Spatial audio augmented reality for positioning and navigation with Airpods

Emiel Nagel - s2523663
Bachelor Creative Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Supervisor: Dr. Le Viet Duc
Critical observer: Dr. Paul Havinga
Date: 21-07-2023

Abstract

Many people listen to music while riding their bikes, which is not only a distraction, but prevents the biker from hearing some of the necessary sounds from traffic. A possible solution would be to substitute the missing sounds with sounds from the headphones or earbuds, that have been given a location through spatial audio. In this research, development of a system that can create such spatial audio is described. It starts off with an introduction of the research, followed by a background research. Possible techniques of achieving spatial audio have been described, as well as the current use of spatial audio and interventions to improve traffic safety for bikers. In the ideation and specification chapters, the functionalities and requirements of the system have been specified.

Based on this, a prototype was developed which was able to create spatial audio, which is described in chapter six. This prototype was then evaluated with seventeen participants using small games to determine the precision with which spatial audio was able to be simulated. The prototype did not manage to fulfill all requirements and some parts of the prototype were unable to be developed. The participants had to find which of three provided locations the played sound came from, as well as which of three played sounds came from a different direction than the other two. At the end, the participants were asked questions to gauge their experiences with using the spatial audio prototype. It was found that, although the prototype was not sufficiently precise and reliable, it did create spatial audio and the participants could hear some directions in the sounds. The participants reported that the system was fun and comfortable to use and could be useful to aid bikers in traffic, given that it is further developed to improve the precision. This resulted in some suggestions for future research, where the quality of the spatial audio of the current prototype could be improved, research could be conducted to specify new criteria and the prototype could be implemented in a bigger system using object recognition and Google Glass to aid bikers in traffic.

In the end, a limited functioning spatial audio system was able to be developed. This research opens doors for future advancements in spatial audio, encourages further exploration to improve the prototype and provides a compilation of knowledge for future spatial audio development.

Acknowledgement

I would like to sincerely thank my supervisor, Dr. Le Viet Duc, for his support and guidance throughout my graduation project. His helpful suggestions and thoughtful discussions were a big part in shaping the outcome of this research. He provided me with advice and tools to be able to continue developing my work.

I am also grateful to the participants of this study for their valuable contributions, without which this research would not have been possible.

Lastly, I want to acknowledge the support of my friends and family, who stood by me during this academic journey.

Table of Contents

List of Figures

Chapter 1: Introduction

Traffic is a dangerous environment where accidents can be fatal, especially for bikers. To safely take part in traffic, it is necessary to be aware of the activities of the other road users. To be aware of the other road users, bikers must gather information through their vision and hearing. In many cases, vision is not enough, for example when a car is coming from behind. Therefore, to safely take part in traffic as a biker, auditory stimuli are essential information to gather.

However, many people listen to music/podcasts, use google maps to navigate and make phone calls while riding their bikes [1]. This extra task can take away the attention of the biker, making biking in traffic unsafe. In the cases of listening to music or making phone calls, these new sources of sound mask or even completely block the sounds necessary to safely participate in traffic [2]. In the worst case, the audio playback devices used, often earbuds or headphones, are specifically designed to block all sound coming from the environment, using features such as active noise canceling. The lack of auditory signals from traffic can be a cause of unsafety and traffic incidents.

To address this problem, a system is proposed where environmental sounds are artificially created within the headphones/earbuds to indicate potential danger. For example, when a biker is crossing the road while listening to music. In this situation, the earbuds will indicate to the user that a truck is taking a right turn by playing the sound of the truck that seems to come from the left side. This way, taking part in traffic while listening to music on earbuds can be made safer.

Since digital sounds are by default unlocatable in 3D space, spatial is used to give virtual locations to sounds. This allows the user to identify the position of the digital sounds [3] and thus to get information about the potential sources of danger, even though the sounds that originate from the sources are blocked.

The aim of this graduation project is to create a system that can place a virtual sound in the real world and test the performance. The intervention must be able to work with earbuds or headphones. Next to this, the intervention must be usable while riding bikes. And lastly, the intervention must also be able to produce accurately locatable sounds in real-time. This means that the biker must be able to identify both the direction and distance of the sound as accurately as possible, without latency, or at least with as little as possible.

The main question is: ‘How can a system be developed that can output spatially locatable sounds accurately in real-time through headphones/earbuds, to aid bikers in traffic?’ There are three corresponding sub-questions:

- What kind of virtual soundscapes should be generated and where in 3D space?
- What kind of headphones/earbuds can give the most suitable 3D effect?
- Which technique for creating spatial audio in real-time is the most useful for this intervention?
- How effective is the system in generating sounds that are locatable in 3D space?

This paper will consist of four parts: Ideation, specification, realization, and user evaluation. The ideation phase will consist of a state-of-the-art research, background research, stakeholder identification, description of relevant concepts and argumentation for the preferred concept to end up with preliminary requirements for the design for the system. The specification phase will convert the preliminary requirements of the ideation phase into a well-defined specification of the envisioned product/service. The realization phase turns the proposed design into a working prototype. And finally, the user evaluation phase will test this prototype for effectiveness and functionality.

Chapter 2: Background research

Starting with the background research. The following section aims to find background information around the topic of spatial audio, to be able to (partially) answer the research sub-questions and to be able to create concepts. The first part of this chapter will be a literature review, covering existing methods of spatial audio, principles behind binaural spatial audio and virtual soundscapes. The second part will be a state-of-the-art research, where similar applications of binaural spatial audio will be identified, as well as the current state of binaural spatial audio. Lastly, a discussion and conclusion section will give a critical review of the background research, as well as try to answer the research sub-questions.

2.1: Literature review

2.1.1 Current methods of spatial audio

When reading literature on the existing methods of spatial audio, it becomes clear that each method can be divided into one of two groups. There are methods using loudspeaker arrays and binaural spatial audio methods, using headphones or earbuds. The sources list essentially one method of creating binaural spatial audio, but multiple techniques of using loudspeaker arrays. As such the methods using loudspeaker-arrays will be covered as one topic.

Binaural spatial audio methods

For starters, the first method that will be discussed is the method for binaural spatial audio. Multiple sources [4, 6] state that binaural spatial audio methods use head-related transfer functions (HRTF). These are functions that describe the difference in properties of sounds that arrive in the left and right ear. The human ear can identify the locations of sounds based on these differences in properties. According to *Nowak et al.* [4] and *Zhang et al.* [6] HRTFs describe the differences in interaural time difference (ITD), interaural level difference (ILD), and spectral cues of the sounds arriving in the left and right ear. These describe the difference in arrival time and volume of the sounds between the left and right ears. *Urviola et al.* [9] define HRTFs as "the ratio between the sound pressure at the ear position and the free-field sound pressure at a reference position." *Lokki et al.* [7] states that the ITD and ILD mostly influence horizontal directional hearing. To include the vertical position and differentiation between forwards and backwards, HRTFs are necessary [7].

The sources [4][5][8] agree that these functions are highly personal and are often created by using two microphones on dummy heads where the ears are normally located. By recording the differences in sounds coming in both microphones, the HRTF can be created. By then applying that HRTF to the sounds from headphones and earbuds, the illusion can be created that the sounds are coming from locations outside of the playback devices. *Chinmay et al.* [5] add to this that head-related transfer functions are based on the principle that once known, they can be used to process any sound to be perceivable as if coming from a 3D space. As such, binaural spatial audio is achieved. *Nowak et al.* [4] emphasizes that it is important that earbuds and headphones are calibrated to create as natural sounds as possible for creating as accurate spatial audio as possible. This is especially necessary given that these playback devices have a tendency to color sounds, making spatial audio harder to recognize.

Next to the coloration of the playback devices, there are other limitations and opportunities to HRTFs. The first is that HRTFs can work with headphones and earbuds by design. *Nowak et al.* [4] reports that this method works with two speakers, as opposed to an array of many more loudspeakers, which is confirmed by *Zhang et al.* [6]. Especially considering the separation of the sounds with earbuds and headphones [4], HRTFs are well suited for earbud applications. *Zhang et al.* [6] also states that this method is therefore much cheaper to set up than the costlier speaker array-based methods. This potential application of HRTFs in earbuds and headphones already makes HRTFs the only method that fits the requirements established in the introduction.

Zhang et al. [6] also reports that next to the costs, using headphones makes the quality of the spatial audio independent of positioning oneself in a so-called sound sweet spot. It allows the user to freely move around with minimal effect on the quality of the audio playback. This also fits the use case of HRTFs described by *Nowak et al.* [4]. Given that the other methods using arrays of loudspeakers require the user to stay in a fixed place [6, 7], spatial audio using head-related transfer functions is much more flexible in its use cases. It confirms that methods using HRTFs are the only method suited for the proposed intervention. However, the sources [4, 6, 7, 8] agree that there are downsides to this flexibility. This includes the need for using a head tracker, to track the orientation of the head and more complicated and processing-intensive audio processing, something which *Urviola et al.* [9] goes more in-depth about.

A final limitation to head-related transfer functions is the required precision. Creating spatial audio through head-related transfer functions comes very precisely. If the equalization and filtering of sounds is off by a slight amount, the effect will not be effective. *Nowak et al.* [4] reports that given the nature of headphones and earbuds, a headphone transfer function will need to be set up to make the headphones sound as clean as possible. They also report that setting up the head-related transfer function comes very precisely and can therefore be costly. It needs to be done in a symmetrically designed room using good quality head models, which can get very expensive. This method of recording is confirmed by other sources [6, 8]. On the other hand, *Urviola et al.* [9] mentions that predetermined HRTFs can be used, negating the costs of setting up an HRTF from scratch. The sources [4, 6, 8] also agree that due to the needed precision, HRTFs are highly individual and one HRTF may not work well for different users. Using a one-size-fit-all HRTF will therefore reduce the accuracy of the spatial audio, depending on the user. Making accurate spatial audio through HRTFs therefore requires each system to have a unique HRTF. It is however unknown whether the reduced accuracy of using only one HRTF is insufficient to the requirements. Research will need to be conducted to confirm this.

Speaker array-based spatial audio methods

As opposed to binaural spatial audio, there is a wide variety of methods using arrays of multiple loudspeakers. Given the overarching similarities and necessity of using loudspeaker arrays, these methods have vastly different requirements from binaural sound reproduction techniques. As such, the speaker array-based methods will be discussed as one topic.

Firstly, it’s important to understand how loudspeaker-arrays create spatial audio.

According to Zhang *et al.* [6], loudspeaker array-based methods generate sound fields containing regions where spatial audio is present. Sound field recording and reproduction uses microphones to record the sound fields of these regions. These recordings are then used to create a system that can create spatial audio in that same spot through speaker arrays [5, 6, 8, 10, 11]. Zhang *et al.* [6] states the following definition: "Spatial sound field reproduction aims to create an immersive sound field over a predefined spatial region so that the listener inside the region can experience a realistic but virtual replication of the original sound field." According to the same article [6], this can be accomplished through careful placement of the loudspeakers at the edges of the spatial audio region and deriving the sounds played by the loudspeakers. *Rajguru et al.* [5] emphasizes that for as accurate as realistic as possible virtual sound reproduction, it is crucial to consider the way in which the sounds are recorded. Multiple sources [5, 6, 8, 10, 11] indicate that recordings are oftentimes performed using 3D microphone arrays.

An extension of this method is so-called multi-zone sound reproduction. According to two sources [6, 12], multi-zone reproduction aims to create a sound environment where multiple listeners can listen to their own audio material at once while not being physically separated from each other. Zhang *et al.* [6] states: "A single array of loudspeakers is used, where sound zones can be placed at any desired location and the listener can freely move between zones." The goal of the method is to maximize the difference between the zones in the room with high and low acoustical energy [6, 12]. *Zhang et al*. [12] describes these as dark (quiet) and bright (loud) zones. The listeners can then sit in the bright zone to hear the spatial audio, or the dark zone to

One outlier to speaker array-based methods uses virtual speaker sources. This method is only mentioned by *Raijguru et al.* [5]. The method uses wave-field synthesis to create virtual sound sources that seem to originate from real objects [5]. The setup is a room where a ring of speakers is hung at ear level. By manipulation of the timing of the sound playback through each speaker, virtual sound sources can be created. According to Rajguru *et al.* [5], the spatial audio system was used in an augmented reality-experience. This system is very impractical compared to the other types of methods and thus not a great fit for the proposed intervention.

A lot of the advantages and disadvantages of head-related transfer function-based methods apply to speaker array-based methods in the opposite way. As opposed to the relatively cheap HRTF-based methods, spatial audio using speaker arrays is by comparison immobile, costly, inflexible and does not work for wearable devices [6, 10]. On the other hand, speaker array setups do not require to track the users head, need less audio processing, and do not need to be set up as precisely, making calibrations work for multiple users [6].

However, these benefits over HRTFs can be compromised by using a powerful enough system. *Song et al.* [13] not only shows that mobile phones are powerful enough to do the calculations, but also propose a method for even faster calculation time. Tracking the users' head is no real limitation, as simple head-trackers are cheap and simple to program. The most impactful limitation is the precision that comes with head-related transfer functions and the fact that they are highly individualistic. It is however possible to overcome this downside, by making the HRTF fit the specific biker.

The spatial audio method best suited for the intervention

Looking at the requirements the spatial audio method should have, the preferred method quickly becomes clear. Given the mobility of the system, as well as the requirement to work with standard earbuds, head-related transfer function-based methods are the only viable method for creating spatial audio. The fact that other methods use speaker-arrays makes it so that they are immobile and impractical to use for bikers. On top of that, they are highly dependent on their surroundings for accurate sound creation. Lastly, when looking at the final requirement of high-accuracy, low-latency sound creation, when calibrated correctly, head-related transfer functions are able to fulfill this requirement as well. This makes head-related transfer function-based methods the method to be used for the intervention of aiding bikers in traffic.

2.1.2 Binaural Spatial Audio

Given that binaural spatial audio, using head-related transfer functions has been determined to be the best suited method for aiding bikers in traffic, this next section will go more in-depth into the workings behind HRTFs. Head-related transfer functions contain all the elements needed to place a virtual sound in 3D space, relative to the listener, using only two speakers. To accomplish this, multiple properties need to be given to the sounds. According to multiple sources [6, 9, 14, 15], these are:

- Interaural Time Difference (ITD), the difference between arrival time of the sound for both ears.
- Interaural Level Difference (ILD), the difference between volume of the sound for both ears.
- Filtering, changing the volume of the individual frequencies that make up the sound, also known as spectral cues.
- Reverberation, the reflections of the sound from objects that prolong the sound after the 'dry' sound has finished.

So, the HRTF takes a directional input relative to the listener's head and uses this to add these properties to the sound that should be played. This results in the simple diagram that can be seen below.

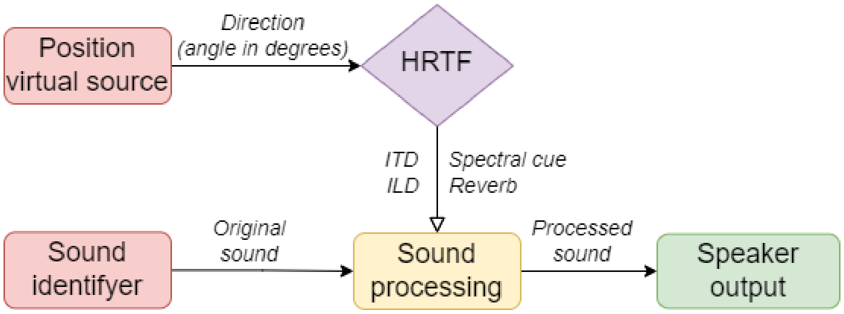


Figure 1: Schematic of simplified HRTF-based spatial audio system.

Since the audio will be placed in 3D space, the direction consists of two components: The angle in the horizontal plane: azimuth, and the angle in the vertical planes: elevation [6][9][14-17]. The azimuth and elevation of the direction of the sound each result in different properties of the virtual sound. The resolutions of the azimuth and elevation in degrees determine how precise the spatial audio can be created. *Zhang et al.* [6] states that the aim is to achieve a resolution of 3-4 degrees for both parameters. *Spagnol et al.* [15] states that elevation has a higher localization blur than the azimuth, namely 1-2 degrees for azimuth and 4 degrees for elevation. Localization blur is the lowest directional change that humans can still perceive. The picture below shows the azimuth and elevation in relation to the direction.

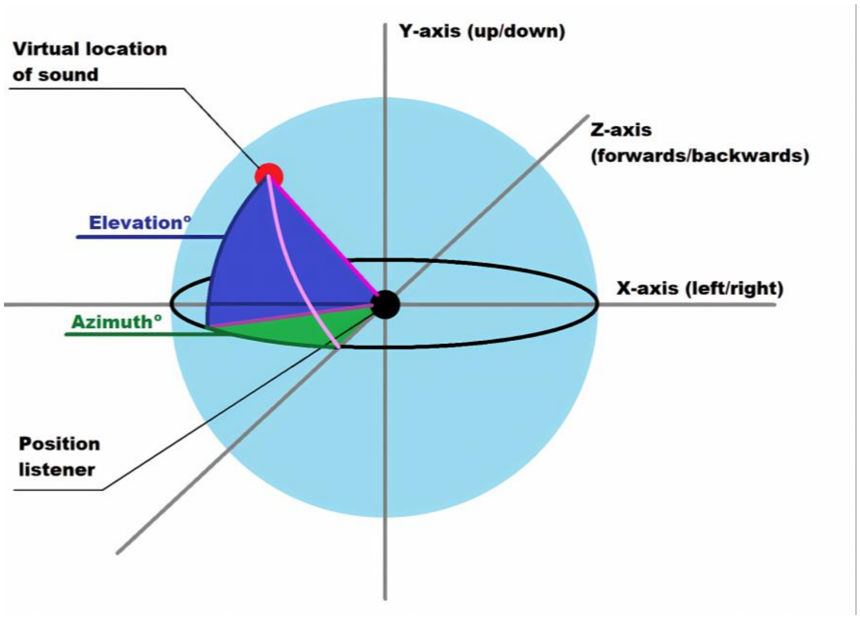


Figure 2: Visualization of azimuth and elevation, inspired by Urviola et al. [9] figure 1.

Parameters influencing azimuth angle.

To replicate the azimuth angle, the most influential parameters are Interaural Time Difference (ITD) and Interaural Level Difference (ILD), see *Buis et al.* [14]. When a sound comes from the left, it will arrive at the left ear earlier, and with a slightly higher volume, than at the right ear. Should a sound come from the front or back, there is no difference between these parameters. Due to small differences between the left and right side of the head, a sound coming from the left will not have the same ITD and ILD as the sound coming from the right at the same angle [15]. *Spagnol et al.* [15] also clarify that ITD and ILD are frequency dependent. Their article explains the following: “ITD is known to be frequency-independent below 500 Hz and above 3 kHz, with an approximate ratio of low-frequency ITD by high-frequency ITD of 3/2, and slightly variable at middle range frequencies.” [15, p. 2]. They also state that effects caused by the shape of the human head make the ILD greatly frequency dependent. This effect is particularly great at frequencies above 1 kHz. However, to differentiate between front and back (azimuth angles between 90 and 270 degrees, relative to the front), ITD and ILD are insufficient. Just like up and down, in this direction, spectral cues will need to be introduced.

Parameters influencing elevation angle.

To give virtual locations above and below the listener as well as forwards and backwards, spectral cues will need to be added to the sound. These are filters that replicate the peaks, local rises, and notches, local dips in the frequency spectrum created by the listener's head (figure 3 contains an example of such filters on white noise). Based on the way the shape of the ear as well as the shape of the head influence the sound, a Head-Related Impulse Response (HRIR) can be created, which can then be used to extract HRTFs. *Onofrei et al.* [17] establishes such HRIRs and goes in-depth into the relationship between the elevation of the sound and the necessary spectral cues. Here they use a KEMAR dummy model and attach so-called pinnae, different models of ears, to generate HRTFs for different ear models [17]. Their graphs show that a higher elevation is correlated with an increase in the frequency of the N1-notch, or the notch with the lowest frequency. The frequency range for these N1-notches is between 3 and 16 kHz [17]. There were however large individual differences to this effect, which again shows the individualistic nature of HRTFs.

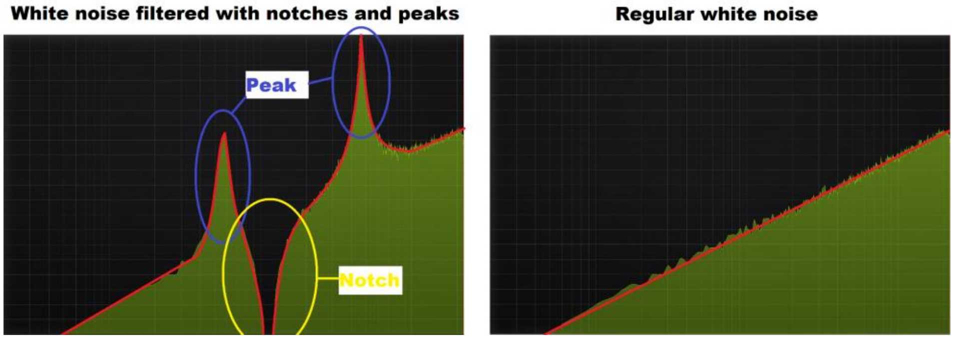


Figure 3: Frequency spectra of white noise filtered through peaks and notches (on the left) and regular white noise (on the right). Here the horizontal axis represents the frequency (Hz) and the vertical axis represents the corresponding loudness (dB). The horizontal axis is a logarithmic scale.

Parameters influencing perceived distance.

Creating a sense of distance in the sound is in principle simpler to achieve. *Kolarik et al.* [19] state that a combination of volume level balancing, low pass filtering and reverberation is mostly used to create a sense of distance for the audio. They claim that each doubling in the distance of the sound will reduce the volume by 6 decibels (dB). The level of reverberation is less affected by distance and only decreases by 1 dB per doubling of distance. This means that sounds further away from the are perceived to have more reverb than sounds that are close to the listener. *Spagnol et al.* [15] adds to this that the main use of reverberation is for distance cues and to simulate different room sizes. Lastly, *Kolarik et al.* [19] explain that a broadband low-pass filter (LPF) at 5, 6 and 6.7 kHz greatly improves the perceived distance of the sound.

Storing Head-Related Transfer Functions

There are different file formats used for storing head-related transfer functions. *Majdak et al.* [36] discusses multiple storing formats used by different databases. CIPIC uses a unique file per listener, where the index of an HRTF relates to the azimuth and elevation the virtual sound is coming from. OpenDaff is like CIPIC, except that it uses keys as indices for the HRTFs as opposed to the directions directly. LISTEN and ARI instead provide all HRTFs at once in a big matrix, while, like the previous two, also storing unique HRTFs for different users. MARL-NYU blends all the previous databases into a single file. Whereas the previous file formats use Matlab formats while storing the data as matrices, SDIF stores the data in mixed test-based and binary formats. And finally, Concert-hall data is stored in .wav audio file formats.

The SOFA file format tries to combine all the positive aspects of the previous formats into one [36]. Different objects relevant in the spatial audio recording (receivers, listeners, emitters, sources, and the room) and their relations are used to describe the measurement setup. This information is then stored as non-zero integers and structured by different measurements. These measurements contain impulse responses, audio characteristics of the setup, of the specific HRTF recording setups and are stored in one big matrix of impulse responses. The information of these formats can be recalled in software and used to process audio to create a spatial effect.

Processing Head-Related Transfer Functions

In order to create an accurate spatial audio experience, audio processing using HRTFs comes very precisely. This is especially the case for the proposed application of aiding bikers in traffic, where mistakes can be fatal. In their article, *Buis et al.* [14] have identified a method of accurately processing HRTFs. Assuming the simplified situation where a sound source *S* can be defined as a flat surface with a number of *N* points spread on top. Their distance d_i is equal and they all produce the same sound $s_t(t)$. They state that the signal that will arrive at the listener's ear can be expressed through the following formula:

$$p(t) = \frac{1}{N} \cdot \sum_{i=0}^{N} \frac{1}{1+d_i^2} \cdot h(\theta i, \varphi i, t) \cdot s_t(t)$$

Formula 1: formula for Head-Related Transfer Function [14]

Here θ is the azimuth of the sound source relative to the user, φ is the elevation relative to the user and $h(\theta i, \varphi i, t)$ is the HRTF belonging to the listener [14]. It boils down to the following: the signal arriving at the eardrums of the listener is the weighted sum of all individual signals generated by each point N at distance d_i, after they went through the HRTF of the listener. Basically, the final signal is the individual sounds coming from the surface of the object added together and manipulated by the HRTF. This means that one virtual sound source is a combination of many different individual sound sources on the surface, allowing for object orientation to be simulated. However, object orientation simulation is not a requirement for the final system and this level of complexity will therefore not be added to the system.

As stated in chapter 2.1.1, head-related transfer functions (HRTF's) are functions that describe the difference in properties of sounds that arrive in the left and right ear. A HRTF can be created by recording the impulse responses of both ears, corresponding to every possible direction a sound could come from [61]. An impulse response is a function that describes the way an environment reacts to an impulse that happens within the environment. For example, when you clap in a room, the sound of the clap would be the impulse and the reverberations and echoes coming after the clap would be the response of the room to the clap sound. The impulse response would be a recording of this echo of the room and thus contain the acoustic characteristics of the room. A head-related impulse response (HRIR) is a recording of the impulse response of the head and ears, following a sound. The creation of a Head-Related Transfer Function involves recording the impulse responses of each ear, following sounds coming from all possibly required directions, see *Onofrei et al.* [17, 61]. The resulting HRTF now contains HRIRs for both ears for all recorded directions. The HRIRs contain information about the Interaural Time Difference (ITD), Interaural Level Difference (ILD) and the frequency spectrum, which, when applied to other sounds, will create the spatial audio effect corresponding to the direction of the recording. These features can either be extracted to get the

ITD, ILD and spectral cues to create spatial audio, or the raw HRIRs can be applied to the sound directly.

The way in which the HRTF data is applied to the sound (using the ITD, ILD and spectral cues, or the raw HRIR), determines the technique with which the sound will be processed. Applying the raw head-related impulse responses uses a sound-processing technique called convolution. Convolution is a tool that can apply the characteristics of one sound onto another. It does this by multiplying the frequency spectra of the original sound with the sound that contains the desired characteristics, the impulse response [56].

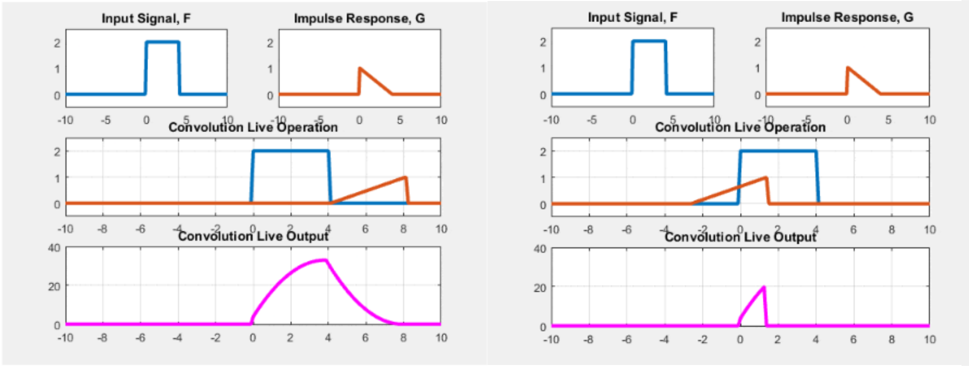


Figure 4: visualization of the process of convolution, see Quincy's Website [57].

This process can simulate what any sound would sound like in the environment to which the impulse response belongs to. For example, the cabinet of a guitar can be recorded and applied to the sound of a synthesizer to make that synthesizer resemble the sound of that guitar. The process of convolution is explained in more detail by Izotope [56] and theProAudioFiles.com [58]. The same technique can be used to create spatial audio. Here, the HRIRs corresponding to the desired direction are selected and through convolution, the selected HRIRs are applied to the sound that should be spatialized. The sound now contains the characteristics of the acoustic properties of the head and ears corresponding to the to-be simulated direction, and spatial audio is created. In a way, the listener listens to spatial audio by listening through the ears of somebody else. *Cuevas-Rodríguez et al.* [37] uses this technique in their spatial audio software library.

Instead of using the raw HRIRs, the ITD, ILD and spectral cues can be extracted from the HRIRs, as described in [59]. These are then applied to the sound through filtering (for spectral cues), volume gain (for the ILD) and time delays (for the ITD). A common method of representing the spectral cues is in parametric form. A filter curve to apply filtering to a sound can be described as a combination of sub-filters that each complete one simple operation: increasing or decreasing the volume of a specific set of frequencies of a sound. If enough of these operations are done in a specific way, the original filter curve can be recreated. According to *Song et al.* [13], an advantage of this approach is a lower processing load on the CPU of the spatial audio device.

Figure 5 shows how these simple operations (the blue, orange, yellow, magenta, and red lines) add up to create a more complex filter curve (the area in turquoise).

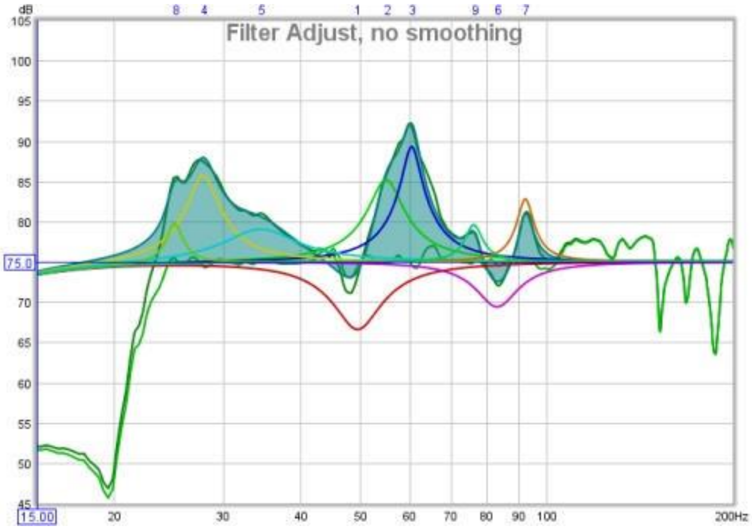


Figure 5: Visualization of parametric filtering, see roomeqwizard.com [60].

Each of these simple operations use parameters, namely filter frequency (the core frequency where the filter is applied), gain (volume increase or decrease caused by the filter) and quality factor (preciseness or slope of the filter), to describe the effect the operation has on the sound, hence the name ‘parametric filtering’, see *Belloch et al.* [20].

An example of this is a simple, real-time method for processing binaural spatial audio using HRTFs for GPU-equipped mobile devices, proposed by *Belloch et al.* [20]. The method makes use of parallel processing to process the audio. *Belloch et al.* [20] state that the HRTF method uses the parametric approach, which allows HRTF filters for unknown azimuth and elevation angles to be interpolated. According to the article [20], two known filters will be averaged out so that a new filter can be created for the unknown azimuth and elevation values. This means that part of the HRTF can be generated during processing, meaning that less data is needed for the HRTF, resulting in lower phone storage usage. The results show that the approach can significantly reduce the computational complexity of the HRTF models while keeping high-quality sound output. The proposed approach also allows for efficient customization of the binaural sound experience by adjusting the parameters of the HRTF models.

Tracking head orientation

In order to place virtual sounds relative to the listener, the orientation of the head must be tracked. For example, if a sound is coming from in front of the listener. If the listener then turns their head to the left, the sound suddenly seems to be coming from the right side of the head. If this is not considered, virtual sound locations will rotate with the head. In the case of the intervention in this paper, this would misinform the biker about traffic and potentially cause accidents. It is therefore crucial to consider the movements of the head.

The movements the human head makes depends on the situation and the specific person. However, *Xiao et al.* [34], who have analyzed head-motion patterns, were able to identify nine different types of typical head motions. These are grouped in different dimensions, namely nods or shakes, small or large movements and fast or slow movements. *Hadar et al.* [35], investigating head movements during conversations, identified five different classes of head movements, namely slow movements, ordinary movements, rapid movements, posture shift, and tremor. They plotted the movements of the heads of different participants and analyzed them as a wave of motions. Here they filtered on different frequencies of head movements, and the amplitudes of those movements. Their research confirmed that the head movements can vary greatly per person. The maximum type of rotational movement they identified for an individual had a mean rotation of 52.6 degrees. Their results also showed that the larger the rotation, the less often the motion occurred.

A device to track the orientation and movement of the head is called a head-tracker. There are many different types of head-trackers. *Al-Rahayfeh et al.* [33] claims there are four:

- Computer-vision-based head movement detection, where a computer can detect the head orientation through image or video input.

- Acoustic-signal-based methods, where the head-direction is detected through the sound of the users’ voice.
- Accelerometer and gyro-sensor based methods, that use movement sensors - accelerometers and gyro-sensors - to detect the user's head movement.
- Hybrid head tracking techniques, which use a combination of the above-mentioned methods.

Each method has different use cases. Whereas the accelerometer and gyro-sensor based methods are attached directly to the users’ head, other tracking methods are fixed in place outside of the users’ body. *Al-Rahayfeh et al.* [33] claims that the acoustic-signal-based method they discussed was very limited in its usability; it was only able to detect three types of movement and susceptible to noise from the outside world. The other methods were more reliable with a combination of them yielding the best results.

Best Devices for Binaural Spatial Audio Playback

The specific earbud/headphone used for audio playback influences the quality of the spatial audio. As stated by *Nowak et al.* [4], the spectral coloration caused by the specific device reduces the accuracy of the spatial audio quality. This implies devices with a flatter frequency response create better spatial audio quality. However, Headphone Transfer Functions (HRTFs), equalizations applied to the sound to negate the spectral coloration caused by the earbuds, can be used to improve the sound quality of the earbuds. Applying this to the audio makes the specific spatial audio playback device less influential on the quality of spatial audio.

Another article [50] found no significant correlation between price of the playback devices, their impedance and whether they are open or closed, and the spatial audio quality. It therefore depends on the devices themselves, rather than their properties, whether they are good at spatial audio playback. Multiple sources [51, 52, 54] mention the ‘Sony WH-1000XM4’ and ‘Apple Air Pods Max’ as the best headphones for spatial audio. Sources [52, 53, 55] mention ‘Sony WF-1000XM4’ and ‘Air Pods’ as the best earbuds for spatial audio listening. This could be due to their spatial audio software. However, one can make the claim that they must have been optimized for spatial audio playback and are therefore generally good at producing spatial audio, regardless of whether they contain spatial audio software.

2.1.3 Soundscapes

To be able to develop a spatial audio system that can be used to aid bikers in traffic, it is useful to determine which soundscapes it should create or enhance. A soundscape is defined as “the quality and type of sounds and their arrangements in space and time.”, see *Southworth* [62]. Basically, we need to know which sounds in traffic are important to spatialize, at which position. *Stelling-Konczak* et al. [1] investigated what sounds are important for safe cycling in which traffic situations, as well as the effect of limited auditory cues on traffic safety. The study found several things. Firstly, listening to music, especially with lyrics, has a bigger negative impact on perception of sound than phone conversations. Furthermore, cyclists that listen to music were more likely to engage in risky behavior and disobey traffic laws. All cyclists in the study got mainly startled by other cyclists and cars, indicating that large vehicles such as trucks and buses were more perceptible to the cyclists. Finally, the study found that, although listening to music was not related to an increase in traffic accidents, it does co-occur with participating in other risky behaviors.

However, *Goldenberg et al.* [63] concluded that frequently listening to music while biking results in a 1.6 to 1.8 times higher chance to be involved in a traffic accident. Both *Stelling-Konczak* et al. [1] and *Goldenberg et al.* [63] found that listening to music causes cyclists to participate in compensatory behaviors, including looking around more often and slowing down the biking speed. *Stelling-Konczak* et al. [1] also suggested that the rise of electric cars, which are more silent than normal cars, might have an impact on traffic safety, in particular an increase in incidents related to the inaudibility of the vehicles.

The (un)safety of vehicles that produce little sounds is confirmed by the Dutch National Institute for Health and Environment (RIVM) [64]. They summarize research findings in the United States, Japan and the Netherlands on the number of traffic accidents involving normal cars versus the more silent hybrid cars. For the research in the United States, in situations where cars were driving slowly, hybrid cars were twice as likely to hit pedestrians and cyclists than regular cars. For the research in Japan and the Netherlands, hybrid cars were only found to be more inaudible than regular cars when their speed was under 20 km/h. This resulted in the hybrid cars being perceived later than regular cars. Lastly, the RIVM states that pedestrians and cyclists will perceive vehicles outside their vision later when these emit relatively low levels of noise, resulting in less time to anticipate and thus higher risk of injury. At speeds under 15 km/h, the time the pedestrians and cyclists must anticipate vehicles is halved for electric and hybrid vehicles, compared to regular vehicles.

All these findings lead to the following determinations for which soundscapes should be generated and where:

- Silent traffic users cause a higher risk for traffic accidents than non-silent traffic users. This means that especially auditory cues from silent traffic users, so cyclists and electric/hybrid vehicles, should be created and enhanced. However, the sounds of noisy road users, so regular cars, trucks, and buses, should not be disregarded.
- The soundscapes of traffic users outside the vision of the cyclist are especially important to create/enhance, given the findings by the RIVM on the anticipation time because of delayed perception [64].
- Slow-moving silent moving vehicles yield the greatest increase in traffic accidents compared to their slow-moving non-silent counterparts. So extra attention should be paid to creating soundscapes for slow-moving silent vehicles.
- It is debatable whether soundscapes of traffic users in the cyclist’s view should be created/enhanced at all. It is implied by the RIVM [64] that silent vehicles within the field of vision of the cyclist are not detected later than their non-silent counterparts. This could indicate that simulating soundscapes within the cyclist’s field of vision will not yield improved detection accuracy. In this case, to ensure clarity for the soundscapes that should be created/enhanced, the soundscapes from within the field of vision should not be generated.
- Lastly, to ensure audible clarity of other soundscapes, the soundscapes of traffic members that move away from the biker (and thus do not pose a risk) should not be generated.

2.2: State of the art

2.2.1 Existing interventions to aid bikers using spatial audio.

Bikers are the road users that are most likely to be killed in a traffic accident. In 2021 there were 207 traffic fatalities involving bikers, the highest among all road types of road users, followed by cars with 175 fatalities [48]. It should therefore come to no surprise that many interventions have been created to make traffic safer for cyclists. However, there were few that used sounds from earbuds or headphones to communicate to the biker.

Berge et al. [21], an article published in September 2022, identified 92 concepts to aid cyclists in traffic. Of these concepts, 36 are wearables for cyclists. Thirteen of these used augmented reality glasses to aid the biker, of which five are already for sale [21]. The remaining concepts used either smartphones, head-up displays, beacons, helmets, or other technology. The study also identified 35 concepts that used attachments to the bike itself, 30 concepts that used attachments to vehicles and 21 concepts that used infrastructural systems [21].

Of the concepts that communicate to the biker in some way, 25 used auditory stimuli. These include signals or buzzers (17 concepts), speech (12 concepts) and lastly one concept that provides auditory signals through the bones of the wearer [21]. As such, the study [21] did not report any concept that used earbuds or headphones to communicate to the biker. A possible explanation for this is that the designers of the concepts did not want to block the sounds coming from traffic. However, it can be assumed that earbuds and headphones will block these necessary sounds. This line of reasoning fits with the concept that provided auditory stimuli through the bones, allowing the biker to still hear sounds from traffic. In any case, the reality is that there are currently very few interventions that help bikers through feedback through earbuds, much less so by using spatial audio.

2.2.2 Current uses of spatial audio as aids

However, spatial audio is already commonly used in entertainment applications such as video games, cinemas, and music. However, given the topic of the graduation project and the difference in requirements, this section will cover the use of spatial audio as tools to aid people, as opposed to (only) entertainment. To identify the current applications of spatial audio as aids, articles will be identified. The selection criteria are:

- The article must cover an intervention that uses spatial audio as an aid.
- The article must be less than 5 years old.

- The article must be peer reviewed.

The criteria regarding the age and peer review are there to ensure that the state of the art covers only the latest interventions with adequate quality. In the end, ten articles have been identified that discuss spatial audio as an aid for people. The articles are displayed in a table below. Next to this, the publishing year, application, and effectiveness of each application are listed in the other columns.

Article	Publishing Year	Function	Effectiveness
Marentakis et al. [30]	2022	Location monitoring	Adding spatial audio to visual feedback systems can reduce visual attention load and increase user placement flexibility compared to the display. [30]
Hu et al. [31]	2022	Aiding visually impaired people	Triple information transfer rate, 40% lower positional error and lower learning effect compared to mainstream speech instructional feedback. [31]
Cuadrado et al. [25]	2021	Increasing child arousal while listening to audio stories	Spatial audio and arousing sound design result in a higher emotional impact in children, with different depending on the age and the type of interaction. [25]
Greenberg et al. [27]	2021	Decreasing stress	Spatial audio can be effective for short-term stress reduction and could be a useful addition to clinical music intervention. [27]
Reynal et al. [32]	2021	Providing feedback on sound source locations in poor visibility	Spatial sound, in combination with vibrotactile feedback makes it easier to determine the sound source. This finding translates to different areas with limited vision. [32]

		environments for airport towers	
Milam et al. [28]	2019	Navigation aid for pilots during simulated helicopter flights in degraded visual environments.	Spatial audio may prevent risk for pilots flying in degraded visual environments. Techniques using the aircraft as reference were more useful than head-referenced techniques. [28]
Gang et al. [22]	2018	Improving drivers' sense of awareness and safety in electric self-driving cars.	The intervention does improve the drivers' awareness in traffic. However, no conclusion could be drawn to an improvement to the sense of trust and comfort of the driver. [22]
Hoffmann et al. [26]	2018	Aiding visually impaired people	Using the system resulted in higher obstacle awareness and less collisions. It quickly proved just as effective as navigating with the currently widely used walking stick. [26]
Johnston et al. [29]	2018	Reducing hypersensitivity in young people	Significantly reduced negative emotional reactions to aversive stimuli. [29]
Spagnol et al. [15]	2018	Electronic travel aids for visually impaired people	No widespread usage, due to high cost, lack of commercial appeal and few functionalities due to limited user evaluation and scientific value. [15]

Table 1: Interventions that use spatial audio to aid people.

Although there are no interventions identified that use spatial audio to aid bikers in traffic, spatial audio is already widely used in other interventions aimed at helping people. In the medical domain, spatial audio is already used to aid visually impaired people. All three identified articles that cover interventions to aid visually impaired people [31, 26, 15] show that spatial audio is an effective tool for this purpose. In some cases [31], the efficiency of using spatial audio surpassed that of conventional methods.

Another use-case of spatial audio is navigation. Three articles [28, 30, 32] cover interventions that use spatial audio as feedback systems convey location information to the user. Although the settings of the interventions are quite different, both show that spatial audio can be utilized effectively. Spatial audio can lower the visual information load and add to, or even replace missing visual stimuli that are normally used for navigation. *Reynal et al.* [32] even stated that this effect can be used in other domains where visibility is limited. This highlights the potential usefulness for the intervention that is going to be built.

Related to reduced cognitive load, some articles show that spatial audio has other mental benefits to listeners. *Cuadrado et al.* [25] show that spatial audio can be used to increase arousal levels in children listening to audio books. *Greenberg et al.* [27] show that spatial audio is effective in reducing short-term stress. Finally, *Johnston et al.* [29] show that spatial audio can reduce negative emotional reactions in young, hypersensitive people to aversive stimuli.

Lastly, closely related to the intervention of this paper, is the intervention that uses spatial audio to increase comfort and awareness of people in self-driving cars. *Gang et al.* [22] finds that spatial audio can be used effectively to increase traffic awareness. However, no conclusion could be drawn to the comfort and trust of the driver in the car. This is especially relevant to this paper as it confirms the usefulness of spatial audio to reach the goal the intervention tries to achieve.

2.2.3 Current state of advancement for Head-Related Transfer Functions

Spatial audio is a rapidly developing field. Each spatial audio engineer develops a new variation of spatial audio that is slightly different from the standard. These novel applications could prove to be very useful for this research, which makes it important to identify the current technological state of spatial audio. This section concerns itself with the state of the art of binaural spatial audio that can operate in real-time. Towards this end, articles will be identified with the following selection criteria:

- The article must cover an intervention that creates binaural spatial audio in real-time.
- The article must be less than 5 years old.
- The article must be peer reviewed.

In the end, ten articles have been identified that discuss novel techniques for spatial audio. The articles are displayed in a table below. Next to this, the publishing year, spatial audio features and effectiveness of each application are listed in the other columns.

Article	Publishing Year	Spatial Audio Features	Effectiveness
Daraban et al. [46]	2022	The paper presents a system that can create binaural spatial audio in real-time. The system can be operated on wearable computers and its main target group is visually impaired people.	The system is accurate in simulating horizontal directional sound. However, it is insufficient at simulating spatial audio with different elevations. [46]
Bruschi et al. [42]	2021	Presents a real-time spatial audio implementation that is specifically designed to simulate moving sound sources through HRIR interpolation.	In comparison to the state of the art, the proposed system delivers better performance. [42]
Bruschi et al. [43]	2021	The paper proposes a method of interpolating head-related impulse responses in real-time for accurate binaural spatial audio.	The system has shown excellent performance when compared to the measured impulse response and another algorithm with the same purpose. [43]
Arend et al. [44]	2021	The paper proposes a method of real-time spherical microphone array data processing through a method using linear filters.	Presented approach is equivalent in performance to conventional methods, while requiring less computing power. [44]

Belloch et al. [20]	2020	The system uses parallel parametric HRTFs to process binaural spatial audio in real-time with low hardware requirements, specifically for mobile phones.	The system could render up to 32 sound sources in real-time, while utilizing mobile-grade CPUs and GPUs. Rendering 16 sound sources has proven to be the most efficient. [20]
Tamulionis et al. [39]	2020	The system uses room impulse responses and HRTFs to create realistic spatial audio. It uses three parallel audio signals to be able to instantly switch in cases of directional change.	The system does not exceed the 30ms latency limit for real-time operation. It is accurate for 3 degrees of azimuth angle change. The system can instantly switch to a new filter when the virtual source changes direction. [39]
Geronazzo et al. [45]	2020	The paper presents a method of creating individualized HRTFs in real-time, by using two images of the user's head and the room properties.	The paper reports improvements in the accuracy of horizontal sound location, when the method is used in addition to more conventional spatial audio methods. [45]
GS et al. [47]	2020	The proposed system uses time-varying FIR filters to create real-time hardware-implemented spatial audio.	The system was successful in creating spatial audio for headphones. [47]
Cuevas-Rodríguez et al. [37]	2019	3DTI Toolkit, a C++ library for creating real-time binaural spatial audio.	The most configurable C++ spatial audio library currently in existence. The only library accepting custom HRTFs. Can operate in real-time and has processing power-decreasing options. [37]

Pausch et al. [38]	2018	The system generates virtual complex acoustic environments through finite impulse response filters.	The system is a platform for industrial-grade hearing aid algorithm evaluation. It operates in real-time and has low hardware requirements. [38]

Table 2: Articles that discuss novel solutions for binaural spatial audio.

Spatial audio using head-related transfer functions has existed for a while. As can be seen, the current progress is mostly aimed at making binaural spatial audio easier to process and more accurate to the listener. Some interventions [20, 46] aim at making binaural spatial audio wearable. To do this, they decrease the processing power of spatial audio while keeping the result accurate enough. This is done through methods such as parallel parametric HRTFs processing [20] or convolution filters [46]. *Arend et al.* [44] simply decreased the required computational power of processing spherical microphone array data through linear filters, while keeping the accuracy.

Other articles aim to make binaural spatial audio more accurate using HRTF interpolation [42, 43]. These systems, developed by the same research group, have shown better performance than the conventional methods, while providing features such as simulating moving sound sources.

Yet other articles have come up with extra novel solutions to make spatial audio even more accurate. *Tamulionis et al.* [39] has developed a method that can instantly switch the post-processing of sounds upon a directional change through parallel audio signal processing. *Geronazzo et al.* [45] has succeeded in making HRTFs more individual using head images of the listener. *GS et al.* [47] was successful in creating real-time, hardware-implemented spatial audio using time-varying FIR filters.

Lastly, yet other interventions aim at easing the development of other spatial audio systems [37, 38]. *Cuevas-Rodríguez et al.* [37] provide an accurate, flexible C++ library, whereas *Pausch et al.* [38] created an industrial-grade system that can evaluate hearing aid performance.

2.3: Discussion and conclusion

The state-of-the-art research looked at interventions that use spatial audio to aid people, as well as recent research into the relevant spatial audio method. It was found that spatial audio is already commonly used in interventions to aid people, especially visually impaired people. It showed that spatial audio can be used effectively to increase spatial awareness as well as aiding in navigation. On top of that, applications already exist that create real-time spatial audio and inspiration can be drawn from these. A spatial audio library has been identified which could be useful for this project as well. Lastly, another article [21] researched current interventions aimed at increasing traffic safety for bikers. It found that, although interventions using sound to aid bikers already exist, none of these use spatial audio. Although, one intervention was identified that provides auditory cues through devices attached to the bones of the listener. From this the following can be concluded: the currently proposed intervention is novel and has great potential to work.

With the knowledge obtained from the background research and state-of-the-art, some of the sub-questions of this research can be answered.

- What kind of virtual soundscapes should be generated and where in 3D space?

Virtual soundscapes of all traffic users outside of the cyclist's field of vision that move towards the cyclist (which could therefore collide with the cyclist) should be generated. Especially soundscapes of electric/hybrid cars and cyclists should be created by the system. And this is especially the case for situations where these road users are moving below 20 km/h.

- What kind of headphones/earbuds can give the most suitable 3D effect?

The background research was unable to identify what kind of headphones/earbuds give the most suitable 3D effect, apart from the flatness of the frequency response. However sources mention 'Sony WH-1000XM4' and 'Apple Air Pods Max' as the best headphones and 'Sony WF-1000XM4' and 'Air Pods' as the best earbuds for spatial audio reproduction.

- Which technique for creating spatial audio in real-time is the most useful for this intervention?

As discussed in the first part of the literature research, multiple different methods of spatial audio have been identified. It became clear, however, that the only methods suitable for creating spatial audio for headphones/earbuds make use of head-related transfer functions. The methods also require information from a head-tracker and use either convolution or parametric filtering, gain and delays to create the spatial audio effect.

- How effective is the system in generating sounds that are locatable in 3D space?

This final sub-question is something that can only be answered through experimental evaluation with participants.

Chapter 3: Methods and Techniques

3.1 Chosen Design Method

The design method used for this research is the Creative Technology design cycle. This design method is a combination of divergence-convergence and spiral models [40]. According to *Mader et al.* [40], it contains four phases: Ideation, Specification, Realization and Evaluation. Here, each of these phases follows and is dependent on each other. However, the method is designed to be able to go back and adjust an earlier phase based on the progress that is made in later phases and newly obtained knowledge.

The method has been chosen for its similarities with the graduation project structure. This structure includes the same phases and allows for an iterative design process. On top of that, the Creative Technology design cycle is a useful method for client projects, such as the graduation project. Clients do not necessarily have all the background knowledge needed for developing a system. Therefore, as the clients themselves often do not exactly know what system they want, the developer will have to perform research and client consultations to find this out. Given the undetermined nature of this design process, an iterative approach is necessary. This makes the Creative Technology design cycle a fitting design process.

3.2 System Requirements

Requirements have already been identified.

3.3 Stakeholder Analysis

Not applicable.

3.4 Concept generation

Based on the literature research and state of the art research in the background analysis and requirements elicitation, current techniques of creating spatial audio will shortly be highlighted again. The technique that would likely perform the best at aiding bikers in traffic will be chosen. With the help of a use-case scenario, functional and non-functional requirements will be set which the prototype should fulfill. At the end, the final concept will be specified. This process will happen in both chapter 4 and 5. The brainstorming method that will be used is called Gap Filling. As explained in a blogpost by *Jouni Halme* [41], gap filling fills the gaps between the current state and the goal. In this case, the gap of which specific techniques of creating spatial audio should be filled based on the requirements.

3.5 Prototype Specification and Realization

To specify and realize the prototype, the decomposition technique will be used. This technique breaks down the product/service into small blocks that together make up the whole system. These blocks can then individually be specified and created, allowing for an easy to understand and easy to implement design process.

3.6 Prototype Evaluation

For evaluating spatial audio systems, multiple sources [42, 43, 46] report the use of both objective and subjective measurements. Objective measurements include measuring the absolute properties of the sound, with which the quality can be determined. Subjective tests include human participants listening to the spatial audio to determine the clarity, accuracy, and experience of the sound [43]. *Daraban et al.* [46] performed three types of measurements on blind or visually impaired people. One with the same direction but different distances, to measure the distance accuracy. One with the same distances, but different directions, to measure the azimuth angles. And lastly, one where neither the direction nor the distances were the same. This method will also be used by this experiment.

To evaluate the performance of the final system, participants will perform tasks while only relying on the developed spatial audio system. The participants will be tested in a gamified setting, where they must answer the questions while wearing the headphones with spatial audio. A gamified setting creates an interest to participate and generates natural motivation to perform well within the participants. The extent to which the participants can complete these tasks is an indicator of the performance of the spatial audio system.

Two types of tasks will be conducted: a task where participants must determine which of three locations produced the sound, to evaluate the directional accuracy. And a task where participants must determine which out of three sounds is located at a different position, to determine the resolution with which spatial audio can be created.

To test any learning effects with spatial audio as an influencing factor on the results, some participants will be trained to use spatial audio. The participants will complete similar tasks as in the evaluation phase but using proven methods of spatial audio. Their performance for these tasks will also be measured to compare to the performance of the prototype.

Chapter 4: Ideation

As stated in the methods, a stakeholder analysis is inapplicable to this research. This comes because the end product does not have a specific purpose, other than accurately producing spatial audio. In the introduction it was said that this system should be part of a bigger system aimed at improving danger awareness and traffic safety for bikers. However, that is only the motivation behind development; it could have many more applications. Next to this, it is unclear whether this product will eventually be transformed into a purchasable product. The only stakeholder is the listener itself. However, requirements for the listening experience have already been identified in the background research. Altogether, this makes the stakeholder analysis inapplicable to this research.

In the research question, several preliminary requirements have already been established: The intervention must be able to work with earbuds or headphones. Next to this, the intervention must be usable while riding bikes. And lastly, the intervention must also be able to produce accurately locatable sounds in real-time.

As explained in the background research, there are multiple ways of achieving spatial audio. The first is using loudspeaker arrays. These can create spatial audio in a room without the need for speakers behind the listener. An extension of this method is to create virtual speaker sources.

The other method of creating spatial audio is binaural spatial audio, a method that works with headphones and earbuds. Using a Head-related transfer function, the perception of direction can be provided to the user. There are multiple processing techniques for applying the HRTFs to create spatial audio.

Based on the preliminary requirements and the background research, this last method has been chosen to be developed. Using speaker-arrays is impractical for bikers and is most effective within rooms. Given how well binaural spatial audio can work with headphones and earbuds, this method is deemed the best option for the prototype. The next section, specification, will specify further how the system will be built and what components it should consist of.

Chapter 5: Specification

The following chapter aims to identify the system requirements for the prototype. These will be found by analyzing a persona and interaction scenario with the spatial audio system. At the end, a time sequence diagram will be developed, which details the sequence of interactions between the persona and the functions of the system. The persona and interaction scenario are based on the upgraded version of the prototype from this project, which can also automatically detect dangers in traffic. Although the persona and interaction scenario do not directly apply to the prototype, the requirements of the prototype should follow the requirements of the system presented in the interaction scenario.

5.1 Persona

PHILIP COOPER

Age : 14

Gender : Man

ABOUT

Philip Cooper is a second year high school student. Although he performs well, he is quite often bored and ends up listening to music in the back of class. Every morning and afternoon he rides his bike across the city, navigating busy traffic situations. He always does this while listening to heavy metal on his earbuds. He sometimes ends up in minor traffic accidents.

PERSONALITY

Figure 6: Persona profile for determining requirements.

5.2 Interaction Scenario

Two scenarios will be presented in which Philip is riding his bike across town to school and gets into a dangerous traffic situation: a worst-case scenario and a scenario when Philip is using the system which will aid him through the dangerous situation. First the worst-case scenario:

Philip is on his way to school by bike. During the ride he listens to music through his earbuds. The earbuds and the sound of music dampen the volume of the surrounding traffic. As a result, he fails to notice the electric car next to him that wants to take a turn to the right to cross his biking lane. The electric car crosses his lane and collides with Philip. Philip is now in the hospital due to the car crash and failure to notice it because of the earbuds.

Now the scenario where Philip uses the spatial audio system:

Philip is on his way to school by bike. During the ride he listens to music through his earbuds. The earbuds and the sound of music dampen the volume of the surrounding traffic. As a result, he fails to notice the electric car next to him that wants to take a turn to the right to cross his biking lane. Normally this would result in a crash, however, Philip makes use of the spatial audio system. The system immediately creates the sound of an electric car coming from the left at 4 meters. Philip recognizes the virtual sound and its location and swiftly turns his head to the left. The sound seems to have rotated around his head, now coming from directly in front of him, from the position of the car. He quickly realizes the impending danger and stops in time to prevent the collision.

5.3 System Requirements

With the help of the interaction scenario, the requirements for the system can now be identified although some of them were already specified in the research question. A distinction is made between functional and non-functional requirements. Functional requirements state what the system must do, the non-functional requirements state how the system will do it. The functional requirements have been categorized according to the MOSCOW method [65] with the categories “Must be able to”, “Should be able to”, “Could be able to” and “Will Not be able to”.

5.3.1 Functional Requirements

The system **must be able to**:

- Spatialize audio in all directions, mainly outside the field of vision of the user.
- This must be done accurately, which 3 degrees of precision.
- Simulate the distance with 1 meter of precision.

- Provide auditory feedback in real-time, the feedback latency must be less than 30 milliseconds.
- Simulate the absolute position of the sound source, therefore sounds must be able to move with the object as the listener’s head turns. Therefore, the movements of the user’s head will need to be tracked.
- Spatialize multiple sounds at the same time.
- Be used by bikers, and therefore be mobile.
- Work with regular earbuds or headphones.
- Be evaluated with participants.

The system **should be able to:**

- Play different sounds of traffic.
- Work with a phone (though other portable devices such as Raspberry Pi would be sufficient for the research question as well).

The system **could be able to:**

- Play sounds with different modes, for example navigation mode and manual positioning.

The system **will not be able to:**

- Detecting sounds/objects in traffic.
- Function as a standalone intervention.
- Aid bikers in traffic.

5.3.2 Non-functional Requirements

The non-functional requirements are linked to the functional requirements and represent how the system will integrate the functional requirements. The non-functional requirements have been ordered in descending order of importance.

The methods of sound processing can either be convolution or parametric filtering (combined with gain and delay effects). Convolution is simpler to implement, as it requires less code, less configuration and the HRIR’s in the HRTF will not need to be parametrized. Parametric filtering does allow for more control and the interpolation of new filters that are not present in the HRTF. However, it is questionable whether this control is necessary for the spatial audio prototype. Therefore, the method of processing the audio will be convolution, chosen for the simplicity of its implementation. To simulate the distance, a combination of reverb, low pass filtering and volume gain will need to be used.

The system does not require a function that identifies sound sources from traffic. However, these sound sources will need to be virtually created to be able to spatialize the audio. Therefore, the system must have some code for a virtual location, which can be moved around, to create the spatial audio. The prototype must be able to create multiple objects of this virtual position to be able to simulate multiple audio sources. To be able to evaluate the prototype with participants, a mode will be created that can position a sound based on a predefined direction and distance.

Given the mobility requirement of the system, the most logical processing device to be used is the mobile phone. However, for testing the accuracy of spatialization of the sounds, since mobility is not a requirement in this case, a laptop or office PC would suffice too. The application will be developed on a windows laptop, and afterwards exported as an android application.

To be able to make the sound rotate with the user's head, a head-tracker will be used. A gyro-sensor/accelerometer-based version would be the most straightforward to implement. Acoustic-signal-based methods are too limited in the directions they can detect to be useful. Computer-vision-based methods would be complicated to set up and it is questionable whether these methods yield better results than gyro-sensor/accelerometer-based methods. Therefore, a gyro-sensor/accelerometer-based version will be developed, given its straightforward implementation.

There are many options for the software with which the prototype could be developed. Possible languages for developing standalone applications are Python, Java, JavaScript, C/C++, C#, Swift, and Ruby. C++ is commonly used for application development due to its high speed of operation, wide functionality, and high number of libraries. C++ also contains many audio processing frameworks, such as JUCE [66], 3D tune-in toolkit [37] and OpenAL [66]. This high speed and the requirement for real-time operation makes C++ the best language for developing the prototype.

Of all frameworks, JUCE was chosen as the framework to develop the prototype with. JUCE works with all relevant operating systems, including Windows, MacOS, Linux, IOS and Android [66]. On top of that, JUCE allows for quick prototyping, is commonly used by audio plugin developers, and contains many DSP processing functions, including convolution and filtering. Although there are libraries, such as the 3D tune-in toolkit and OpenAL, that are used specifically for spatial audio development, JUCE is chosen over these as audio processing framework. This choice was made because the spatial audio libraries have already built the systems for spatializing audio and as such, using these frameworks would mean reusing a spatial audio system, instead of developing a new one. It is also possible to develop the software from the ground up, however this would be too complicated and error-prone for this graduation project. Next to this, the flexibility, wide functionality, and ability to quickly prototype an audio system made JUCE a great choice as a framework to develop the spatial audio prototype on. The chosen IDE for developing the C++ code is Visual Studio, given its compatibility with JUCE and the ease of setting up and managing projects.

The system must contain an interface with which users can select different play modes and sounds. It allows for the options to position the sound manually and automatically and must require a mode that can be used to evaluate the prototype with. The prototype will be able to play different sounds, based on input from the GUI. These sounds will be sounds from traffic. In the end, four different sounds were chosen: a truck, an electric car, a biker bell and pedestrians having a conversation.

To be able to be used with headphones or earbuds, the prototype will make use of binaural spatial audio, which uses head-related transfer functions to function. These can be found online in the form of one large .sofa or separate .wav files. The first file type requires the implementation of a library that can read .sofa. The second file type does not require such a library and is much more flexible for development, albeit more chaotic due to the use of separate .wav files for each direction. Given the complexity of setting up an extra library with the audio framework, the chosen filetype of storing the impulse responses will be a collection of .wav files, as opposed to a single .sofa file.

5.4 Functional Architecture and Time Sequence Diagram

The requirements translate to the following concept that can be seen in the schematic below:

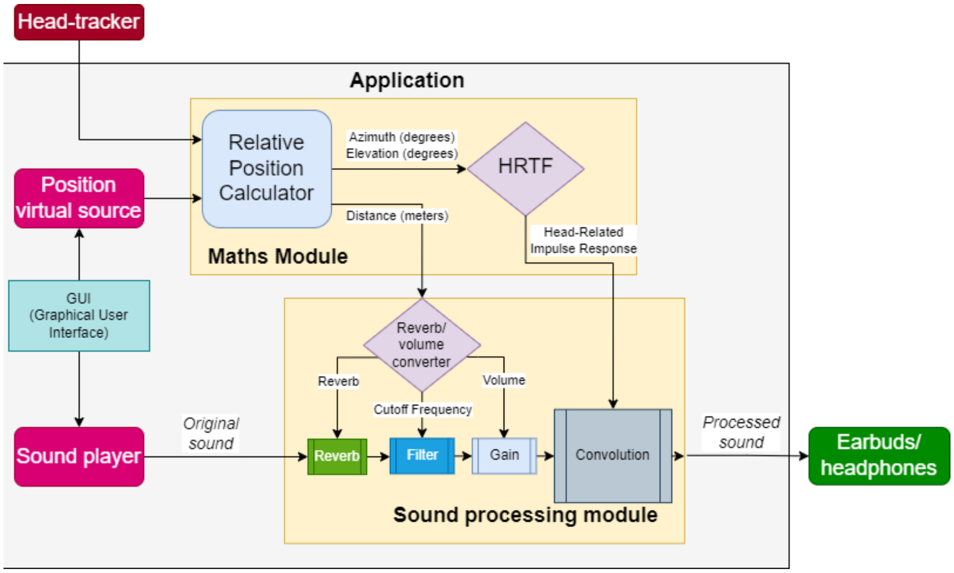


Figure 7: Schematic of final concept of intervention.

Through the graphical user interface, the user can select what sound should be played at which position and with which play mode. These get sent to the sound player, and the position of the virtual source module. The user can add another sound through the GUI which will receive its own menu. The virtual location, together with the head orientation from the head-tracker, gets sent to a math module that calculates the relative position of the source compared to the head. The direction gets sent to the HRTF, which sends the corresponding HRIR to the sound processing module. The raw distance gets sent to the sound processing module as well, which contains a reverb/volume converter to determine the values of the reverb, low-pass filter, and gain. The sound processing module contains a chain of effects, including a reverb, low-pass filter, and gain for simulating the distance. The convolution processing module adds direction to the sound based on the HRIR that is put into it. The final processed spatial sound gets sent to the earbuds or headphones, where the user can listen to the spatial audio.

Below is the time sequence diagram that belongs to this interaction:

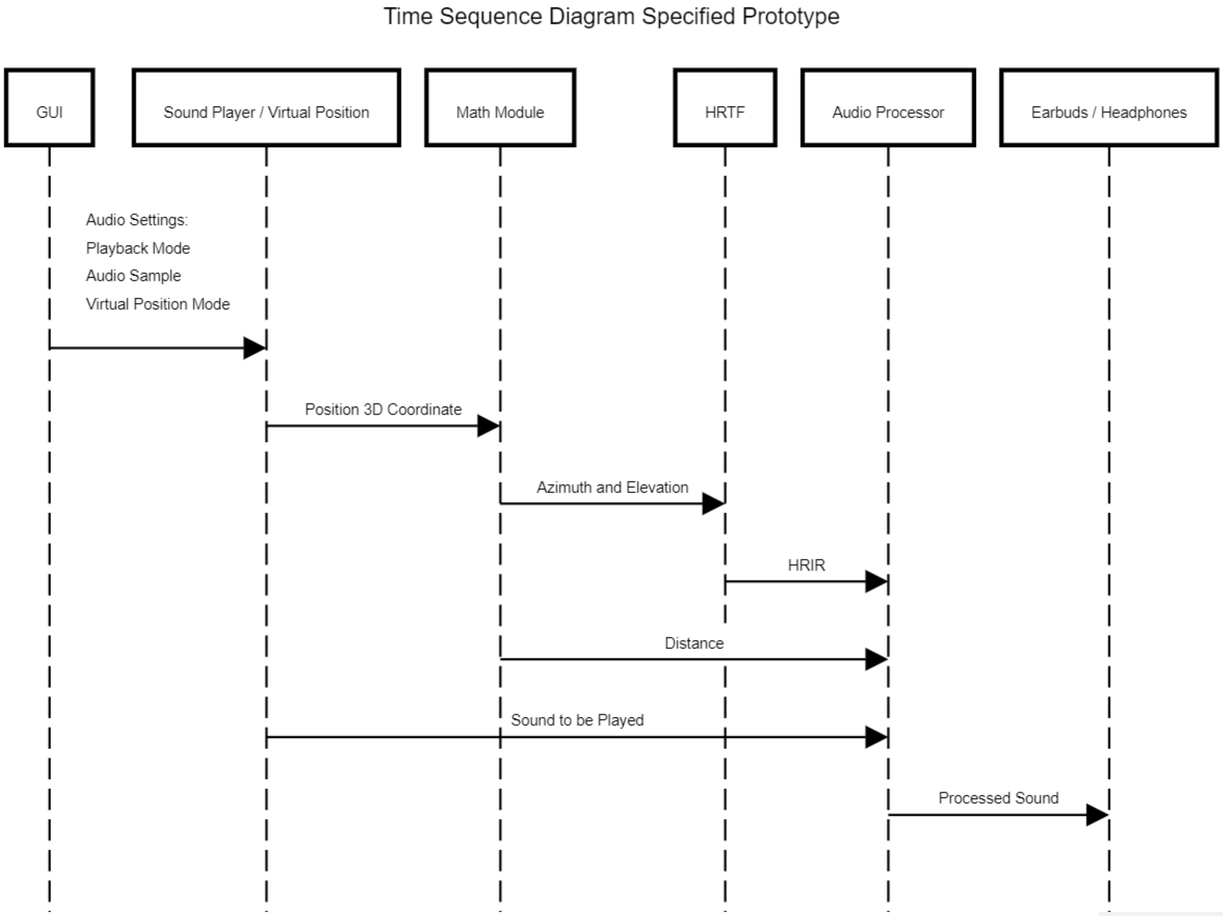


Figure 8: Time Sequence diagram of interaction with prototype.

Chapter 6: Realization

The following chapter will describe the realization of the spatial audio prototype. First, the chapter will outline the developmental process, the prototype itself, its functionalities and structure. After this the individual subsystems will be discussed in detail, followed by a description of integration of these subsystems.

6.1 Developmental Process

Development of the prototype took place on a windows laptop in Microsoft Visual Studio C++, as well as Arduino. The development took place mostly from the researcher's home. The process did not proceed linearly, although at no point were two features developed at the same time. There have been a significant number of problems along the way, mostly in the form of setting up the project, as well as errors that needed to be worked around. However, most of these errors were able to be solved, except for exporting the application as an APK file and running the exported program on Raspberry Pi.

Tools used for the development.

To develop the prototype, the following tools were used (categorized by type):

Physical devices:

- Acer Aspire 7 Windows 10 Laptop.
- Windows 10 home pc.
- Raspberry Pi 3 1GB Ram.
- SEEED XIAO BLE SENSE [68].
- Adafruit FLORA Accelerometer/Compass Sensor - LSM303 [69].

Software:

- Microsoft Visual Studio (C++).
- Arduino.
- JUCE [66].
- Android Studio.
- Wine [70].

Guides:

- JUCE documentation.

- YouTube.
- Online forums such as StackOverflow.com.
- C++ tutorial websites such as cplusplus.com.

Uncompleted Features

During the implementation process, some features were unable to be developed, due to the number of errors present. These features were: converting the C++ application into an APK file through Android Studio, making the .EXE file from Visual Studio run on Raspberry Pi through Wine [70], and setting up the Serial communication from Arduino to C++ Windows. For this last feature, working code has been developed that will retrieve the data from Arduino, when run separately from the rest of the software. However, when the ‘SerialReader.h’ file is implemented to work together with the rest of the Windows code, it conflicts with one of the running threads. On the other hand, the main functionalities, the spatial audio system itself, was successfully implemented.

6.2 Prototype

At the end a prototype was developed that was able to add spatial audio to selected audio samples. The prototype consists of a head-tracker, with its own software, and a main program that runs on the computer. Six subsystems were created that operate together to make the prototype work. The diagram below shows a small schematic of how some of the subsystems interact with each other:

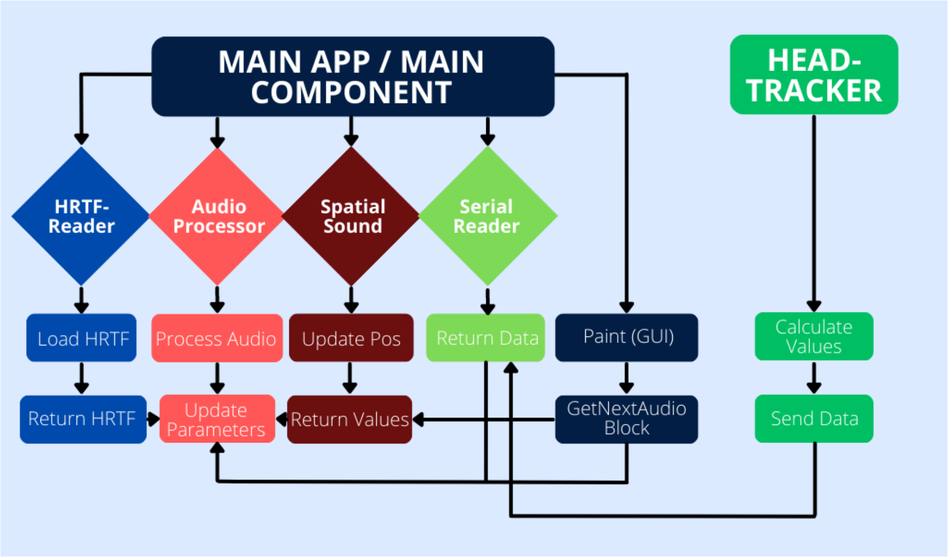


Figure 9: Interaction chart of global prototype.

There is a Main application handler, which does nothing but creating a screen and running the Main Component subsystem. This subsystem is standard code for developing an audio application in JUCE and is unaltered. The Main Component is also standard for JUCE applications; however, this class has been heavily edited to contain the necessary interactions. All other subsystems were created from scratch. Subsystems HRTF-Reader, Audio Processor, Spatial Sound and Serial Reader are called from the Main Content Component subsystem. The Head-Tracker is its own device and only sends directional data to the Main Content Component through the Serial Reader. In total, the prototype consists of 2801 lines of code; 2358 for the main application and 443 for the Head-Tracker.

6.2.1 First Subsystem: Main App / MainComponent

The first subsystem, the MainComponent subsystem, consists of three files: Main.cpp, MainComponent.cpp and ‘MainComponent.h.’ Main.cpp contains the basic startup code for the JUCE application. It creates the window where the user can interact with the software and takes care of starting and closing the prototype. ‘MainComponent.h’ declares the class that will contain all the interactions, the MainComponent class. In ‘MainComponent.h’, all the methods and variables of the MainComponent class are declared. The file MainComponent.cpp uses the MainComponent class from ‘MainComponent.h’ and fills the methods with code and interactions.

The MainComponent class has three main methods: ‘paint’ and ‘resize’, which fill the GUI with information, and ‘getNextAudioBlock’, which does all the processing of the audio and calls the most important functions. There is also ‘run’, which runs the communication thread between the messenger and audio threads. Besides that there are many methods for determining the interaction with the sliders in the GUI and updating the play modes and virtual sound position. Within the MainComponent class, there is also the class ‘ReferenceCountedBuffer’, which is used for storing audio buffers in the communication thread. Below is a schematic of the functionalities of this subsystem and its interactions with the other subsystems:

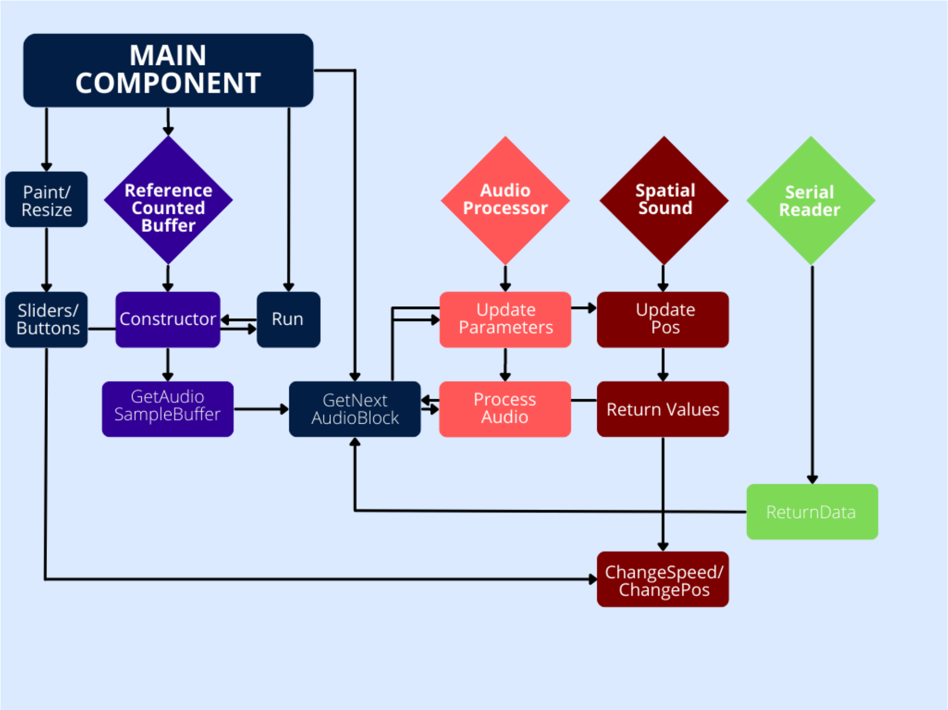


Figure 10: Interaction chart of Main Component subsystem of prototype.

6.2.2 Second Subsystem: Audio Processor

To create the spatial audio effect, the Audio Processor subsystem processes the desired audio files to add a virtual location. It does this through convolution, gain, reverb, and low-pass filtering. The file 'AudioProcessor.h' contains three classes: 'Distancer', Convolver, and 'AudioEngine', which manages the other two classes. To code this class, inspiration was taken from the following tutorial: https://docs.juce.com/master/tutorial_dsp_convolution.html, and the JUCE documentation: https://docs.juce.com/master/index.html.

The Distancer class simulates the distance of the sound, which it does this through a combination of volume gain, reverb and a 12db/octave low-pass filter. It contains the methods 'prepare' and 'changeVariables' to change the parameters of the processor modules. The method 'process' then processes the audio block and simulates distance.

Commented [1]: convert into reference

Commented [2]: convert into reference

The Convolver class handles the processing of the audio to simulate the azimuth and elevation. It contains a convolution module that applies the HRIRs of the HRTF onto the sound. It contains the methods ‘prepare’ and ‘loadIR’ to load a new HRIR into the convolver, which it calls from ‘HRTFReader’. The method ‘process’ then processes the audio block and simulates distance.

The AudioEngine class acts as the mother class of Distancer and Convolver. It calls all their events and controls what data goes where, based on what it gets from MainComponent.cpp. The methods ‘prepare’ and ‘updateParameters’ call the methods ‘prepare’ and ‘changeVariables/loadIR’ respectively from the Distancer and Convolver classes. The method ‘renderNextBlock’ calls the ‘process’ methods from Distancer and Convolver and passes an audio block for them to process. The AudioEngine class is called from MainComponent, and its functions are called from ‘getNextAudioBlock’. Below is a schematic of the functionalities of this subsystem and its interactions with the other subsystems:

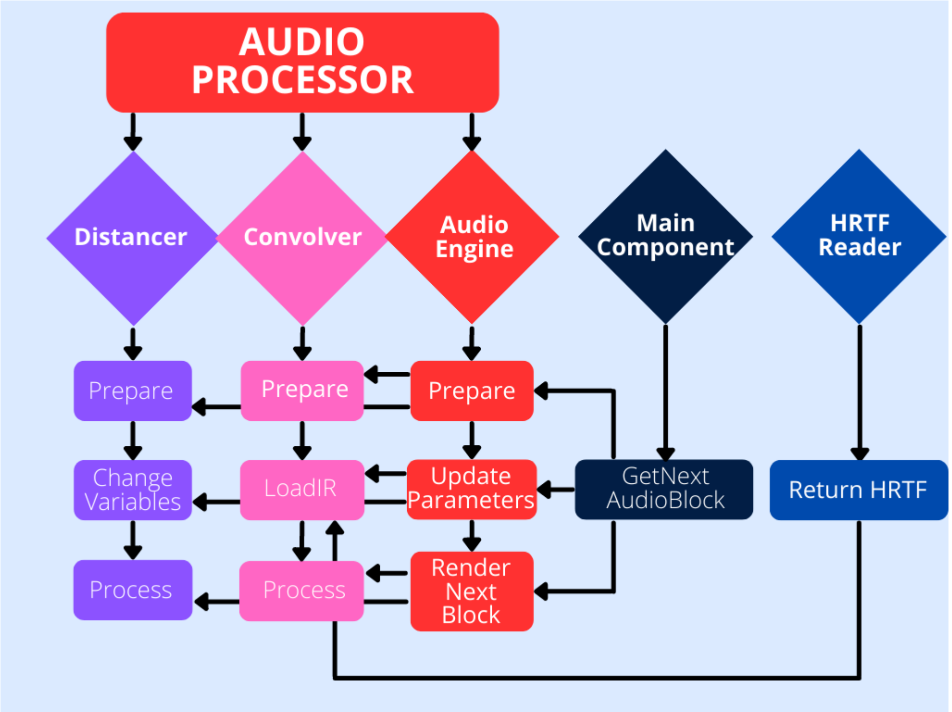


Figure 11: Interaction chart of Audio Processor subsystem of prototype.

6.2.3 Third Subsystem: Spatial Sound

The Spatial Sound subsystem creates a virtual location that the sound can simulate. The subsystem can be controlled via the GUI and heavily interacts with the MainComponent subsystem. It consists of two files: ‘SpatialSound.h’ and ‘PositionMath.h’, each with their respective classes. ‘SpatialSound.h’ contains the class SpatialSound. This class has several variables for the position, the speed, and direction of the virtual location. The class acts as a controller for the positioning and movement of the virtual location. It also calls the class PositionMath in ‘PositionMath.h’ to convert the 3D coordinate into an azimuth, elevation, and distance. Method ‘updatePos’ moves the sound around, based on the parameters.

Methods ‘changePos and ‘changeSpeed give a new speed to the virtual position. And the class returns its contents to MainComponent through methods ‘returnDirectionCoefficients’ and ‘returnPositionCoefficients’. In PositionMath the methods ‘calculateAzimuth’, ‘calculateElevation’ and ‘calculateDistance’ each convert the 3D coordinate into their respective directional coefficient. The last two classes, ‘calculateRotationPos’ and ‘returnDist’, are used for calculating the circular motion that can be simulated. Together, SpatialSound and PositionMath form the mathematical component of the prototype. Below is a schematic of the functionalities of this subsystem and its interactions with the other subsystems:

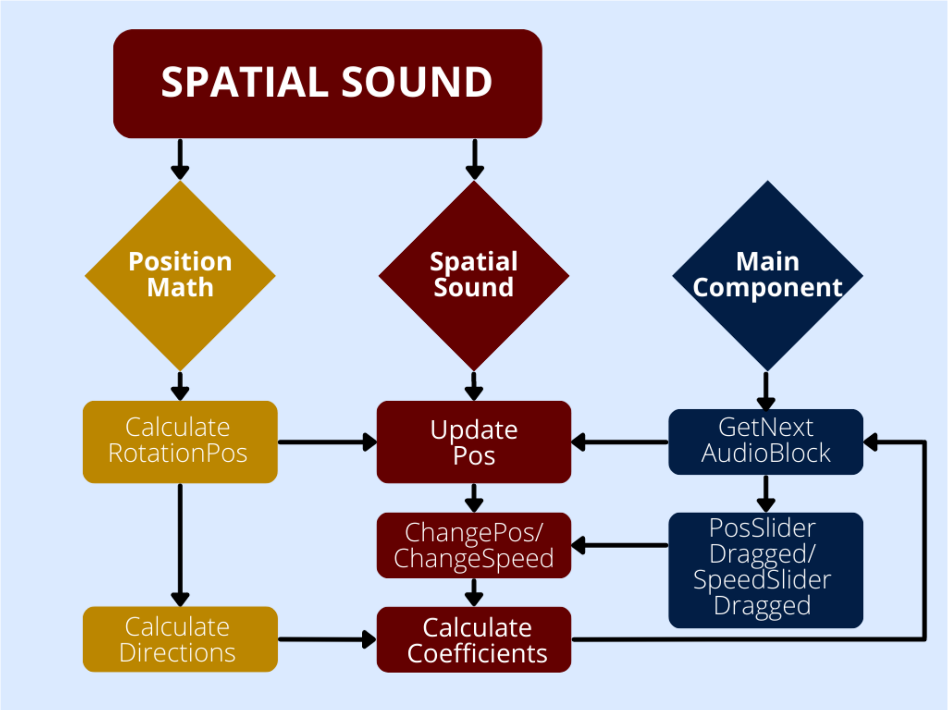


Figure 12: Interaction chart of Math subsystem of prototype.

Commented [3]: Add figure description

6.2.4 Fourth Subsystem: HRTF Reader

As the head-related transfer function is stored in many separate .WAV files instead of one .sofa file, a subsystem will need to be made that can load all these files into one place and pass the right HRIR to the audio processor. This is what the HRTF Reader subsystem does. It is stored in the 'HRTFReader.h' file and consists of two classes: HRTFReader and HRTFDataStruct.

HRTFReader is the class that loads all the .WAV files from the corresponding folder and returns the correct HRIR when asked. The method 'loadHRTF' loops through all file names of the .WAV files, then asks 'returnDirection' for the corresponding azimuth and elevation. It then creates a corresponding object that contains the HRIR, as well as the corresponding azimuth and elevation. The method 'returnDirection' determines the azimuth and elevation from the filename of the HRIR. It splits the filename up into substrings and converts the substrings containing the azimuth and elevation into floats, which it then returns. The method ‘returnImpulseResponse’, called by the Audio Processor, loops through the directions of all of the stored HRIR objects and finds the position of the HRIR closest to the desired direction. It then returns the HRIR at that position to the Audio Processor.

The HRTFDataStruct is the object that is created and read by HRTFReader. It stores the HRIR together with its directions in one place and loads the HRIR from the filename passed through its constructor. With the method ‘returnDirection’ it returns the direction of its HRIR, and with the method ‘returnImpulseResponse’ it returns its HRIR to the HRTFReader, which then sends it to the Audio Processor. Below is a schematic of the functionalities of this subsystem and its interactions with the other subsystems:

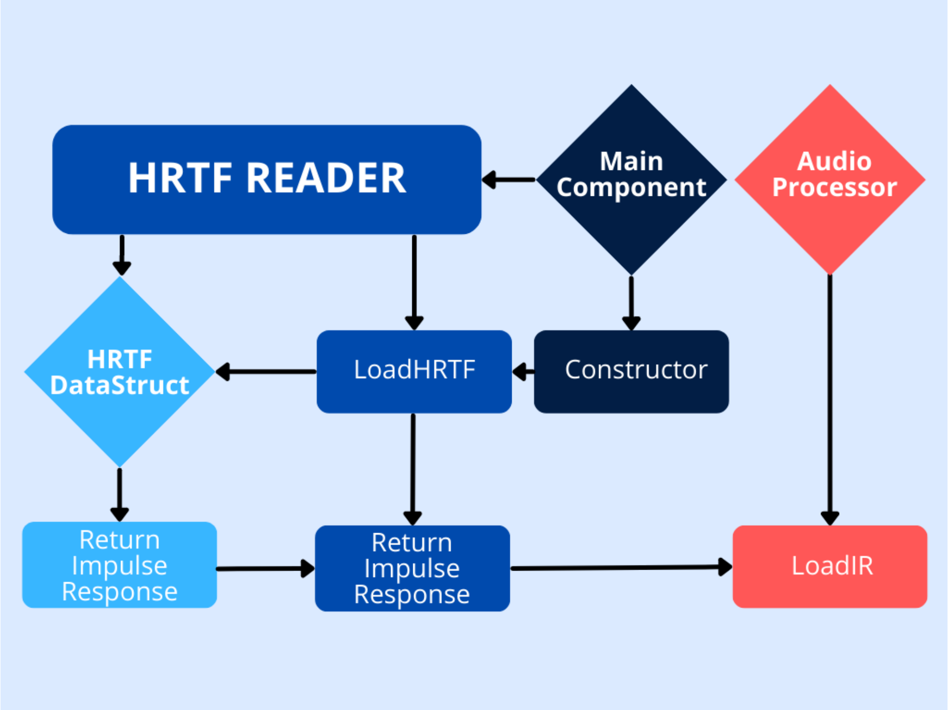


Figure 13: Interaction chart of HRTF Reader subsystem of prototype.

6.2.5 Fifth Subsystem: Serial Reader

This subsystem acts as a gate to receive the data from the Head-Tracker. It opens the Serial port to which the Head-Tracker is connected and passes the data to the MainComponent class. The Serial Reader is spread out over two files: ‘SerialClass.h’ and ‘SerialReader.h’. ‘SerialClass.h’ is a piece of code that establishes a connection with the port to which the Head-Tracker is attached. It allows for both reading and writing to the Head-Tracker, although in this prototype, only the reading functionality is used. The code from ‘SerialClass.h’ is from an Arduino tutorial on Serial communication to window [71].

‘SerialReader.h’ contains the SerialReader class that acts as a wrapper around ‘SerialClass.h’. It forms a bridge between the SerialClass.h file and the MainComponent class. SerialReader limits the functionality of SerialClass.h to establishing and breaking down a Serial connection and only reading the data coming from the Head-Tracker. It also translates the received data from the string filetype into floats. This allows the MainComponent class to only must call two lines of code in order to receive the data from the Head-Tracker. Below is a schematic of the functionalities of this subsystem and its interactions with the other subsystems:

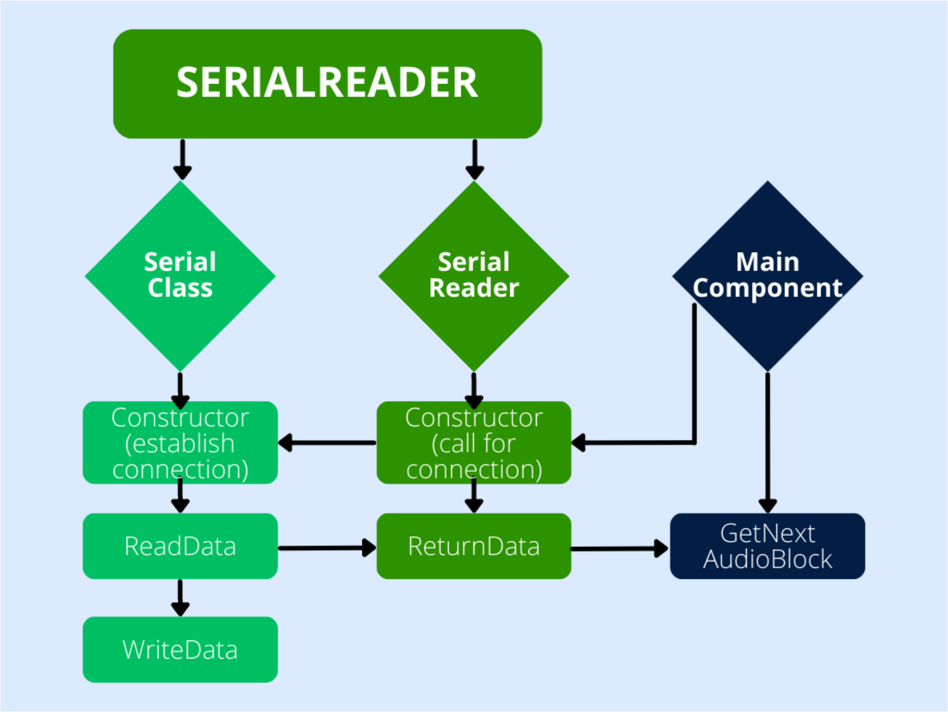


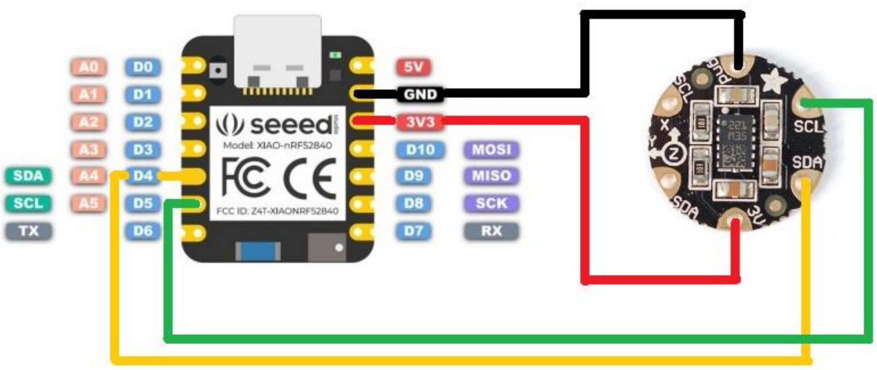
Figure 14: Interaction chart of Serial Reader subsystem of prototype.

6.2.6 Sixth Subsystem: Head-Tracker

The last subsystem is the Head-Tracker. The Head-Tracker can track the direction and position of the user through an accelerometer and magnetometer. It calculates the azimuth, elevation, x-coordinate (left and right) and z-coordinate (forwards and backwards), and then sends these values through a USB-cable to the main application.

Figure 15: Image of Head-Tracker.

The Head-Tracker consists of two modules: the SEEED XIAO BLE SENSE and Adafruit FLORA Accelerometer/Compass Sensor - LSM303. The SEEED XIAO-microcontroller runs the code and contains the accelerometer that does the measurements. The FLORA-sensor senses the direction of earth's magnetic field. They are soldered to each other with four wires, then attached to a glasses frame without glasses, in such a way that the two sensors are parallel to the glasses. This way, the sensors will always determine the original head-orientation, without major directional offsets. The wires connect the 3V, Ground, SDA and SCL pins of the SEEED XIAO and FLORA Compass together, as can be seen in the schematic below:



SEEED XIAO Image Source: https://wiki.seeedstudio.com/XIAO_BLE/

Adafruit FLORA Compass Image Source: https://kamami.pl/en/retired-products/203913-flora-accelerometercompass-sensor-lsm303-v10.html

Figure 16: Wiring schematic of the Head-Tracker.

The Head-Tracker code consists of two files: ‘TrackerCode.ino’ and ‘DataProcessing.h’. ‘TrackerCode.ino’ runs the main code. It consists of a setup and loop, which are used to set up and loop the rest of the program. The loop does two things: it measures and calculates the angles and positioning, and it sends the data through the Serial to the main program. The azimuth and elevation are calculated in different, but very similar functions. First the accelerometer and magnetometer are recorded. These values are then sent to the ‘DataProcessing.h’ code to smooth out the data. The smoothed-out data is then used to calculate the angles. To calculate the position comes more difficult. First the acceleration in both horizontal directions is measured. This is then transformed in the right direction based on the azimuth. The last step is to multiply the value for calibration.

As mentioned earlier, the ‘DataProcessing.h’ file removes the fast changes from the data it receives. It does this by averaging out the new value with the past 100 values by the same sensor. It applies this process to all the values that are recorded. This process does cause a slight delay in the calculation of the angle, although it should not be noticeable. Below is a schematic of the functionalities of this subsystem and its interactions with the other subsystems:

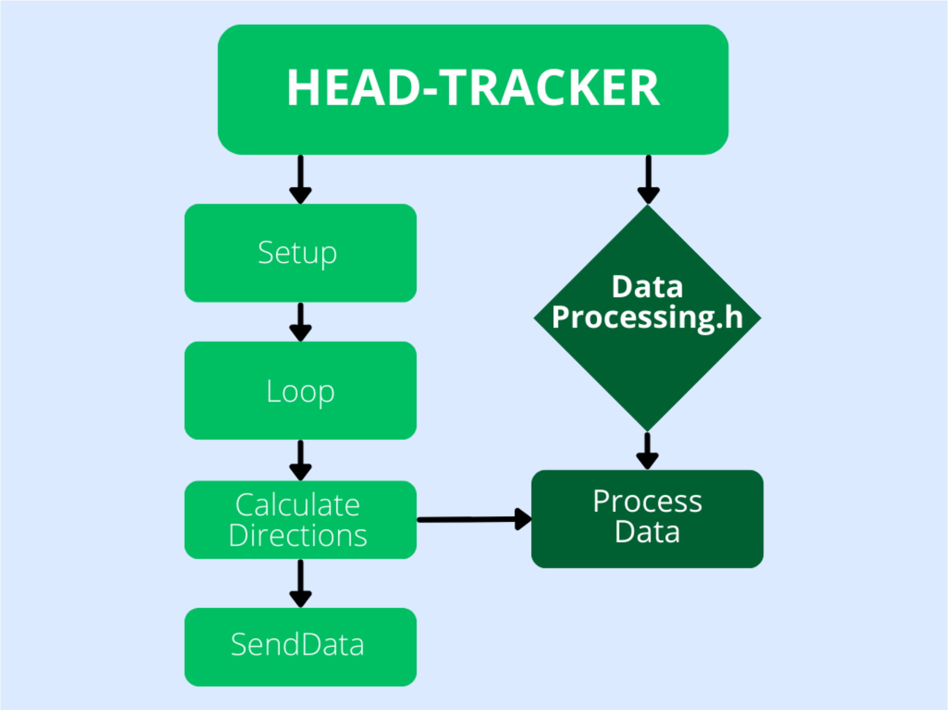


Figure 17: Interaction chart of Head-Tracker subsystem of prototype.

6.3 Integration Process

The integration process was simple. Since most of the subsystems were built on top of each other and by default made to work with each other, the integration process was already completed. However, this was not the case for the Head-Tracker. Since the Head-Tracker exists as its own device, the SerialReader subsystem needed to be developed to integrate it with the rest of the prototype. On top of that, the Spatial Sound subsystem had already been built before the rest of the application was developed. Because of this, the code had to be modified slightly to work well with the rest of the code.

There were difficulties however when attempting to export the application to make it mobile. For these two attempts were made. The first was to export to an APK file using Android Studio to make the program work on the phone. This ultimately failed due to the complexity of setting up the project and removing bugs. The second attempt was to make the code work on Raspberry Pi through a Windows emulator called Wine. This way the prototype would still be considered mobile and would be able to work with headphones or earbuds. However, when executing the .EXE file in Wine, the emulator consistently failed to load the necessary libraries to be able to execute the application. As such, the system was unable to be made mobile.

Finally, it was not possible to get the Serial communication between the Head-Tracker and the main program to work properly. When running the code, it is possible to retrieve the first 2 messages. However, the communication thread comes into conflict with the Serial Reader class. The reason for this is unknown and was not able to be solved in time.

6.4 Functional System Test

Most of the prototype's requirement fulfillments will be analyzed in the user evaluation. However, the processing latency is something that must be evaluated within the prototype itself. The requirement for the latency is that it is lower than 30 milliseconds. To evaluate the latency, code has been added to the prototype that returns the time it took to process the audio. Using the C++ library ‘Chronos’, two-time measurements are taken in each program loop: the current time in milliseconds before and after the audio processing. By calculating the difference between the two-time measurements, the processing time can be found.

The first of these measurements was placed at the beginning of the ‘GetNextAudioBlock’ method. The second measurement was placed at the end of the same method. The reason for this is twofold: Firstly, the audio thread and messenger thread (which also contains the GUI), operate separately from each other. Interactions with the GUI will have no effect on the processing latency. The GUI will therefore not have to be considered when determining the latency. Given this separation of threads, the fact that the audio thread handles everything related to audio, that the audio thread directly passes the processed audio to the audio output and that there is no delay in this last step, it is assumed that the playback latency is equal to the processing latency.

Secondly, ‘GetNextAudioBlock’ contains all processes that happen on the audio thread. Placing the measurements at the start and end of ‘GetNextAudioBlock’ should encapsulate all operations on the audio thread and therefore measure the full playback latency. It should also be noted that because the system could not be ported to a mobile device, the latency evaluation will take place on an Acer Aspire 7, Windows 10 laptop, with an Intel Core i7-9750H at 2.60GHz and 16GB of DDR4 ram. After the test was run, 1026 measurements were taken. The resulting dataset had the following properties:

Number of Measurements	Standard Deviation	Mean	Median	Lowest Value	Highest Value	Frequency Highest Value
1026	2,055901	4,598049	5	1	15	1

Table 1: Dataset description of the processing latency of the prototype.

These results will be further analyzed in the discussion.

Chapter 7: Evaluation

The following chapter describes the evaluation process of the prototype. It starts off with an overview of the evaluation, then covers the results and ends with the analysis of the acquired results.

7.1 Overview of the Evaluation

The goal of the evaluation was to determine the quality of the spatial audio the system can create. What was tested is the extent to which participants can hear the direction in the sound. Both accuracy and resolution of the spatialization are measured. This was done for the Azimuth, Elevation and Distance of the sounds. There are five key areas that will be measured: The precision (accuracy and resolution combined) of simulating the azimuth, elevation, and distance. The performance of the prototype compared to the already existing systems. The potential existence of a learning curve when using spatial audio systems. The effect of the type of sound on the perception of spatial audio. And lastly, the personal experience of the participants with using the system.

To be able to reach answers in these areas of interest, the questions will be randomized as much as possible. The sounds that were evaluated for one directional coefficient were still evaluated in combination with the other coefficients. For example, when testing the azimuth, the evaluation and distance were randomized. This was to determine any effects the directional coefficients might have on the perceptibility of each other.

Four different sounds were evaluated: a truck, an electric car, a biker bell, and pedestrians holding a conversation. The sounds were chosen to represent real sounds in traffic. The reason for using different sounds is to determine whether a certain sound is easier to be heard than other sounds. Which sound was played at which question was randomized as well.

Who made the sound.

The evaluation consisted of two games. The first game is called ‘Who made the sound’, the second game ‘The odd one out’. Each game was evaluated with questions starting at low locational preciseness and ending in high preciseness. The goal of ‘Who made the sound’ was to determine the accuracy of the direction of the spatialized audio. In this context, accuracy means how accurate the sound was simulated within a certain range of possible directions. The higher the accuracy, the less doubt there is with regards to the position of the sounds.

The participants were presented with a sound and three possible sources of where the sound might have come from. It was up to the participants to determine which of the three sources produced the sound. To give a location to these sources, a ring has been created on which the sound positions can be placed. Below are pictures of the two rings that were used in the evaluation. The first ring was made of wood and is used for physical evaluations. The second ring was digital and is mainly used for online evaluations. It was programmed in openprocesing.org and the code is in appendix D.

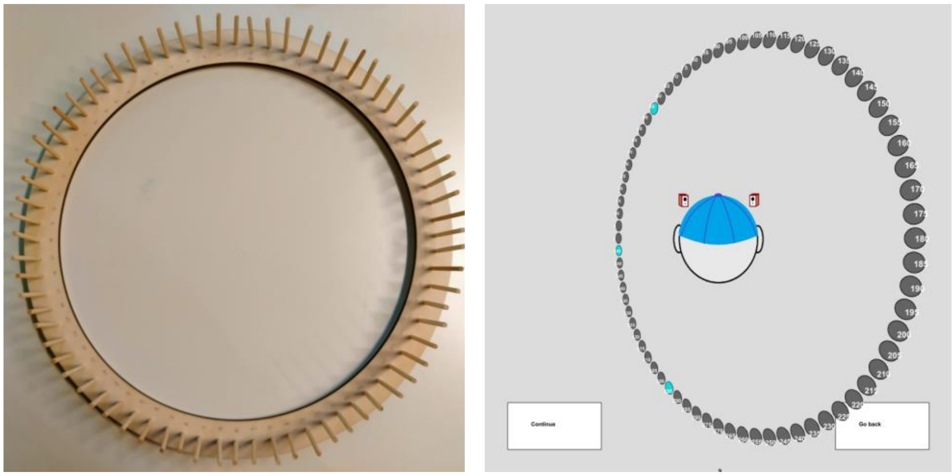


Figure 18: Pictures of both rings used in the evaluation.

Each of the rings had 72 possible locations the sound might have come from, each with 5 degrees from the neighboring locations. For evaluating the azimuth, the participant was asked to imagine themselves in the center of the ring. In the case of the physical ring, three white cylinders were placed on the possible locations. In the case of the digital ring, the possible locations were lit up in cyan. The distance between the possible locations determines the accuracy that is being tested. For example, when testing the azimuth with 60 degrees accuracy, the possible locations were placed apart 60 degrees from each other. When testing for 5 degrees accuracy, the locations were placed right next to each other on the ring. Each azimuth test also had a random elevation and distance, meaning the effect of one directional coefficient on the azimuth could be tested. Eight sounds were presented, testing accuracies within possible location differences of 120, 60, 20, 20, 10, 10, 5 and 5 degrees in azimuth.

The same procedure was used for evaluating the accuracy of the elevation, with one difference: the ring was flipped on its side and rotated in the direction of the azimuth of the sound, with 0 degrees being right in front of the participant. This was done to consider the different directions of the sounds, as each test in the elevation evaluation has different azimuths and distances. If the ring was not rotated, the listener would get incorrect positions for the elevations, since the elevation is dependent on the azimuth. This does not apply to the azimuth, as this value is independent of the elevation. It works the same as earth's coordinate system, where horizontal lines are all in the same direction, but vertical lines rotate compared to each other. Again, eight sounds were presented, testing accuracies within possible location differences of 120, 60, 20, 20, 10, 10, 5 and 5 degrees in elevation.

To evaluate the distance, no ring was used. Instead, the participant was asked to imagine possible different distances the sound could have come from. For example, the participants were told that a sound could have come from either 1, 15 or 30 meters away. The participant then had to determine which of these three distances the sound came from. Small markers were placed to provide the participant with some extra perspective. Besides a given distance, each sound was simulated with a different azimuth and elevation. Here, six sounds were presented, testing accuracies within possible location differences of 15, 15, 10, 10, 5, 5 meters.

The odd one out

The second test, 'The odd one out', simply presented the user with three sounds. Two of them are the same, one has a slightly different location. It was up to the user to figure out which one is different. The purpose of this test was to determine the resolution with which positions can be simulated. The resolution in this case means the smallest audible difference in sound location the prototype can simulate. Since the participants do not have to focus on finding the locations of the sounds, determining small directional deviations becomes easier. For both the azimuth and elevation, four questions were presented, each testing a resolution of 4, 3, 2 and 1 degrees respectively.

For the distance, eight questions were presented, testing a distance resolution of 4, 4, 3, 3, 2, 2, 1 and 1 degrees. This doubling of questions was since the distance is suspected to be easier to differentiate close to the participant than further away. So, each question tested the distance nearby and far away. The distances tested were between one and thirty meters away.

Different systems

Not only the prototype was evaluated; three spatial audio systems were put to the test as well. This was done to evaluate how the prototype performs compared to the standard. The existing systems were evaluated before the prototype. Whether the participant will evaluate an existing system or not is different for each participant. Through this, any learning effects can be determined, meaning the participant got higher scores after already having evaluated a system by getting used to spatial audio. If such an effect were to be observed, it would be a favorable outcome for the prototype since it means the users would get better at using it over time.

The existing systems that were evaluated were 'Anaglyph' [72], 'BL Spatializer' [73] and 'dearVR MICRO' (made by Sennheiser) [74], in combination with 'AirMusic' [75] to simulate the distance. The systems exist as VST plugins, meaning they work as audio effects in so-called Digital Audio Workstations, or DAWs for short. DAWs are pieces of software that allow for creating, arranging, and processing of music and audio samples [76]. To prepare the sound files of the above-mentioned systems, the sounds were placed in the DAW, in this case FL Studio 21 and the spatial audio plugins were loaded with the right settings. The sounds were then exported and now contained the spatial audio effect created by each system. In total, 114 sounds were processed and exported. Because of the amount of work that goes into preparing that many sounds, only one set of sounds was created for each system. This set was used by all participants that evaluated that system.

Questionnaire

At the end of each evaluation, the participants were asked a series of questions. This was done to determine their experiences with using the prototype, as well as to find whether they had any experience with listening to spatial audio. The questions that were asked were the following:

1. Have you listened to spatial audio before this experiment?
2. What did you expect to hear before participating?
3. How did the evaluation live up to your expectations?
4. What was positive / what positively surprised you about using the system?
5. What was negative / what negatively surprised you about using the system?
6. Did you experience any discomfort when using the system?
7. Do you think the system can be useful to aiding bikers in traffic?
    a. Why do you think so?
8. Did you enjoy the evaluation?
9. Do you have any final remarks you want to add?
10. Would you like to take part in the contest?

Contest

Originally, each participant was invited to participate in a contest where they would use the navigation functionality of the prototype. The participants who would have performed the best would be given a gift card for online shopping worth 10 euros. The reason for organizing such a contest was to incite enthusiasm in the participants to participate in the evaluation. However, the contest was unable to proceed since the navigation feature was not completed in time. On top of that, the amount of work writing the report was too large.

Participants and tools

In total seventeen participants were included in the research, with the only exclusion criterion being that the participant must be older than 18 years old. All participants are acquainted with the researcher and were recruited through WhatsApp and personal contact. No demographics or personal information have been recorded since it was found unlikely that age or gender would have any major impact on the results. The evaluation took place both in-person at the green benches in front of the UT-Flexoffice in the Zilverling building on the campus of UTwente, and online through Discord. The tools used for the evaluation procedure were:

- An Acer Aspire 7 laptop with an Intel Core i7 processor, 16 GB ram and Windows 10 as operating system.
- A pair of HIFIMAN Sundara headphones.
- One of the rings (either physical or digital).
- Google Sheets (for recording data).
- FL Studio 21 (for preparing the already existing systems).
- Oppo A5 smartphone for recording the data.

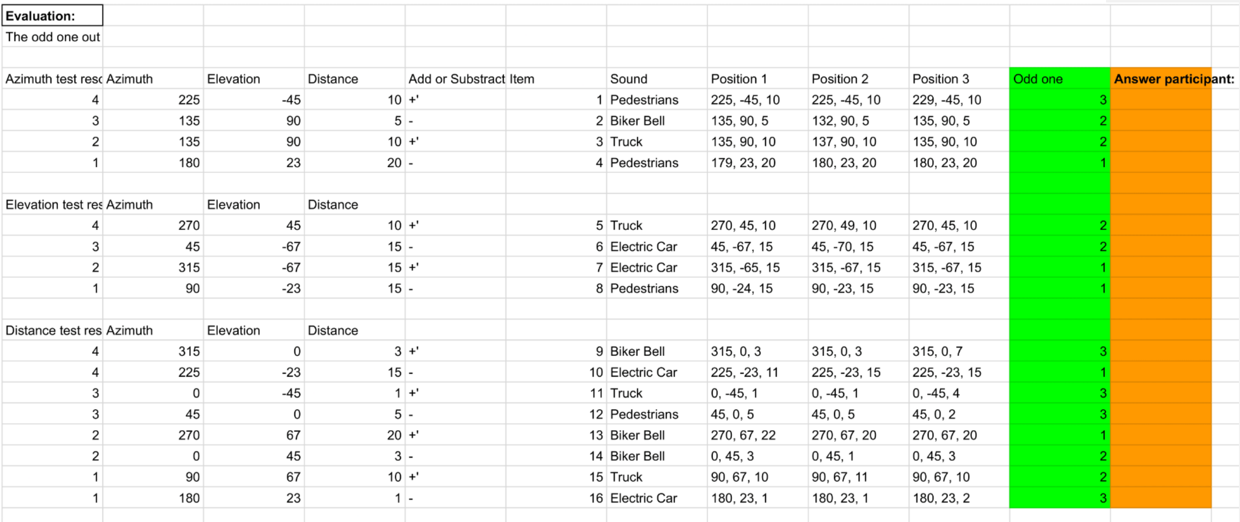
Evaluation Reflection

Although the evaluation process took a while to get fully started, after the third participant, the evaluations went smoothly. In the beginning, the researcher had little experience with the evaluation process, and for that reason replacing the sound positions in ‘Who made the sound’ went slightly slow and chaotic. During the first evaluation it was also unclear whether the evaluation should be continued. On top of that, the functionality of the prototype to play the ‘The odd one out’ game, made it so that the first three participants were unable to be evaluated on the spatial audio resolution. After the first day, this functionality was added, and the resolution could be evaluated. ‘The odd one out’ did already work for the already existing systems, however, which was evaluated with the second participant. The second half of the ‘Who made the sound' evaluations were done with the digital ring, instead of the physical one, because of the higher speed of switching the possible locations. After the first three evaluations were completed, the questions for testing the accuracy of the distance were changed. For this reason, the data on the accuracy of the distance of the first three evaluations is not included in the results.

All participants reported no discomfort when using the systems and found participating in the evaluation enjoyable. They unanimously agreed that the spatial audio prototype was able to simulate the general direction of a sound. However, the more precise sounds were reportedly difficult to hear. Nearly all participants agreed that the prototype in its current state does not perform well enough to be used to aid bikers in traffic, although it could if the prototype was developed further. This indicates that the prototype works in the basics but needs more refinement to be able to be used properly.

7.3 Data Preparation

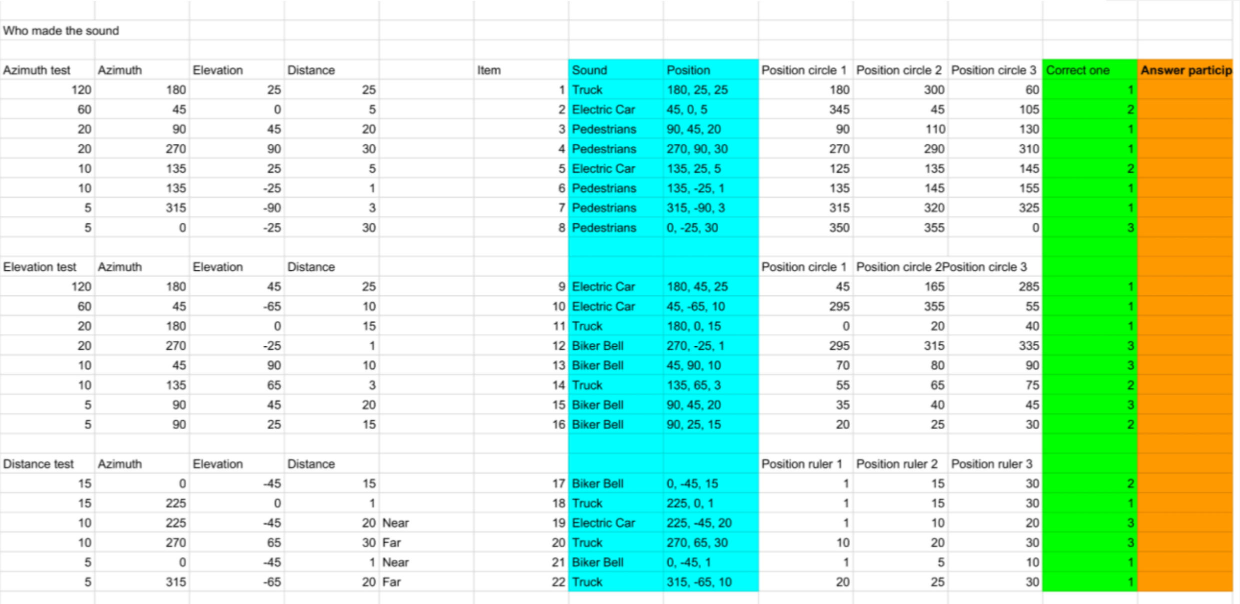
To prepare the data, different sheets were created for each participant. All individual sheets can be viewed in appendix C. To create the questions, a list of possible numbers for each variable was created. Variables include: the azimuth, elevation, and distance of the sound, but also which of the answers is correct and which sound is played. To create a new data file, the previous one was copied, and all the variables were randomized. Then the positions of the sounds were filled out according to the preciseness of the question and the newly randomly obtained position. For evaluating an already existing system, the same data sheet was used as the previous ones. Below is an example of what a sheet looks like for a participant after it has newly been created:



Evaluation:

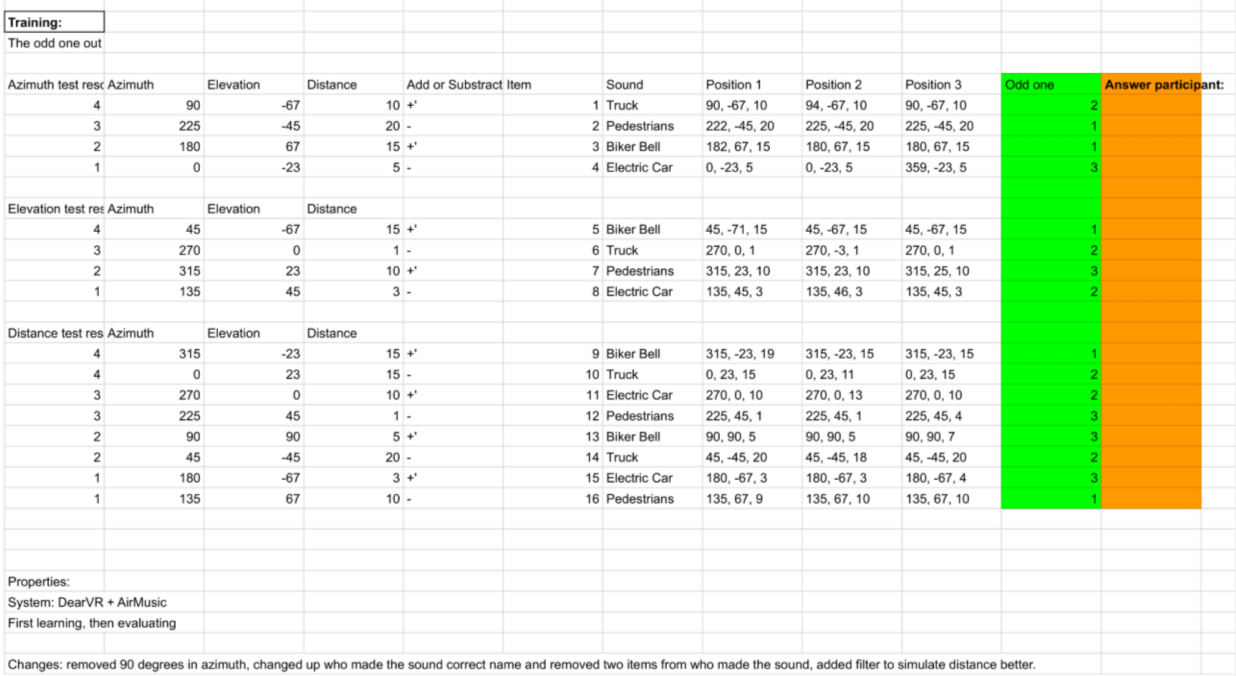
The odd one out

Azimuth test resc	Azimuth	Elevation	Distance	Add or Substract	Item	Sound	Position 1	Position 2	Position 3	Odd one	**Answer participant:**
4	225	-45	10	+'	1	Pedestrians	225, -45, 10	225, -45, 10	229, -45, 10	3	
3	135	90	5	-	2	Biker Bell	135, 90, 5	132, 90, 5	135, 90, 5	2	
2	135	90	10	+'	3	Truck	135, 90, 10	137, 90, 10	135, 90, 10	2	
1	180	23	20	-	4	Pedestrians	179, 23, 20	180, 23, 20	180, 23, 20	1	
Elevation test res	Azimuth	Elevation	Distance								
4	270	45	10	+'	5	Truck	270, 45, 10	270, 49, 10	270, 45, 10	2	
3	45	-67	15	-	6	Electric Car	45, -67, 15	45, -70, 15	45, -67, 15	2	
2	315	-67	15	+'	7	Electric Car	315, -65, 15	315, -67, 15	315, -67, 15	1	
1	90	-23	15	-	8	Pedestrians	90, -24, 15	90, -23, 15	90, -23, 15	1	
Distance test res	Azimuth	Elevation	Distance								
4	315	0	3	+'	9	Biker Bell	315, 0, 3	315, 0, 3	315, 0, 7	3	
4	225	-23	15	-	10	Electric Car	225, -23, 11	225, -23, 15	225, -23, 15	1	
3	0	-45	1	+'	11	Truck	0, -45, 1	0, -45, 1	0, -45, 4	3	
3	45	0	5	-	12	Pedestrians	45, 0, 5	45, 0, 5	45, 0, 2	3	
2	270	67	20	+'	13	Biker Bell	270, 67, 22	270, 67, 20	270, 67, 20	1	
2	0	45	3	-	14	Biker Bell	0, 45, 3	0, 45, 1	0, 45, 3	2	
1	90	67	10	+'	15	Truck	90, 67, 10	90, 67, 11	90, 67, 10	2	
1	180	23	1	-	16	Electric Car	180, 23, 1	180, 23, 1	180, 23, 2	3	



Who made the sound

Azimuth test	Azimuth	Elevation	Distance		Item	Sound	Position	Position circle 1	Position circle 2	Position circle 3	Correct one	**Answer particip**
120	180	25	25		1	Truck	180, 25, 25	180	300	60	1	
60	45	0	5		2	Electric Car	45, 0, 5	345	45	105	2	
20	90	45	20		3	Pedestrians	90, 45, 20	90	110	130	1	
20	270	90	30		4	Pedestrians	270, 90, 30	270	290	310	1	
10	135	25	5		5	Electric Car	135, 25, 5	125	135	145	2	
10	135	-25	1		6	Pedestrians	135, -25, 1	135	145	155	1	
5	315	-90	3		7	Pedestrians	315, -90, 3	315	320	325	1	
5	0	-25	30		8	Pedestrians	0, -25, 30	350	355	0	3	
Elevation test	Azimuth	Elevation	Distance					Position circle 1	Position circle 2	Position circle 3		
120	180	45	25		9	Electric Car	180, 45, 25	45	165	285	1	
60	45	-65	10		10	Electric Car	45, -65, 10	295	355	55	1	
20	180	0	15		11	Truck	180, 0, 15	0	20	40	1	
20	270	-25	1		12	Biker Bell	270, -25, 1	295	315	335	3	
10	45	90	10		13	Biker Bell	45, 90, 10	70	80	90	3	
10	135	65	3		14	Truck	135, 65, 3	55	65	75	2	
5	90	45	20		15	Biker Bell	90, 45, 20	35	40	45	3	
5	90	25	15		16	Biker Bell	90, 25, 15	20	25	30	2	
Distance test	Azimuth	Elevation	Distance					Position ruler 1	Position ruler 2	Position ruler 3		
15	0	-45	15		17	Biker Bell	0, -45, 15	1	15	30	2	
15	225	0	1		18	Truck	225, 0, 1	1	15	30	1	
10	225	-45	20	Near	19	Electric Car	225, -45, 20	1	10	20	3	
10	270	65	30	Far	20	Truck	270, 65, 30	10	20	30	3	
5	0	-45	1	Near	21	Biker Bell	0, -45, 1	1	5	10	1	
5	315	-65	20	Far	22	Truck	315, -65, 10	20	25	30	1	



Training:

The odd one out

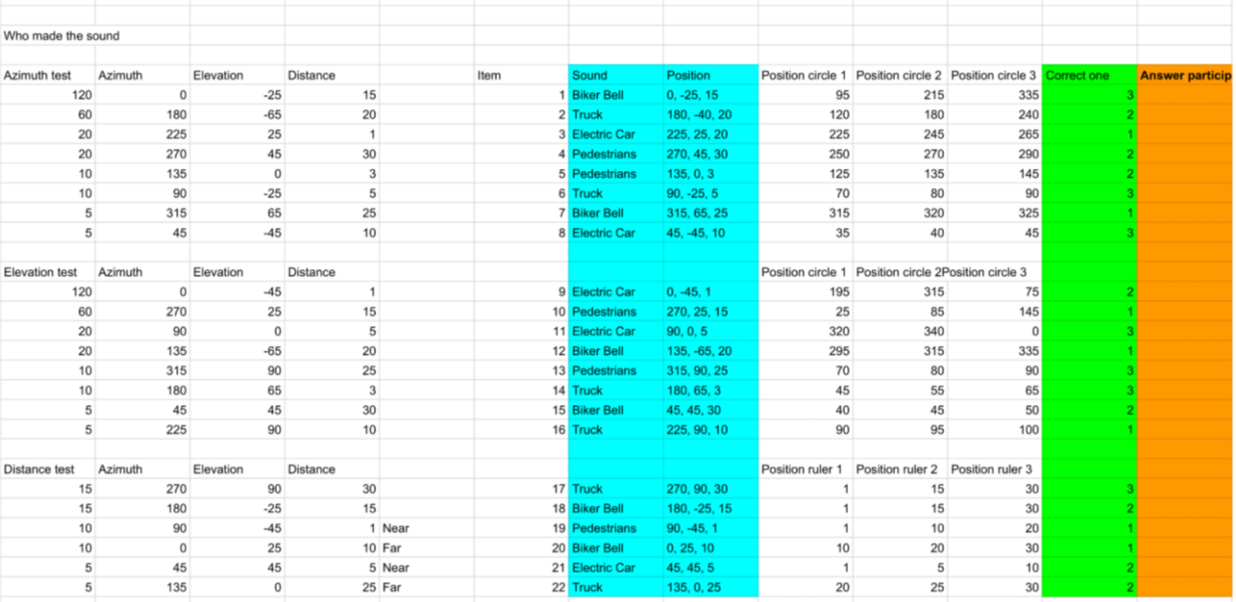
	Azimuth	Elevation	Distance	Add or Substract	Item	Sound	Position 1	Position 2	Position 3	Odd one	**Answer participant:**
Azimuth test res	Azimuth	Elevation	Distance	Add or Substract	Item	Sound	Position 1	Position 2	Position 3	Odd one	**Answer participant:**
4	90	-67	10	+'	1	Truck	90, -67, 10	94, -67, 10	90, -67, 10	2	
3	225	-45	20	-	2	Pedestrians	222, -45, 20	225, -45, 20	225, -45, 20	1	
2	180	67	15	+'	3	Biker Bell	182, 67, 15	180, 67, 15	180, 67, 15	1	
1	0	-23	5	-	4	Electric Car	0, -23, 5	0, -23, 5	359, -23, 5	3	
Elevation test res	Azimuth	Elevation	Distance								
4	45	-67	15	+'	5	Biker Bell	45, -71, 15	45, -67, 15	45, -67, 15	1	
3	270	0	1	-	6	Truck	270, 0, 1	270, -3, 1	270, 0, 1	2	
2	315	23	10	+'	7	Pedestrians	315, 23, 10	315, 23, 10	315, 25, 10	3	
1	135	45	3	-	8	Electric Car	135, 45, 3	135, 46, 3	135, 45, 3	2	
Distance test res	Azimuth	Elevation	Distance								
4	315	-23	15	+'	9	Biker Bell	315, -23, 19	315, -23, 15	315, -23, 15	1	
4	0	23	15	-	10	Truck	0, 23, 15	0, 23, 11	0, 23, 15	2	
3	270	0	10	+'	11	Electric Car	270, 0, 10	270, 0, 13	270, 0, 10	2	
3	225	45	1	-	12	Pedestrians	225, 45, 1	225, 45, 1	225, 45, 4	3	
2	90	90	5	+'	13	Biker Bell	90, 90, 5	90, 90, 5	90, 90, 7	3	
2	45	-45	20	-	14	Truck	45, -45, 20	45, -45, 18	45, -45, 20	2	
1	180	-67	3	+'	15	Electric Car	180, -67, 3	180, -67, 3	180, -67, 4	3	
1	135	67	10	-	16	Pedestrians	135, 67, 9	135, 67, 10	135, 67, 10	1	

Properties:

System: DearVR + AirMusic

First learning, then evaluating

Changes: removed 90 degrees in azimuth, changed up who made the sound correct name and removed two items from who made the sound, added filter to simulate distance better.



Who made the sound

Azimuth test	Azimuth	Elevation	Distance		Item	Sound	Position	Position circle 1	Position circle 2	Position circle 3	Correct one	**Answer particip**
120	0	-25	15		1	Biker Bell	0, -25, 15	95	215	335	3	
60	180	-65	20		2	Truck	180, -40, 20	120	180	240	2	
20	225	25	1		3	Electric Car	225, 25, 20	225	245	265	1	
20	270	45	30		4	Pedestrians	270, 45, 30	250	270	290	2	
10	135	0	3		5	Pedestrians	135, 0, 3	125	135	145	2	
10	90	-25	5		6	Truck	90, -25, 5	70	80	90	3	
5	315	65	25		7	Biker Bell	315, 65, 25	315	320	325	1	
5	45	-45	10		8	Electric Car	45, -45, 10	35	40	45	3	
Elevation test	Azimuth	Elevation	Distance					Position circle 1	Position circle 2	Position circle 3		
120	0	-45	1		9	Electric Car	0, -45, 1	195	315	75	2	
60	270	25	15		10	Pedestrians	270, 25, 15	25	85	145	1	
20	90	0	5		11	Electric Car	90, 0, 5	320	340	0	3	
20	135	-65	20		12	Biker Bell	135, -65, 20	295	315	335	1	
10	315	90	25		13	Pedestrians	315, 90, 25	70	80	90	3	
10	180	65	3		14	Truck	180, 65, 3	45	55	65	3	
5	45	45	30		15	Biker Bell	45, 45, 30	40	45	50	2	
5	225	90	10		16	Truck	225, 90, 10	90	95	100	1	
Distance test	Azimuth	Elevation	Distance					Position ruler 1	Position ruler 2	Position ruler 3		
15	270	90	30		17	Truck	270, 90, 30	1	15	30	3	
15	180	-25	15		18	Biker Bell	180, -25, 15	1	15	30	2	
10	90	-45	1	Near	19	Pedestrians	90, -45, 1	1	10	20	1	
10	0	25	10	Far	20	Biker Bell	0, 25, 10	10	20	30	1	
5	45	45	5	Near	21	Electric Car	45, 45, 5	1	5	10	2	
5	135	0	25	Far	22	Truck	135, 0, 25	20	25	30	2	

Interview	
Question:	**Answer:**
Have you listened to spatial audio before this experir	
What did you expect to hear before participating?	
How did the evaluation live up to your expectations?	
What was positive / what positively surprised you abo	
What was negative / what negatively surprised you a	
Did you experience any discomfort when using the s	
Do you think the system can be useful to aiding biker	
-Why do you think so?	
Did you enjoy it?	
Do you have any final remarks you want to add?	
Would you like to take part in the contest?	

Figure 18: Example of resulting dataset after preparation.

7.4 Results and Statistical Analysis

This section will cover the results and statistical analysis of the five areas of interest. First, the accuracy and resolution of the audio spatialization that was achieved with the prototype will be determined. This will be done by counting the number of correct answers for each question testing the same preciseness. For example, all the correct answers from the azimuth that tested 20 degrees accuracy will be counted together. A binomial test will be applied to the number of correct answers to the question. The probability of getting at least the number of correct answers on that question will be determined with the following formula:

$$\sum_{n=C}^{n=N} \left(\frac{1}{x}\right)^{n} \cdot \left(\frac{x-1}{x}\right)^{N-n} \cdot \left(\frac{N!}{n!\,(N-n)!}\right)$$

Formula 2: formula for calculating the binomial probability [77].

In this formula, N is the total number of items in the test, X is the number of choices for each item (in this case 3 for each question) and C is the number of correct responses [77]. The null hypothesis (H_0) for each question is that there is no significant difference between the measured amount of correct answers chance. The alternative hypothesis (H_1) is that there is a significant difference between the measured number of correct answers and chance.

For this test, an alpha of 0.05 is chosen. If the probability of acquiring the result is below 0.05, meaning below 5%, it can be assumed that the achieved number of correct answers was not due to chance, but through the performance of the prototype. If the probability is higher than 0.05, it is unknown whether the results are due to chance. The threshold 0.05 was chosen as it is the industry standard threshold in statistics for accepting or rejecting an observed effect [78].

This resulted in the following data:

Test: Who made the sound, **Accuracy**	Number of questions	Number of correct answers	Probability of at least the number of correct answers	Statistically significant (< 0.05)
Azimuth, 120 degrees	17	12	0.0019	Yes
Azimuth, 60 degrees	17	6	0.52	No
Azimuth, 20 degrees	34	14	0.21	No
Azimuth, 10 degrees	34	13	0.32	No
Azimuth, 5 degrees	34	13	0.19	No
Elevation, 120 degrees	17	7	0.32	No
Elevation, 60 degrees	17	13	0.00034	Yes
Elevation, 20 degrees	34	16	0.067	No
Elevation, 10 degrees	34	13	0.33	No
Elevation, 5 degrees	34	9	0.85	No

Distance, 15 meters	28	19	0.00020	Yes
Distance, 10 meters	28	14	0.0503	No
Distance, 5 meters	28	11	0.34	No

Table 2: Table containing all accuracy measurements.

Test: The odd one out, **Resolution**	Number of questions	Number of correct answers	Probability of at least the number of correct answers	Statistically significant (< 0.05)
Azimuth, 4 degrees	14	5	0.52	No
Azimuth, 3 degrees	14	6	0.31	No
Azimuth, 2 degrees	14	3	0.89	No
Azimuth, 1 degree	14	3	0.89	No
Elevation, 4 degrees	14	3	0.89	No
Elevation, 3 degrees	14	6	0.31	No
Elevation, 2 degrees	14	4	0.74	No
Elevation, 1 degree	14	6	0.31	No
Distance, 4 meters	28	20	0.000043	Yes
Distance, 3 meters	28	20	0.000043	Yes
Distance, 2 meters	28	14	0.0503	No
Distance, 1 meter	28	10	0.46	No

Table 3: Table containing all resolution measurements.

The second area of interest is the performance of the prototype compared to already existing systems. The same method of the prototype analysis will be used to evaluate the existing systems. The systems will not be evaluated individually, but. The performance of the individual systems against each other is not of interest to this research; it is expected that these systems have already been evaluated and performed to minimum industry standard levels. Having evaluated three different systems is still of use, since one system might still outperform the other. Evaluating different systems gives a more unbiased result.

This resulted in the following data:

Test: Who made the sound, **Accuracy**	Number of questions	Number of correct answers	Probability of at least the number of correct answers	Statistically significant (< 0.05)
Azimuth, 120 degrees	8	2	0.80	No
Azimuth, 60 degrees	8	5	0.088	No
Azimuth, 20 degrees	16	5	0.66	No
Azimuth, 10 degrees	16	3	0.94	No
Azimuth, 5 degrees	16	9	0.049	Yes
Elevation, 120 degrees	8	1	0.96	No
Elevation, 60 degrees	8	8	0.015	Yes
Elevation, 20 degrees	16	8	0.13	No
Elevation, 10 degrees	16	6	0.45	No
Elevation, 5 degrees	16	4	0.83	No
Distance, 15 meters	12	4	0.60	No
Distance, 10 meters	12	9	0.0039	Yes

Distance, 5 meters	12	5	0.37	No

Table 4: Table containing all accuracy measurements of existing systems.

Test: The odd one out, **Resolution**	Number of questions	Number of correct answers	Probability of at least the number of correct answers	Statistically significant (< 0.05)
Azimuth, 4 degrees	7	3	0.42	No
Azimuth, 3 degrees	7	5	0.045	No
Azimuth, 2 degrees	7	5	0.045	No
Azimuth, 1 degree	7	3	0.42	No
Elevation, 4 degrees	7	4	0.17	No
Elevation, 3 degrees	7	2	0.73	No
Elevation, 2 degrees	7	3	0.42	No
Elevation, 1 degree	7	2	0.73	No
Distance, 4 meters	14	5	0.52	No
Distance, 3 meters	14	9	0.017	Yes
Distance, 2 meters	14	6	0.31	No
Distance, 1 meter	14	8	0.058	No

Table 5: Table containing all resolution measurements of existing systems.

The third area of interest is whether there is an improvement in the performance of participants when evaluating a second system, compared to the first system. This way, any potential learning curves can be established. To evaluate this, the answers where only the prototype has been evaluated will be compared with the answers of evaluating an already existing system beforehand.

This results in the following data:

Test: Who made the sound, **Accuracy**	Number of questions	Number of correct answers	Probability of at least the number of correct answers	Statistically significant (< 0.05)
Azimuth, 120 degrees	9	8	0.00096	Yes
Azimuth, 60 degrees	9	3	0.62	No
Azimuth, 20 degrees	18	9	0.11	No
Azimuth, 10 degrees	18	7	0.39	No
Azimuth, 5 degrees	18	3	0.97	No
Elevation, 120 degrees	9	5	0.14	No
Elevation, 60 degrees	9	7	0.0083	Yes
Elevation, 20 degrees	18	9	0.11	No
Elevation, 10 degrees	18	7	0.39	No
Elevation, 5 degrees	18	6	0.59	No
Distance, 15 meters	14	10	0.004	Yes
Distance, 10 meters	14	10	0.004	Yes
Distance, 5 meters	14	7	0.15	No

Table 6: Table containing all accuracy measurements of the prototype after evaluating an existing system.

Test: The odd one out, **Resolution**	Number of questions	Number of correct answers	Probability of at least the number of correct answers	Statistically significant (< 0.05)
Azimuth, 4 degrees	7	1	0.94	No
Azimuth, 3 degrees	7	2	0.74	No
Azimuth, 2 degrees	7	1	0.94	No
Azimuth, 1 degree	7	2	0.74	No
Elevation, 4 degrees	7	1	0.94	No
Elevation, 3 degrees	7	3	0.43	No
Elevation, 2 degrees	7	4	0.17	No
Elevation, 1 degree	7	2	0.74	No
Distance, 4 meters	14	9	0.017	Yes
Distance, 3 meters	14	9	0.017	Yes
Distance, 2 meters	14	6	0.31	No
Distance, 1 meter	14	4	0.74	No

Table 7: Table containing all resolution measurements of the prototype after evaluating an existing system.

Test: Who made the sound, **Accuracy**	Number of questions	Number of correct answers	Probability of at least the number of correct answers	Statistically significant (< 0.05)
Azimuth, 120 degrees	8	5	0.088	No

Azimuth, 60 degrees	8	4	0.26	No
Azimuth, 20 degrees	16	6	0.45	No
Azimuth, 10 degrees	16	4	0.83	No
Azimuth, 5 degrees	16	8	0.12	No
Elevation, 120 degrees	8	2	0.80	No
Elevation, 60 degrees	8	6	0.020	Yes
Elevation, 20 degrees	16	8	0.13	No
Elevation, 10 degrees	16	7	0.26	No
Elevation, 5 degrees	16	2	0.99	No
Distance, 15 meters	16	12	0.00079	Yes
Distance, 10 meters	16	8	0.13	No
Distance, 5 meters	16	9	0.049	Yes

Table 8: Table containing all accuracy measurements of the prototype without evaluating an existing system.

Test: The odd one out, **Resolution**	Number of questions	Number of correct answers	Probability of at least the number of correct answers	Statistically significant (< 0.05)
Azimuth, 4 degrees	7	4	0.17	No
Azimuth, 3 degrees	7	3	0.43	No
Azimuth, 2 degrees	7	2	0.17	No
Azimuth, 1 degree	7	2	0.17	No

Elevation, 4 degrees	7	1	0.94	No
Elevation, 3 degrees	7	4	0.17	No
Elevation, 2 degrees	7	1	0.94	No
Elevation, 1 degree	7	3	0.43	No
Distance, 4 meters	14	11	0.00069	Yes
Distance, 3 meters	14	11	0.00069	Yes
Distance, 2 meters	14	7	0.15	No
Distance, 1 meter	14	7	0.15	No

Table 9: Table containing all resolution measurements of the prototype without evaluating an existing system.

The fourth area of interest is whether there was an effect on the type of sound and the ability to recognize its position. To evaluate this, the number of correct answers per sound is established and a two-tailed binomial test is used to determine whether there is a significant difference from the other sounds. The formula used for this evaluation is the following:

$$z = \frac{\hat{p}1 - \hat{p}2}{\sqrt{\hat{p}(1 - \hat{p}1)(\frac{1}{n1} + \frac{1}{n2})}}$$

Formula 3: formula for two-tailed z-test [79, 80].

Here, $\hat{p}$ is $\frac{n1\hat{p}1+n2\hat{p}2}{n1+n2}\hat{p}_1$ and $\hat{p}_2$ represent the percentage of correct answers and n_1 and n_2 represent the total amount of answers for both categories. In this case, the specific sound and the average categories are used. For each sound, the null hypothesis (H_0) is that there is no difference in spatial perceptibility between the specific sound and the average. The alternative hypothesis (H_1) is that there is a significant difference in spatial perceptibility between the specific sound and the average. The significance of the two-tailed test is again set to <0.05. This When looking at the ‘Commonly Used Critical Values z0 from the Standard Normal Distribution from’ section on page 8 of the ‘Binomial Probability Distribution table’ from San-José state university [80], the critical range for the Z-scores can be seen. For a significance of 0.05, this means that the critical range for the resulting Z-scores is greater than 1.96 or smaller than -1.96.

This results in the following data:

Sound	Total number	Number correct	Ratios correct/total
Biker Bell	231	109	0.4718614719
Electric Car	216	102	0.4722222222
Pedestrians	213	103	0.4835680751
Truck	234	94	0.4017094017
Total	894	408	0.4563758389

Table 10: Table containing the number of correct answers per sound.

	Total number	Ratio
Biker Bell:	231	0.472
All	894	0.456
p̂	0.4592853333	
p̂1-p̂2	0.016	
p̂(1-p̂)	0.2483423159	
z	0.4350038567	

Table 11: Table testing the z-value for the Biker Bell sound compared to the average.

	Total number	Ratio
Electric Car:	216	0.472
All	894	0.456
p̂	0.4591135135	
p̂1-p̂2	0.016	
p̂(1-p̂)	0.2483282952	

z	0.4234879181	

Table 12: Table testing the z-value for the Electric Car sound compared to the average.

	Total number	Ratio
Pedestrians:	213	0.484
All	894	0.456
p̂	0.4613875339	
p̂1-p̂2	0.028	
p̂(1-p̂)	0.2485090775	
z	0.7366677429	

Table 13: Table testing the z-value for the Pedestrians sound compared to the average.

	Total number	Ratio
Truck:	234	0.402
All	894	0.456
p̂	0.4447978723	
p̂1-p̂2	-0.054	
p̂(1-p̂)	0.2469527251	
z	-1.47982032	

Table 14: Table testing the z-value for the Truck sound compared to the average.

The fifth area of interest is the results from the questionnaire. The results will be analyzed qualitatively, meaning that groups of similar answers will be established and the number of answers in these groups will be counted. Some answers were categorized in two groups and some answers were blank. Therefore, the amount of answer categories for one question does not reflect the amount of answers given by the participants. The last question, whether the participants would like to participate in the contest, was excluded, due to its irrelevance for the data. This results in the following eight tables:

Have you listened to spatial audio before this experiment?	
Category:	Answer Count:
Yes, YouTube	5
Yes, games	3
Yes, real life	1
Yes, science museums	1
Yes, cinema	2
Yes, technology stores	1
Yes, at home	1
Yes, air pods	1
Yes, virtual reality	1
Yes, 4D songs	1
Probably	1
No	1
Yes, music	2

What did you expect to hear before participating?	
Category:	Answer Count:
I was curious to the accuracy	1
Different sounds and locations	8
Expected it to not be accurate	1
The sound would move	2
Noises/dings in different directions	1
Traffic sounds	3
Sounds on headphones	2
No clue	2
More general questions than this specific	1

How did the evaluation live up to your expectations?	
Category:	Answer Count:

What was positive / what positively surprised you about using the system?	
Category:	Answer Count:

Good enough	1
Harder than expected/Difficult	4
Was interesting/fun	2
Good difficulty	2
Expected different sounds	1
Had none	2
Positively surprised	1
Would be better if I were in the ring	1
Was close to expectations	1
Vertical direction was most interesting	1
Don't know	1
Went well once in the flow	1
Vertical and front-back were hardest	1
Expected larger differences	1

It works	2
Interface looks cool	1
Was fun!	2
Differences were interesting	2
How it works is interesting	2
Even small differences were audible	1
Distance was easily heard	2
The thorough testing	1
General direction was very audible	1
I was really able to locate sounds	2
Azimuth was easily heard	1
Elevation was easily heard	1
The number of correct answers	1
Didn't expect something	1
I thought I would be better at it	1
The beautiful test environment	1

When the sound was showcased	1

What was negative / what negatively surprised you about using the system?	
Category:	Answer Count:
Sound too soft	1
Directions did not work sometimes	1
The difference in what I heard vs what it should have been	1
Not a lot	1
I can't hear above me	1
Sometimes I heard the sound at the opposite direction because of my hearing	1
How difficult it was	5
Did not feel like sounds came from the right directions	1
Electric car and truck did not work well for me	1
Distance was hard	1
Azimuth was hard	1

Did you experience any discomfort when using the system?	
Category:	Answer Count:
No	13
Bit sweaty	1
No clue you dragged me into this	1
The pedestrians sound was annoying	1
Headphones were nice	2
Volume was good	1
Maybe wearing headphones in general was discomforting	1

Number of incorrect answers	1
Didn't expect a lot	1
Feels like it’s very different per person	1
I'm deaf	1
Especially small differences were hard	2

Do you think the system can be useful to aiding bikers in traffic?	
Category:	Answer Count:
Yes, but preciseness needs improvement	2
Distance is incorrect, will only be distraction	1
Yes, since you can hear the locations	1
Don't know how though	1
If it works, yes	2
Yes, if developed further	2
Yes	6
Yes, but only if you're already listening to music	1
Not really, maybe for identifying left/right	1
Yes, for some people, since some people won't hear traffic sounds	1
Maybe	1

Why do you think so?	
Category:	Answer Count:
Yes, it can substitute missing sounds when listening to music	2
Yes, because even a general direction is useful	1
Yes, because high preciseness is not necessary	1
Don't know, just a guess	1
Yes, it can give directions or warnings	3

Did you enjoy it?	
Category:	Answer Count:
Yes	14
Was interesting how it works	1
It was hard	1
It was chill	1
It was fun to interact with	1
But I was not good at it	2
It took quite long	1

Do you have any final remarks you want to add?	
Category:	Answer Count:
Thanks for the cookies	1
It was fun	2
Great experience	1
It is cool	1
I don't know how, but it works	1
No	8
Sometimes I heard the sound come from opposite direction	1
Interesting to see where this leads	2
We did not have a baseline	1
Distance had to be heard by comparing to others	1
What's the point of the pictures	1
Maybe create prefabs for people	1
Success with writing your report!	1

Tables 15 - 25: Qualitative categorization of responses from participants to each question.

Chapter 8: Discussion

Now that the results have been presented and statistically analyzed, the discussion will interpret these results. The interpretation of results will be divided into the five areas of interest, as discussed in the evaluation section. It will also include an evaluation of the results of the functional system test This will be followed by the limitations, fulfilled criteria, and finally future work.

8.1 Interpretation of Results

Precision of Directional Perception

As stated in chapter 7, the directional perception includes both the accuracy and resolution of a sound. The degree value for the accuracy states the directional distance the closest neighbors are away from the correct sound. When looking at table 2, the results for the accuracy of the azimuth show that for 120 degrees, the number of correct results is significantly correct. This indicates that participants were overall able to correctly perceive the azimuth with 120 degrees of accuracy. This cannot be said about the other accuracy values. Simulating the horizontal direction therefore seems to be accurate for at best 120 degrees. Given the danger traffic presents, it is questionable whether any mistakes should be allowed to be made at all. Twelve out of eighteen correct answers may therefore still not be sufficient for aiding bikers in traffic.

The results for the accuracy of the elevation show an interesting effect. There is no significant indication that participants were able to hear an accuracy of 120 degrees. However, for 60 degrees, with 19 out of 28 answers correct, there is a significant effect. An explanation could be that the participants needed to get used to the elevation to answer it properly. However, it is unknown whether this is an effect. It is inconclusive whether 120 degrees accuracy is harder to distinguish than 60 degrees, or whether one of the results is incorrect. For the precise measurements there was no significant effect. It is also questionable whether the elevation is an important factor to simulate, given that traffic mostly happens at around eye level.

The accuracy of distance perception shows that participants were able to distinguish a sound difference of up to 15 meters. However, as with the accuracy of the azimuth, distance is an important coefficient to simulate. And given the danger traffic presents, one might argue that 100% correct answers should be considered the norm.

For the resolution in table 3, it becomes quite clear that even 4 degrees was not significantly distinguished for both the azimuth and elevation. It appears that that level of resolution was not achieved by the prototype. With regards to the distance however, both 4- and 3-meters distance were audible to the participants. Although not significant, 2 meters also showed a high likelihood of being perceivable. Here we can conclude that participants were able to hear a difference in distance of up to 3 meters.

Performance Compared to Existing Systems

When looking at the accuracy of the azimuth for the existing systems in table 4, there seems to be no statistically significant number of correct answers for any of the questions, apart from testing 5 degrees. With a probability of 0.049, although statistically significant, it is still questionable whether the high number of correct answers is caused by the prototype or chance. This is especially the case given the low statistical significance of the previous questions testing lower accuracy.

For the accuracy of the elevation, a similar effect as with the prototype was observed. In both cases, the question testing 120 degrees was answered poorly, whereas the 60 degrees one was answered well. This might mean that 60 degrees is easier to distinguish than 120 degrees. It might also mean that the participants needed to get into some kind of flow when hearing the elevations. However, this dilemma cannot be answered yet.

The accuracy of the distance shows the result that 10 meters is easier to distinguish than 15 meters. This is an odd result, as it can be assumed that distinguishing distances further apart from each other should be easier, due to a bigger difference in the sound. All in all, it seems that for accuracy, the prototype performed similarly as the existing systems.

The same is the case for the resolution in table 5, where the existing system did not measure any significant results, apart from the 3 meters question. This is very similar to the prototype with roughly the same results. All in all, this shows that the prototype performed to a similar level in simulating both the accuracy and resolution of a virtual location.

Potential Learning Curve

When comparing the performance of the participants after having evaluated an already existing system in tables 6 and 7, versus the performance with the prototype being the first system in tables 8 and 9 , some differences can be seen. The results where participants have evaluated existing systems, show that the participants quite easily identified the azimuth at an accuracy of 120. Lower accuracy values were still difficult to determine. For the tests where existing prototypes have not been evaluated, there were no questions that were answered statistically significantly, meaning that it could not be determined whether the azimuth was able to be identified. This means that the participants with experience performed slightly better at hearing the accuracy of the azimuth.

For both types of evaluation, the elevation was performed very similarly. In both test situations, the 120 degrees accuracy results were outperformed by the 60 degrees accuracy results. No significant differences were found between either test situations.

The accuracy for the distance shows an interesting effect. For both systems, 15 meters accuracy seems to have been correctly heard by participants. For the evaluations that included existing systems, 10 meters was heard to a similar extent as 15 meters. 5 meters was unable to be distinguished significantly enough. The opposite is true for evaluations that did not include existing systems. Here 10 meters was not identified significantly correctly, whereas 5 meters was, although barely. It is unknown whether these measurements represent reality; it might be caused by a too small sample size.

For the resolution of the azimuth and the elevation, there was no difference in performance since no results were significant. Regardless of the type of evaluation, participants were not significantly able to identify 4 degrees and below. For the distance, the performance was the same as well. For both types of evaluation, up to 3 meters resolution seems to have been able to be heard by participants. In short, there seems to have been almost no difference between the performance of the participants that have experience with evaluating spatial audio systems and those that do not. A learning effect was therefore not observed.

Effect of Sound type on Spatial Perceptibility

From tables 10 to 14, the effect of the type of sound on the spatial perceptibility can be seen. The Z-values for each sound indicate how far the results from that specific sound deviate from the average. This means that the higher the Z-value, the better, and the lower the Z-value, the worse the performance was for questions with those sounds. In the results, a critical region has been established, indicating which Z-values are likely corresponding to an effect and which are likely caused by randomness. The critical region states that Z-scores should either be greater than 1.96 or smaller than -1.96, to be considered statistically significant. None of the measured Z-values fall within this critical region meaning that no sound caused a statistically significant difference in the results. This indicates that there likely is no effect of the type of sound on the spatial perceptibility.

Personal Experience of Participants

From the results of the questionnaire, tables 15 - 25, the experience of the participants when using the system can be seen. All participants reported that the evaluation was enjoyable and almost all reported that there was no discomfort when using the prototype. Many reported that the prototype did work, and directions could be heard in the sounds. However, many participants reported difficulty with hearing the exact directions of the sounds. This was also seen in the results. There were differences in the individual experiences of the participants, where some found certain directional coefficients hard to perceive, whereas others were quite good at those, but worse at other coefficients. Upon the question “Do you think the system can be useful to aiding bikers in traffic?”, most participants answered ‘yes’, as long as it is further developed, and the locations are easier to identify. These results indicate that the prototype works in the basics and is not viewed as an obstruction. However, for its intended purpose, the prototype needs more refinement. Some participants do report that the level of precision during the evaluation is not necessary for successfully aiding bikers in traffic. They thought that only playing the general direction would already be helpful enough.

Functional System Test Results

In table 1 in chapter 6.4, the results of the playback latency of the system were shown. The mean playback latency was 4.59 milliseconds, and the median was 5 Ms. When using the system, the average playback latency lies far below the 30 Ms maximum criterion. This indicates that the spatial audio will be perceived as being real-time. There might however still be outliers, where the playback latency was higher. Of the 1026 measurements, the highest playback latency was 15 milliseconds, which was measured only once. That is still half the max 30 Ms criterion. The system latency was consistently low enough for real-time spatial audio to be heard, even in rare moments where the playback latency peaked.

8.2 Limitations

This study was subject to several limitations: first, the prototype could not be evaluated in its intended environment. The prototype evaluation was unable to be performed on the phone. However, this most likely would not have made a difference for the perceptibility of the spatial audio, since the mathematical calculations would have been the same for either the phone or the laptop; the perceptibility of spatial audio is independent of the device if the audio processing is identical. For the playback latency, the device used does make a big difference. It is highly likely that, given the lower processing power of phones compared to laptops, the playback latency would have been larger. Especially when the device plays music at the same time.

Another limitation was the fact that the prototype could only play one sound at the same time. Had it been evaluated while processing multiple sounds, the results would have been different. For starters, the playback latency would have been lower, due to the higher processing demand. On top of that, the perceptibility of the spatial audio would have been worse, since other sounds would be playing at the same time, making the individual sounds harder to distinguish. On the other hand, it might have been easier for participants to identify individual sounds, since the other sound could act as a reference. Further research will have to be done to find out this effect.

The location of the spatial audio sounds was static. This means that the participants were entirely dependent on their ability to hear locations. However, if the sounds were moving, similarly to real-life traffic scenarios, the participants might have been able to compare the location of each sound with their previous locations. Moving sounds might therefore be easier to identify than static sounds, meaning the performance in traffic would be higher than what was evaluated in this research. Again, further research will need to be done to find out whether this is true.

The headphones used, HIFIMAN Sundara, are producer-grade headphones worth 350 euros, meaning their audio quality is much higher and cleaner than the playback devices that would likely be used by cyclists in traffic. In chapter 2.1.2, it was explained that playback devices with flatter frequency responses likely perform better at playing spatial audio. An evaluation comparing different playback devices would be useful to determine for which devices the quality of the spatial audio would be sufficient to be used to aid bikers in traffic.

The final limitation is the fact that the Head-Tracker was unable to be integrated with the main system. Had this been achieved, tests could have been conducted where the response of the system to rotating the head could have been measured. This is an arguably important factor, as bikers will likely turn their heads. To make the system useful to bikers, the absolute location of danger should be provided to the user. If the biker turns their head and the danger now seems to be coming from the right side of their bike, since the audio did not rotate with their head, the system might instead become a source of misinformation and therefore danger. Evaluating how the system responds to head rotations is therefore necessary. This is something that will need to be evaluated in further research.

8.3 Fulfilled Criteria

Based on the evaluation and limitations, it can be determined to what extent the prototype fulfills the functional system requirements set in chapter 5.3.1. There are several requirements that the prototype fulfills. The prototype can spatialize audio in all directions, both in and outside of the vision of the user. This is evidenced by the questionnaire, where almost all participants reported that the system does produce spatial audio and general direction were able to be heard. The system does work with headphones, whether it works with all headphones or whether it works with earbuds has not been tested yet. But this is entirely determined by the playback devices themselves, as the software has been shown to work with stereo playback devices with two drivers. The system was evaluated with participants. It can play different sounds, something that has also been tested. The sounds have little influence over the performance of the system, meaning that all sounds work about equally well for creating spatial audio. The system can play sounds with different modes, including a manual mode and two evaluation modes, one for each evaluation game.

However, the prototype was unable to achieve the accuracy of 3 degrees and 1 meter simulated precision that was set in the requirements. On the laptop, the playback latency was low enough that the experience should be real-time. Though, it can be questioned whether this still applies if the prototype was ported to a mobile device and was able to play multiple sounds at the same time. Although the Head-Tracker works by itself, it was not possible to get the SerialReader to work in tandem with the rest of the code, meaning that data could not be received by the main application. Therefore, the requirement of moving the sound with head rotations was not fulfilled. As hinted at earlier, the prototype can also not play and spatialize multiple sounds at the same time. And given that the processing device is a laptop, the prototype is not as mobile as a mobile phone. Although it can technically still be used by bikers, this is highly impractical, and it is questionable whether this requirement can be considered fulfilled.

A discussion can be had about whether all requirements are sufficient. The questionnaire showed an interesting result: according to the participants, high accuracy spatial audio playback does not have to be achieved to serve its purpose to improve traffic safety for bikers. Providing an indication of just the general direction would already be sufficient to give the bikers an indication of the danger around them. Lowering the precision with which spatial audio feedback is created should therefore be considered. Although, given the results from the evaluation, the prototype would likely not fulfill the new requirements either.

8.4: Future Work

Based on the discussion, a few suggestions can be made for future research with building a spatial audio system to aid bikers in traffic. The biggest downside of the current prototype is the fact that it cannot produce very clearly locatable spatial audio in a consistent manner. Both consistency and precision are necessities to make the prototype sufficiently safe to aid bikers in traffic. The locations simulated by the prototype are hard to distinguish and do not work the same for everyone. Therefore, the first suggestion for future work would be to improve the accuracy and resolution with which the prototype can simulate sound. This can be done by experimenting with different Head-related transfer functions, different types of audio processing or even customizing the Head-related transfer function to the specific user.

The second suggestion for future research would be to conduct research to find out what auditory cues are required to safely participate in traffic. Currently, not a lot of research has been conducted on the effects of the properties of audio cues on the ability to safely take part in traffic. This would not only be useful to determine what precision the current prototype should aim for, as that is a currently unclear factor. It also could be useful information for road design, to make traffic and intersections safer to manage for bikers.

Lastly, the prototype could be improved to make it usable to aid bikers in traffic and integrate it into the bigger system to improve traffic safety for bikers. All the features that are currently missing should be added to the prototype; making the application work on mobile phones and making the prototype able to spatialize multiple sounds at the same time. The features that were suggested to be in the final system to aid bikers in traffic should be added as well; namely object recognition and visual feedback by google glass. The entire system can then be evaluated by bikers in traffic or a simulation of traffic.

Chapter 9: Conclusion

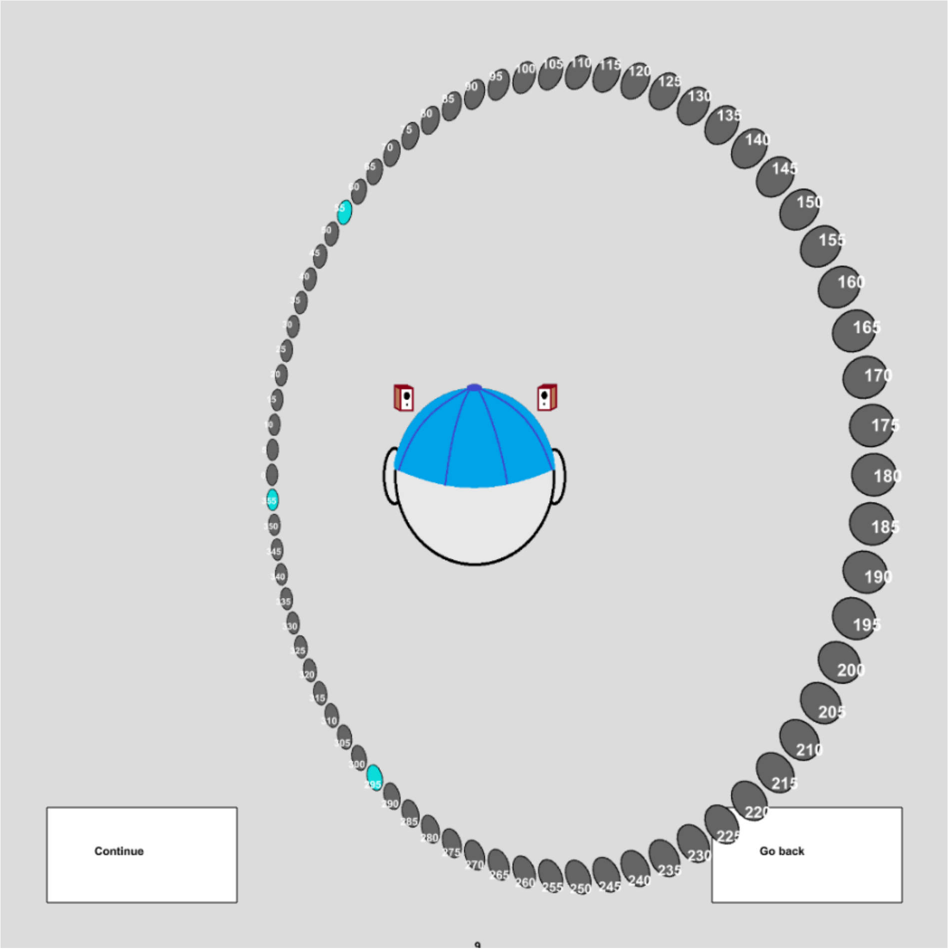
Many people nowadays listen to music while riding their bikes in traffic. This causes them to miss necessary sounds from traffic, resulting in potentially dangerous situations. A suggestion was made to substitute the missing sounds with spatial audio sounds from the headphones or earbuds. The goal of this thesis was to develop such a system that can create spatial audio with sufficient performance so that it can be used to aid bikers in traffic.

It was found that the best method to create such a spatial audio system would be to create Binaural spatial audio, making use of so-called Head-related transfer functions. Currently there exist no such systems to aid bikers in traffic through spatial audio, but based on other applications of spatial audio, it has great potential to work. Then, a final idea and criteria were established, based on which a prototype was developed. This prototype was then evaluated with seventeen participants using two games. It was found that the prototype was able to create spatial audio to the same extent as already existing systems, although without the necessary performance to aid bikers in traffic. However, the participants had a great experience using the prototype and suggested that perhaps the performance criteria are too strict. According to them, the prototype could work to aid bikers in traffic, given that the spatial audio quality is improved. Improving the spatial audio prototype, along with finding new criteria for the spatial audio were then suggested as topics for future research.

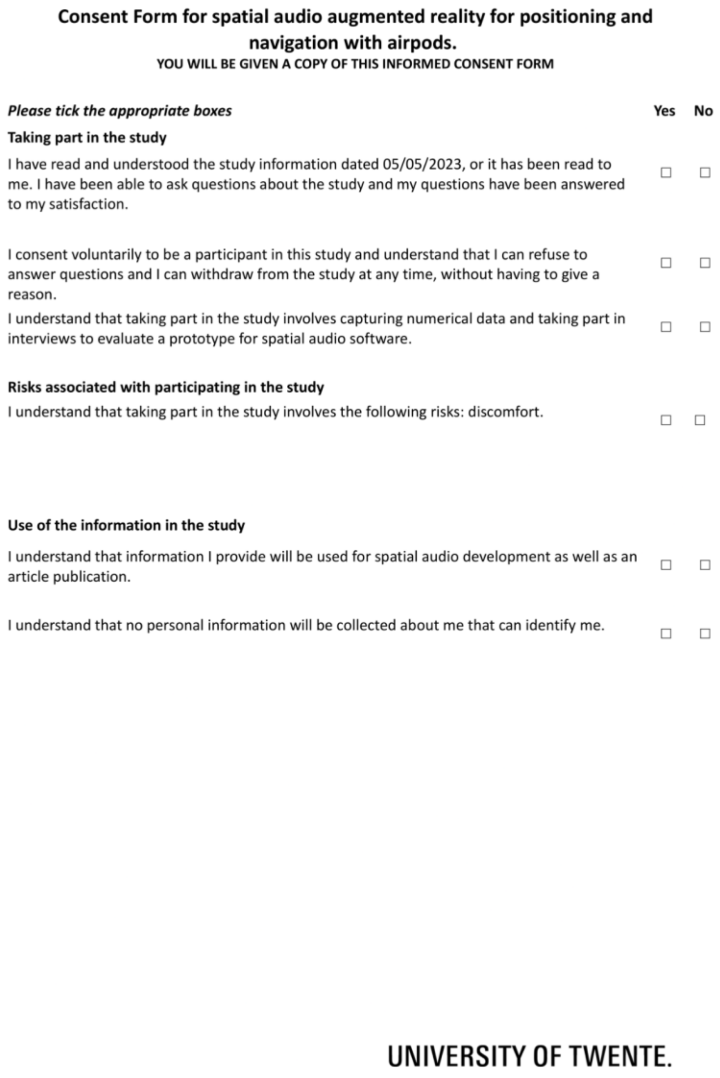
This work is a first step towards building a system that can improve biker safety in traffic. A method with which spatial audio can be created. Future research can be conducted to improve the performance of this prototype and add more features to make it suitable to aid bikers in traffic. Moreover, this paper presents information on methods to develop spatial audio systems and shows that one method works. Developers can use this work as reference for developing their own spatial audio applications. All in all, this research opens doors for future advancements in spatial audio, encourages further exploration to improve the prototype and provides a compilation of knowledge for future spatial audio development.

Appendix

A: Images of Prototype and Evaluation Tools


95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
Continue
Go back

B: Information Letter and Consent Form



Consent Form for spatial audio augmented reality for positioning and navigation with airpods.

YOU WILL BE GIVEN A COPY OF THIS INFORMED CONSENT FORM

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information dated 05/05/2023, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	☐	☐
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	☐	☐
I understand that taking part in the study involves capturing numerical data and taking part in interviews to evaluate a prototype for spatial audio software.	☐	☐
Risks associated with participating in the study		
I understand that taking part in the study involves the following risks: discomfort.	☐	☐
Use of the information in the study		
I understand that information I provide will be used for spatial audio development as well as an article publication.	☐	☐
I understand that no personal information will be collected about me that can identify me.	☐	☐

UNIVERSITY OF TWENTE.

Signatures

____________________   ____________________   ________
Name of participant [printed]   Signature   Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

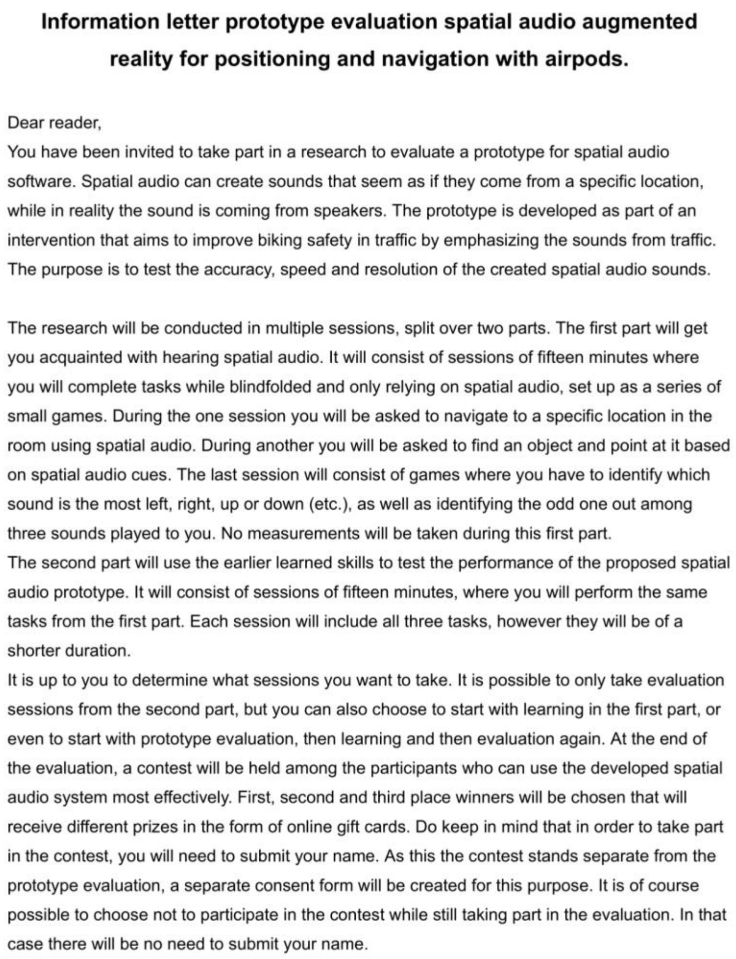
Emiel Nagel
Researcher name   Signature   Date

Study contact details for further information: Emiel Nagel, e.nagel@student.utwente.nl

Contact Information for Questions about Your Rights as a Research Participant

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee Information & Computer Science: ethicscommittee-CIS@utwente.nl

UNIVERSITY OF TWENTE.



Information letter prototype evaluation spatial audio augmented reality for positioning and navigation with airpods.

Dear reader,
You have been invited to take part in a research to evaluate a prototype for spatial audio software. Spatial audio can create sounds that seem as if they come from a specific location, while in reality the sound is coming from speakers. The prototype is developed as part of an intervention that aims to improve biking safety in traffic by emphasizing the sounds from traffic. The purpose is to test the accuracy, speed and resolution of the created spatial audio sounds.

The research will be conducted in multiple sessions, split over two parts. The first part will get you acquainted with hearing spatial audio. It will consist of sessions of fifteen minutes where you will complete tasks while blindfolded and only relying on spatial audio, set up as a series of small games. During the one session you will be asked to navigate to a specific location in the room using spatial audio. During another you will be asked to find an object and point at it based on spatial audio cues. The last session will consist of games where you have to identify which sound is the most left, right, up or down (etc.), as well as identifying the odd one out among three sounds played to you. No measurements will be taken during this first part.
The second part will use the earlier learned skills to test the performance of the proposed spatial audio prototype. It will consist of sessions of fifteen minutes, where you will perform the same tasks from the first part. Each session will include all three tasks, however they will be of a shorter duration.
It is up to you to determine what sessions you want to take. It is possible to only take evaluation sessions from the second part, but you can also choose to start with learning in the first part, or even to start with prototype evaluation, then learning and then evaluation again. At the end of the evaluation, a contest will be held among the participants who can use the developed spatial audio system most effectively. First, second and third place winners will be chosen that will receive different prizes in the form of online gift cards. Do keep in mind that in order to take part in the contest, you will need to submit your name. As this the contest stands separate from the prototype evaluation, a separate consent form will be created for this purpose. It is of course possible to choose not to participate in the contest while still taking part in the evaluation. In that case there will be no need to submit your name.

At any point during the evaluation you can choose to withdraw from taking part. In this case, please send an email to the main researcher. His contact information can be found at the bottom of this document. You can also choose to erase the collected information from your evaluation sessions from the researcher's archive.
The data that will be collected numerical data (numbers) about the precision and accuracy with which you have completed the tasks using the spatial audio prototype. These will be used to further improve the prototype. It is important to note that failure to successfully complete the tasks is not an indication of your own abilities, but rather the usefulness of the spatial audio system itself. At the end, anonymous interviews will be held as well, with regards to the comfort and experience of using the spatial audio system.
The data will be stored on the researcher's laptop and destroyed after the prototype has been completed, around the 7th of July. Only the researcher will have access to this data. It is possible that the data will be published in a paper describing the development of the system. However, your personal information will not be collected and you will not (and cannot) be linked to the prototype evaluation results. Besides the article, the data will not be spread to other sources and will be destroyed after the system development has been completed.

The benefits of participation are the opportunity to play games using spatial audio, which could result in a fun time as well as improving your skills. There will of course also be coffee and freshly baked cake at the end of each session. And finally, at the end of the research there will be a contest, where prizes can be won.
The risks of participation are few. No personal data will be collected and the sessions will be set up in such a way that there is no risk of injury. Using the spatial audio device might cause discomfort, in which case please notify the researcher and the research will immediately be stopped. Data on your performance will be collected and published, however this is done completely anonymously and can not be linked back to you. This research has been reviewed by the Ethics Committee Information and Computer Science.

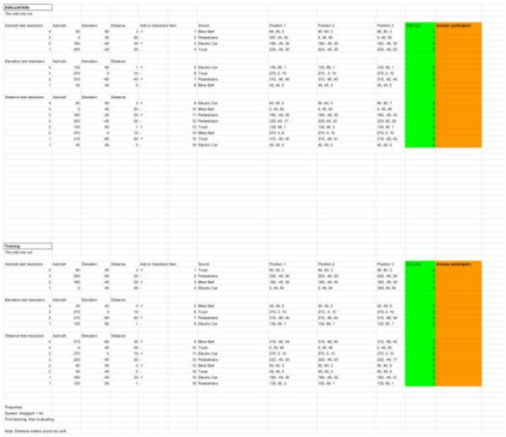
If you have any questions or concerns, please contact the researcher.
Contact details:
Emiel Nagel,
e.nagel@student.utwente.nl
0628879888

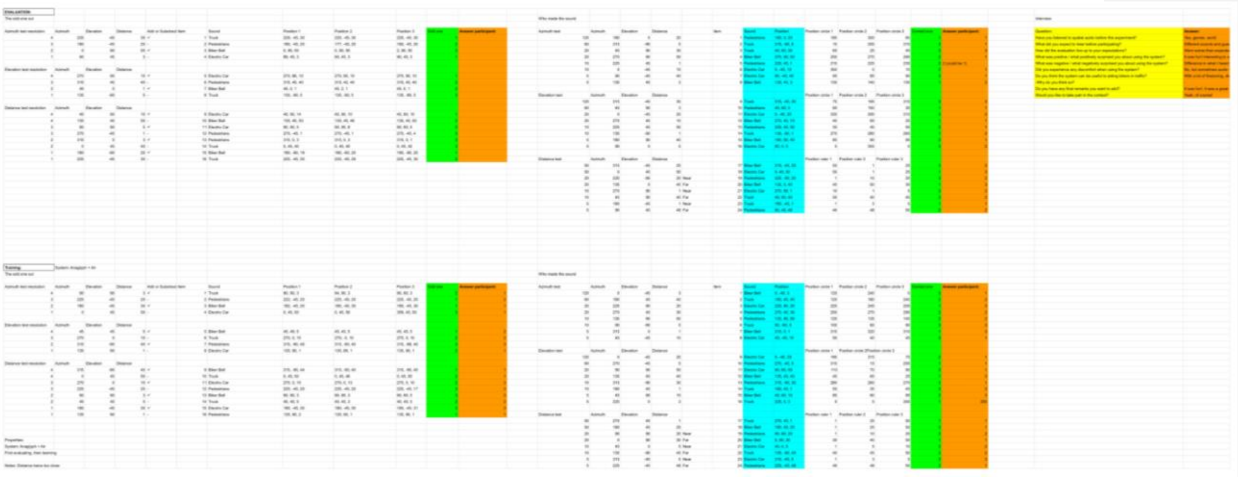
Contact Information for Questions about Your Rights as a Research Participant

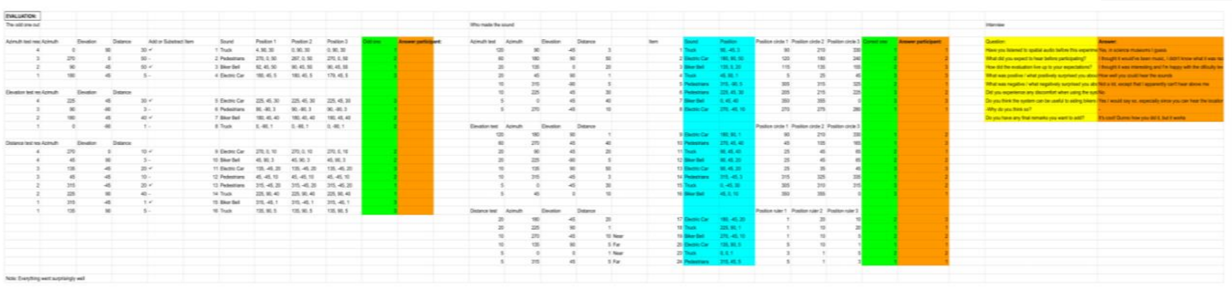
If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee Information & Computer Science: ethicscommittee-CIS@utwente.nl

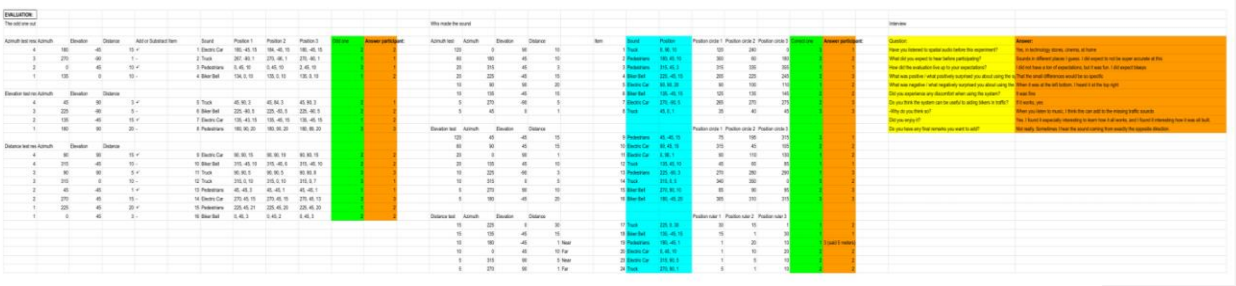
C: Results of User Evaluation

Link to sheets file: https://docs.google.com/spreadsheets/d/19V0NX0lVgJsEFl9eK8gxuuz7zF-bKdBDeXuVcu886hM/edit?usp=sharing









EVALUATION:

The odd one out

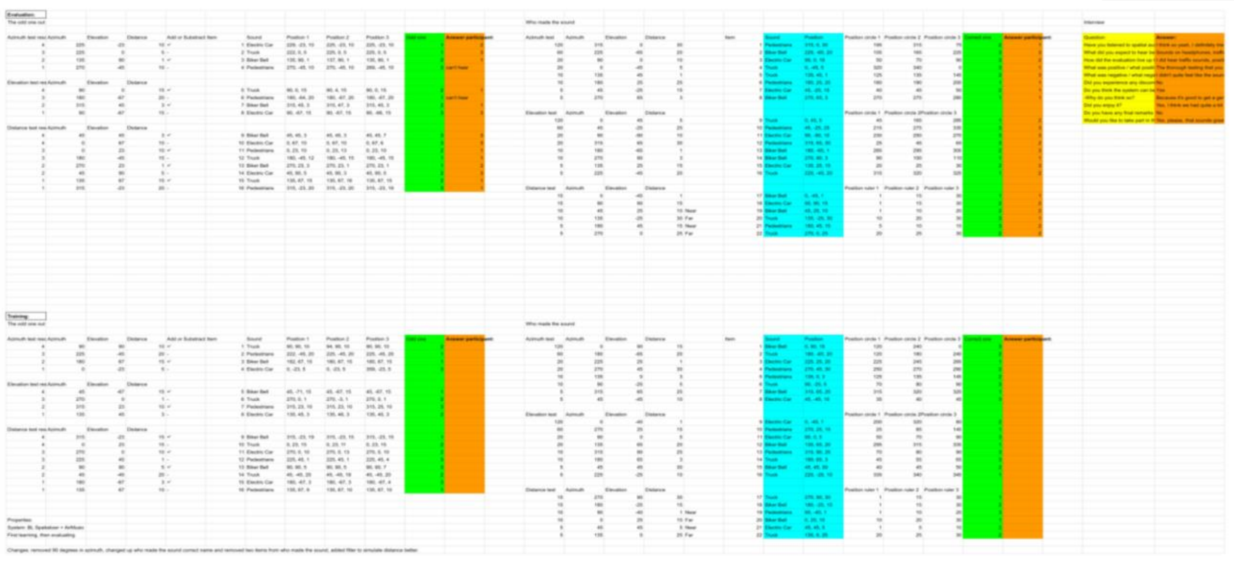
Azimuth test resu	Azimuth	Elevation	Distance	Add or Substract Item		Sound	Position 1	Position 2	Position 3	Odd one	Answer participant:
4	0	45	1	✓	1	Biker Bell	0, 45, 1	4, 45, 1	0, 45, 1	2	2
3	135	-45	10	-	2	Truck	132, -45, 10	135, -45, 10	135, -45, 10	1	can't tell
2	270	0	15	✓	3	Pedestrians	270, 0, 15	270, 0, 15	272, 0, 15	3	2
1	180	-90	10	-	4	Electric Car	179, -90, 10	180, -90, 10	180, -90, 10	1	2
Elevation test res	Azimuth	Elevation	Distance								
4	135	-45	3	✓	5	Truck	135, -45, 3	135, -41, 3	135, -45, 3	2	3
3	225	-90	20	-	6	Electric Car	225, -87, 20	225, -90, 20	225, -90, 20	1	2
2	45	90	15	✓	7	Biker Bell	45, 90, 15	45, 92, 15	45, 90, 15	2	can't tell
1	180	90	5	-	8	Pedestrians	180, 90, 5	180, 90, 5	180, 89, 5	3	2
Distance test res	Azimuth	Elevation	Distance								
4	45	45	20	✓	9	Pedestrians	45, 45, 20	45, 45, 24	45, 45, 20	2	2
4	90	0	1	-	10	Electric Car	90, 0, 1	90, 0, 1	90, 0, 5	3	3
3	90	45	3	✓	11	Biker Bell	90, 45, 6	90, 45, 3	90, 45, 3	1	1
3	270	-45	5	-	12	Truck	270, -45, 5	270, -45, 2	270, -45, 5	2	2
2	315	90	10	✓	13	Biker Bell	315, 90, 10	315, 90, 10	315, 90, 12	3	2
2	0	45	15	-	14	Electric Car	0, 45, 15	0, 45, 15	0, 45, 13	3	3
1	315	-45	10	✓	15	Truck	315, -45, 11	315, -45, 10	315, -45, 10	1	1
1	225	90	15	-	16	Pedestrians	225, 90, 15	225, 90, 14	225, 90, 15	2	2

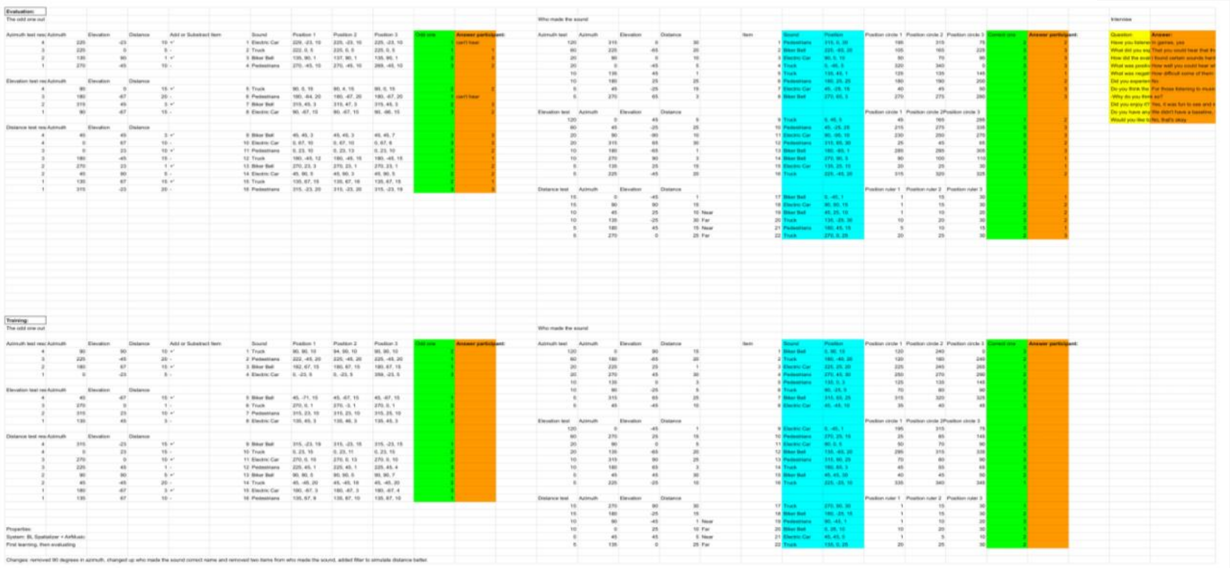
Who made the sound

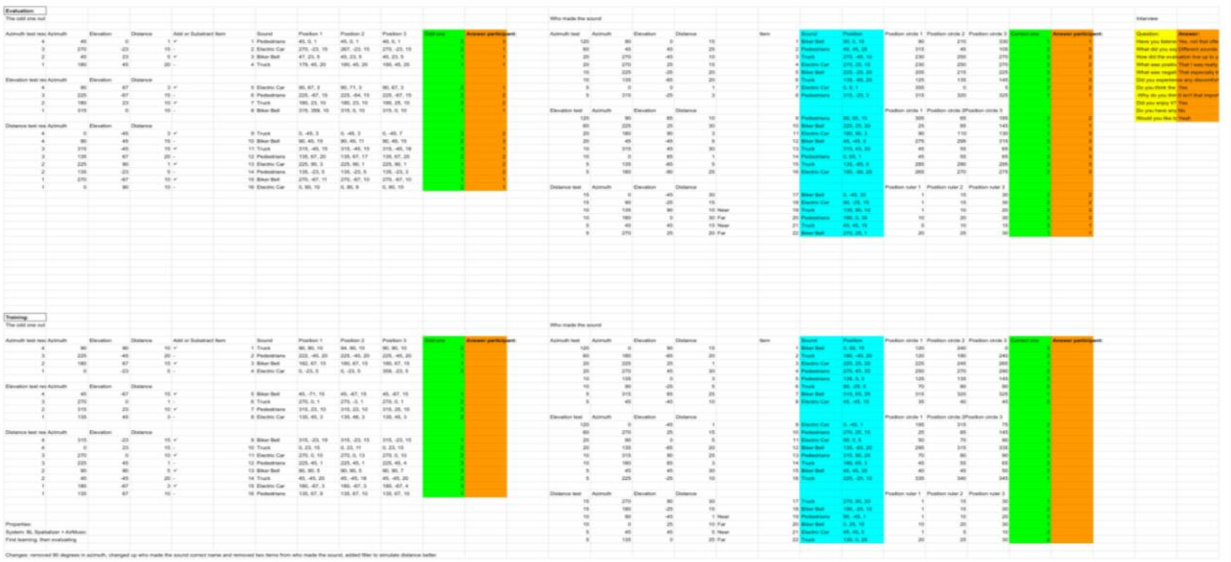
Azimuth test	Azimuth	Elevation	Distance		Item	Sound	Position	Position circle 1	Position circle 2	Position circle 3	Correct one	Answer participant:
120	225	0	10		1	Pedestrians	255, 0, 10	255	15	135	1	1
60	270	90	15		2	Biker Bell	270, 90, 15	150	210	270	3	1
20	180	-45	10		3	Pedestrians	180, -45, 10	160	180	200	2	1
20	90	-45	15		4	Electric Car	90, -45, 15	90	110	130	1	2
10	315	45	5		5	Electric Car	315, 45, 5	305	315	325	2	1
10	45	45	3		6	Biker Bell	45, 45, 3	35	45	55	2	3
5	135	90	20		7	Truck	135, 90, 20	125	130	135	3	3
5	0	-90	1		8	Truck	0, -90, 1	350	355	0	3	2
Elevation test	Azimuth	Elevation	Distance					Position circle 1	Position circle 2	Position circle 3		
120	315	0	20		9	Pedestrians	315, 0, 20	120	240	0	3	3
60	45	-45	5		10	Truck	45, -45, 5	315	15	75	1	2
20	180	-90	15		11	Biker Bell	180, -90, 15	250	270	290	2	3
20	135	90	3		12	Electric Car	135, 90, 3	50	70	90	3	2
10	0	45	10		13	Biker Bell	0, 45, 10	45	55	65	1	2
10	270	-45	15		14	Pedestrians	270, -45, 15	315	325	335	1	3
5	225	45	1		15	Truck	225, 45, 1	35	40	45	3	1
5	90	90	10		16	Electric Car	90, 90, 10	85	90	95	2	1
Distance test	Azimuth	Elevation	Distance					Position ruler 1	Position ruler 2	Position ruler 3		
15	225	0	30		17	Biker Bell	225, 0, 30	15	30	1	2	2
15	0	45	15		18	Electric Car	0, 45, 15	15	30	1	1	1
10	180	90	1	Near	19	Truck	180, 90, 1	1	10	20	1	2
10	135	-45	10	Far	20	Pedestrians	135, -45, 10	20	10	30	2	1
5	270	90	1	Near	23	Truck	270, 90, 1	5	1	10	2	2
5	315	-45	25	Far	24	Electric Car	315, -45, 5	25	20	30	1	1

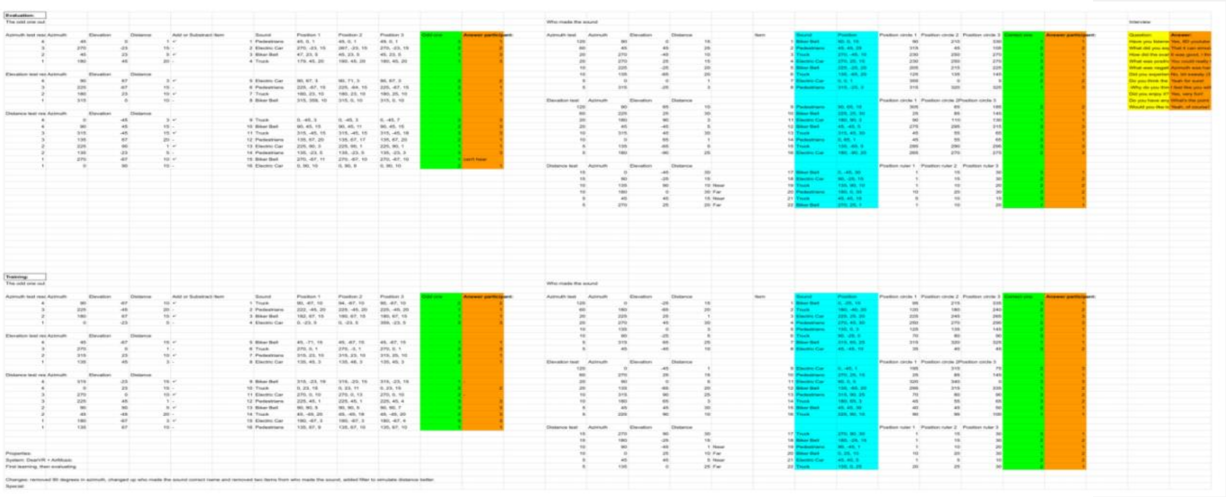
Interview

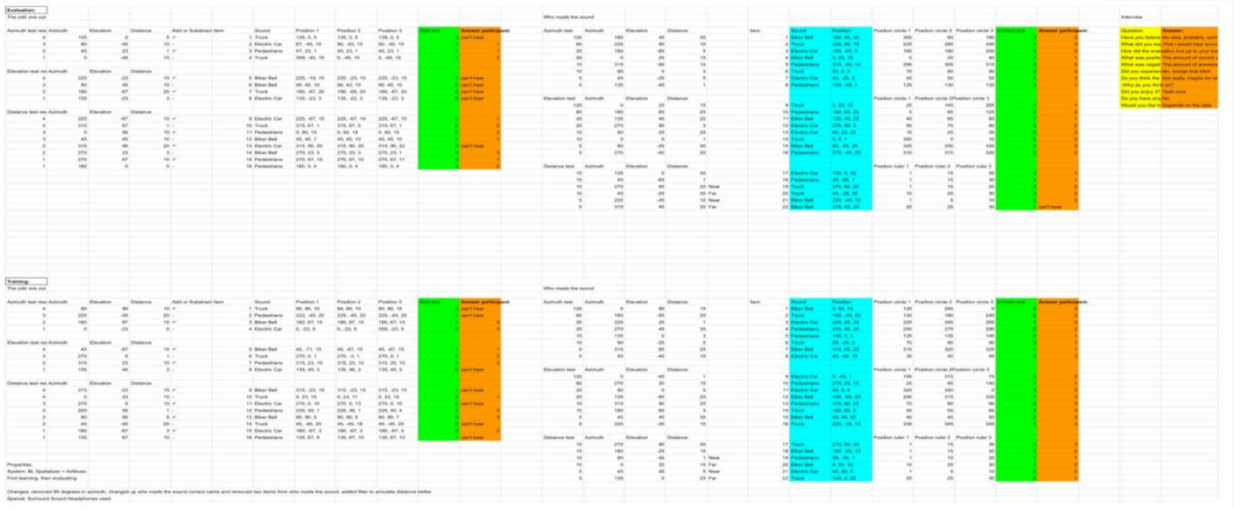
Question:	Answer:
Have you listened to spatial audio before this experim	Once maybe, on yout
What did you expect to hear before participating?	I expected that it woul
How did the evaluation live up to your expectations?	I really had no expect
What was positive / what positively surprised you abo	It's an interesting con
What was negative / what negatively surprised you ab	It was a bit difficult wh
Did you experience any discomfort when using the sy	No
Do you think the system can be useful to aiding biker	I think it would be use
-Why do you think so?	
Did you enjoy it?	Yuh, it was fun! It was
Do you have any final remarks you want to add?	I would just be interes
Would you like to take part in the contest?	Sure why not, I'll see i

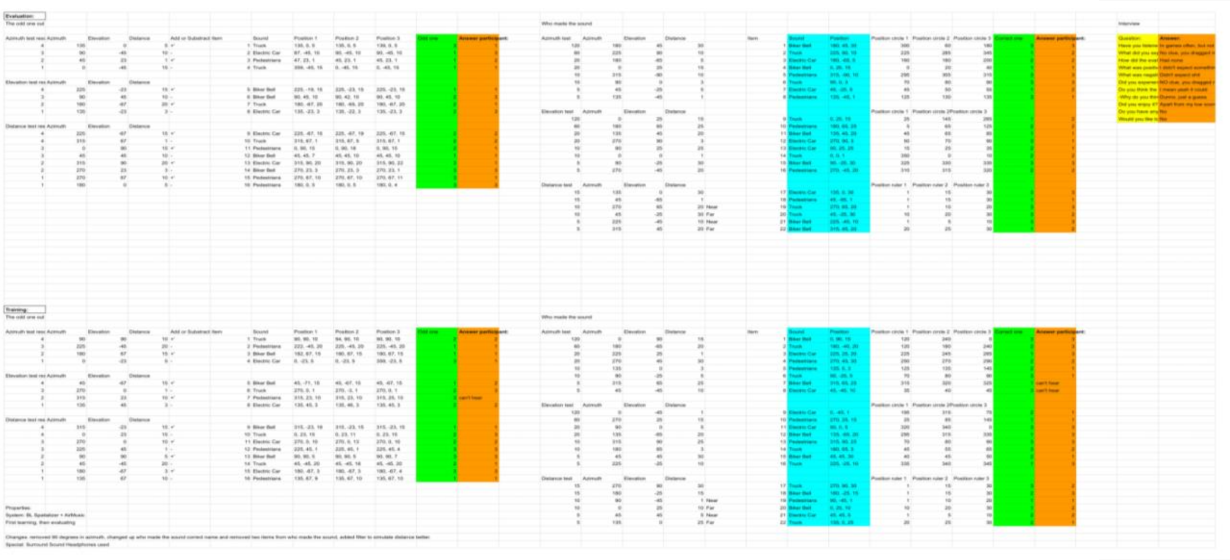


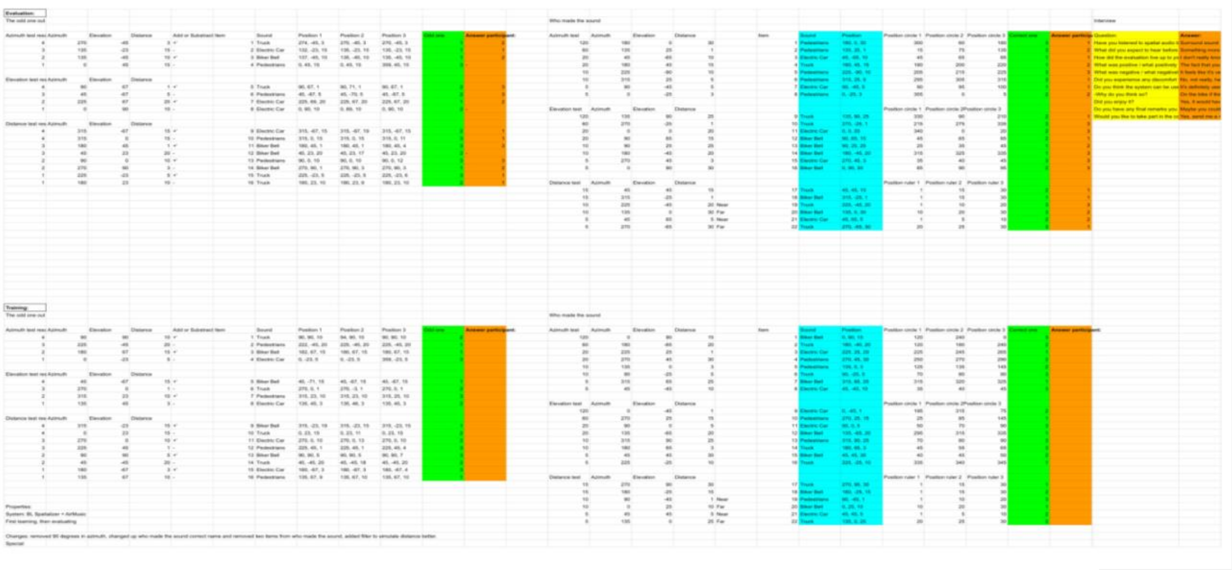


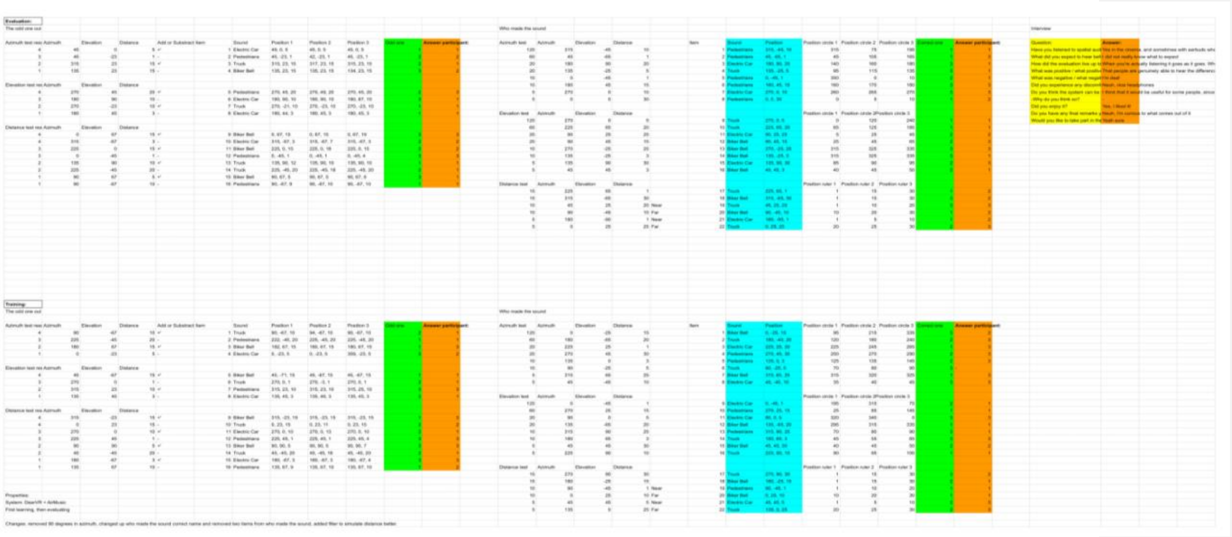


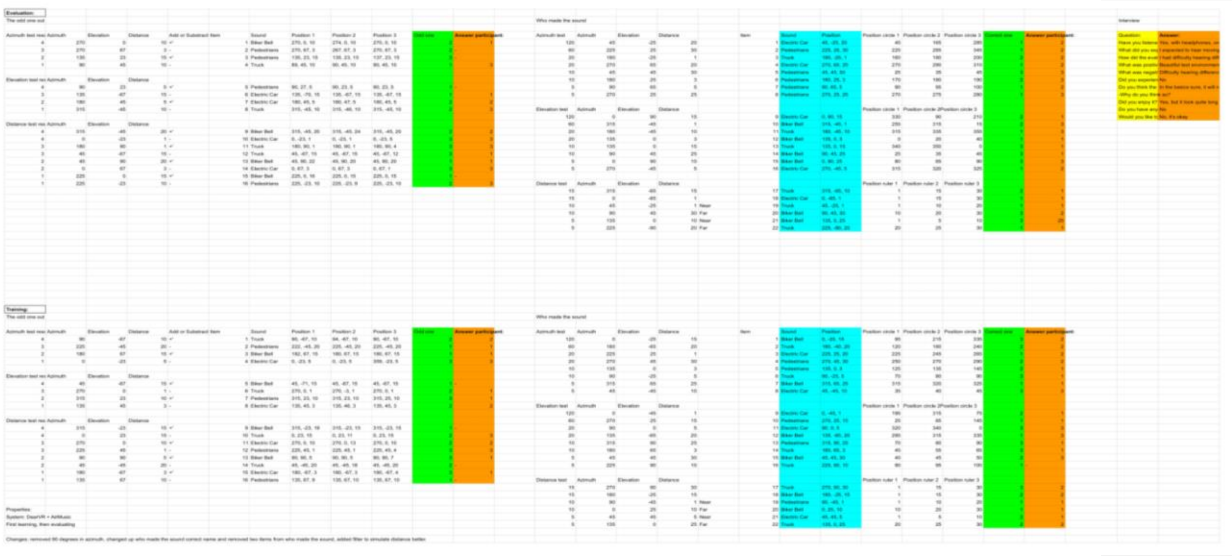


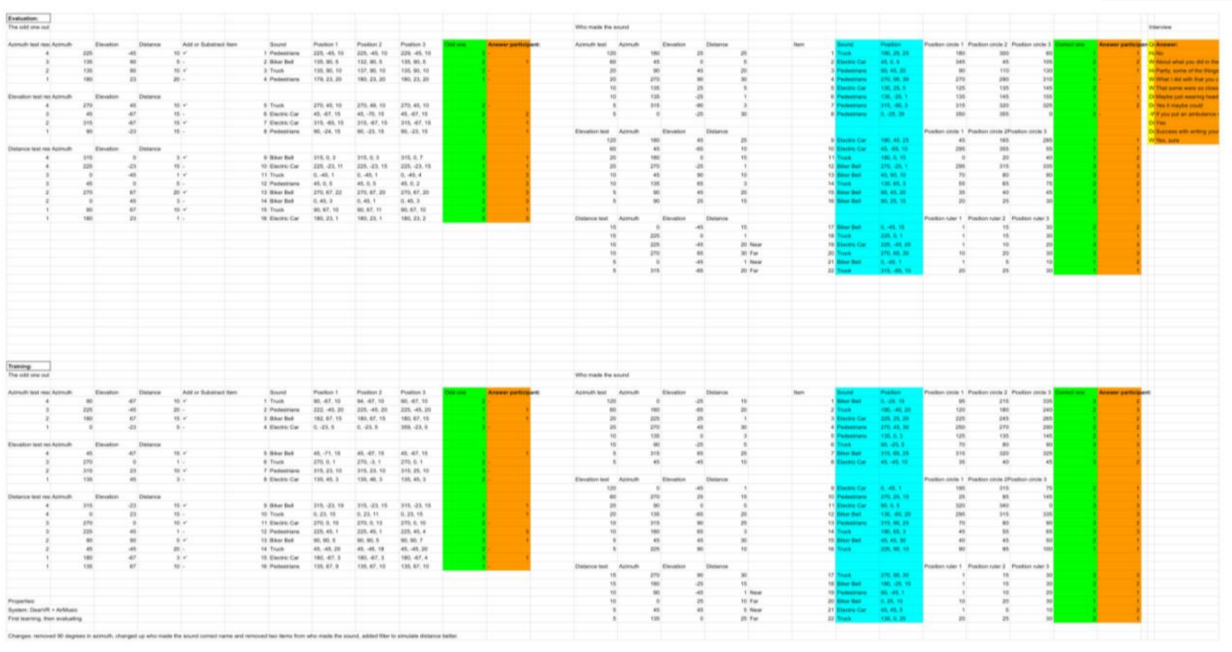












D: Example of Evaluation Code

Link to code: https://openprocessing.org/sketch/1957011

Login email: emielnagel01@gmail.com

Password: Penpran2002!

```
class button {
constructor(xPos, yPos, width, height, text)
{
        this.xPos = xPos;
        this.yPos = yPos;
        this.width = width;
        this.height = height;
        this.text = text;
}

displayButton()
{
        translate(this.xPos, this.yPos)
        fill(255);
        stroke(0);
        rect(0, 0, this.width, this.height);
        fill(0);
        text(this.text, 50, 50);
        translate(-this.xPos, -this.yPos)
}

checkMouseClicked(x, y)
{
        if ((x > this.xPos + 500 && x < this.xPos + 500 + this.width) && (y > this.yPos + 500 && y < this.yPos + 500 + this.height))
        {
                return true;
        }
        else
```

```
				{
					return false;
				}
			}
		}

		function setup() {
			createCanvas(1000, 1000, WEBGL);

			myFont = loadFont('arialbd.ttf');
			textFont(myFont);

			imgNormal = loadImage('Head.png');
			imgForwards = loadImage('HeadForwardsEdited.png');

			degreeToRadians = PI/180;
			radius = 400;

			backGroundColor = color(220, 220, 220);
			defaultColor = color(100, 100, 100);
			highLightColor = color(10, 220, 220);

			azimuths = [190, 45, 190, 270, 45, 135, 90, 90];

			azimuthsTest = [10, 270, 90, 135, 315, 190, 45, 225];

			elementColor = [];
			for (let index = 0; index < 72; index++)
			{
				elementColor.unshift(defaultColor);
			}

			soundPositions =
			[[180, 300, 60],
```

```
        [345, 45, 105],
        [90, 110, 130],
        [270, 290, 310],
        [125, 135, 145],
        [135, 145, 155],
        [315, 320, 325],
        [350, 355, 0],
        [45, 165, 285],
        [295, 355, 55],
        [0, 20, 40],
        [295, 315, 335],
        [70, 80, 90],
        [55, 65, 75],
        [35, 40,          45],
        [20, 25, 30]];

       soundPositionsTest =
[[95, 215, 335],
[120, 180, 240],
[225, 245, 265],
[250, 270, 290],
[125, 135, 145],
[70, 80, 90],
[315, 320, 325],
[35, 40, 45],
[195, 315, 75],
[25, 85, 145],
[320, 340, 0],
[295, 315, 335],
[70, 80, 90],
[45, 55, 65],
[40, 45, 50],
[90, 95, 100]];
```

```
        previousColor = [];

        soundNumber = 0;

        continueButton = new button(-450, 350, 200, 100, 'Continue');
        backwardsButton = new button(250, 350, 200, 100, 'Go back');
}

function draw() {
        background(backGroundColor);
        if (soundNumber < 8)
        {
                image(imgNormal, -100, -100, 200, 200);
        }
        else
        {
                image(imgForwards, -100, -100, 200, 200);
        }
        continueButton.displayButton();
        backwardsButton.displayButton();
        changeColor();
        rotateForwards();
        for (let index = 0; index < 72; index++)
        {
                degree = index * 5;
                xPos = -sin(degree * degreeToRadians) * radius;
                yPos = -cos(degree * degreeToRadians) * radius;
                translate(xPos, yPos);
                element(degree, index);
                translate(-xPos, -yPos);
        }
        rotateBackwards();
        fill(0);
        text(soundNumber, 0, 500);
```

```
}

function element(degree, index) {
	fill(elementColor[index]);
	circle(0, 0, 30);
	rotateBackwards();
	translate(0, 0, 20);
	fill(255);
	text(degree, -10, 5);
	translate(0, 0, -20);
	rotateForwards();
}

function changeColor() {
	for (let index = 0; index < 3; index++)
	{
		elementColor[previousColor[index]] = defaultColor;
	}
	previousColor = [];
	for (let index = 0; index < 3; index++)
	{
		var colorIndex = soundPositions[soundNumber][index] / 5;
		previousColor.unshift(colorIndex);
		elementColor[colorIndex] = highLightColor;
	}
}

function rotateForwards()
{
  if (soundNumber > 7)
  {
    rotateZ(90 * degreeToRadians);
    rotateX((azimuths[soundNumber - 8] + 90) * degreeToRadians);
  }
```

```
}

function rotateBackwards()
{
    if (soundNumber > 7)
    {
        rotateX((-azimuths[soundNumber - 8] - 90) * degreeToRadians);
        rotateZ(-90 * degreeToRadians);
    }
}

function mouseClicked()
{
            if (continueButton.checkMouseClicked(mouseX, mouseY) == true && soundNumber < 15) {soundNumber += 1;}
            if (backwardsButton.checkMouseClicked(mouseX, mouseY) == true && soundNumber > 0) {soundNumber -= 1;}
}
```

References

[1] A. Stelling-Konczak, G. P. van Wee, J. J. F. Commandeur, and M. Hagenzieker, “Mobile phone conversations, listening to music and quiet (electric) cars: Are traffic sounds important for safe cycling?,” *Accident Analysis & Prevention*, vol. 106, pp. 10–22, Sep. 2017, doi: 10.1016/j.aap.2017.05.014.

[2] D. de Waard, K. Edlinger, and K. Brookhuis, “Effects of listening to music, and of using a handheld and handsfree telephone on cycling behaviour,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 14, no. 6, pp. 626–637, Nov. 2011, doi: https://doi.org/10.1016/j.trf.2011.07.001.

[3] J. C. Middlebrooks, “Sound localization,” *The Human Auditory System - Fundamental Organization and Clinical Disorders*, vol. 129, pp. 99–116, 2015, doi: https://doi.org/10.1016/B978-0-444-62630-1.00006-8.

[4] P. Nowak, V. Zimpfer, and U. Zölzer, “3D virtual audio with headphones: A literature review of the last ten years,” *3D Virtual Audio with Headphones: A Literature Review of the Last Ten Years*, Mar-2018. [Online]. Available: https://www.researchgate.net/profile/Patrick-Nowak/publication/330412558_3D_Virtual_Audio_with_Headphones_A_Literature_Review_of_the_Last_Ten_Years/links/5c3f0c2b4

58515a4c7299990/3D-Virtual-Audio-with-Headphones-A-Literature-Review-of-the-Last-Ten-Years.pdf. [Accessed: 26-Mar-2023].

[5] C. Rajguru, M. Obrist, and G. Memoli, “Spatial soundscapes and virtual worlds: Challenges and opportunities,” *Frontiers in Psychology*, vol. 11, 2020, doi: https://doi.org/10.3389/fpsyg.2020.569056.

[6] W. Zhang, P. Samarasinghe, H. Chen, and T. Abhayapala, “Surround by sound: A review

of spatial audio recording and reproduction,” *Applied Sciences*, vol. 7, no. 5, p. 532, 2017, doi: https://doi.org/10.3390/app7050532.

[7] T. Lokki, L. Savioja, R. Vaananen, J. Huopaniemi, and T. Takala, “Creating Interactive

Virtual Auditory Environments," *IEEE Computer Graphics and Applications*, vol. 22, no. 4, pp. 49–57, Aug. 2002, doi: 10.1109/MCG.2002.1016698.

[8] B. Rafaely, V. Tourbabin, E. Habets, Z. Ben-Hur, H. Lee, H. Gamper, L. Arbel, L. Birnie, T. Abhayapala, and P. Samarasinghe, "Spatial Audio Signal Processing for binaural reproduction of recorded acoustic scenes – review and Challenges," *Acta Acustica*, vol. 6, p. 47, Oct. 2022, doi: https://doi.org/10.1051/aacus/2022040.

[9] A. Urviola, S. Sakamoto, and C. D. Salvador, "Ear centering for accurate synthesis of near-field head-related transfer functions," *Applied Sciences*, vol. 12, no. 16, pp. 82–90, Aug. 2022, doi: https://doi.org/10.3390/app12168290.

[10] J. Hong, J. He, B. Lam, R. Gupta, and W.-S. Gan, "Spatial Audio for Soundscape Design: Recording and reproduction," *Applied Sciences*, vol. 7, no. 6, p. 627, Jun. 2017, doi: 10.3390/app7060627.

[11] A. Alexandridis, A. Griffin, and A. Mouchtaris, "Capturing and reproducing spatial audio based on a circular microphone array," *Journal of Electrical and Computer Engineering*, vol. 2013, pp. 1–16, Mar. 2013, doi: https://doi.org/10.1155/2013/718574.

[12] W. Zhang, T. D. Abhayapala, T. Betlehem, and F. M. Fazi, "Analysis and control of multi-zone sound field reproduction using modal-domain approach," *The Journal of the Acoustical Society of America*, vol. 140, no. 3, pp. 2134–2144, Sep. 2016, doi: https://doi.org/10.1121/1.4963084.

[13] Y. Song, W. Tu, R. Hu, X. Wang, W. Chen, and C. Yang, "An efficient method using the parameterized hrtfs for 3D audio real-time rendering on mobile devices," *Advances in Multimedia Information Processing – PCM 2017*, pp. 663–673, May 2018, doi: 10.1007/978-3-319-77383-4_65.

[14] G. Buis, T. Goela, L. Pyla, and P. Zeilstra, "Augmented Reality And Spatial Audio For Positioning And Navigation." University of Twente, Enschede, 04-Nov-2022.

[15] S. Spagnol, G. Wersényi, M. Bujacz, O. Bălan, M. Herrera Martínez, A. Moldoveanu, and

R. Unnthorsson, “Current use and future perspectives of spatial audio technologies in Electronic Travel AIDS,” *Wireless Communications and Mobile Computing*, vol. 2018, pp. 1–17, Mar. 2018, doi: https://doi.org/10.1155/2018/3918284.

[16] R. H. Y. So, N. M. Leung, J. Braasch, and K. L. Leung, “A low cost, non-individualized surround sound system based upon head related transfer functions: An Ergonomics Study and prototype development,” *Applied Ergonomics*, vol. 37, no. 6, pp. 695–707, Nov. 2006, doi: https://doi.org/10.1016/j.apergo.2006.01.001.

[17] M. G. Onofrei, R. Miccini, R. Unnthorsson, S. Serafin, and S. Spagnol, “3D ear shape as an estimator of HRTF notch frequency,” Jun. 2020, doi: 10.5281/zenodo.3898720.

[18] F. Brinkmann, A. Lindau, S. Weinzierl, G. Geissler, S. van de Par, M. Müller-Trapet, R. Opdam, and M. Vorländer, “The Fabian head-related transfer function data base,” *DepositOnce*, 09-Feb-2017. [Online]. Available: https://depositonce.tu-berlin.de/items/3b423df7-a764-4ce1-9065-4e6034bba759. [Accessed: 04-Apr-2023].

[19] A. J. Kolarik, B. C. Moore, P. Zahorik, S. Cirstea, and S. Pardhan, “Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss,” *Attention, Perception, & Psychophysics*, vol. 78, no. 2, pp. 373–395, Nov. 2015, doi: 10.3758/s13414-015-1015-1.

[20] J. A. Belloch, G. Ramos, J. M. Badia, and M. Cobos, “An efficient implementation of parallel parametric HRTF models for binaural sound synthesis in Mobile Multimedia,” *IEEE Access*, vol. 8, pp. 49562–49573, Mar. 2020, doi: 10.1109/ACCESS.2020.2979489.

[21] S. Berge, J. de Winter, and M. Hagenzieker, “Support systems for cyclists in automated traffic: A review and future outlook.” *SHAPE-IT: Supporting the Interaction of Humans and Automated Vehicles: Preparing for the Environment of Tomorrow,* Sep. 2022.

[22] N. Gang, S. Sibi, R. Michon, B. Mok, C. Chafe, and W. Ju, “Don't be alarmed,”

Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 237–246, Sep. 2018, doi: https://doi.org/10.1145/3239060.3265636.

[23] V.-T. Tran and W.-H. Tsai, “Audio-Vision Emergency Vehicle Detection,” *IEEE Sensors Journal*, vol. 21, no. 24, pp. 27905–27917, Nov. 2021.

[24] I. Kwan and J. Mapstone, “Interventions for increasing pedestrian and cyclist visibility for the prevention of death and injuries,” *Cochrane Database of Systematic Reviews*, vol. 2009, no. 4, 2006.

[25] F. Cuadrado, I. Lopez-Cobo, T. Mateos-Blanco, and A. Tajadura-Jiménez, “Arousing the sound: A field study on the emotional impact on children of arousing sound design and 3D audio spatialization in an audio story,” *Frontiers in Psychology*, vol. 11, May 2020, doi: https://doi.org/10.3389/fpsyg.2020.00737.

[26] R. Hoffmann, S. Spagnol, Á. Kristjánsson, and R. Unnthorsson, “Evaluation of an audio-haptic sensory substitution device for enhancing spatial awareness for the visually impaired,” *Optometry and Vision Science*, vol. 95, no. 9, pp. 757–765, Sep. 2018, doi: 10.1097/OPX.0000000000001284.

[27] D. M. Greenberg, E. Bodner, A. Shrira, and K. R. Fricke, “Decreasing stress through a spatial audio and immersive 3D environment: A pilot study with implications for clinical and medical settings,” *Music & Science*, vol. 4, Mar. 2021, doi: https://doi.org/10.1177/2059204321993992.

[28] L. Milam, E. Akins, H. Williams, B. Simpson, and H. G. Jones, “Spatial audio cueing AIDS pilot navigation during simulated helicopter flight in degraded visual environments,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, no. 1, pp. 1634–1638, Nov. 2019, doi: https://doi.org/10.1177/1071181319631498.

[29] D. Johnston, H. Egermann, and G. Kearney, “The use of binaural based spatial audio in the reduction of auditory hypersensitivity in autistic young people,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 19, p. 12474, Sep. 2022, doi:

10.3390/ijerph191912474.

[30] G. Marentakis, “Spatial Audio for multimodal location monitoring,” *Interacting with Computers*, vol. 33, no. 5, pp. 564–582, May 2022, doi: https://doi.org/10.1093/iwc/iwac009.

[31] X. Hu, A. Song, Z. Wei, and H. Zeng, “Stereopilot: A wearable target location system for blind and visually impaired using spatial audio rendering,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1621–1630, Jun. 2022, doi: 10.1109/TNSRE.2022.3182661.

[32] M. Reynal, J.-P. Imbert, P. Aricò, J. Toupillier, G. Borghini, and C. Hurter, “Audio focus: Interactive spatial sound coupled with haptics to improve sound source location in poor visibility,” *International Journal of Human-Computer Studies*, vol. 129, pp. 116–128, Sep. 2019, doi: https://doi.org/10.1016/j.ijhcs.2019.04.001.

[33] A. Al-Rahayfeh and M. Faezipour, “Eye Tracking and head movement detection: A state-of-art survey,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 1, pp. 2100212–2100212, Nov. 2013, 10.1109/JTEHM.2013.2289879.

[34] Xiao B, Georgiou P, Baucom B, Narayanan SS. Head motion modeling for Human Behavior Analysis in Dyadic interaction. IEEE Transactions on Multimedia. 2015May13;17(7):1107–19, 10.1109/TMM.2015.2432671.

[35] Hadar U, Steiner TJ, Grant EC, Rose FC. Kinematics of head movements accompanying speech during conversation. Human Movement Science. 1983;2(1-2):35–46, https://doi.org/10.1016/0167-9457(83)90004-0.

[36] Majdak P, Iwaya Y, Carpentier T, Nicol R, Parmentier M, Roginska A, et al. Spatially oriented format for acoustics: A Data Exchange Format representing head-related transfer functions [Internet]. NYU Scholars. 134th Convention of the Audio Engineering Society; 1970 [cited 2023Apr17]. Available from: https://nyuscholars.nyu.edu/en/publications/spatially-oriented-format-for-acoustics-a-datA-exchange-format-re.

[37] Cuevas-Rodríguez M, Picinali L, González-Toledo D, Garre C, de la Rubia-Cuestas E, Molina-Tanco L, et al. 3D tune-in toolkit: An open-source library for real-time binaural spatialisation. PLOS ONE. 2019Mar11;14(3), doi: https://doi.org/10.1371/journal.pone.0211899.

[38] F. Pausch, L. Aspöck, M. Vorländer, and J. Fels, “An Extended Binaural Real-Time Auralization System With an Interface to Research Hearing Aids for Experiments on Subjects With Hearing Loss.” https://journals.sagepub.com/, 21-Aug-2018, pdf: https://doi.org/10.1177/233121651880087.

[39] M. Tamulionis and A. Serackis, "Listener Movement Prediction based Realistic Real-Time Binaural Rendering," *2019 IEEE 7th IEEE Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*, Liepaja, Latvia, 2019, pp. 1-4, doi: 10.1109/AIEEE48629.2019.8977123.

[40] A. Mader and W. Eggink, “Creativity,” in *DS 78: Proceedings of the 16th International conference on Engineering and Product Design Education (E&PDE14), Design Education and Human Technology Relations, University of Twente, The Netherlands, 04-05.09.2014*, 2014, https://research.utwente.nl/en/publications/a-design-process-for-creative-technology.

[41] J. Halme, “16 brainstorming techniques for boosting innovation,” *info.orchidea.dev*, 13-Aug-2021. [Online]. Available: https://info.orchidea.dev/innovation-blog/brainstorming-techniques-for-boosting-innovation. [Accessed: 05-Apr-2023].

[42] V. Bruschi, S. Nobili and S. Cecchi, "Real Time Binaural Synthesis of Moving Sound Sources over Loudspeakers," 2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA), Bologna, Italy, 2021, pp. 1-9, doi: 10.1109/I3DA48870.2021.9610886.

[43] V. Bruschi, S. Nobili and S. Cecchi, "A Real-Time Implementation of a 3D Binaural System based on HRIRs Interpolation," 2021 12th International Symposium on Image

and Signal Processing and Analysis (ISPA), Zagreb, Croatia, 2021, pp. 103-108, doi: 10.1109/ISPA52656.2021.9552125.

[44] J. M. Arend, T. Lübeck, and C. Pörschmann, “Efficient binaural rendering of spherical microphone array data by linear filtering,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2021, no. 1, Nov. 2021, doi: https://doi.org/10.1186/s13636-021-00224-5.

[45] M. Geronazzo, J. Y. Tissieres and S. Serafin, "A Minimal Personalization of Dynamic Binaural Synthesis with Mixed Structural Modeling and Scattering Delay Networks," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 411-415, doi: 10.1109/ICASSP40776.2020.9053873.

[46] M. Daraban, K. Lengyel, A. Faluvegi and C. Amariei, "Real-time Spatialized Sound Generator on Embedded Wearable Platform," 2021 IEEE 27th International Symposium for Design and Technology in Electronic Packaging (SIITME), Timisoara, Romania, 2021, pp. 71-76, doi: 10.1109/SIITME53254.2021.9663671.

[47] S. GS, B. K. Acharya, B. Ali, D. S. P. and S. David S., "Real-Time Hardware Implementation of 3D Sound Synthesis," 2020 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Ha Long, Vietnam, 2020, pp. 232-235, doi: 10.1109/APCCAS50809.2020.9301667.

[48] “Netherlands: Road Fatalities, by mode of transport 2021,” *Statista*, 13-Apr-2023. [Online]. Available: https://www.statista.com/statistics/523252/netherlands-number-of-road-fatalities-by-mode-of-transportation/. [Accessed: 17-Apr-2023].

[49] E. Aggius-Vella, M. Gori, C. Campus, B. C. J. Moore, S. Pardhan, A. J. Kolarik, and N. Van der Stoep, “Auditory distance perception in front and rear space,” *Hearing Research*, vol. 417, p. 108468, Apr. 2022.

[50] Y. Soeta and S. Ooi, “Spatial audio quality factors for headphone evaluation,” *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, Nov. 2021, doi:

10.1109/I3DA48870.2021.9610844.

[51] J. Kanter, “RS recommends: The best headphones with spatial audio for immersive tunes,” *Rolling Stone*, 07-Mar-2023. [Online]. Available: https://www.rollingstone.com/product-recommendations/electronics/best-headphones-with-spatial-audio-1234593481/. [Accessed: 21-Apr-2023].

[52] A. Schiffer, “The best headphones with spatial audio features,” *Major HiFi*, 08-Jul-2022. [Online]. Available: https://majorhifi.com/the-best-headphones-with-spatial-audio-features/. [Accessed: 21-Apr-2023].

[53] T. Lee, “10 spatial audio wireless earbuds that deliver excellent surround sound,” *LyreBeats*, 21-Nov-2022. [Online]. Available: https://lyrebeats.com/10-spatial-audio-wireless-earbuds-deliver-excellent-surround-sound. [Accessed: 21-Apr-2023].

[54] J. Blake, “The best headphones for experiencing spatial audio,” *Acoustic Nature*, 08-Oct-2022. [Online]. Available: https://acousticnature.com/journal/best-headphones-spatial-audio. [Accessed: 21-Apr-2023].

[55] D. Folami, “5 best earbuds that support spatial audio in [year],” *Oscarmini*, 18-Dec-2021. [Online]. Available: https://oscarmini.com/best-earbuds-that-support-spatial-audio/. [Accessed: 21-Apr-2023].

[56] G. Brown, “The basics of convolution in Audio Production,” iZotope, https://www.izotope.com/en/learn/the-basics-of-convolution-in-audio-production.html (accessed Jul. 21, 2023).

[57] Q. A. Flint, “Discrete convolution visual,” Quincy’s Website, https://quincyaflint.weebly.com/academic-material/discrete-convolution (accessed Jul. 21, 2023).

[58] P. Mantione, “The basics of Impulse Responses & Convolution (+ plugins, software & hardware),” Pro Audio Files, https://theproaudiofiles.com/impulse-responses-and-convolution/ (accessed Jul. 21, 2023).

[59] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, “Extracting the frequencies of the pinna spectral notches in measured head related impulse responses,” *The Journal of the Acoustical Society of America*, vol. 118, no. 1, pp. 364–374, Jul. 2005. doi:10.1121/1.1923368.

[60] “EQ Window,” EQ window, https://www.roomeqwizard.com/help/help_en-GB/html/eqwindow.html (accessed Jul. 21, 2023).

[61] F. C. Tommasini *et al.*, “Individual head-related impulse response measurement system with 3D scanning of pinnae,” *Proceedings of Meetings on Acoustics*, vol. 28, no. 1, Sep. 2016. doi:10.1121/2.0000519.

[62] M. Southworth, “The sonic environment of cities,” *Environment and Behavior*, vol. 1, no. 1, pp. 49–70, Jun. 1969. doi:10.1177/001391656900100104.

[63] C. Goldenbeld, M. Houtenbos, E. Ehlers, and D. De Waard, “The use and risk of portable electronic devices while cycling among different age groups,” *Journal of Safety Research*, vol. 43, no. 1, pp. 1–8, Feb. 2012. doi:10.1016/j.jsr.2011.08.007.

[64] E. N. G. Verheijen and J. Jabben, “Effect of electric cars on traffic noise and safety.” RIVM, Bilthoven, 2010, https://www.rivm.nl/bibliotheek/rapporten/680300009.pdf.

[65] “Moscow prioritization,” What is MoSCoW Prioritization? | Overview of the MoSCoW Method, https://www.productplan.com/glossary/moscow-prioritization/ (accessed Jul. 21, 2023).

[66] “Home,” JUCE, https://juce.com/ (accessed Jul. 21, 2023).

[67] OpenAL, https://www.openal.org/ (accessed Jul. 21, 2023).

[68] M. Orlassino, M. Edenhauser, G. B. Moon, N. Boirot, and G. Sindjic, “Seeed studio Xiao NRF52840 sense - tinyml/tensorflow lite- IMU / microphone - bluetooth5.0,” Seeed Studio XIAO nRF52840 Sense - TinyML/TensorFlow Lite- IMU / Microphone - Bluetooth5
- Seeed Studio,
https://www.seeedstudio.com/Seeed-XIAO-BLE-Sense-nRF52840-p-5253.html
(accessed Jul. 21, 2023).

[69] A. Industries, “Flora Accelerometer/Compass Sensor - LSM303,” adafruit industries blog RSS, https://www.adafruit.com/product/1247 (accessed Jul. 21, 2023).

[70] “What is wine?,” WineHQ, https://www.winehq.org/ (accessed Jul. 21, 2023).

[71] “Arduino and C++ (for windows),” Arduino Playground - CPPWindows, https://playground.arduino.cc/Interfacing/CPPWindows/ (accessed Jul. 21, 2023).

[72] “Anaglyph, 3D to your ears,” Anaglyph, http://anaglyph.dalembert.upmc.fr/ (accessed Jul. 21, 2023).

[73] Vst, “Blue Lab Audio - Spatializer,” FREE VST PLUGINS, https://freevstplugins.net/blue-lab-audio-spatializer/ (accessed Jul. 21, 2023).

[74] Dear Reality, “DearVR micro,” Dear Reality, https://www.dear-reality.com/products/dearvr-micro (accessed Jul. 21, 2023).

[75] N. Fonseca, “Air: Music edition,” Sound Particles, https://soundparticles.com/products/air-music-edition (accessed Jul. 21, 2023).

[76] “What is a DAW and what can you do with it?,” Creativity First - Our Passion for Music Inspires, https://www.steinberg.net/tutorials/what-is-a-daw/ (accessed Jul. 21, 2023).

[77] “Work out how likely an assessment result could have been achieved by chance ,”

Commtap,
https://en.commtap.org/additional-resource/work-out-how-likely-assessment-result-could-have-been-achieved-chance (accessed Jul. 21, 2023).

[78] J. Miller and R. Ulrich, “The quest for an optimal alpha,” *PLOS ONE*, vol. 14, no. 1, Jan. 2019. doi:10.1371/journal.pone.0208631

[79] “Test statistic,” Wikipedia, https://en.wikipedia.org/wiki/Test_statistic (accessed Jul. 21, 2023).

[80] “https://www.sjsu.edu/people/saul.cohn/courses/stats/s0/BinomialProbabTable.pdf.” San José State University, San José,
https://www.sjsu.edu/people/saul.cohn/courses/stats/s0/BinomialProbabTable.pdf.