EVALUATING THE SPATIAL SONIFICATION OF THE MOLECULAR STRUCTURES OF AMINO ACIDS USING MULTIPLE CONCURRENTLY SOUNDING SOURCES

Danyi Liu

Leiden Institute of Advanced Computer Science
Snellius Gebouw, Niels Bohrweg 1
Leiden, 2333CA, The Netherlands
d.liu.7@liacs.leidenuniv.nl

Edwin van der Heide

Leiden Institute of Advanced Computer Science
Snellius Gebouw, Niels Bohrweg 1
Leiden, 2333CA, The Netherlands
e.f.van.der.heide@liacs.leidenuniv.nl

ABSTRACT

We have designed an interactive form of sonification in which the listener navigates through the molecular structures of amino acids over the network of carbon atoms. We use pitch and density as the two main features for the sound design of the four common chemical elements (H, C, N, O). We use multiple concurrently sounding sources to spatially sonify the atoms around a certain carbon atom of the amino acids. The main goal of this paper is to evaluate this sonification approach. We cover the design, execution and evaluation of two separate cycles of experiments that aim to evaluate our sonification approach and get insight in factors that may influence the individual performance of the concurrently sounding source identification and localization.

1. INTRODUCTION

In the context of auditory display and sonification, sound has been used to represent complex data, enhance visualizations, as well as support the understanding of subjects in an educational context. The sonification of the molecular structures of amino acids forms a case study to research concurrently sounding sources that are spatially distributed We have designed an interactive form of sonification in which the listener navigates over the network of carbons [1]. We started with the structural formulas of amino acids because of their relatively easy structures compared to the three-dimensional structures of proteins or other molecules. We have transformed these structures spatially so that they have become flat and using bond angles of either 90 or 180 degrees, and identical bond lengths (see Figure 1). We use a four-speaker setup (see Figure 2) to simplify the sound localization task so that the speakers always correspond to the actual direction of the sound sources. In order to enable the navigation in the structures, specific rules have been made. Listeners are only able to navigate through the molecule by moving from one carbon atom to its neighboring carbon atom(s), and so on. We have performed two cycles of experiments to evaluate, and further develop, our sonification approach. We do this to explore the potential of identification and localization of multiple concurrently sounding sources in the horizontal plane. In experiment 1, only the atoms directly connected to the current carbon position are sonified (we call this the first layer). For experiment 2, we have developed a more extended version of the sonification approach in which the atoms behind the directly connected atoms are sonified as well (we call this the second layer).

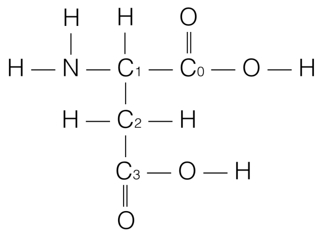


Figure 1: A structural formula example

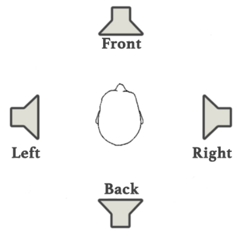


Figure 2: Speaker setup

Sonification approaches are based on the human's auditory system, which is highly capable of deriving the three auditory dimensions that are commonly used in auditory display: loudness, pitch and timbre, from what we hear [2]. With these primary features, humans are able to separate and identify different sound sources, each with their own characteristics. In our design, the developed sounds are used to represent the type and position of the atoms from molecular structures. Four sounds are designed to represent the four elements common in amino acids: carbon (C), hydrogen (H), oxygen (O) and nitrogen (N). We use pitch as the main feature for the four sounds because the differences are easily perceivable and distinguishable. Each sound consists of four partials. We intentionally use the term partials (and not harmonics) because we synthesize a non-harmonic spectrum. The intervals between the partials are identical for each of the four sounds. The partial ratios are: 1 : 1.5 : 2.2 : 3.2. We use octave intervals between the four sounds. The pitch of the four sounds corresponds to the weight of each element, the lighter the element, the higher the assigned pitch. In experiment 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 440 Hz and the first partial of hydrogen is 880 Hz. The spectra of the four sounds are shown in Figure 3. Changes to this have been made in experiment 2, which will be explained in section 3. 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 [3]. We hope to achieve a similar improvement in the ability to localize a sound by using only four frequency partials for each of the sounds. We realize that with these four partials we cover a substantially smaller frequency range than Hartman.

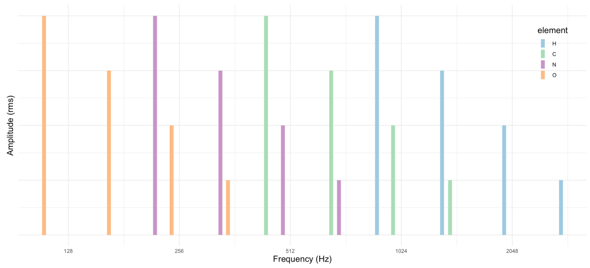


Figure 3: Frequency components for each element

Unlike time-based melodies or other sequentially played sounds, our research focuses on concurrently sounding sources that have a continuous nature. We play the combination of elements simultaneously (each sound originating from its own speaker) and want the elements to be easily and quickly recognizable. In order to do this, we give every sound its own irregular amplitude pattern whereby the density depends on the atom type. When two or more identical atoms are being played (each originating from its own speaker position) they share the same pitch, but each atom has its own irregular (and thereby asynchronous) amplitude pattern. Like this we aim to avoid the merged perception of two or more identical atoms. We use noise sources in combination with comparators to create these irregularly timed impulses structures and apply different densities for the different atom types. We use a second noise source to randomize the amplitude of the generated impulses. For each of the sounds, the impulses are sent to a combination of four resonant bandpass filters, each with a high q factor. This way the bandpass filters act as the oscillators/resonators of the individual partials. 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 [1]. We believe it is intuitive to associate a faster pattern with a lighter atom.

The main goal of our sonification design is to enable the listeners to localize and identify the surrounding atoms in an environment of concurrently sounding sources. We used molecular structures as a case study but a similar approach can be applied to other contexts where sounds are used to represent multiple objects or attributes. We have designed two cycles of experiments to evaluate our approach step by step and see whether our design choices and assumptions can be substantiated. In previous evaluations of sonification applications, users 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 [4]. One of the ten tasks they described is an identification task, which can be used to investigate the ability of sounds to be perceived and recognized uniquely. 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 participants' performance of identification with different combinations of sound sources and matching them with corresponding elements and positions.

2. EXPERIMENT I

The first experiment has been conducted to investigate up to which extent the participants can learn and remember the mappings between the sounds and elements and identify each atom in different constellations of multiple concurrently sounding sources. Only first-layer of sounds are presented in this experiment, consequently, up to four sounds are positioned around the participant at the same time. An important aspect of experiment 1 is to investigate the immediacy - the time it takes to recognize the sonified elements that are surrounding the listener. With our approach we create environments in which all the sounds are present in parallel. By doing this we hope to achieve immediacy. The irregular structure is experienced as a kind of granular-like texture. This approach not require participants to remember a concrete sound or a specific rhythmical pattern and compare with each other.

Bruce and Walker designed five experimental training conditions for the participants to experience between the pretest and the posttest of a 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 promt or an interactive presentation with both voice-over and visual explanation. The study showed that practice with feedback may be more effective compared with other situations [5]. We have chosen to divide experiment 1 in three stages: a pretest, a practicing session with feedback of correct answers, and 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 the posttest. According to what Bruce and Walker concluded, we would hypothesize that participants would be able to learn and comprehend our approach and perform better with practice and feedback. Calculation of the effect size is also necessary, in order to measure the amount of gain when comparing pretest and posttest results [6].

2.1. Materials

From each direction there could be positioned one of five possible options, including four elements and none. The total amount of possible combinations can be up to 625 (5^4). However, it is not realistic to implement all of the possible combinations during the experiment. The sonification approach is designed for the interactive navigation in molecular structures of amino acids, thus we looked through possible combinations of directly connected atoms of carbon atoms among the structures of the 20 natural amino acids. From this we have chosen 14 possible combinations (see Figure 4), which have been used for the experiment.

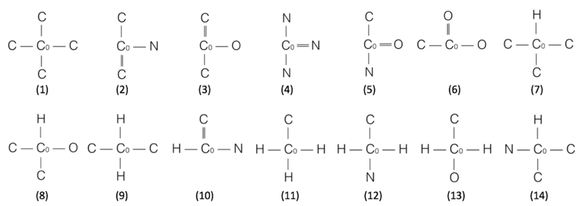


Figure 4: 14 possible structures used in Experiment I

[1]Recording examples of each element can be found at: `https://surfdrive.surf.nl/files/index.php/s/QuaO9KYnnL1JUel`

The irregular impulses are generated by differently colored noise in combination with a comparator with a variable threshold, which results in random impulses. In order to avoid the auditory differences of generating a same element in real-time while taking the experiment, we decided to use recordings of all possible combinations. So that each participant will hear exactly same auditory results in a randomized order.

We would rather not ask the participants to finish the questions as soon as possible, in case it may bring anxiety. Therefore we chose to use two different playback durations: four seconds and eight seconds. This enables us to compare the participants' performance between the two different durations.

2.2. Procedure

The experiment consisted of four phases. Firstly, the participants were introduced to the four different sounds representing four elements H, C, N, O. They were told that the perceived frequency irregular pattern had been mapped to the weight of each element. They could press the keys of h, c, n, o on the keyboard to playback the corresponding sounds. After they felt they were able to recognize the sounds, they would start the pretest. They were aware that sounds would come from four surrounding speakers, and there would be up to one sound source on each speaker. Additionally, they were allowed to change the head orientation during the experiment.

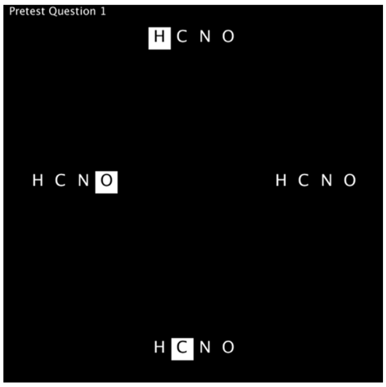


Figure 5: A screenshot from the user interface (Experiment I)

During the pretest, 28 recordings of 14 structures of both 4-second and 8-second durations were played back in a randomized order varied from participants. After a structure was played the participants were asked to indicate, in a simple screen-based interface, for each speaker which element they heard (H, C, N, O or none) originating from that position (see Figure 5). The elements were displayed in the order from light to heavy (H, C, N, O). After indicating the corresponding atoms they had to press ENTER to go to the next structure. If they did not hear any sound coming from one speaker or they were unable to identify the sound, they could leave it blank, which would be marked as nothing ("-").

The pretest is followed by a training session would be a training part after the pretest. 18 training questions were prepared in this part and the participants would get feedback upon giving their answers. 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. Localization task was added later. In the last six questions, participants were told how many sound sources they would hear and were asked to point out their directions and name each atom. Participants took the posttest after the training, which included the same 28 recordings but in a differently randomized order. 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?

2.3. Results

27 participants joined the experiment. There were 17 male and 10 female participants. Most participants were between 20 and 30 years old. One participant was 39 and one 46 years old. The participants were all students and staff from Leiden University. Most of them have basic knowledge about chemistry or biology, but none of them have a professional background in the field of chemistry or biology, or have done related research. We mentioned before that for each of the 14 presented structures in the pre- and the posttest, we record the answers given for each of the 4 directions. In order to calculate a correctness score per presented structure we use the following point system: every correctly identified element in a given direction counts for 0.25 points. This leads to a total score per question of 0, 0.25, 0.5, 0.75 or 1.

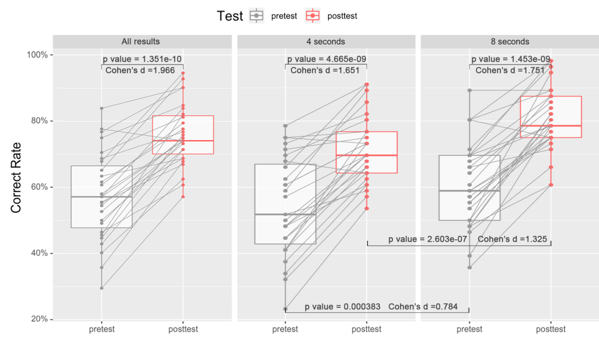


Figure 6: Correct rate of all participants (Experiment I)

Figure 6 shows a general result from all participants in both pretest and posttest. We have implemented a paired t-test on the correct rate of all the participants. The p-value is 1.351e-10, which is below the significant level 0.05. Therefore we can conclude that there is a significant increase in the correct rate between the pretest and the posttest. The mean correct rate for the pretest is 57.2% and for the posttest is 75.9%, and the effect size is 1.966, which put the difference in gain between the two tests at the 98_{th} percentile. It is indicated that the participants' performance in the posttest was a lot better. Since individual training part was around 5 to 7 minutes, it was concluded that people were able to learn this sonification approach relatively fast. Meanwhile, we reviewed the correct rate of the 4-second and 8-second recordings separately (see Figure 6). In the pretest, the p-value of 0.000383 shows a significant difference between the results of the 4-second and 8-second recordings though, the effect size is 0.784 at medium level. In the posttest, the p-value between the results of the 4-second and 8 second recording is 2.603e-07 and the effect size is 1.325. This seems to imply that the duration difference may influence slightly more on the participants' performance of identification and localization in the posttest. We also compared different recording durations between the pretest and the posttest individually. For the results of the 4-second recordings, the mean difference between the two tests is 17.4%, the p-value is 4.665e-09 and the effect size is 1.651. For the results of the 8-second recordings, the mean difference between the two tests is 19.9%, p-value is 1.453e-09 and the effect size is 1.751. There is a statistically significant change of the correct rate after the practice for both durations of recordings.

2.3.1. *Elements*

The table in Figure 7 shows that there is a significant difference of correct rate for all the elements between the pretest and the posttest. Column '-' represents for the situation where no sound was played. It can be observed in the table that the correct rate of nitrogen is relatively low as well as the p-value for the difference between two tests. The correct rates for hydrogen (70.7%) and oxygen (63.9%) are already high in the pretest. In the stacked bar chart (figure 7b), the x-axis represents the elements that were played including none and the y-axis shows the result of the identified element ('-' shows the situations where no 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 mainly misidentified as nitrogen (14.5%) or nothing (11.4%). In the posttest, hydrogen's correct rate reaches to 89.2% and it was mainly mistaken for carbon (7.5%). The correct 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 correct rate increased to 77.8% and misidentification rate as nitrogen decreased to 10.5%. The correct 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%). Inference of possible explanations will be discussed in the later section.

	H	C	N	O	-
P-value	0.0001207	2.449e-08	0.002381	9.564e-05	0.01457
Mean(pre-test)	70.7%	47.2%	43.3%	63.9%	81.2%
Mean(post-test)	89.2%	70.9%	59.5%	77.8%	88.9%

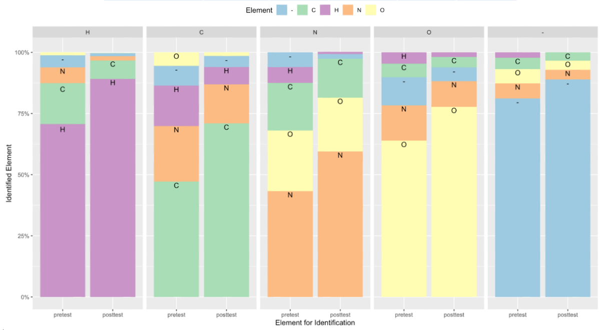


Figure 7: Correct rate of each element (Experiment I)

Figure 8 gives more detailed information about the duration influence on element identification. Participants performed better with 8-second recordings in general. It can be observed that there were more times of nothing heard in 4-second recordings, especially in the pretest. It could be that because of the short duration the participants may not have had enough 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.

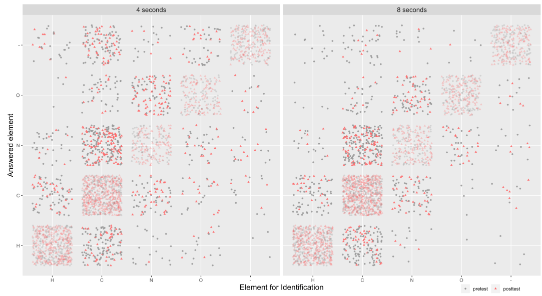


Figure 8: Distribution of element identification with 4-second and 8-second recordings (Experiment I)

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

Figure 9: Distribution of element identification with position differences (Experiment I)

2.3.2. *Directions*

From figure 9, we can see that the front direction in the 4-second recordings was often wrongly identified as nothing, mainly when it's carbon or oxygen, while the back direction in both 4- and 8-second recordings was sometimes wrongly identified as nothing. 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 for oxygen a lot at back speaker, and for carbon and oxygen at right speaker. In general, wrongly identified points of the posttest were less than the ones of the pretest and participants performed better with 8-second duration recordings from both front and back speakers. The performance of front and back speaker was much worse in the pretest.

2.3.3. *Structures*

Figure 10 shows the error rate of each structure independently, and most structures' error rate in the pretest is lower than in the posttest, the only exception is structure 6. 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 structures 2, 3, 4, 5, 6, 10 (see Figure 4). But the error rate of structure 4 and 10 are higher even in the posttest (see Figure 10). This implies that the identification of nitrogen only 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 structures 1, 7, 9 11, there is a transformation from four carbon atom to the combination of carbon and hydrogen atoms (see Figure 4). Together with figure 9, we found that it would be easier and 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. And it would be easier and faster to recognize one hydrogen atom and three carbon after the practice (structure 7). This may give an indication that the sound of hydrogen is easier to learn and remember than the sound of carbon.

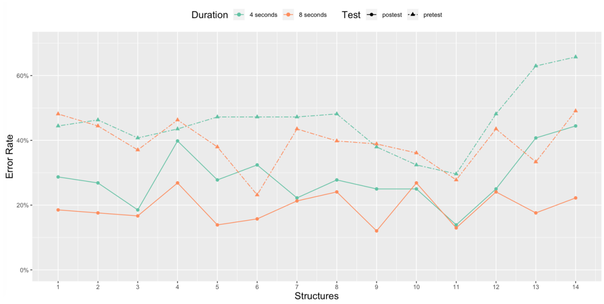


Figure 10: Error rate of each structures (Experiment I)

2.4. Discussions and Conclusion

The average correct rate of the 8-second recordings in the posttest was 80.7%, which implies that with enough playback duration provided, it is possible to achieve a relatively high correct identification and localization of the first layer of sounds (up to four concurrent sound sources). During the individual interviews, we found that most of the listeners had identified the elements according to their pitch differences. The irregular patterns, where each type of element has its own density, play an important role in the separation of concurrently sounding sources, and help to avoid the merged perception of two or more identical atoms. However, density is not the most intuitive feature for the listeners to distinguish the different element types from each other. There were three participants (no. 5, 16, 22) who mentioned that they were unable to perceive the pitch differences and found the density differences more distinct, which they described as the 'speed' of each sound. They perceive certain sounds as 'faster' (more dense) while the other sounds are '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.

The results show that the highest sound and lowest sound, hydrogen and oxygen, are easier, and faster, to be identified in both the pretest and the posttest. Without the highest or lowest sound(s) as a reference, it becomes harder to identify carbon and nitrogen alone which have pitches in between the highest and lowest pitches. It might be confusing for the listeners to identify whether it's one of the middle two pitches or the lowest/highest one. Or when there were several rent sound sources, it becomes harder to distinguish the ones whose pitches are in between. When we looked through the identification results of each participant manually, 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 (see Figure 4).

3. EXPERIMENT II

In experiment 2, we are interested to take our sonification approach a step further and add more sounds sources by adding the second layer of atoms. We want to know how many atoms (= sounds) listeners are able to recognize and localize maximally. 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.

Additionally, we have implemented a number of changes in our sonification design, based on the results we got from experiment 1. We have raised the pitch for hydrogen and carbon sounds by one octave (see Figure 11),), so that there is a now two-octave interval between the carbon and nitrogen atoms. We hope this helps to correctly identify the elements and avoid the confusion that we have seen in experiment 1[2].

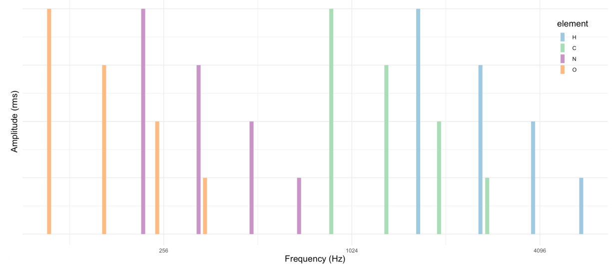


Figure 11: Frequency components for each element (Experiment II)

Besides the increased pitch interval we have added some timbral changes. First of all, we have increased the differences between the sounds by finetuning the q factors of the bandpass filters for the individual partials of the individual sounds. Secondly, regarding 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 a too long period between two consecutive impulses of the sound and thereby resulting in a bit more continuity. Lastly, while working with reverb allows us to create different sensation of distance for the elements in the first layer and the elements in the second layer, the reverb also blurs the sounds a bit time and therefore it becomes a little more difficult to distinguish the sounds from each other especially when many atoms are present. We 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

[2]Recording examples of each element on 2 layers can be found at: `https://surfdrive.surf.nl/files/index.php/s/pJSkLBcmhOm8Ik6`

and thereby make them directly audible. This more impulse-based attack makes it easier to detect and localized the individual sounds.

As mentioned before, experiment 2 aims to evaluate how well the listeners are able to identify and localize two layers of sounds surrounding the listening position. We use different loudness settings for the direct sound on layer 1 and layer 2. With this experiment we want to evaluate up to which extent our approach enables the participants to distinguish the layer positions from each other.

We have designed the experiment in such a way that we use two different approaches to how we present the elements of layer 1 and layer 2 and compare these to conditions to each other. In condition 1, only the first layer sounds are played from the start and only 10 seconds later the second layer sounds are added as well. A detailed description is given in section 3.2. We hypothesize that participants would perform better in condition 1, resulting in a significant difference in the ability to identify all the sound sources between the two conditions. A within-subjects designs is used, with the advantage that individual differences in subjects' overall level of performance are controlled [7]. For example, some subjects may more skilled at 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 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.

3.1. Materials

From the 14 structures used in the previous experiment, we chose the structures (1,2,6,7,8,14 in Figure 4) with lower average error rate in the posttest test. We extended these structures and added the second layer based on combinations that are found among in amino acids. This resulted in 8 structures that were used for experiment 2. All the sounds were generated in Pure Data in real time [3].

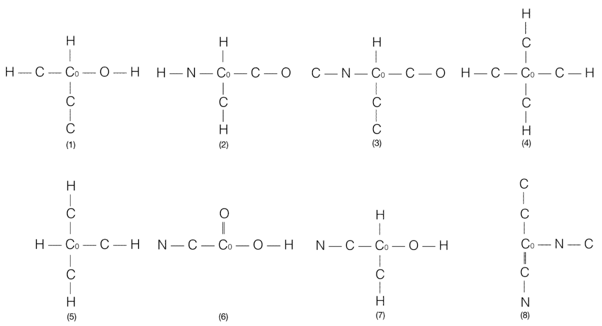


Figure 12: 8 possible structures used in Experiment II (2 layers)

3.2. Procedure

The experiment consisted of four phases. Firstly, an introduction to the four sounds was given to the participants identical to experiment 1. After they felt they were able to recognize the sounds, they would start the training phase. They were aware that sounds would come from four surrounding speakers, and there would be up to two sound sources on each speaker. Then, the participants had two different conditions of sound tests. In condition 1, 8 sets of sounds were played in a randomized order. Participants were told that a maximum of 8 sound sources would be positioned around, and the first layer sounds would be played first. After 10 seconds, the second layer of sounds would be added up. In condition 2, same 8 sets of sounds were played in a randomized order. Participants were told that all sound sources would be played simultaneously for 20 seconds and each direction would contain up to two layers of sounds. During the time a structure was played the participants were able to choose in a similar interface which element they heard (H, C, N, O or none) originating from each direction and layer (see Figure 13).

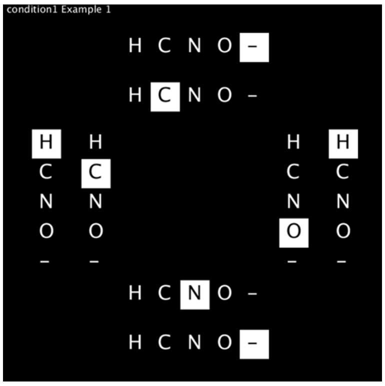


Figure 13: A screenshot from the user interface (Experiment II)

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, participants are given three examples at the beginning to get familiar with the interface and how the sounds were played. During the whole experiment, participants were allowed to change the head orientation during the experiment.

3.3. Results

35 participants joined the experiment. There were 19 female and 16 male participants. All participants were in 20s and 30s, except one participant was 47 years old. . The participants were all students and staff from Leiden University. Most of them have basic knowledge about chemistry or biology, but none of them have a professional background in the field of chemistry or biology, or have done related research.

In general, participants performed better in condition 1. A paired t-test is implemented on the correct rate of all the participants. The p-value is 1.051e-05, which is below the significant level 0.05. This indicates a significant difference between the two conditions. Since first layer sounds were played separately in condition 1, the average correct rate for first layer sounds identification in condition 1 is 79.2%, and 63.6% for condition 2. The p-value is 2.283e-07 and the effect size is 1.09, which indicates that the participants performance for the first layer sounds is a lot better. However there seems to be no significant difference between the results for the second layer sounds comparing the two conditions,

[3] Recording example of structure 1 and 2 can be found at https://surfdrive.surf.nl/files/index.php/s/oiJ2cglXFOKKUqe

and the p-value is 0.1347. The average correct rate for second layer sounds in condition 1 is 46% and in condition 2 is 43.2%.

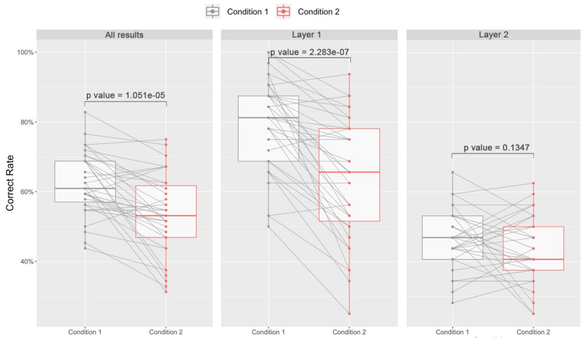


Figure 14: Correct rate of all participants (Experiment II)

3.3.1. *Elements*

For both conditions the participants performed better for the sounds positioned on the first layer than second layer (see Figure 15). In condition 1, the correct rates for all the identified elements positioned on first layer are all above 72%, especially the correct rate of hydrogen and oxygen reach 82%. There was less chance to misidentify nitrogen with oxygen or confuse carbon with nitrogen. In condition 2, all the sounds were played in parallel, and participants can identify the first layer sounds relatively well and the overall correct rate for all elements on the first layer is above 55%.

However, participants had similar performance for the second layer sounds with the ones in condition 1. It was more difficult for the participants to identify and localize the sounds from the second layer for both conditions, the average correct rate for second layer sounds is around 44.6% when we combine the results for both conditions. The correct rate of both hydrogen(35.1%, 31.6%) and carbon(22.9%, 21.1%) are low. More than half of hydrogen atoms were not heard (marked as no sound) in condition 1 or mistaken as on the first layer in condition 2.

	H	C	N	O	-
Condition1 - Layer1	83.4%	77%	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%

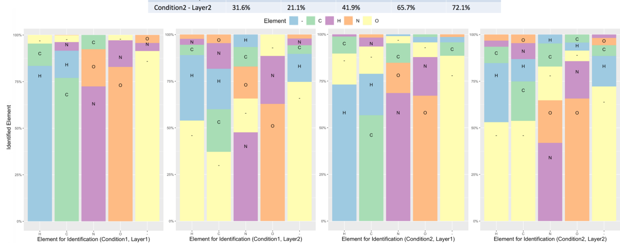
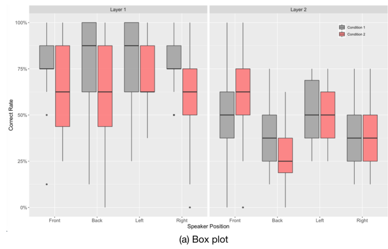


Figure 15: Correct rate of each element (Experiment II)

3.3.2. *Directions*

From Figure 16, we can see that in general the participants performed better for the front and left speakers. The average correct rates 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 (28.3%) in condition 2 (see Figure 16b).

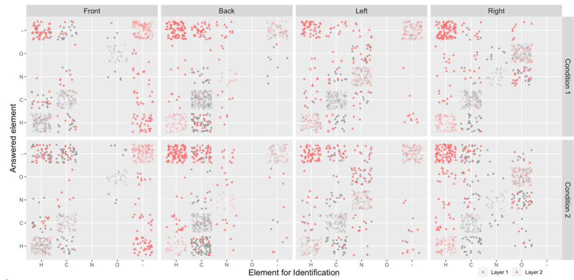
The performance for the different directions gets influenced both by the presented element structures and possible differences in our abilities to localize and distinguish the sounds from each other. We assume that its more difficult to distinguish front versus back from left versus right. As for the hydrogen sound from front speaker, it was confusing for the participant to localize which layer it was on. Both the hydrogen and carbon on the second from back speaker were difficult to be heard. First layer carbon from back speaker was mostly misidentified as hydrogen in both conditions.



(a) Box plot

	Front	Back	Left	Right
Con1/L1	78.2%	77.9%	80.7%	80%
Con1/L2	51.4%	37.1%	52.9%	42.5%
Con2/L1	63.8%	63.8%	70%	57.1%
Con2/L2	57.9%	28.3%	45.4%	41.38%
Condition 1	64.7%	57.5%	66.8%	61.3%
Condition 2	60.8%	46%	57.7%	49.3%
Layer 1	71%	70.8%	75.36%	68.6%
Layer 2	54.6%	32.7%	49.1%	41.9%

(b) Table



(c) Dot plot

Figure 16: Correct rate of each direction from (Experiment II)

3.3.3. *Observations from Training*

In the training session, there were four structures that include six or seven sound sources playing in parallel. The listeners were required to point out their directions, corresponding layer and element type, for all the sound sources they heard. The result and order that each participant answered were recorded. Figure 17 shows the average correct rate of each sound source and average order that each participant pointed out a sound source they heard. Most of participants were able to point out at least 5 sound sources correctly and it's possible to point out 6 or 7 sound sources. Additionally, the sounds from left were earlier to be pointed out. In structure Q13, more than half of the participants said they heard the nitrogen sounds from all directions, but the left and right ones were easier to be identified. In structure Q14, 10 participants could hear the carbon positioned on the second layer from left, rest of the participants were unable to hear it even after a hint was given. Oxygen positioned on the first layer from left was the fastest one to be identified in structure Q15, while the other oxygen positioned on the second layer from front was harder to be heard and became the last one to be identified. In structure Q16, second layer hydrogen from right can be very lastly heard while only four participants were able to hear second layer carbon from the back with a hint was given.

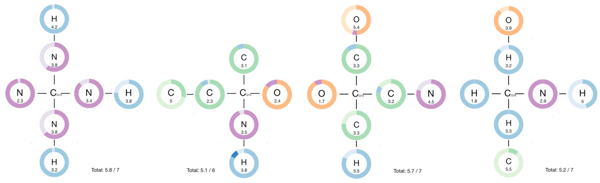


Figure 17: Average correct rate and order of the sound sources (Experiment II, Training phase, Q13-Q16)

3.4. Discussions and Conclusion

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 negatively influence their performance.

The correct 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 used reverb. In contrast, the used reverb settings 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 for hydrogen atoms and second layer hydrogen atoms were often not heard. Combined with the dot plot (see Figure 16c) and the observation from each participants detailed raw result, this usually occurs when there was a hydrogen atom at the second layer, such as the combination of C-H in a structure. It leads to the hydrogen atom creating the illusion of being at the first layer and thus not being identified at the second layer. It seems difficult to separate the hydrogen and carbon sounds when they are coming from the same direction. Similarly, when a first layer hydrogen from the front is combined with a first layer carbon from the back (structure 1, 2, 3 in Figure 12), only hydrogen sound is identified as first layer sound. Based on these results and the participants' individual feedback during the training session, we think that auditory masking may happen, 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, and 2) left and right sounds might mask or make it more difficult to identify the front and back sounds. However, further research on masking effects is still needed.

4. OVERALL CONCLUSION

It can be concluded from experiment 1 that our sonification approach is learnable and people are able to learn it relatively quickly. Although the carbon and nitrogen sounds were difficult to distinguish in experiment 1. The changes in experiment 2: the increased pitch interval between the nitrogen and carbon sounds, and the added more articulated attack seem to improve the performance of element identification for this experiment. While in experiment 1, the average correct rate in the posttest (8-second) for carbon is 71.6% and for nitrogen was 63.4%. In experiment 2, the average correct 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 experiment 2 (see Figure 15). Although the participants and the test materials for both experiments were different, we do believe that the results seem to indicate that it becomes more intuitive for the participants to identify each sound without the need of other sounds as reference.

Experiment 2 shows that when more than four sound sources are presented, it becomes harder to identify the second layer of sounds. However, our second experiment shows that it is still possible to distinguish 6 or 7 sound sources within the chosen time frame. Some participants have indicated that the used time frame was a little tight.

With our setup and experiments we have developed sonification systems that use concurrently sounding sources in a spatial configuration and applied a systematic approach to evaluate its qualities and limitations. By doing this we hope to make a meaningful contribution both with what we have achieved and with the reasoning behind the proposed approach.

5. LIMITATIONS AND FUTURE DEVELOPMENT

In both experiments, we have used a limited number of chemical structures that are identical to the ones found in amino acids. Because of the limited materials that we choose to offer, certain element (combinations) are only at 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 the different sounds, both regarding sounds that share a speaker and sounds that are spatially separated using different speakers. The four-speaker setup was a challenging choice when representing multiple rent sound sources, especially with distance differences. It is interesting for us to test rent sound sources separation and localization on other sound systems in the future.

6. REFERENCES

[1] D. Liu and E. van der Heide, "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. (ICAD2019)*. Newcastle, UK: Georgia Institute of Technology, 2019, pp. 140–156.

[2] J. G. Neuhoff, "Perception, cognition and action in auditory displays," *The sonification handbook*, pp. 63–85, 2011.

[3] W. M. Hartmann, "Localization of sound in rooms," *The Journal of the Acoustical Society of America*, vol. 74, no. 5, pp. 1380–1391, 1983.

[4] A. A. A. Ibrahim, F. M. Yassin, S. Sura, and R. M. Andrias, "Overview of design issues and evaluation of sonification applications," in *2011 International Conference on User Science and Engineering (i-USEr)*. IEEE, 2011, pp. 77–82.

[5] B. N. Walker and M. A. Nees, "Brief training for performance of a point estimation sonification task," in *Proceedings of the 25th International Conference on Auditory Display. (ICAD2005)*. Limerick: Georgia Institute of Technology, 2005.

[6] R. O. York, *Statistics for human service evaluation*. SAGE Publications, 2016.

[7] D. M. Lane, D. Scott, M. Hebl, R. Guerra, D. Osherson, and H. Zimmer, *Introduction to statistics*. Citeseer, 2017.