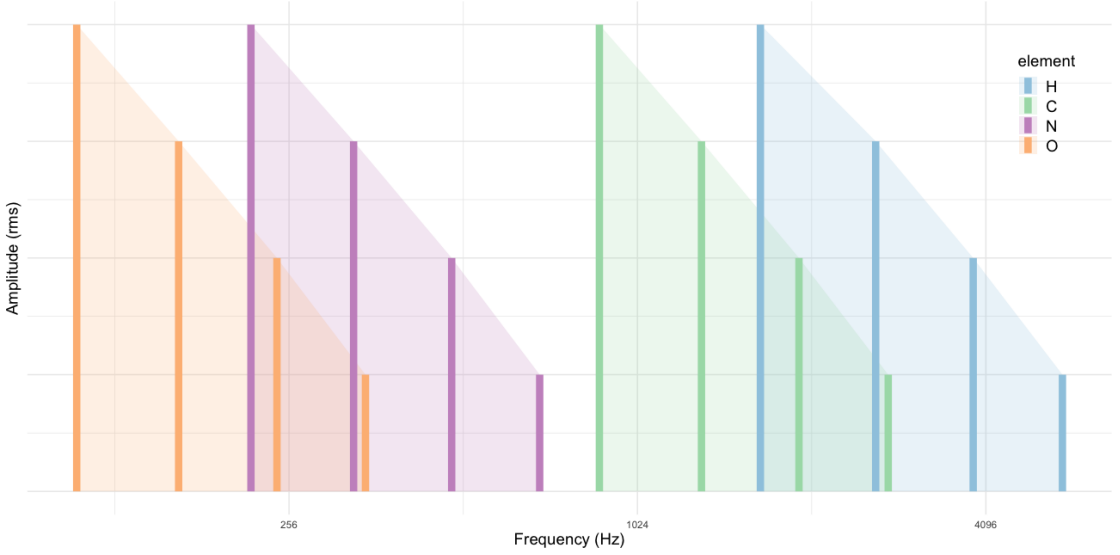

Figure 6.1: Frequency components for different elements. The x-axis represents frequency in Hz, and the y-axis represents amplitude in rms. The graph shows the frequency components of four certain sounds representing four elements, highlighted using different colors. This representation emphasizes the 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.

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.

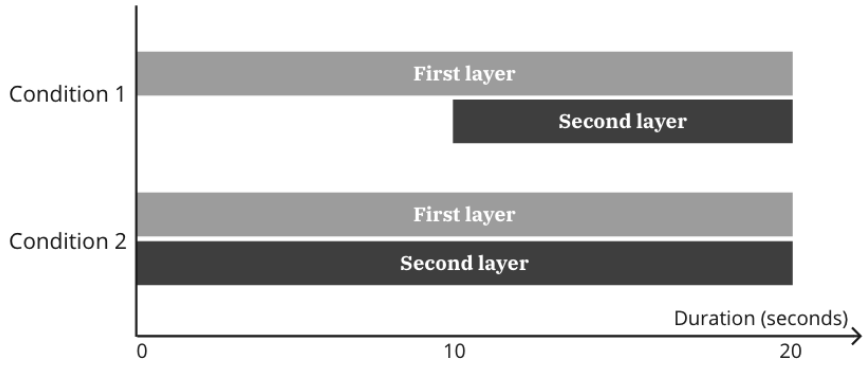

Figure 6.2: Visualization of the sequential presentations of sound sources in two conditions. The x-axis represents duration in seconds. The graph illustrates distinct timing patterns in the experimental setup.

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

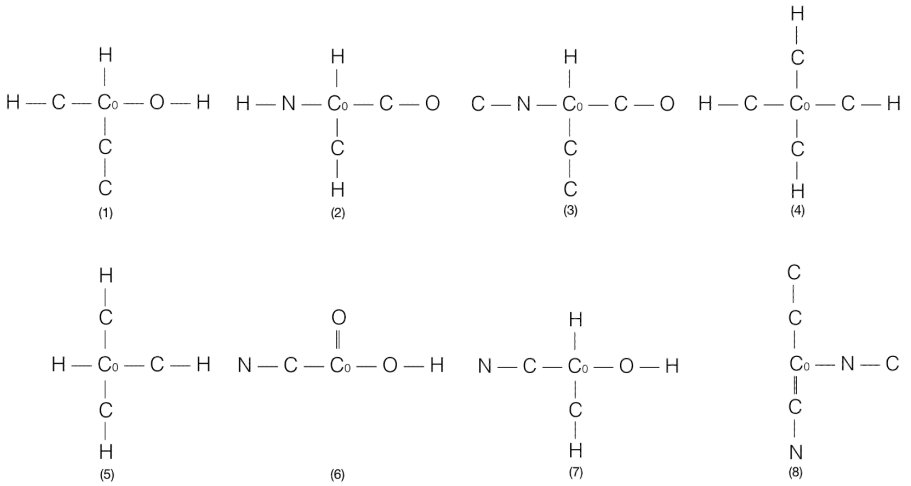

Figure 6.3: 8 structural formulas for Validation 2 (2 layers). Structure 1 is an extension of structure 8 from Validation I. Structure 3 is derived from structure 14, Structure 4 is derived from structure 1, Structure 6 is derived from structure 6, Structure 8 is derived from structure 2, Structure 9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

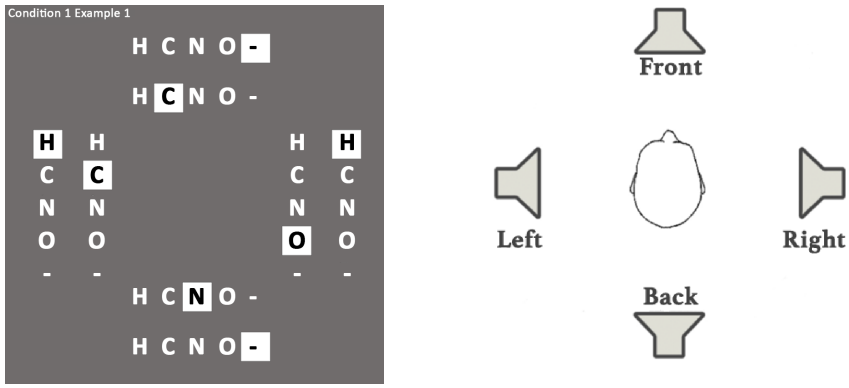

Figure 6.4: A screenshot of the user interface for the participant to indicate the sounds they heard during the experiment of Validation 2 (2 layers). In the user interface, the up and down directions correspond to the front and back speakers, while the left and right directions correspond to the left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

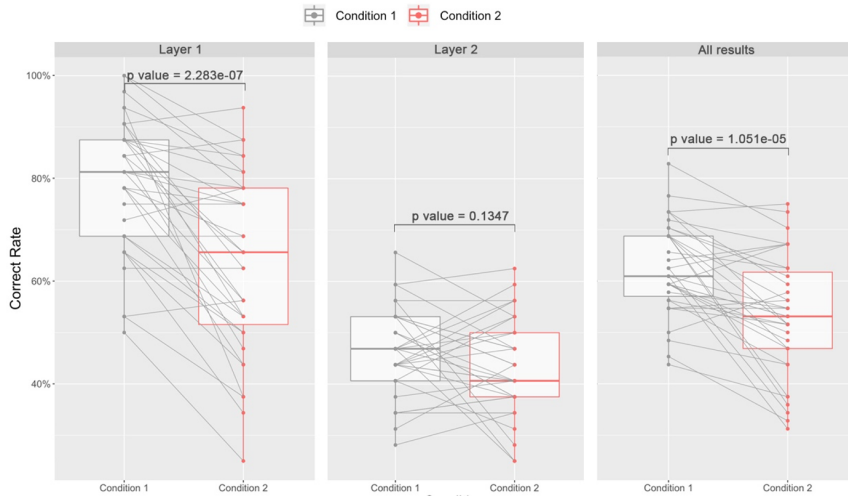

Figure 6.5: A visual comparison of correctness rate between condition 1 and condition 2 for different layers of sounds. The data points displayed as circles to illustrate the individual changes of the average correctness rate.

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

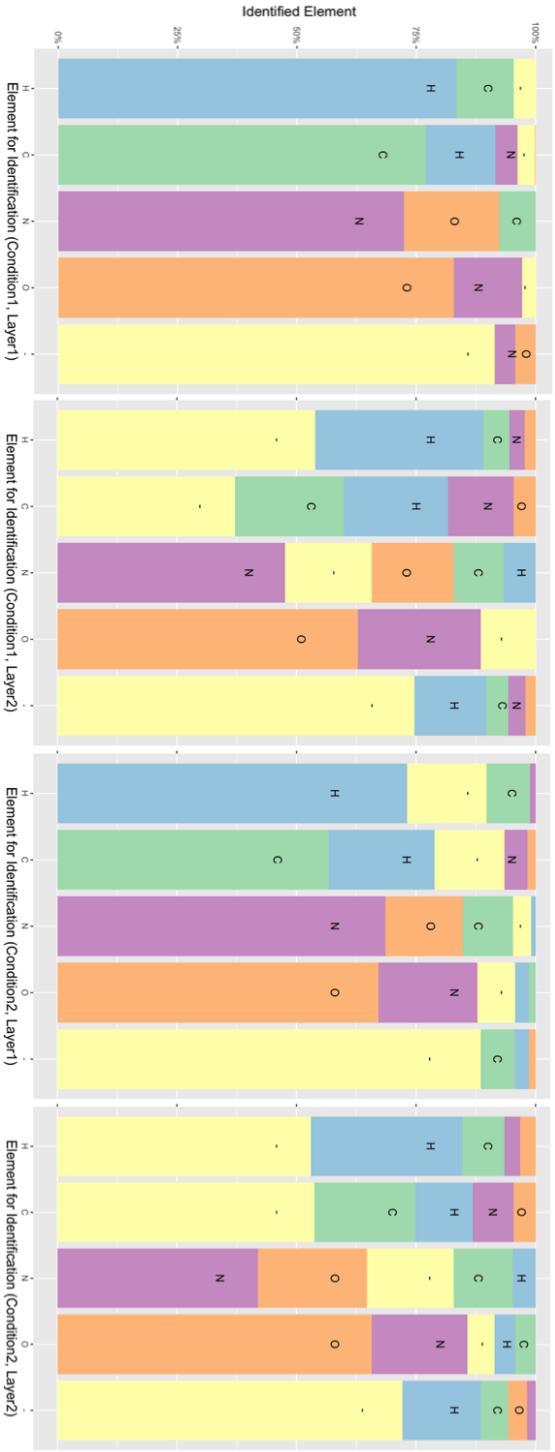

Figure 6.6: Stacked bar plot of element identifications. The y-axis displays the percentage of identifications, with the bars stacked in descending order from highest to lowest percentage. The different colored segments within each bar represent the proportions of correct and misidentifications of different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

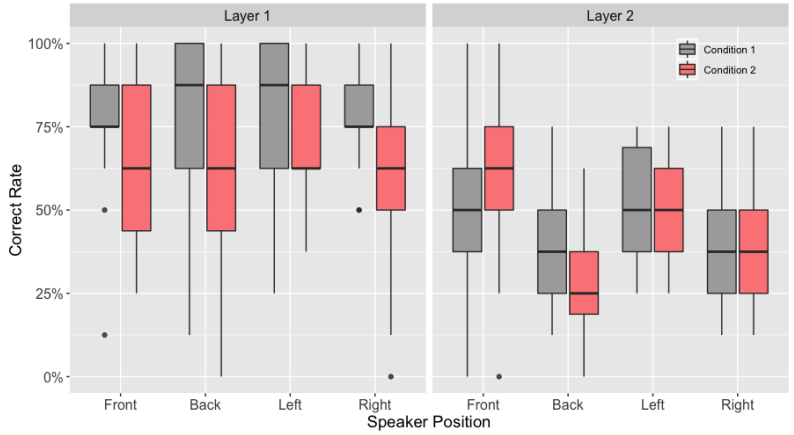

Figure 6.7: Boxplots with whiskers representing the distribution of correctness rates (%) for four directions (front, back, left and right) in both conditions. The boxplots provide an overview of the variations in correctness rates across conditions, directions, and two layers of the sounds. The whiskers 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.

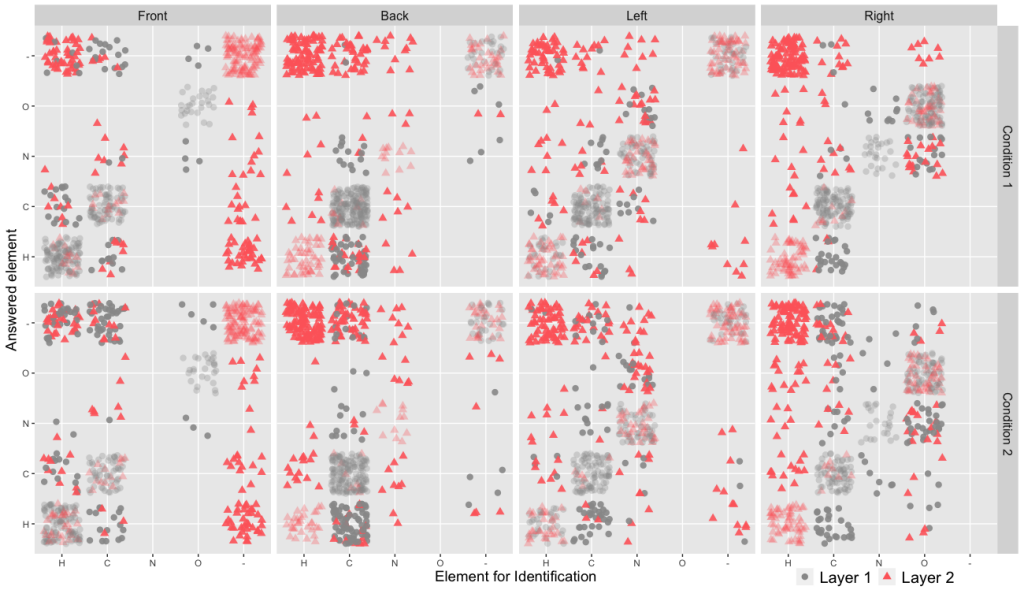

Figure 6.8: Distribution plot illustrating the accuracy of participants' identification of target elements in both conditions, from four directions (front, back, left and right). The x-axis represents the elements to be identified (H, C, N, O, -), and the y-axis represents the elements that participants 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.

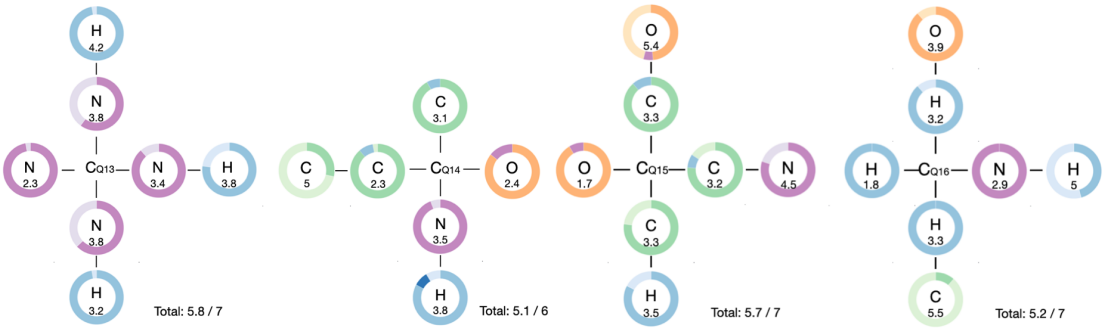

Figure 6.9: Visual representation of the correctness rate identifications for questions 13 to 16 during the training phase (see Appendix C, training session). The size of each colored segment represents the proportion of correct identifications for each element, with larger segments indicating higher 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.