INTERACTIVE AUDITORY NAVIGATION IN MOLECULAR STRUCTURES OF AMINO ACIDS

A CASE STUDY USING MULTIPLE CONCURRENT SOUND SOURCES REPRESENTING NEARBY ATOMS

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 are interested in sonifying the molecular structures of amino acids. This paper describes the context and the first design choices for our approach. So far, we believe an amino acid molecule is too complex to be perceived at once. Therefore, we have designed an interactive form of sonification in which the listener navigates through the molecule over the network of carbon atoms. We describe our different approaches and discuss the topic of immediacy: the time it takes to recognize the structure surrounding the listener's position while navigating. Furthermore, we touch upon the question how many atoms we can sonify simultaneously and the role auditory masking plays in this context. To overcome auditory masking, we propose to use irregular but easy to recognize sounds. We conclude with an interest in a three-dimensional navigation environment using general molecular structures for further research and development.

1. INTRODUCTION

In our daily lives we are used to navigate through sound environments consisting of multiple sources that not only indicate their positions but also communicate information to us. In laboratory environments, listeners are often presented with rather simple auditory stimuli and listening tasks in order to learn more about our spatial perception. Many studies researched the localization of diverse sound stimuli in the form of single sound sources positioned at various azimuths and elevations [1, 2, 3, 4]. However, relatively few studies focused on our ability to localize two or more concurrent sound resources [5, 6]. In this paper, we illustrate and discuss the approach we have taken to develop an interactive sonification system using multiple sound sources that are spatialized in the horizontal plane around the listener. We are using a simple four-speaker setup in which the positions of the speakers correspond to the directions of the sound sources (see Fig.1). We are currently interested in sonifying the structural formulas of amino acids because of its relatively easy structures. In the future we aim to sonify RNA structures including folding.

Our ability to perceive a sound's direction and estimate the origin of a sound is called sound localization. This works through a process known as binaural hearing. In horizontal plane, our localization relies on a combination of multiple acoustic cues: a) interaural time/phase differences (ITD/IPD), b) interaural intensity differences (IID) and c) the spectral shape [7]. An enormous amount of research has been done on spatial hearing and the ability of a human to localize sound, both using headphones, as well as in free-field setups with loudspeakers. Stevens and Newman conducted experiments in the open air in 1936. Sounds were produced by a speaker which could be moved noiselessly over a circle in the horizontal plane. They concluded that noise was localized more easily than any of the pure tones [1]. Later, Hartmann tested and compared the performance of localizing continuous pure sine tones, broadband noise and complex signals in a room. The result indicated that azimuth judgement became more precise when the spectral density of the sound increased [2]. Lokki et al. did an auditory navigation experiment in 2000 in which the subjects were asked to move in a virtual space with arrow keys of a keyboard and find a point-shaped sound source with a random-position [3]. The sound reproduction equipment was a headphone. They tested three different factors: a) audio stimuli with different spectra including pink noise, artificial flute and recorded anechoic guitar, b) different panning methods for the positioning of the sound, and c) different acoustical conditions: direct sound, combined with early reflections, combined with reverb. The results proved that noise is the easiest stimulus to localize, and reverberation complicates the navigation. Letowski et. al pointed out that sound sources producing impulse sounds (e.g., firearms) are easier to be localized than sources emitting continuous or slowly rising long tones in closed spaces (rooms) [4]. These studies have investigated different aspects that may affect the localization accuracy of single sound sources. On the other hand, Brungart et al. conducted an experiment in which 14 different continuous but independent noise sources were turned on in a sequence within a geodesic sphere consisting of 277 speakers [6]. Each time when a new source was added, the listener was asked to localize it. They found that localization accuracy was modestly better for the sounds with rapid onsets than 1-second ramp onsets. Additionally, accuracy declined as the number of sources increased but was still higher than expected on the basis of chance when all 14 sources were on.

Sound localization is only one possible aspect of sonification. In our study, the sounds represent the type and position of the atoms around us. It is important that the sonification is easy to learn and understand in an intuitive way. 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. Several approaches are distinguished from each other such as the used of earcons, auditory icons, parameter mapping sonification (PMSon) and model-based sonification (MBS) [8]. All of these approaches are based on the human's auditory system, which derives three auditory dimensions that are commonly used in auditory display: loudness, pitch and timbre [9]. With these primary features, humans are able to separate and identify different sound sources, each with their own characteristics. While auditory icons are meant to represent events directly, earcons are synthesized sounds which require a learning process to relate the indirect sound to a specific meaning. When a continuous data stream is involved, it is effective to use PMSon with predetermined relations between the chosen auditory features and the information the data contains. MBS often uses a dynamic model that can include interaction, and utilizes sound to help to analyze a specific data task. Additionally, Carlie showed that the auditory system is sensitive to differences in the duration of a sound larger than 10ms, generally the smallest detectable change increases with the duration of the sounds [10]. This brought us to the idea that duration could also be used as a parameter for identifying different sounds sources. In order to be able to localize and identify the multiple surrounding atoms as fast as possible, our decisions for the sound design were affected by the features mentioned above. We will explain our choices in detail in Section 3.

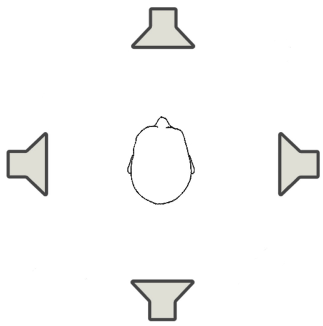

Figure 1: Positions of four speakers setup.

2. INTERACTION DESIGN

The visual field of the human eye has a limited arc while sounds is perceived omnidirectional. Sounds could reveal the existence of something that is difficult to be seen. The three-dimensional structures of proteins attract us, especially the folded parts where amino acids interact with each other. The aim of our research is to sonify multiple surrounding objects simultaneously in the horizontal plane, and to test whether they can be perceived, localized and identified by means of interactive navigation. Due to the complexity and inherent high dimensional order of proteins, we chose to start with exploring the structural formulas of different amino acids in two dimensional schematics. Unlike written chemical formulas, the structural formulas provide a geometric representation of the molecular structure. To simplify the localization task, we have transformed the formulas into flat graphical ones with identical bond angles of either 90 or 180 degrees, and identical bond lengths (see Fig.2). We are aware that this is an extreme simplification of the actual structure but it simplifies the sound spatialization in such a way that the speakers always correspond to the actual directions of the sound sources and we don't need to create phantom source locations in between the speakers. It relates more to how a molecule is drawn on paper than to its spatial shape in three dimensions.

2.1. Speaker Setup

Different from the common quadraphonic speaker setup, we place the four speakers around us to the front, left, back and right (see Fig.1). The physical position of each speaker always corresponds to the position (or direction) of the sonified atoms. We don't need to create phantom source locations in between the speakers and thereby we avoid potential negative effects of spatialization techniques. We sonify the atoms that are connected to a certain carbon atom that forms the (imaginary) center of the speakers and is not audible itself. Detailed sonification and localization implementations will be explained in Section 2.2.1.

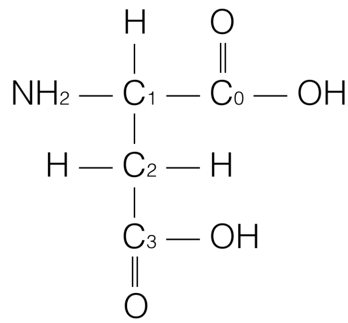

Figure 2: The structural formula of Aspartic acid.

2.2. Interactive Navigation of Structural Formula

In the past decades, structural biology developed into dealing with the molecular structure of biological macromolecules, like proteins, made up of amino acids or nucleic acids. Atoms are organized in a complex ordered 3D manner and form a macromolecule. Grond et al. developed SUMO, an open source software environment to sonify structure data contained in PDB files[1]. They implemented acoustic signatures for each amino acid, where different amino acids had different sounds, and parameterized earcons were used to distinguish pairwise distances and conformation differences of amino acids [11]. SUMO shows how sonification can be complementary to visually displaying macromolecules. Two years later, Grond et al. combined visualization, sonification and interaction in their application to represent the possible secondary structures of an RNA sequence. The application was designed to turn RNA structures into auditory timbre gestalts according to the shape classes they belong to, on the different abstraction levels [12]. Thereby, it became possible for the users to quickly compare structures based on their sonic representation. Additionally, the users were able to learn the meaning of the sound by selecting the visual pieces and playing back the corresponding sound. Compared with sonifying the structures as a whole part in [11], such interactions provide an interesting and effective way for the users to discern the meaning of the sounds.

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

In previous studies, sound has been used to enhance the existing structural visualization of static data. Is it conceivable for the listeners to follow the structures when the visuals are removed? What kind of method could help the listeners to learn the meaning of the sounds when there are multiple concurrent sounds? In our design, we would like to only use sound to represent the structural formulas of amino acids. The listeners are able to navigate the structures by moving over the carbon atoms in the molecule with the arrow keys on the keyboard. The navigation task provides opportunities for the listeners to explore the structure and take notice of the surrounding environment on a step by step basis. Meanwhile it allows the listeners to focus on a part of the molecular structure. We assume that such an interactive design would help the listeners to learn the meaning of the sounds and understand the molecular structures.

2.2.1. Navigation Rules

It is necessary to find an accessible way for the listeners to navigate through the structures and not get lost. The 20 natural amino acids contain amine (-NH_2) and carboxyl (-COOH) functional groups, with different R groups (side chains). The common elements are carbon (C), hydrogen (H), oxygen (O), nitrogen (N), while other elements like sulphur (S) and selenium (Se) are found in the R groups of specific amino acids. There is a carbon chain attached to the central carbon atom called C_1 (see Fig.2), which is next to the carboxyl group. Starting from the central carbon, there are several carbon atoms connected and forming the skeleton structure. Therefore, we chose for a navigation method where the user is able to explore the structure by moving from one carbon atom to its neighboring carbon atom(s). The starting point is numbered as C_0, which is the carbon part of the -COOH group and connects directly to the central carbon (see Fig.2). In this case, the user can not move to the right, but only to the left where C_1 is located. If there is an attempt to move into a direction that is not a carbon atom, a short alarm sound will be played as feedback.

2.2.2. Concurrent sound sources implementation

The various elements (atoms) that are connected to the current carbon position will be sonified independently. The -NH_2 and -OH groups are exceptions to this rule and will be sonified as independent groups. In our first stage, only the four atoms/groups connected directly to the current carbon position, will be sonified. For example, when the listener stands on C_0, only -OH, =O and C_1 will be audible (see Fig.2). In this way, the listeners can learn the information conveyed by the sounds and audibly observe the structures by navigating. In our next stage we decided to sonify one more layer of atoms; the atoms connected to the first layer of sonified atoms and in positioned the same direction. In this stage, the groups will be decomposed into single atoms (see Fig.3). Accordingly, N connected to C_1 and H connected to -O are audible (see Fig.3). Thus up to eight atoms will be audible at the same time.

In the future, we would like to sonify even larger areas. For example, all of the atoms in a row of a carbon atom could be sonified simultaneously. When the listener stands on C_1, not only the two layers of atoms connected with it will produce sounds, the O connected with C_3 and the H connected with -O will also be audible (see Fig.3). When the listener moves to C_0, the same atoms in this horizontal row will still be heard but the changes of the surrounding sounds could imply the listener's position changes, and give evidence of how the atoms in this row are positioned. Furthermore, we will consider the use of spatialization techniques to realize phantom sound source locations and work with depth in the sound. For now, we have specifically chosen to make the speaker positions correspond to the location of the intended sound source positions and avoid possible negative side effects that the spatialzation could bring.

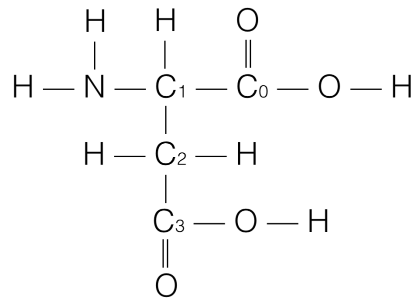

Figure 3: The structural formula of Aspartic acid for the larger area sonification.

3. SOUND DESIGN

In recent decades information sonification in the fields of chemistry and biology, mainly focuses on DNA sequences and macromolecular structures. Many different choices have been made to sonify and represent objects (e.g. amino acids, proteins, nucleotides) and events. For example, a) single note is mapped directly to string data derived from a DNA sequence [13, 14], b) short musical phrases are formed by the Morse code of the amino acids, nucleotides and nucleotide pairs [15, 14], c) parameterized earcons help the users to distinguish similar but different structures such as amino acids. Different parameters in a sound synthesizer can be mapped to the different features of an object or event [11, 12, 16], and d) pre-recorded samples are used as auditory icons to represent events extracted from simulation progress [17]. In these studies, sonification was utilized often to enhance the visual display of complicated structures. However, it remains unclear whether the listeners are able to recognize and comprehend the sounds without the visual input.

For our approach it is essential that the (interacting) listeners can both identify and localize the atoms purely by means of sound. This brings us to the question how the atoms should sound? For atoms there are no metaphorical approaches that are already familiar to us in daily life and therefor auditory icons are not applicable in our context. Therefore we considered earcons as a conceivable way to establish a mapping stratagem between the atoms and their sonic representation. Earcons are defined as short, structured musical messages, where different musical properties of sound are associated with different parameters of the data being communicated [8]. The relations between the earcons and the atoms are supposed to be understood and acquired by the listeners. The goal of our sound design is to be able to easily recognize and distinguish the different sounds from each other, even if they sound simultaneously. We have used Pure Data, a graphical programming language for real-time interactive multimedia processing, for both the interactive navigation and the real-time sound synthesis.

3.1. Sound Synthesis Techniques

We have experimented with different approaches regarding how to sonify the different atoms and how to deal with time (i.e. use rhythmical structures or not). The aims of our sonification are to represent as many surrounding atoms as possible (as many concurrent sounds as possible) and to be able to localize and identify the atoms in as little time as possible. We started with different drum samples because the timbre of different parts from a drum set (e.g. bass drum, snare drum, hi-hat) is easy to be distinguished and such percussion sound is short and easy to localize. In our first early prototype, hydrogens produced closed hi-hat sounds every 400ms, carbons produced snare drum sounds every 1.6s, oxygens and groups generated bass drum sounds every 3.2s. However, the drum samples might be distracting since the listeners can recognize them and may have problems to relate them with chemical elements.

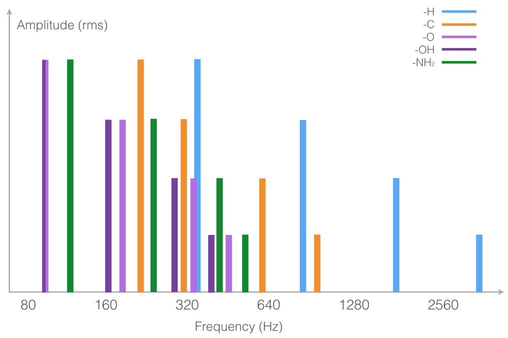

Figure 4: Frequency components for each element.

Then we tried filtered white noise with different amplitude envelopes. The central frequency of the bandpass filter is inversely related to the number of protons in the atom. The fewer protons, the higher filter frequency. This means that the sound that represents hydrogen has the highest frequency setting and the oxygen sound has a lower filter frequency than the carbon sound. The amplitude envelope enables different durations and loudness developments for each of the elements. The oxygen sound is the longest. While the single atoms have a clear and sharp start, the groups have a longer attack time. For example, the frequencies of a single oxygen atom and the -OH group are the same, but -OH has a slower attack time and longer duration at the sustain level. The filtered noise sounds are more abstract than the drum samples. We use pitch as the main feature in this design because the changes are easily perceivable and distinguishable. Hartman examined a tone with a fundamental frequency of 200Hz and 11 harmonics up to 5800Hz and concluded that the mixing of components within a single critical band plays a significant role in localization [2]. Therefore, we decided to add three more bandpass filters for each representation of an atom, resulting in a richer spectrum with four frequency partials, in order to improve the ability to localize the sounds. As shown in Fig.4, the frequency components made up for hydrogen are much higher, which are 352Hz, 877Hz, 1811Hz, 2941.1Hz. As a group, -OH relates to oxygen and the frequency components of -OH are slightly lower than oxygen. Both of them start with 100Hz, then oxygen develops with 201Hz, 350Hz, 461.1Hz and -OH includes 173Hz, 331Hz, 401Hz.

The main problem of this sonification approach is that it is hard to separate the sounds from each other when two or more of the same elements are played together. The similar frequency components produced from identical atoms may cause frequency masking. Also, merging may happen if they are positioned in a row (meaning in the same direction).

3.2. Sound Composition

When there is a complicated sound environment containing multiple concurrent sound sources, Brungart et al. used a sequential localization process to examine localization accuracy in 360 degrees. Each time, the subjects were asked to localize one newly activated sound source, but the previous played sources would remain. The sound sources were physically localized with 277 independently-addressable speakers which formed a geodesic sphere. Furthermore, each source was separated by 45 degrees from all the other sources. Brungart et al. pointed out that this method could avoid that sources originated from same direction, as well as help to reduce proximity-dependent effects of the individual maskers on the target [6]. Our approach does involve multiple sound sources played in parallel. The various frequency components contribute to be able to segregate one object from the others. Nevertheless, there are only four speakers representing four directions in our research, sound sources could be positioned in a row and produced from one same speaker. Later we will discuss other approaches to solve the merging problem when sources are concurrent and even played on one speaker. All of the approaches mentioned below started with the implementation of only sonifying the directly connected atoms and groups of the current carbon position (we call this the first layer). Afterwards we have extended some of the approaches and sonified also the atoms behind the directly connected atoms (we call this the second layer).

3.2.1. Rhythmical Pattern

Several researches have focused on melodic patterns in the field of sonification and auditory display, but there is little relevant research on rhythmical patterns. Rhythmical patterns could be regarded as a sound character to enhance and help the listeners to distinguish and localize multiple sound sources played simultaneously.

Firstly, we divided 4 speakers as 4 beats in a bar, and play a counter-clockwise sequence (front - left - behind - right) with a fixed tempo. This way the sounds will be played sequentially[2]. We implemented the envelope and duration differences mentioned in Section 3.1, combined with the bandpass filter groups. We would like to investigate whether sequenced nature could help the listeners to distinguish the different elements. This approach is a way to solve the problem of the overlapping sounds. However, it takes 2.4 seconds to finish a bar which might be a bit long for the listener to recognize and remember the sounds. It is still possible after several times of repetition but we would like to accelerate the process to achieve a faster and intuitive recognition of the different sounds in a (near) simultaneous way. Therefore, we tried another approach: Besides the envelope and duration differences, we assigned different repetition speeds to different elements. But the position always determines the beat where the sound starts to

[2] A binaural recording example of navigating in the structural formula of Aspartic acid with rhythmical pattern I can be viewed at: https://www.dropbox.com/s/p05lv10fg91equi/Rhythmical_pattern_1_Aspartic_1layer.wav?dl=0

play[3]. For example, when the listener stands on C_1 (see Fig.2), the hydrogen sound repeats at 600 bpm and synchronous to the first beat of the bar. The sound that represents -NH_2 repeats at 45 bpm is synchronous to the second beat in the bar. The carbon sounds repeat at 80 bpm synchronous to both the third and the forth beat. When all four speakers start to play sounds together, it is clear and direct for the listeners to note the similarities and dissimilarities among them. One of the disadvantages of this approach is that each element has an independent and distinct speed that can affect listeners to perceive different tempi at the same time. In addition, the sound results can be chaotic and annoying when there are various elements sonified together.

3.2.2. Bouncing Pattern

We also tried loops of a bouncing pattern to create a more interesting pattern for the listeners to identify. Imagine a ball is lifted at a certain height and then released, when it hits a surface it will create a sound, lose some potential energy and bounce into the air again, but lower than the original height. It keeps bouncing until it stops. As for the atoms, they could be balls falling from different height and have various bouncing patterns. Like hydrogen falls at a lower height and produces shorter bounces. Each element has a different bouncing speed and duration. A decay envelope is used to control the decrease in bounce period[4]. The bouncing pattern might be complicated and confusing at some point compared with the previous approaches of rhythmic pattern. The impact sound at the starting point of each loop is always clear, whereas further bounces quickly speed up and become rather intensive. Another problem is that when there are atoms of a same element that generate sounds, the bouncing pattern is also the same. Such bouncing sounds could be mixed up together and challenging for the listeners to separate one from the other, even though they are coming from different speakers. Furthermore, this approach will sound rather confusing when a larger area of the structure is sonified.

3.2.3. Irregularly Triggered Bandpass Filter Banks

The bouncing patterns brought us to the idea of a granular structure sound. In order to create a more continuous but irregular pattern, we used colored noise in combination with a comparator with a variable threshold as a way to generate random impulses with random amplitudes. The signal changes vary a lot from white, pink and brown noise. By choosing between different types of noise varying the threshold we can generate different impulse patterns with different desired densities. We chose to give the lighter elements an intensive but (light) pattern and the heavier elements and groups a more extensive pattern with a larger range of amplitude changes. Due to the irregular signal impulses, the all the sounds have their own non repetitive structure. This means that two or more identical atoms still have their own irregular structure. We use the impulse patterns as input signals for banks with four bandpass filters that we used before and mentioned in Section 3.1. Now, even when there are multiple sound sources generated together, the differences are still recognizable[5]. The main difference is that the irregular structure is experienced as a kind of granular-like texture. This makes it easy to recognize the sounds and the listeners are not required to remember the rhythmical patterns and compare them with each other. Now we can play the different sounds in parallel and they can all be identified simultaneously. We have found a way to avoid the merging problem that we had before.

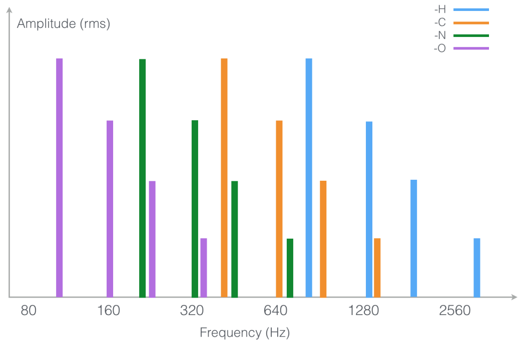

Figure 5: Frequency components for each element.

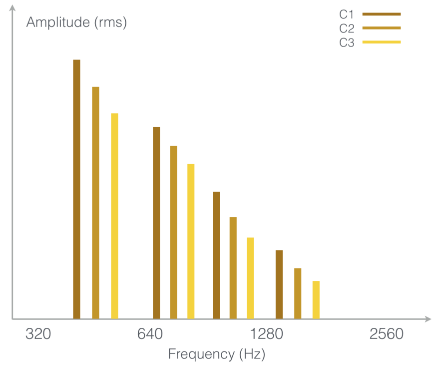

Figure 6: Different frequency components for the same element.

Now that we have achieved this we are curious to know if we can sonify even more atoms in parallel by sonifying the second layer around the carbon atom. Now we don't have to sonify the groups anymore since their individual atoms will both be played. The frequency settings of the filterbanks can be seen in Fig.4. Reverb is employed to enhance the sensation of distance of atoms in the second layer. The amplitude of the direct sound of the atoms from the second layer is one third of the ones from the first layer while the amount of reverb is the same. When the listener stands on C_1, C_2 and C_3 are both sonified (see Fig.3). On one hand, the distance determines the loudness and the sound of C_2 is louder than C_3. On the other hand, the q value of the bandpass filter of C_3 is slightly higher than C_2. The C_3 has more resonance and becomes less sharp and intensive. This is likely to solve the problem that the more intensive sound may mask a less intensive sound. In

[3] A binaural recording example of navigating in the structural formula of Aspartic acid with rhythmical pattern II can be viewed at: `https://www.dropbox.com/s/14jkq9urf515k83/Rhythmical_pattern_2_Aspartic_1layer.wav?dl=0`

[4] A binaural recording example of navigating in the structural formula of Aspartic acid with the bouncing pattern can be viewed at: `https://www.dropbox.com/s/bx9nhybgbswoqz4/Bouncing_pattern_Aspartic_1layer.wav?dl=0`

[5] A binaural recording example of navigating in the structural formula of Aspartic acid with ITBPFB can be viewed at:`https://www.dropbox.com/s/gf7qrcte9z4pwhu/ITBPFB_1_Aspartic_1layer.wav?dl=0`

our previous design, some frequencies were too low or too close to each other, which may have had a negative effect on separation and localization when two layers of objects are sonified simultaneously[6]. The frequency components have been adjusted and we have started to use a fixed interval size between the atoms and expanded the used filter frequencies in order to use a wider range (see Fig.5). There is an octave between two elements, for example oxygen is increased to 110Hz, nitrogen starts with 220Hz, carbon has 440Hz and hydrogen gets 880Hz. While oxygen and nitrogen remain with a less dense pattern, the resonance of the bandpass filters for these two elements is higher than for hydrogen and carbon[7]. In order to make the differences easily perceivable when two or more identical elements are positioned in the same direction we have chosen to give the elements of the second order a slightly higher pitch. The difference is small enough so that it is clearly identified as the same atom but larger enough to be able separate the sounds from each other and avoid merging. Fig.6 shows an example of different frequency components of the same carbon elements. There is a fixed ratio between two neighboring atoms. For example, if there are three carbon atoms positioned in a row at the same direction, the closest carbon is made up of 440Hz, 661Hz, 973Hz and 1389Hz and louder than other carbon atoms. The second carbon consists of 484Hz, 727.1Hz, 1072Hz and 1528Hz and the third carbons frequency components increase at the same ration of 1.1. However, it remains unknown what the maximum number of layers is that we can segregate.

4. CONCLUSIONS FUTURE WORKS

In this paper, we have discussed several different approaches to implement the spatial and interactive sonification of amino acids. We have personally evaluated the sound results in a research by design kind of approach. We are aware that part of our work could have been more detailed but have chosen to focus on the experimentation with the different approaches. We started with the concept of earcons in order to achieve the immediacy of sound recognition and localization. Unlike conventional earcons, such as time-based melodies or other sequentially played sound samples, our research focuses on concurrent sounds. We started with using fixed sound samples for the first rhythmical patterns and changed to real-time synthesized sound using banks of bandpass filters. While the repeating rhythmical patterns and bouncing patterns may take a longer learning time, the irregular impulses allow for a faster and simultaneous recognition of the atoms without a separation period. Currently, we combine frequency and irregular density as two main features for our sonification, to help the listeners to identify multiple simultaneous sound sources. By doing this we have expanded our approach that started with earcons toward a model-based sonification. It would be our next step to play an even larger area of concurrently sounding atoms. We already found that making light variations in frequency, density and loudness may (partially) solve the merging problem of multiple identical atoms coming from the same direction. The sound changes are regarded as auditory feedback from the interactive navigation, which may influence the localization accuracy and improve the segregation. In addition, it would be possible to realize a richer spectrum but avoid auditory masking.

Since all of the approaches mentioned above require a learning progress for the listeners to understand the mappings, further experimental investigations are considered to evaluate 1) whether the sounds properly represent the different elements, 2) whether the sounds are intuitive for the listeners to be recognized, and 3) whether the navigation could help to identify and localize multiple concurrent sources. Our main goal is to find out how complex a structure could be while still perceivable and recognizable. We will invite listeners to participate in usability and evaluation tests.

In order to simplify the localization task at present, we are using a particular 4-speaker setup in combination with the flat structural formulas. However, the molecular structures are three-dimensional, and the bond lengths and angles vary from one to another. It would be a logical step to represent the structures in a three- dimensional auditory environment. Setups consisting of more speakers in combination with different spatialization techniques will be considered. Bond lengths and angles could be included in the parameters used for the spatialization. Meanwhile, we are thinking how we can include active head movement in our research, which has proven to reduce front/back confusion and improve localization in elevation [18, 19].

[6] A binaural recording example of navigating in the structural formula of Aspartic acid with ITBPFB can be viewed at:`https://www.dropbox.com/s/ylgqw9p3u2nuvwr/ITBPFB_2_Aspartic_2layer.wav?dl=0`

[7] A binaural recording example of navigating in the structural formula of Aspartic acid with ITBPFB can be viewed at:`https://www.dropbox.com/s/tdxf0949uetdutd/ITBPFB_3_Aspartic_2layer.wav?dl=0`

5. REFERENCES

[1] S. S. Stevens and E. B. Newman, “The localization of actual sources of sound.” *The American journal of psychology*, 1936.

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

[3] T. Lokki, M. Grohn, L. Savioja, and T. Takala, “A case study of auditory navigation in virtual acoustic environments.” Georgia Institute of Technology, 2000.

[4] T. R. Letowski and S. T. Letowski, “Auditory spatial perception: Auditory localization,” Army Research Lab Aberdeen Proving Ground MD, Tech. Rep., 2012.

[5] P. L. Divenyi and S. K. Oliver, “Resolution of steady-state sounds in simulated auditory space,” *The Journal of the Acoustical Society of America*, vol. 85, no. 5, pp. 2042–2052, 1989.

[6] D. S. Brungart, B. D. Simpson, and A. J. Kordik, “Localization in the presence of multiple simultaneous sounds,” *Acta Acustica united with Acustica*, vol. 91, no. 3, pp. 471–479, 2005.

[7] A. N. Popper, R. R. Fay, and A. N. Popper, “Sound source localization,” pp. 272–316, 2005.

[8] T. Hermann, A. Hunt, and J. G. Neuhoff, *The sonification handbook*. Logos Verlag Berlin, Germany, 2011.

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

[10] S. Carlile, “Psychoacoustics,” *The sonification handbook*, pp. 41–61, 2011.

[11] F. Grand and F. Dall Antonia, "Sumo. a sonification utility for molecules." International Community for Auditory Display, 2008.

[12] F. Grond, S. Janssen, S. Schirmer, and T. Hermann, "Browsing rna structures by interactive sonification," in *Proceedings of the 3rd Interactive Sonification Workshop*, no. Human Interaction with Auditory Displays, 2010.

[13] N. Munakata and K. Hayashi, "Basically musical," *Nature*, vol. 310, p. 96, 1984.

[14] M. D. Temple, "An auditory display tool for dna sequence analysis," *BMC bioinformatics*, vol. 18, no. 1, p. 221, 2017.

[15] X. Shi, Y. Cai, and C. Chan, "Electronic music for biomolecules using short music phrases," *Leonardo*, vol. 40, no. 2, pp. 137–141, 2007.

[16] A. Tek, M. Chavent, M. Baaden, O. Delalande, P. Bourdot, and N. Ferey, "Advances in human-protein interaction-interactive and immersive molecular simulations," in *Protein-Protein Interactions-Computational and Experimental Tools*. IntechOpen, 2012.

[17] B. Rau, F. Frieß, M. Krone, C. Muller, and T. Ertl, "Enhancing visualization of molecular simulations using sonification," in *2015 IEEE 1st International Workshop on Virtual and Augmented Reality for Molecular Science (VARMS@ IEEEVR)*. IEEE, 2015, pp. 25–30.

[18] W. R. Thurlow and P. S. Runge, "Effect of induced head movements on localization of direction of sounds," *The Journal of the Acoustical Society of America*, vol. 42, no. 2, pp. 480–488, 1967.

[19] M. Kato, H. Uematsu, M. Kashino, and T. Hirahara, "The effect of head motion on the accuracy of sound localization," *Acoustical science and technology*, vol. 24, no. 5, pp. 315–317, 2003.