Sound Swarm

Experience sound from the inside.

Graduation report – Creative Technology
Wouter Westerdijk, s1458272
Supervisor: dr. ir. Edwin Dertien

University of Twente, July 2017

Abstract

This research provides explorative design suggestions concerning the experience of the potential art installation of *Sound Swarm*. The research uses scientific literature on human auditory localisation and perception in combination with practical experiments conducted in a virtual simulation. The results are findings in important variables concerning the installation, the types of potential experiences, and further design suggestions.

The important variables are: type of composition that is being used; method of dividing the composition over the sound sources; amount of sound sources; speed of the sound sources; movement behaviour and freedom of the sound sources; volume of the sound sources; and the size of the room in cubic meters. These variables influence the overall experience, which is divided into two categories: swarm-like experience, which is more focused on the swarm aspect as unique value point, and includes a composition with higher tempo, divided over frequency ranges with less speakers, random movement behaviour with quick changes and a higher speed; and music-like experience, which is more focused on a new, but pleasant method of listening to music. The composition is more complex, divided over musical aspects with more speakers that are denser around the audience and less dense close to the audience. The movement behaviour is focused on rotational motion with lower speed.

In general, to optimise the auditory experience, it is advised to have stationary subwoofers near the audience for frequencies below 200 Hz. Movement should be focused on rotational changes concerning the front, back, left and right of the audience. Fluctuation in elevation and distance should be big.

Furthermore, two design possibilities are provided that show potential based on literature and practical research. These designs are described on page 57.

Acknowledgement

Christine Maas – The subject of this research has been initiated by the Dutch artist Christine Maas. Thanks to her, the concept of the art installation of *Sound Swarm* exists.

Edwin Dertien – Supervisor of this research and representative of the University of Twente in relevance to this research. He especially aided with analysing what aspects to research and provided motivation when needed.

Research participants – They subjected themselves and enabled some of their time for the progression of this research.

Thanks to all the authors of the scientific literature that has been used in this research as a fundamental groundwork regarding most design decisions and experiment designs. Also thanks to the creators of the game engine Unity®, which has provided the option of building a virtual simulation.

Sound Swarm

Experience sound from the inside.

Contents

List of Figures

1. Introduction

The term 'swarm' is generally used to describe a group of multiple insects of the same species (e.g. bees) that together seem to form an autonomous cloudlike creature. Just like a swarm consists of individual insects, a song consists of multiple individual layers (e.g. instruments) as well. However, a song is commonly only experienced as a whole, without mindfulness of the individual layers.

Christine Maas is a Dutch artist who wants to develop an art installation to experience a song as if it were a cloud of individual sounds. She contacted the University of Twente to help her with a concept idea for this new art installation. This installation will be all about experiencing sound and music, that should be achieved by dividing the source music over multiple speakers so that each speaker contributes its own unique part. The installation will be referred to as *Sound Swarm* ('Geluidszwerm', the original name in Dutch).

The concept idea of *Sound Swarm* consists of a room with multiple speakers that can theoretically move randomly in every direction, filling the atmosphere. The audience will be placed in the middle of the room, so the speakers will be all around them. This should create the experience as if the audience is inside the song that the speakers are producing together. However, because the speakers move, different parts of the song will in their turn get closer to the audience's ears and thus draw more attention.

This report will focus on the design of the art installation. The design needs to be verified before the actual construction of *Sound Swarm*. Both theoretical research and practical experiments will be used to optimise design decisions. The practical experiments will be conducted in a virtual environment. The design should be optimised for providing the best and most intense experience of being inside a song and being able to hear the different song particles in their spatial location.

At the end of this report, several things must become clear. The amount, placement and movement freedom of the speakers should be optimised to provide the best auditory spatial experience. Besides that, it must be clear what aspect of a music composition will be used to divide the composition over the individual speakers. These two requirements should provide enough knowledge to answer the main question of this project:

What is the best setup design of Sound Swarm *in order to provide the most exquisite experience of being inside a music composition?*

2. *Analysis*

The fundaments of the design of *Sound Swarm* are determined by the laws of physics, available technology and previously conducted research. The different design aspects of the art installation will be explored in 2.1. These aspects will contribute to understanding what can be researched and what needs to be tested with practical research. Aspects that can be examined using background research will be further explained in 2.2.

Problems that can be ran into during the design phase of *Sound Swarm* might also already have occurred to other (art) installations that have been designed in the past. The developers of those installations might have found intuitive solutions that can be useful to this project as well. For some problems, it might be the case that recent technology can also provide the solution. The state-of-the-art, including similar installations and relevant new technologies, will therefore be covered in this report in 2.3.

2.1. Design Aspects

There are many things to keep in mind when designing this installation. To keep a clearer overview, the project can be divided into seven different design aspects. Due to possible time constraints, the first five design aspect will have priority.

2.1.1. Visuals

The most important part of the *Sound Swarm* is the ability of the human brain to locate sound sources. Even though the brain is able to determine a sound source quite adequate, especially on the horizontal plane, localisation of sound sources is significantly improved when there is also visual input [1]. Adding visuals, such as LEDs, to the (active) speakers would thus increase the ability of the users to locate the sound sources and therefore potentially increase the experience.

2.1.2. Dispersion of frequencies

The speakers will create soundwaves. These soundwaves will each have their own amplitude, frequencies and dispersion. A big hollow room will also create echoes; unwanted second-hand sound sources that can influence the experience (negatively). However, echoes always take longer to reach the ears than direct sound does. The human brain also takes this into account, thus echoes barely influence the accuracy of auditory localisation [2]. Besides that, it is important to keep in mind the sound dispersion and the amplitude. Too narrow dispersions in the wrong direction or too small amplitudes may result in the user unable to hear the speaker and missing a part of the composition.

2.1.3. Movement of sources

The speakers will move around randomly or a specific order and routine. The concept idea states that the speakers would all be moving randomly in all axes, but realistically that could result in collisions with each other or with the audience. In the concept idea, the speakers are able to float, but that would be unrealistic. Other things that should be kept in mind is the minimum and maximum distance that each speaker should be able to travel, and in which directions. With moving sound sources, the Doppler effect should also be kept in mind, so the speed of movement cannot be too high [3].

2.1.4. Division of original audio

The idea is to divide a (music) composition over multiple speakers. There are multiple ways to divide music. It can be divided by instruments, frequency ranges, or a combination of the two. More defiant ways could be by melody, harmony (chords/arpeggios), rhythm, or timbre. Another way is including more than one composition (probably preferably the same *BPM*[1]) so there would be more than one song playing simultaneously, either in the same scale or not.

2.1.5. Grouping of sources

Not all frequencies can be localised with equal brain computation effort [4]: for the horizontal plane, sound sources with frequencies lower than 200 Hz cannot be located precisely; frequencies lower than 80 Hz are impossible to localize. It should therefore be considered to have different speakers for different frequency ranges. Since localisation is not precise for frequencies lower than 200 Hz, a design choice would be to have stationary subwoofers to produce the low frequencies. The human ability of auditory localisation will be explored more detailed in section 2.2.

Another design choice is the amount of speakers and whether or not to combine them into multiple small groups. The speakers within a group will flock together. The groups will be based on similar sound aspects. The amount of speakers will highly depend on the sound quality, frequency range and sound dispersion of the speakers.

2.1.6. Distribution of audio files and corresponding speakers

In order to make the art installation work properly, the music composition needs to be divided properly and every part must reach its corresponding speaker. In case the composition would be broken up into different small frequency ranges, it would be possible to write a program that uses a bandpass filter and render every chosen frequency range separately. Other division choices would probably mean that the original audio samples need to be divided separately.

2.1.7. Implementation and controlling of floating speakers

As described in 2.1.3, it is highly unlikely that the speakers would actually float. A solution would be to hang every speaker on a small wire from the ceiling. To do so requires a grid pattern to provide as much freedom in the horizontal plane as possible. Each speaker would need a motor to shorten or lengthen the wire so the speakers can move up and down. That would mean that every speaker needs to be controlled individually, which would require a lot of electronics.

2.2. Human auditory system

As can be concluded from 2.1, the physics of soundwaves play an important role in the design of *Sound Swarm*. Just as important is the way humans would perceive these soundwaves, since the perception of sound is the key to the degree visitors would experience the art installation. The research conducted in this section is largely based on a research review that has been done regarding *Sound Swarm* as well [5].

[1] BPM: *beats per minute*: musical term to describe the speed of a song.

2.2.1. Auditory localisation

The human ability to localise auditory input is critical for *Sound Swarm* in order to create the best experience. Any auditory input given by the speakers that does not contribute to- or tricks the human auditory localisation is wasted, as it most likely will not provide improvement to the experience. This means that the design of *Sound Swarm* will be shaped largely by the human ability to localise sound sources.

The studies about the human ability of localising sound sources used in this research divide the three-dimensional space into two perpendicular two-dimensional planes: the horizontal plane and the vertical plane. In Figure 1 this is visualized for better understanding. The horizontal plane includes everything to the left and the right of the listener's head, whereas the vertical plane includes everything above and beneath. The two planes meet directly in the front and the back of the listener's head. The human brain uses different techniques to calculate the source location in both planes, this will be taken into account in the design decisions.

As accurate auditory localisation would presumably lead to a more intense experience in terms of the art installation, it may be best to analyse the following findings for soft design constraints or requirements.

Estimating the location of a sound source requires three parameters: the angles relative to the head on both the horizontal and vertical plane, and the actual distance between the observer and the source. However, findings suggest that distance and height localisation are less accurate than horizontal localisation in terms of rotation around the head [6 - 8]. To better understand the human ability of localising sound sources, the techniques relevant to the horizontal and vertical planes will be explained more thoroughly in the next two subsections.

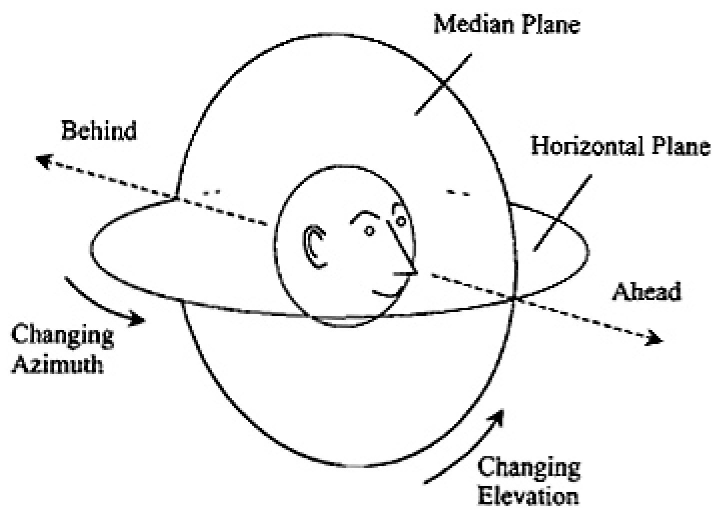


Figure 1: Visualization of the two planes in auditory localisation. The median plane is often referred to as vertical plane. Figure is from Plack's *The Sense of Hearing: Second Edition* **[2]**.

2.2.1.1. Auditory localisation on the horizontal plane

The human brain uses the differences in amplitude and phase of the soundwaves that are perceived by both ears to determine the position of the sound source on the horizontal plane [9, 10]. These differences are called interaural level differences (ILDs) and interaural phase differences (IPDs) respectively, and the basis of this theory was founded by Lord Rayleigh in 1907 [11]. This theory is often referred to as the 'duplex' theory and it suggests that ILDs are mainly used for high frequencies and IPDs for low frequencies regarding the auditory localisation on the horizontal plane. For sinusoids in the range of 1.5 kHz to 3 kHz, localisation performance is worst, the frequencies are believed to be too high for usable IPDs and too low for usable ILDs [12].

An ILD occurs when a sound source from one side would be shadowed by the listener's head. The ear on the far side would perceive a smaller amplitude than the ear on the short side. Since low frequencies can more easily diffract around the listener's head, ILDs are unreliable below 500 Hz [2]. According to Moore [13], the ILD for pure tones played directly to the side of the head may be less than 1 dB at 200 Hz and 20 dB at 6 kHz. Humans require an ILD of at least 1-2 dB in order to detect it [14]. This means that frequencies lower than 200 Hz cannot accurately be localized using ILD.

2.2.1.2. *Auditory localisation on the vertical plane*

To locate a sound source on the vertical plane, humans require the outer ear, specifically the pinna [15, 16]. In a paper by Makous and Middlebrooks [8] many studies about how vertical localisation exactly works are being discussed, but there is no clear conclusion. However, most theories discussed are based on Batteau's model [15] and "interpret the pinna as producing changes in the *spectrum* of the sound source that reaches the tympanic membrane" [8]. This is because Batteau suggests that the convolutions of the pinna would create echoes lasting only a few microseconds. These echoes would provide the brain information to calculate the source location.

2.2.2. *Spatial perception*

Speakers that produce higher frequencies will most likely perceived to be higher in the air than speakers that produce lower frequencies. The illusion of tones occupying a spatial location in correlation to their 'height' of pitch can be used to strengthen the localisation on the vertical plane [16, 17]. This can be combined with the finding of Roffler and Butler [16] that including frequencies of above 7 kHz would increase accuracy in vertical localisation.

2.2.2.1. *Auditory distance estimation*

For humans in order to determine the distance between themselves and the sound source, a reference sound or experience is needed. The human ability to determine the distance between itself and the sound source is described by Middlebrooks and Green [8] as "certainly not very good". When a listener would try to localise an unreferenced or new sound, additional information is needed in order to determine the distance, as the sound could be quiet because of the big distance, or simply because of the low volume of the sound source (or vice versa). Despite the true distance, humans tend to perceive soft sounds as being far away and loud sounds as being nearby [2]. Zahorik [18] stated that in rooms with reflective walls or other reverberant environments, the human ability to determine the distance of a sound source is slightly increased. This is because the brain can use the reflected sound as a cue to calculate the distance of the original sound source. The direct sound and its reflections form a direct-to-reverberant ratio. Nonetheless, this provides only coarse information about the distance, as humans can only detect changes in distance greater than a factor of two using this cue alone [18].

2.2.2.2. *Visuals in auditory localisation*

Visual input and feedback have an important and significant influence of localisation, and vice versa. Locating sound sources while being blindfolded leads to less accuracy in localisation than with vision [19], meaning that localisation accuracy improves when the listener can also see the target [20, 21]. The visual influence becomes even clearer when the visual input is just an *illusion* of the sound source, a phenomenon that occurs while watching TV, for example. The sounds are spatially associated with the events on the screen, rather than the true sound source that is the speaker [22]. The opposite may also occur, when visual input is so blurred or unclear that visual localisation is impossible, sound may capture vision [23]. This finding suggests that both visual and auditory senses can influence each other, which may occur when either one supplies an inherent advantage in spatial processing. The human brain thus varies between the senses, based on their reliability and added value to localization [23 - 26].

2.3. State-of-the-Art

Recent technological progression may provide solutions or options to take different design approaches. These relevant technologies will be covered and evaluated for usefulness in 2.3.2. Relevant art installations that have already been built in the past will be examined in 2.3.1. For the sake of a realistic virtual design, the current development of virtual reality technologies will also be explored in 2.3.3, since it plays an important role in the overall design phase of *Sound Swarm*.

2.3.1. Similar installations

Dividing sound over different speakers as art installation has been done before. Different artists each put their own touch on such installations. This subsection will cover the artists and installations that show most resemblance with the concept idea of *Sound Swarm*.

Florian Hecker – Chimerization

In the art installation *Chimerization*, Hecker uses three speakers that hang from the ceiling [27]. The speakers are arranged in a triangular configuration, as shown in Figure 2. The speakers play a reading of an experimental libretto by Iranian philosopher Reza Negarestani. However, the reading is presented in English, German and Farsi and each language is played by one of the speakers. Hecker has more installations like Chimerization; stationary speakers from the ceiling that produce sound to experiment with the physiology and psychology of soundwaves and how the visitors perceive them.

Figure 2: *Chimerization* (2012) by Florian Hecker. Copyright the artist, courtesy Sadie Coles HQ, London. Picture from The Wire **[27]**.

Janet Cardiff – The Forty Part Motet

The installation consists of forty speakers that are strategically placed throughout a large room, as shown in Figure 3. Each speaker represents a voice from a choir. Eight different choirs sing a part of *Spem in Alium*, a complex song with forty different harmonies. Each single choir is based baritone, alto, tenor and soprano. After 11 minutes of singing, there are 3 minutes of human sounds from the choir, such as speaking and coughing. This should bring personality to the installation as well. The goal of *The Forty Part Motet* was to enable the audience to experience a piece of music from the viewpoint of the singers, rather than the traditional viewpoint in front of the choir [28].

Figure 3: *The Forty Part Motet* by Janet Cardiff.

Tristan Perich – Microtonal Wall

Perich's creation is an excellent example of how individual uniqueness combined can produce something that is perceived as an autonomous phenomenon on its own. The installation consists of 1,500 individual one-bit speakers that are varied continuum of pitch. Each speaker on its own sounds like a single tone, but all speakers combined would let the audience perceive a totality of white noise. As shown in Figure 4, the speakers are placed on a flat wall. This means that wherever the audience is listening, the perception of each speaker would never be equidistant, except for an infinite distance.

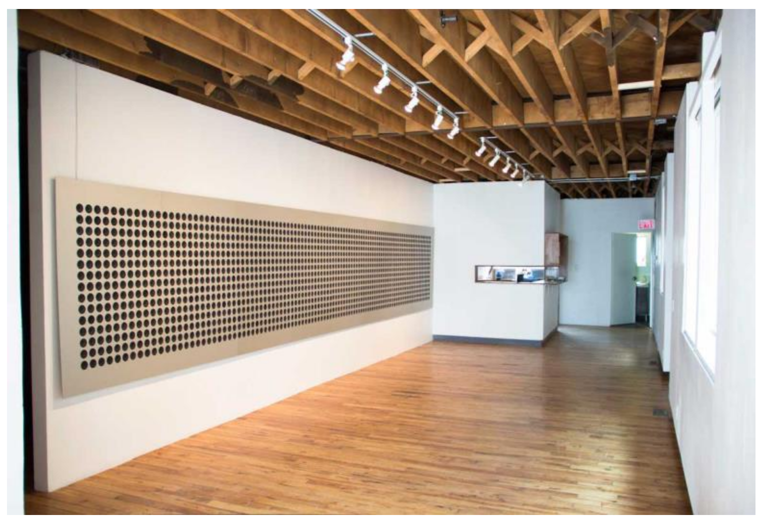


Figure 4: *Microtonal Wall* by Tristan Perich. Installation at InterAccess Gallery, Toronto, 2012. Courtesy the artist. Picture from MoMA.

Nye Parry – The Exploded Sound and Significant Birds

The Exploded Sound would be the existing art installation that shows most resemblance to the concept idea of *Sound Swarm*. It consists of sixty speakers suspended from the ceiling, each contributing a tiny proportion of the overall sound of the installation. The goal of the installation is, according to Parry [29], "exploring the sound of a choir or orchestra from the inside, not just walking among the instruments, but entering the very stuff of sound itself, the 'partials' that make up the complex musical sounds you hear." Parry went through a number of research phases prior to *The Exploded Sound* and *Significant Birds*, both at the Lansdown Centre of Electronic Arts at Middlesex University and at CRiSAP, University of Arts.

Figure 5: *The Exploded Sound* by Nye Parry in the Jacopic Gallery. Picture from Divergence Press **[30].**

Parry also wrote a paper about the technical and aesthetical aspects of both installations [30]. In this paper, he describes his process of new findings and how they were used and adapted to his installations. If problems may occur during the design phase of *Sound Swarm*, Parry's paper and the used references could bear the solutions.

Lucio Capece – "Epoché" Flying Speakers

The art work *Epoché* by Lucio Capece consists of three balloons each lifting a single speaker, as can be seen in Figure 6. Capece used a wide range of sound sources to represent the activity in the surroundings of human daily life. In a small part of the art piece he also used pure sine waves, with which he tries to represent and suggest the focus and suspension of the attention to the world's activity and its invitation to perceive things how people perceive them. Capece wrote on his blog that he considers "that we live at the moment in a society overloaded of information and that the greatest part of the music and arts works are focused on the artist activities and ideas, instead of the listener's experience. My work in terms of perception intends to go to how things start being what they are" [31]. In order to more easily catch the audience's attention, he made use of two of the oldest and most basic psychoacoustic techniques.

Figure 6: Live presentation of Epoché in London, January 2015. Photo by Andrej Truu.

In contrast to the previously mentioned installations Capece's *Epoché* contains speakers that are both moving and floating around in the room. This is similar to the desired speaker behaviour described in Maas' concept idea.

2.3.2. Technology

The available technology that is examined in this subsection does not necessarily need to be useful for the design of *Sound Swarm*. However, it is relevant to the project and may theoretically be an aid to different approaches.

Levitating Speaker – *levitating-speaker.com*

A round speaker design that float a few centimetres above its corresponding platform. The Levitating Speaker proves that a floating speaker is possible using magnetic forces. However, it is unlikely that this approach would be realistic when scaled up to multiple speakers covering several meters.

Ambisonics – *ambisonic.net*

A surround sound system that would overcome major problems of quadraphonic systems. Ambisonics contain a speaker-independent representation of a sound field called *B-format*. This can be decoded to the listener's speaker setup. Ambisonics are said to provide a solid technical foundation and many advantages over normal surround sound systems, but it has not (yet) been a commercial success. Ambisonics may be used to create virtual sound sources in the room instead of having to place a physical speaker in that spatial position. This would solve problems stated in 2.1.7, but it would also cause the lack of visual localisation, which might be essential for an improved spatial auditory localisation experience, as explained in 2.2.2.2.

2.3.3. Virtual Reality

The important factor of recent developments within virtual reality is the quality of accurate spatial sound representation.

Omnitone – *popsci.com/google-gives-new-spatial-vr-audio-omnitone*

A new platform by Google that allows the listener to perceive virtual spatial sound sources. Google called it "a key element for an immersive virtual reality experience".

Google VR (Spatial Audio) – *developers.google.com/vr*

The Spatial Audio is a tool that is part of the Google virtual reality software development kit. It uses the main audio cues that is used by the human brain to localise sounds. It is a mainly passive plugin containing algorithms to calculate auditory spaciousness more thoroughly.

Oculus Virtual Reality Audio – *developer.oculus.com/documentation/audiosdk/latest*

Seemingly similar to the previously mentioned tool by Google.

3. Methods and Techniques

In 3.1.2, Maas' concept idea will be explained more thoroughly. To make the definition of optimal auditory experience more concrete and objective, the human perception will be left out in the theoretical analysis. Instead, it will be assumed that the optimal experience can be achieved if the installation is designed in such a way that the human auditory localisation is able to distinguish all the separate sound sources, as explained in the first paragraph of the literature research in 2.2.

The methods and techniques phase roughly follows the design process introduced by Mader and Eggink that consists of three phases: ideation, specification and realisation [32]. Since the design process is optimised for ICT-based designs to solve problems in the daily life or niche markets, and not for designing an art installation with a predetermined problem, the research will slightly deviate from this design process in the subsections. In Ideation (3.1) Maas' concept idea will be examined to find the key research questions, which will be used to prepare for actual experiments. A way to design setups for conducting experiments will be more thoroughly explained and discussed in Specification (3.2). In Realisation (3.3) the design setup(s) from the specification-phase will be applied to find answers to the key questions found in the ideation-phase.

3.1. Ideation

Before being able to design any kind of research experiments or prototypes, a thorough analysis of the problem needs to be conducted in order to find the requirements and constraints of the art installation *Sound Swarm* and factors that might influence both the research and installation. The stakeholders and their potential influences will be analysed in 3.1.1. Maas' concept idea will then be analysed. This concept idea already forms the fundament of a solution to a deeper problem. To increase the value and effectiveness of this research, the underlying problem needs to be understood first. Analysing Maas' concept idea might bring the underlying problem to the surface. This is important, because the problem is the reason for Maas to initiate her solution in the form of her concept idea. However, it might be that this solution does not cover all the requirements to fully provide a solution. In other words, the concept idea on its own might only provide the scope of the art installation, rather than the objective, which can be found by dissecting the concept idea and examining the underlying problem of the solution. The information found in 3.1.2 will form the basis of the research experiments and design, which will be explained in section 3.3.

3.1.1. Stakeholders

This subsection will examine two kinds of stakeholders: the stakeholders of the research and its results and the stakeholders of the actual art installation. Both groups of stakeholders are important to keep in mind during the research process. The stakeholders of the research results are considered the direct stakeholders of this report. Next to that, the potential user groups are also examined.

3.1.1.1. Research stakeholders

Christine Maas has contacted the University of Twente to ask for help with designing her concept idea of *Sound Swarm*. Since the University of Twente is involved as middleperson, it will benefit from a positive outcome. This makes the university a stakeholder, even though it is not directly related to the art installation itself. A negative outcome may have a negative influence on the image of the university.

Maas is considered the most important stakeholder regarding the research about her concept idea of *Sound Swarm*. Results of this research on its design will most likely influence the final and physical construction. The research is meant to provide a solid fundament to design and build the art installation on.

3.1.1.2. Art installation stakeholders

As the concept idea of the art installation is initiated by Christine Maas, she is considered to be the most important stakeholder for the installation itself as well. For realising the art installation, Maas might request third-parties to back her with resources. Any third-party involved will automatically become a stakeholder. A third-party stakeholder can back the installation with resources like materials (such as speakers) or money. In return, the third-party may receive more exposure towards potential customers that are visiting the art installation, or a percentage of the entrance fee.

It is debatable whether the user groups (visitors) are stakeholders as well. With products that can enhance the life of its users, these users can be considered stakeholders since they profit from a positive outcome (successful outcome) and, most importantly, unsuccessful outcome is a setback for the potential user groups as well. With an art installation such as *Sound Swarm*, visitors may profit from a successful outcome, but an unsuccessful outcome does not have to be a setback for the potential visitors, as they might not know what they are missing. Therefore, the user groups (visitors) are considered partial stakeholders.

3.1.1.3. Potential user groups

The art installation *Sound Swarm* is a project initiated by the Dutch artist Christine Maas for an exhibition for *Nieuw Dakota*, a wharf designed for temporary art exhibitions in Amsterdam-Noord. On the website, *Nieuw Dakota* describes itself as an initiative to improve international art and culture and support new talents, art expressions and products such as extraordinaire expositions and new insights concerning international arts and culture [33]. The content on the website also shows prove that the organisation aims for innovative and educational art. This would mean that the involvement of technology is almost inevitable.

Such exhibitions generally attract more open-minded visitors that are interested in technology and expressions. Most of the time the audience would be young adults or middle-aged people who want to keep up with today's innovations. As the concept idea of the *Sound Swarm* includes an innovative perspective on experiencing music set up in a way such that it also includes education and expression, it would be likely to assume that the described audience would be the most potential user group.

3.1.2. Concept idea

The concept idea of *Sound Swarm* as initiated by Christine Maas is, simply put, to have a large room with multiple floating speakers that independently and randomly move across the room. The idea also states that *Sound Swarm* will move around between multiple exhibitions, meaning that the room cannot be considered a permanent component. The speakers in the room all produce a unique part of a single sound composition such as a music track. Combining all the sound particles that the speakers produce would recreate the original composition. In the middle of this room the audience is located. Speakers producing their own unique sound particles will randomly pass the audience. The speakers that are closer to the audience will be in the foreground, dominating the perception of the song. This leads to the goal of the art installation: a unique perception of music. However, this perception has not been experienced yet and thus the goal of the art installation cannot be determined. So, rather than trying to interpret this idea as the fundamental basis of the construction of *Sound Swarm*, it may be better to rewrite it into a single question:

What kind of listening experience will be created when a single sound composition is divided into multiple unique sound particles that independently and randomly move within a large room?

This question exposes the purpose of the art installation and its underlying problem: the experience of perceiving sound using multiple sound sources that each produce a unique particle of that sound while randomly moving around a room is unknown, thus the art installation *Sound Swarm* should be constructed to enable the existence of the experience.

3.1.2.1. Requirements

A qualitative analysis of the concept idea can provide more information about the requirements and constraints of the design of *Sound Swarm*. A requirement analysis technique that is mostly used in business managements in order to prioritise certain requirements is called the MoSCoW analysis [34]. This analysis can help prioritising certain features that a product *must have*, *should have*, *could have* and *won't have* with its corresponding priority in descending order. This tool can (partly) be used to dissect and find out the priority of features concerning the final experience of *Sound Swarm*, as is done in Table 1. The analysis will not consider *won't have*s as this part would include requirements that will not be in the project or product, but are considered extra features for the future, which is irrelevant to this research. More importantly, since a MoSCoW analysis is mainly used to identify the priority of requirements rather than the requirements themselves, the requirements need to be predetermined in order to determine which of those are *won't have*s. In Table 1 the MoSCoW analysis technique is used to actually determine the requirements based on the underlying problem of Maas' concept idea.

Must	In order to be able to call the experience of *Sound Swarm* a success, the experience must include…
	• An auditory perception of depth; • An auditory perception of movement; • Soundwaves with multiple frequencies within the range of what is on average perceived as 'low frequencies' and 'high frequencies'.
Should	High-priority features that should be included in the experience are…
	• A visual perception of chaos, preferably similar to the visual perception of a swarm; • A positive perception on the auditory level (i.e. frequency combinations that are perceived as pleasant to listen to); • Soundwaves with multiple frequencies within the range of the average human ability to perceive (approx. 20 Hz – 20,000 Hz).
Could	If time and resources permit, it is desired that the experience also includes…
	• A visual stimulation of the auditory perception.

Table 1: Experience requirements using the MoSCoW analysis technique.

With the experience requirements prioritised, it has become clearer what the actual art installation should cover. As is described in the previous section, the only required physical components are a large room and multiple speakers. These components, however, require variables that are not well-defined by the concept idea. The components and their variables and specifications are listed in Table 2. None of the variables have specific predefined values. Some qualitative information provided in the concept idea can be used to at least start in the right direction regarding some variables.

Component	**Variable**	**Specification**
Room	Dimensions (height, depth, width)	Not specified, but realistically large
Room	Shape	Not specified, but presumably bar-shaped with the height being significantly smaller than the depth and width
Room	Materials	Not specified
Speaker	Amount	Not specified, but at least 2
Speaker	Speed	Not specified, but presumably within the range of: not moving < *speed* < noticeable Doppler effect
Speaker	Volume	Not specified
Speaker	Source composition	Not specified, but focus on music
Speaker	Division of sound	Not specified

Table 2: Given components with their variables and specifications.

3.1.2.2. Constraints

One of the biggest constraints, already addressed in 2.1.3 and 2.1.7, is the fact that the concept idea states that the speakers are able to randomly float in every direction throughout the whole room. To try to realise this both realistically and artistically is enough work to be a whole research on its own. Maas therefore explicitly said that this research should not focus on that constraint, but rather focus on the possibilities of the art installation if it were possible to have floating and randomly moving speakers.

The movement of the speakers will most likely not be totally random. This is because the placement of the audience has to be kept in mind: in order not to hit anyone with a floating speaker, the speakers must either avoid the space where the audience is placed, or have enough elevation to move over the heads of the audience without the possibility of hitting anyone. The latter will most presumably lead to a more elaborate auditory experience, since the sound particles will then also come from above.

The audience will be placed in the middle of the room. It is yet unknown of how many people the audience is going to consist. Most likely, it will consist of an amount that can be defined by common understanding as 'small group'. *Sound Swarm* will not continuously operate, instead Maas wants to define a beginning and ending. During the performance, the audience will be seated and the speakers will float throughout the room. Before and after, the speakers should be static, so the audience can either enter or leave the room without having to watch out for flying speakers.

3.1.3. Knowledge to be acquired

As discussed in 3.1.2.1, none of the variables to be researched have predefined quantitative values. It would be nearly impossible to test every variable on its own. More important is to analyse the variables for correlations. All the variables together form the overall experience of *Sound Swarm*, meaning that if an optimal experience exists, then there should be at least one configuration with all variables having the optimal value. Analysing these values may lead to discovering certain correlations between variables, which could make the optimisation of the experience more agile and therefore less challenging to research. Based on 3.1.2, the most important variables will be analysed in a theoretical environment to try to find their influencing factors and correlations. These variables are listed in Table 3, together with a corresponding critical question. Correct answers to these questions should lead to the optimal experience, disregarding some minor variables.

Some of these variables have already been roughly examined in the design aspects of 2.1. Due to time constraints and the risk of complications during the research, the sound sources will be assumed to be omnidirectional. This means that the dispersion of sound, mentioned in 2.1.2, as well as the impedance of speakers, will not be taken into account. The amplitude of the soundwaves will be referred to as the volume. The movement of sources in 2.1.3 will be divided into two individual variables: speed and movement. The latter will include the movement freedom and behaviour. Also, the amount of speakers and grouping of speakers, both mentioned in 2.1.5, will be considered to be individual variables.

Variable	*Critical question*
Amount	How many speakers?
Speed	How fast do the speakers need to go?
Movement	How much should the speakers move?
Grouping	(How) should the speakers be grouped?
Division	How should the source composition be divided over the speakers?
Volume	How much decibel Sound Power Level should each speaker produce?
Room	What should be the dimensions of the room?

Table 3: Important variables and their critical questions.

3.1.3.1. Variables analysis

The relevant variables that are derived from Maas' concept design, should be analysed to better understand possible correlations or potential value ranges.

Room size, amount and movement

A hypothesis is that there is a certain amount of speakers to create the optimal auditory experience. However, the amount seems to also be dependent on the size of the room. Take N as the optimal amount of speakers for a room with a volume of X m^3. This means that if the speakers were to be equally divided over the room, each speaker would have its own space of movement of X/N m^3. Say the outcome would be $X/N = M$ m^3. Now the question is whether the amount N or the movement space M is the decisive factor concerning the optimal auditory experience. The answer depends on whether the size of the room X really matters. This can be determined by the volume of the sound source(s). If the room size were to be infinite, at some point all sound sources with a rational volume will be inaudible to an observer in the middle of the room. Anything outside the audible range is useless regarding the auditory experience, which means that the room should end at that point, thus the room has a rational maximum size. However, Maas states in her concept idea that *Sound Swarm* is meant to travel between multiple exhibitions, focusing on *Nieuw Dakota* first. Where exactly the art installation will end up is unclear, so the dimensions of the room cannot be foreseen.

Volume

Realistically, the volume of a single speaker should be high enough so the sound source at the end of the room is just detectable (V_{min}) and low enough so visitors will not find it unpleasant when the sound source is close to their head (V_{max}). However, due to the built of the human ear, sensitivity to perception differs per frequency range. A soundwave with a frequency of 100 Hz will only be audible at 40 dB or louder, while a frequency around 3 kHz will still be audible at 0 dB [35]. If the original sound would be divided in frequency ranges, then this information should be kept in mind while designing the installation. For this research, all speakers will have the same volume. This means that the optimal volume V of a single speaker should be between V_{min} and V_{max}.

The maximum volume V_{max} can be derived from looking at the table provided in the book *The Scientist and Engineer's Guide to Digital Signal Processing* [35]. The table can be found in Appendix A. This table states that a normal (human) conversation is around 60 dB SPL[2]. The average rock concert is rated at 110 dB, while the OSHA[3] limit for industrial noise is set at 90 dB. Considering the potential user groups (see 3.1.1.3) and their motivation to visit an installation like *Sound Swarm*, it will be likely to set V_{max} for a single speaker at around 70 to 80 dB.

The minimal volume V_{min} is partially dependent on the room size. The bigger the room, the higher the volume, so sound sources at the end of the room are still just audible. According to the *Inverse Square Law*, the sound intensity will decrease close to 6 dB per doubling of the distance [36]. If a sound source with a minimum distance to the audience at 1 meter has a sound pressure level of 80 dB, then at 100 meters the audience perceives a sound pressure level of 40 dB, which is the minimum loudness for 100 Hz to still be just audible [35]. This means that with this setup the room size can be 200x200x100m.

The amount of speakers also influences the volume of all speakers combined, V_{total}. If the installation would contain one speaker, then $V_{max} = V_{total}$. For an increase in speakers, given that all speakers produce the same volume, the formula to calculate the total volume V_{total} in dB is as follows:

$$\Delta L = 10 \cdot \log(n) \text{ or } n = 10^{(\Delta L/10)}$$

This formula shows a 3 dB increase per doubling of the sound sources, meaning that an increase in the amount of speakers should decrease V_{max} of all individual speakers in order to keep V_{total} equal to V_{max} for one speaker.

Speed

For research purposes, the speed will be a constant variable that is applied to all speakers in the design. Visitors should get enough time to be able to focus on sound sources, yet at the same time the sources should also have a certain velocity such that the installation will remain interesting and dynamic. Moving sound sources, as with *Sound Swarm*, will inevitably be affected by the Doppler effect [37, 38]. According to its formula:

---

[2] dB SPL: Decibels Sound Power Level, a common logarithmic scale to express volume. 10^{-16} Watts/cm^2 equals 0 dB SPL.

[3] OSHA: Occupation Safety and Health Administration, agency of the United States Department of Labor.

$$f = \left(\frac{c + v_r}{c + v_s}\right) f_0$$

where f_0 is the actual frequency, c the velocity of the waves in the medium, v_r and v_s are the velocities of the receiver and source respectively; changes in frequencies will already occur when $v_r \mid v_s > 0$. This means that the slightest movement will already result in a change of observed frequencies.

Common sense implies that small changes in velocity do not noticeably change the observed sound-waves. For example, running randomly around a stationary speaker setup that is playing music, does not (noticeably) change the pitch of the song. However, sirens on an ambulance driving towards an accident are perceived higher in pitch when the ambulance is driving towards the observer, and slides down in pitch when the ambulance has passed.

According to Doppler, a relative velocity difference between the observer and source of 40 m/s will change the perception of a note C to D, if the observer and source are moving towards each other [37]. A relative difference of 5.4 m/s is described by Doppler as the threshold for observers with absolute hearing. Velocity differences higher than 5.4 m/s will change the observed frequencies of a composition such that an observer with absolute hearing is able to notice a change in pitch [37]. However, untrained ears will most likely not notice this difference, since the frequency changes are proportional. A composition affected by the Doppler effect will not sound off by itself, only when compared to non-affected tones.

Division of sound

The idea of *Sound Swarm* is that a (music) composition will be divided over several sound sources that each produce their own unique sound part. In 2.1.4, most possible methods of dividing the composition are already listed. It will depend on Maas and her choice of composition which method will have the preference. Most likely, the sound will be divided into instruments or frequency ranges, or a combination of the two. The division of sound is highly dependent on the composition to be divided.

As described in 2.2, certain frequencies or frequency ranges may have different characteristics concerning spatial location. These characteristics should be kept in mind while researching the design of *Sound Swarm*, especially when the choice would be to divide the composition in certain frequency ranges. The biggest advantage of using frequency ranges as division method is that there are practically infinite possibilities to divide the composition, and it is applicable to any kind of sound source.

Grouping

While the (music) composition can be divided into equally important or present sound particles, there is also the possibility to subdivide the composition into certain groups. The sound sources of a group may then flock together. This could be done by, for example, grouping the sources per instrument and then dividing the instruments into frequency ranges. Grouping is not mentioned in Maas' concept idea, but it may create more depth in the installation.

3.1.3.2. Correlations analysis

To better understand Maas' concept idea, a theoretical simulation will be imagined and written down. This simulation will not contain any numerical values, merely ideas of what kind of values each variable would contain and how these values would correlate to each other. During the theoretical simulation, it needs to be clear which variables are more likely to be affected by other variables, and which variables will only affect others.

Since the room will not be a permanent component of *Sound Swarm*, the installation might need to adapt to different room sizes. This means that *room size* is a variable that is not affected by other variables. For this simulation, the room size will change. Since the analysis is only about correlations, it does not matter whether the room size increases or decreases. To keep the perception of the installation consistent, the variable most likely to be affected will be the *amount*. If the *amount* does not change accordingly, then the *movement* will, providing either more or less movement freedom per speaker. This will most likely change the perception of *Sound Swarm*.

However, the amount seems to not only be affected by the room size, but also the *division of sound*. This holds true if the installation can consist of only unique sound particles. Since it is mentioned in Maas' concept idea that the sound sources should be unique, the amount of sources will also be highly dependent on the method of dividing the original composition. On the other hand, it must be said that it will also most likely be possible to adapt the division method to divide the composition into a specific amount of unique sound sources, especially if the division of sound will be based on frequency ranges.

It is the question whether the exact amount of sound sources is more important than the method of sound division, assuming that there can only be unique sound sources. By changing the amount of speakers, the composition needs to be divided differently, so all sources remain unique and no sound particle is leftover. The other way round, by changing the division of sound, the amount of sound sources will most likely also change, meaning there is a different amount of speakers needed. It will thus come down to which of the two variables is considered more important. This variable will hugely affect the other. For the sake of this analysis, both variables are considered equally important to each other.

An overview of how variables will most likely interact in certain scenarios is given in Table 4. All variables given in the table are assumed to have values that combined provide the perfect experience of *Sound Swarm*, theoretically. By changing one variable, either increasing or decreasing, other variables should change correspondingly in order to keep the perfect experience. The table is subject to the presumption that there can only be unique sound particles, and that an increase in the division of sound will lead to more subdivisions (groups) instead of adding more sound sources to already existing groups. Without presuming the latter, there would be no clearance on the correlation between grouping and other variables.

Room size	**Amount**	**Div. of sound**	**Movement**	**Volume**	**Grouping**
INCREASE	INCREASE	INCREASE	-	*DECREASE*	INCREASE
INCREASE	-	-	INCREASE	-	-
-	INCREASE	INCREASE	*DECREASE*	*DECREASE*	INCREASE
-	*DECREASE*	*DECREASE*	INCREASE	INCREASE	*DECREASE*

Table 4: Examples of how the most important variables may correspond to each other.

The amount of *movement* is a direct result of the correlation between room size and amount of sources, according to the hypothesis of $X/M = N$ m^3, as discussed in 3.1.3.1. The change of *volume* is based on keeping the V_{total} equal in all scenarios. That means that with an increase of the amount of sources, the volume of each individual speaker needs to decrease, and vice versa. Note that *speed* is not mentioned in Table 4, as there seems to be no correlation between the speed of the speakers and other variables.

In case of *grouping*, the total volume of individual speakers flocking in one group is treated with one V_{min} and V_{max}, as if it were one individual speaker. This is because the hypothesis of grouping derives from the idea that dividing a unique sound particle produced by one sound source, and distributing this over more sound sources that flock together, would create more depth in the experience. However, since this is not mentioned in Maas' concept idea, it is important to find out whether grouping will affect the experience at all.

An overview of how the variables seem to correlate with each other according to this analysis is given in Table 5. Arrows from one variable to another mean that the variable at the origin of the arrow influences the variable the arrow is pointed to. The overview does not suggest whether the correlation between variables is positive or negative.

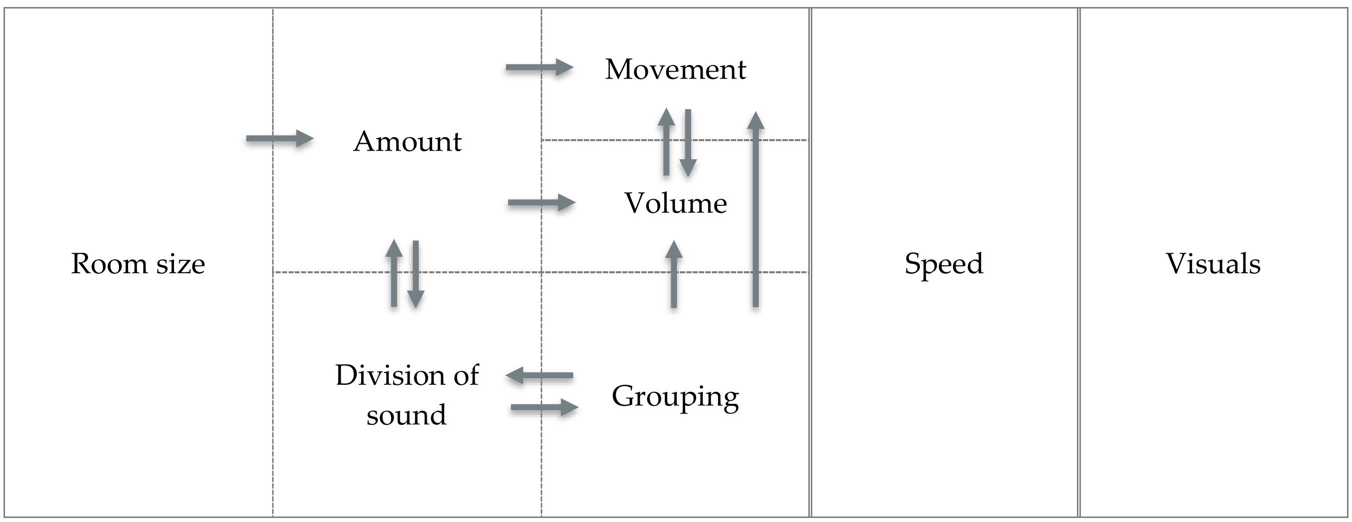


Table 5: Overview of correlations between variables.

3.1.4. Concept simulation

For further research about the variables given in Maas' concept idea, experiments need to be conducted. To comply with the given requirements and constraints, the experiments need to be conducted in a stable environment such that irrelevant variables remain constant throughout the experiment. An important requirement regarding this research is that the sound sources need to be able to move randomly in all axes, without any constraints. This is explicitly mentioned by Christine Maas.

The most feasible option to conduct experiments with, is to create a digital *Sound Swarm* in a virtual environment. A virtual simulation would allow quick design changes, while keeping irrelevant variables precisely constant. As the simulation can be ran on a laptop or even multiple computers, there is no geographical constraint concerning participation in experiments. Most importantly, a simulation would allow to bypass constraints from the physical world, which would allow sound sources to move randomly in all directions.

However, there might also be several downsides. Since it is a virtual simulation, it can be quickly concluded that the experience given by a digital *Sound Swarm* will never be fully accurate. For example, the ability to localise sound sources will not only be depending on the human ears, but also mainly on the algorithm that calculates the auditory spaciousness of the sound sources and the quality of the audio output. Besides that, there will also not be any somatic sensation of the soundwaves, which might very likely change the perception of the overall experience. Therefore, it is important to recreate a simulation that behaves as realistically as possible, auditory-wise.

3.2. Specification

To get a clearer overview of what the simulation should achieve and how it could achieve this, the specification requirements are divided into different levels, based on *Software Requirements: Styles and Techniques* [39]. The method has been invented by Soren Lauesen and it is focused on software-based projects. However, the technique used to dissect requirements top-down can be very useful for this project as well. Lauesen describes four requirement levels, which will all be used in this research, each to a certain degree. This makes it easier to analyse the design of the simulation and especially the reasoning behind it.

Lauesen uses the goals of a project to form a basis for all the requirements involved. With each level, the requirements become more detailed and concrete. The four levels described by Lauesen are: goal-level, domain-level, product-level, and design-level [39]. While these levels are meant to describe requirements on different levels to pick the requirement statement that provides most clearance, this research will use all four levels to identify concrete requirements based on more abstract ones. The goal- and domain-level requirements will be used to describe the experience specification of *Sound Swarm* in 3.2.1, which will contain quite abstract requirements. The results from the experience analysis will then be used to specify the functions of the simulation in 3.2.2, also making use of the product- and domain-level requirements. Once the specifications of the experience and its corresponding functions are listed, the requirements and constraints of the actual simulation software will be analysed in 3.2.3 to provide the best match.

3.2.1. Experience specification

It is important to first analyse the goals of the simulation and how the simulation must behave in order to fulfil these goals. The goals of the simulation may be different than the goals of the actual art installation, since the latter will not be used for research purposes. Initially, the goal-level and domain-level requirements are used to examine how the product would gain customer-value, and how this will be implemented in its functionality, respectively [39]. However, the requirements concerning the goal and functionality of the simulation are not to satisfy customers, but to create a credible experiment environment to satisfy this research. Therefore, the experience specification will slightly deviate from the goal- and domain-level requirements analysis as Lauesen describes it.

3.2.1.1. Goal level

The simulation needs to be able to provide a certain experience that should be perceived as realistically as possible. The simulation will be built according to Maas' concept idea, described in 3.1.2. Since there is no existing version of *Sound Swarm*, either physical or virtual, the experience created with the simulation cannot be compared for verification. It is important that attention will be paid to the variables that determine the acoustic environment, and that the simulation will be verified for its accuracy.

The auditory experience is most important. The simulation will also have visual representation, but the credibility of the acoustic environment is more important than the visual environment. This is because the simulation will most likely be used to mainly test the variables that influence the auditory localisation.

It is also desired that the simulation enables quick design changes, while keeping variables that are irrelevant to the current experiment constant. This way, the simulation can provide a solid environment for testing differences between values of only one variable. The combination of both goal level requirements would make a reliable research environment with the ability to quickly change design settings. In Table 6, the two goal level requirements are written down in a single sentence for clarification.

R1	The virtual simulation needs to provide a realistic environment, especially concerning the auditory experience.
R2	The virtual simulation needs to enable quick design changes, while keeping irrelevant factors constant.

Table 6: Goal level requirements concerning the simulation of *Sound Swarm*.

3.2.1.2. Domain level

The virtual simulation of *Sound Swarm* needs to provide functionalities that support the goal level requirements. The two requirements listed in the previous subsection both satisfy a different goal. From Table 6, R1 should be met to provide a realistic simulation that is necessary for (re)creating a desired virtual experience. This goal also affects participants and stakeholders, while R2 should be met to maintain credibility and usability concerning this research only.

For goal level requirement R1, providing a realistic environment, the simulation should provide both an auditory and visual experience. As stated before, the focus of this research is mainly on the auditory experience. This means that it is important that the simulation contains a realistic acoustic environment and a reliable algorithm to compute auditory spaciousness. Next to that, there needs to be a visual aspect that suggests that the observer is in a digital room with speakers.

To satisfy the goal level R2, the simulation software needs to support relevant parameters that can be manually changed. It is also very likely that there need to be multiple design setups with small variations, so the software needs to have the ability to quickly save and load different scenarios. For parameters that are non-existent in the simulation by default, the software needs to support the ability to declare new parameters. This could be done by coding.

3.2.2. Function specification

To further specify what the virtual simulation should look like, the experience specification needs to be translated into concrete function designs. These will be described in this subsection.

3.2.2.1. Product level

To better understand the simulation, the domain level requirements will be applied to the virtual components of the installation. These components combined will cover all variables discussed in 3.1.3. The main components of the installation are the room and the speaker(s).

As explained in 3.2.1, the room is an important aspect of the simulation. Its importance within the research itself may be relatively small, but to create a realistic experience, the room has a significant influence in both the auditory and visual aspects. Details on the visual experience have a low priority, but in general, an observer needs to be able to look around and confirm that he or she is looking at a virtual room. This means that there need to be walls, a roof and floor with textures. When a sound is being played in that room, the acoustics should also enable the observer to confirm that the sound is being played inside a relatively large, empty room. The room should thus contain a realistic reverb and auditory spaciousness.

For the speakers, the most important experience is the auditory experience. While there should be visual representation of an object that can be interpreted as a sound source, details on visuals have a low priority. It is more important that the acoustics of the speakers are realistic. This means that the output volume of the speakers represents a sound pressure level of around 70 decibels. Also, the Doppler effect should be applied.

To enable quick and reliable design changes, the software should support coding in which variables can easily be changed. This way, speakers can be programmed to move as desired with a velocity that can be altered easily. The speakers must be able to continuously move in either pre-programmed or random directions without: disappearing outside of the room; suddenly discontinuing; or falling into repetition, if it should move randomly. The speakers should be able to act as independent sound sources, meaning every speaker in the room is a unique sound source. Specific sound files could thus be loaded into an individual speaker and easily be substituted again, if desired.

3.2.2.2. Design level

There are no specific requirements of what the virtual environment should exactly look or sound like. The design for the room and sound sources, concerning both visual and auditory perception, should comply with the product level specifications.

3.2.3. Simulation environment

There are several ways to develop a virtual environment. Some options provide a lot of customisability and control to the developer, but are very resource-consuming. This can be done by using an object-oriented programming language such as C++ or Java. It provides full control to the programmer and everything can be adjusted to meet all the requirements. On the other hand, it is also possible to use a pre-existing software environment that is already focused on creating, manipulating and rendering virtual objects. This will be less resource-consuming, but may lead to less control of meeting all the requirements.

Due to resource constraints, programming the entire virtual simulation is not an option regarding this research project. Since the simulation of *Sound Swarm* contains virtual 3D-objects, movement in time, and interactivity, the most suitable option is to use an existing software program that is optimised for game developers. Such a software program is referred to as a game engine, which provides a user interface onto a backend rendering engine. This way, a developer does not need to get into the backend processes that provide the actual rendering, for example. Instead, these detailed and specialised tasks are already programmed in the core of the game engine. However, since this is only the fundamental coding, a game engine must allow the developer to program certain aspects of the game design, such as interactivity and movement.

For this research, the choice of game engine is Unity® 5.4, developed by Unity Technologies[4]. The software is widely used by both professionals and amateurs, making the online community a big and helpful one. Unity® provides a reliable game engine with the ability to add scripts, which can be written in C# or UnityScript – a language specifically designed for Unity, based on JavaScript. Because of prior experience with Unity®, and the fact that it seems to support all features needed to create a virtual *Sound Swarm*, this game engine is the first choice regarding this research.

3.2.3.1. Possibilities and limits

Since soundwaves are vibrations through a medium, the somatic sensory system is able to perceive certain frequencies as well. Especially lower frequencies are often also felt through the whole body. This perception will not be possible using a virtual environment, as the observer is not actually in the room.

Another difference is the quality of the speakers. In a simulation, the sound sources are treated as perfect speakers, whereas physical speakers are always submitted to impedance, which changes relative to the output frequencies. This may result in volume fluctuation over a change in frequency.

3.2.3.2. Important variables

Unity does not work with real-life units by default, instead it works with ratios. A distance of 1 in Unity can be anything in real-life. To prevent complications concerning the distance, the units in Unity will be treated as meters, meaning a distance of 1 in Unity represents a distance of 1 meter in real-life.

Visuals

The variables relevant for the visual perception are the field of view, near and far clipping planes, stereo separation. Except for the latter, all variables will be left on the default values.

Audio source

The audio sources will be omnidirectional to prevent complications concerning the direction of the sources. All sources will have the same priority and volume ratio. The volume roll-off will be logarithmic, with the minimum distance at 0 and the maximum distance at 20. This means that the sound will attenuate between these distances in meters, which would comply with the volume variable analysis in 3.1.3.1. To increase the sense of virtual surround sound, the head-related transfer functions are turned on.

Audio environment

The materials for the walls of the room are set as coarse concrete block, with a tiled floor and plastered ceiling, provided by the Google VR plugin. The proportion of direct sound that is being reverbed is set at 50%, with a gain of -20 decibels.

3.3. Realisation

Based on the information provided in the specification section, an actual virtual environment can be built. After the fundaments of the environment have been built and verified, several experiments can be

[4] Unity Technologies: http://www.unity3d.com/.

conducted using different scenarios, which will be described in 3.3.2. The results of the experiments will be provided and analysed in 3.3.3.

3.3.1. Building simulation

The first thing the simulation needs is a room. As described in the concept idea, analysed in 3.1.2, the room will not be a permanent component of the art installation, as it will most likely temporarily display in multiple exhibitions. However, the concept idea also states that the room needs to be realistically large. To ensure there is enough room for the speakers to move, a dimension of 20 meters in length and width and 4 meters in height has been chosen. Four omnidirectional light spots have been added to provide a natural and equally spread lighting.

The visual design of the speakers is based on an example speaker that Maas once brought with her. The speaker is shaped like a flattened ball, with a height and width of 0.3 meters, and a depth of 0.2 meters. All speakers contain at least a script for movement, which will be more explained in the subsections of 3.3.2, and an audio source component. In Figure 7, a screenshot of the simulation is shown to demonstrate the differences when adding textures. The textures do not show much detail, yet the visual experience has already increased.

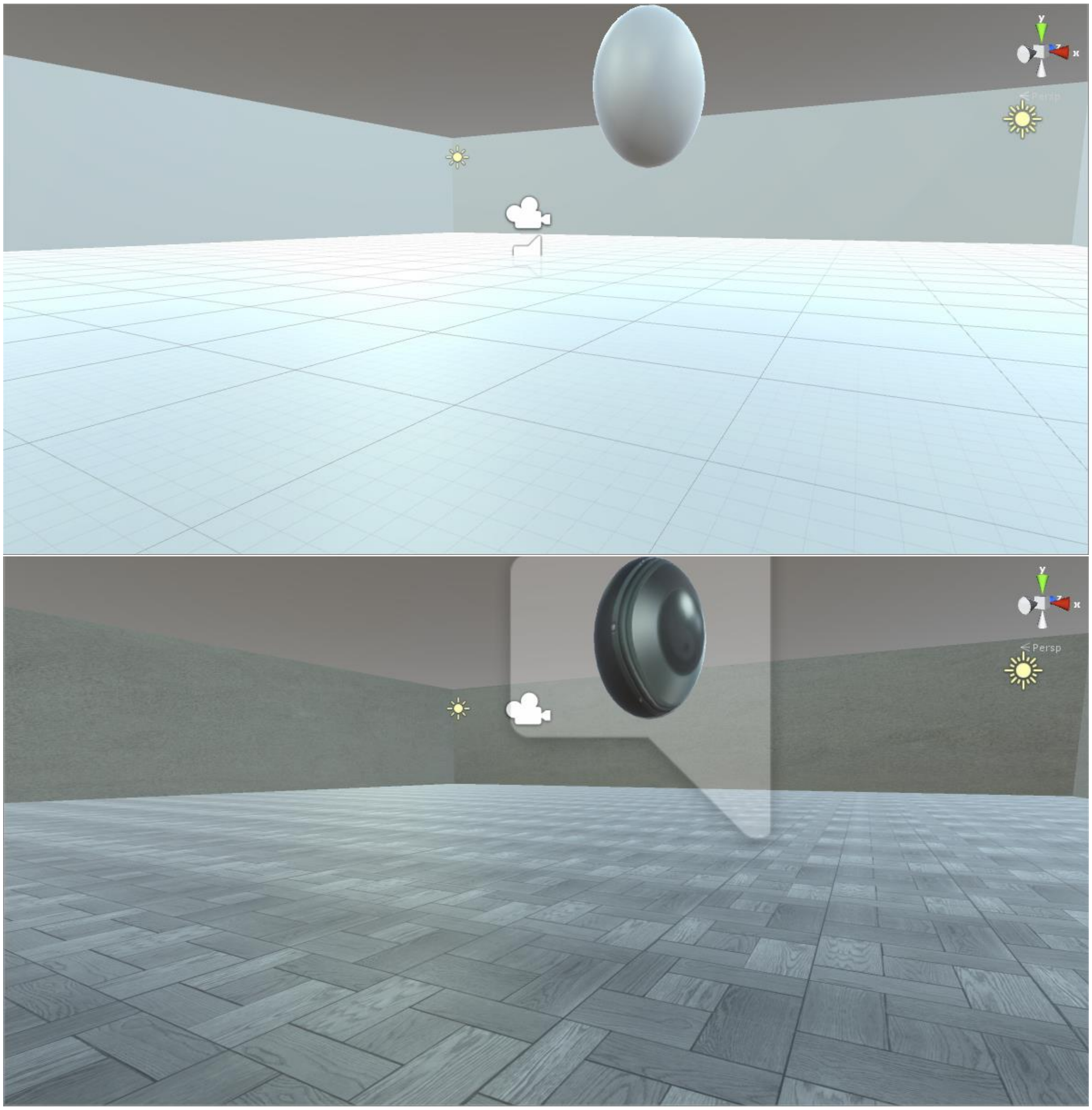


Figure 7: Screenshot of the virtual simulation in development.

3.3.2. Experiments

The simulation will be used to conduct a series of experiments to better understand the experience of the digital *Sound Swarm*. An important part will be to verify the reliability of the simulation and to what extend results of the experiments conducted using the virtual simulation can be used for the physical installation. Due to time constraints and the research phase, the experiments are not aimed at large amounts of quantitative data or statistical significance, but more at exploration and qualitative feedback. The main reason for this, is the fact that the concept idea does not contain concrete requirements, meaning that the research conducted in this project should aim at explorative consultancy regarding the build of *Sound Swarm*.

This research will focus on exploring setup designs for *Sound Swarm* to find the optimal auditory experience. However, as with all forms of art and human perception, the degree of satisfaction will most likely differ per person. An auditory experience that deviates more from the auditory perception that most people will acknowledge as *normal* or *common* will presumably have a higher level of satisfaction, but also the possibility of being perceived with a higher level of dissatisfaction. This might decrease if the auditory experience deviates less from *common* auditory perceptions.

Since an experience is a personal and individual impression of subjective perceptions, it is not effective to solely test the virtual *Sound Swarm* for the experience participants would achieve. Instead, the key of the experience needs to be objectified to enable better experiment designs. The hypothesis has been mentioned a few times already, mainly as motive for the background research on auditory localisation in 2.2. The hypothesis states that all relevant sound sources in the art installation should be distinguishable using auditory localisation. If this is not the case, then the sound sources that are indistinguishable have no purpose of flying solo. Therefore, the fundamental concept of the experiments is that the sound sources should be distinguishable.

However, since concrete and absolute results need a large sample size to verify statistical significance – which is not applicable to this research – the experiments need to contain a combination of both quantitative results that are relevant to the previously mentioned hypothesis, and qualitative results that are relevant to the personal experience. This means that this research cannot be used as a solid design fundament, but rather as a reliable source of explorative information.

The experiments will be conducted using the simulation of *Sound Swarm*. Every experiment will be conducted in a different scenario, each scenario providing different speaker setups. The room and all the important variables discussed in 3.2.3.2 will remain the same. This will prevent undesired variable changes and biased results. The experiments and corresponding scenarios will be explained in the next subsections.

These experiments together form the conducted practical research. Volunteers who participate the research are asked to complete all experiments in one session. The order of the subsections explaining the experiments in detail equals the order of experiments to conduct. The research setup consists of two laptops, an external screen and headphones (Sennheiser HD 477). One laptop will run the simulation and is connected to the external screen and headphones, while the second laptop shows a Google Forms questionnaire with all necessary information and questions for the participant. A screenshot of the first section of this form, which is shown at the very beginning of the research, is included as Figure 14 in Appendix B. The laptop running the simulation will be used by the researcher, and the screen of the laptop is turned in such a way that the participant cannot see what is on the screen. The external screen, however, is pointed towards the participant. During the first two experiments, which will be explained more thoroughly in the next subsections, the participant cannot see the simulation. Therefore, the simulation is run on the laptop screen of the researcher. For the other experiments, the simulation will be run on the external screen, meaning the participant can also get visual input. The participant can look around using a computer mouse.

3.3.2.1. Validation

This experiment is to find out how reliable the environment of the virtual *Sound Swarm* is and how significant the results of further experiments would be, compared to a physical installation. Most important to find out is how well participants can localise the sound sources and how they perceive the room acoustics. Therefore, the simulation is used to create a test scenario. This scenario contains one speaker continuously playing a short music composition. A script with ten different positions, each linked to a specific key on the keyboard, enables the speaker to instantly jump to the new location once one of the ten keys is pressed. While the ten positions are initially random, they will remain the same for every participant. The script is included in Appendix C.

Participants will only see a top-down view of the room represented by a grid of 20x20 squares. Next to it is a single row of four squares on top of each other, representing the height of the room. This is programmed in Processing[5], and participants should use it to click where they think the speaker should be located. The code and a screenshot, Figure 15, are both included in Appendix C. Beforehand, participants are well-informed about how to use the grid. They are asked to localise the sound source, solely using auditory input. To keep resemblance with the answer grid, it is not possible for participants to rotate in the simulation. Once they have inserted their estimate, the researcher will press the next key and the speaker instantly jumps to a different position. This will be repeated for ten different positions.

Afterwards, participants are being asked whether they have encountered any problems while trying to locate the sound. Also, for qualitative feedback purposes, participants are asked if they think that the acoustics of the room come across as realistic, given the room size. While the latter is a closed-ended question, which can bias some participants by putting words in their mouths [40], a better solution has not been found. An alternative would be to ask what the room size would be, based on the acoustics, but it does not seem likely that the participants would correctly guess the dimensions, and it is also not clear what the margin of error would be. For this question, it is most important that the participants are already steered towards thinking about the room acoustics, such as reverb, and compare that with experiences from the real world. To engage the participants into thinking more thoroughly about it, the question explicitly states to elaborate their answer. A screenshot of the questions concerning validation is included in Appendix C as Figure 16.

3.3.2.2. Grouping

Similar sound sources can be flocked together, to possibly create more depth in the auditory experience. The hypothesis is that subdividing the sound, and flocking together the speakers that produce similar sound characteristics, would pull apart the composition more. This should create an experience of increased depth and dynamics. The grouping experiment is also based on the idea that indistinguishable sound sources are useless to the *Sound Swarm* experience, as they will be perceived as a single source. Would that be the case, then it is more sufficient to take the sound sources out. For the grouping experiment, this means that there are two different scenarios. The first scenario contains four speakers in every corner of the room, moving in a plus-shaped pattern, while also moving up and down. A composition is divided into four sound particles, each particle consists of two similar instruments regarding frequencies: drums and percussion, female choir and electronic organ, piano and synth, and bass strings and bass

[5] Processing Foundation: https://processing.org/.

pad. Scenario 2 contains the same division, but for every sound particle there are five speakers. Every speaker outputs a specific frequency range of a sound particle. The frequency ranges have been divided in such a way that all five sources cover the total frequencies evenly, when the logarithmic scale is divided by five. In every corner, the same instruments are being played, but now with five different speakers. The speakers in each corner are distributed in a plus-shape, as can be seen in Figure 8. The middle speaker of every five speakers moves up and down, while the other speakers continuously move towards the speaker in the middle and back to their starting position with a speed of 1 m/s.

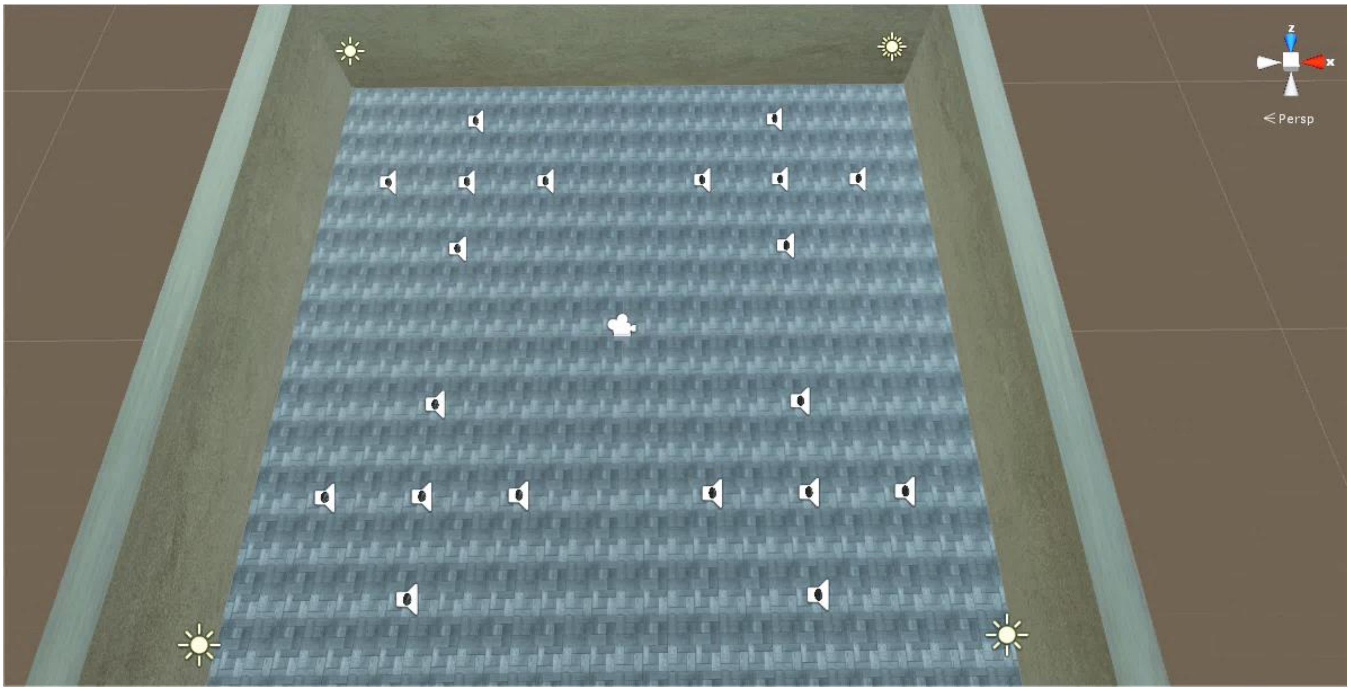


Figure 8: Overview of the speaker setup starting the second scenario of the grouping experiment.

Similar to the previous experiment, participants can only use their auditory system to perceive the simulation. After each scenario, participants are asked how many sound sources they heard. Once they have answered both questions, they are asked if they heard any other differences between the scenarios. This way, quantitative data is being gathered to verify whether there is an actual difference in sound sources, and qualitative data that is more focused on single experiences. Since the hypothesis that more sound sources are only useful when they can be distinguished is based on absolutes (they're either localisable or not), and the ideology of *Sound Swarm* is mainly based on experience and perception, both answers can be valuable to this research. In Appendix D in Figure 17, a screenshot of the questions regarding the grouping experiment is included.

3.3.2.3. Division of sound

This experiment consists of two different scenarios that are almost identical. Each scenario contains eight speakers that move randomly across the room. The script that engines the movement is included in Appendix E. The speed of all speakers is set at 4 m/s. Both scenarios play the same composition. However, in the first scenario this composition is divided into the following instruments: bass pad, drums, female choir, percussion, piano, electronic organ, bass strings and a synth. In the second scenario, the same composition is divided into eight frequency ranges, using the same division method as the previous experiment. This time the whole composition together is divided in eight frequency ranges, instead every instrument solely divided in five frequency ranges.

Before the first scenario starts, participants are informed that they can now also see the simulation and look around with the mouse. Since the audience of the actual installation is supposed to be seated, participants are not allowed to change the coordinates of the camera. They will also be informed that they will hear two different setups. Next to that, they are asked to pay attention to any differences and think about which one they would prefer. Finally, participants are also asked to pay attention to the speed of the speakers. A screenshot of the questions to be asked is included in Appendix E as Figure 18 and Figure 19.

After each scenario, participants will be asked to directly describe their experience and whether or not this experience would positively increase by increasing or decreasing the amount of speakers. After both scenarios, participants are asked whether or not the speakers would move too fast or too slow and if this would apply to both scenes. Finally, participants will be asked to elaborate which one of the two scenarios has their preference, and if they have other remarks.

3.3.2.4. Experience

This experiment gives an example of what sort of experience *Sound Swarm* may deliver. The scenario consists of twenty speakers, using the same sound sources as with the grouping experiment in 3.3.2.2. However, for this experiment, all speakers float randomly across the room. The speed of the speakers is 4 m/s. The speakers are not grouped or flocked. Since the speaker movement should behave similar to the previous experiment, the same script is used to engine the speakers. Participants are allowed to look around the room without changing the coordinates of the camera. In Figure 9, a screenshot is shown as an example of what a participant might see.

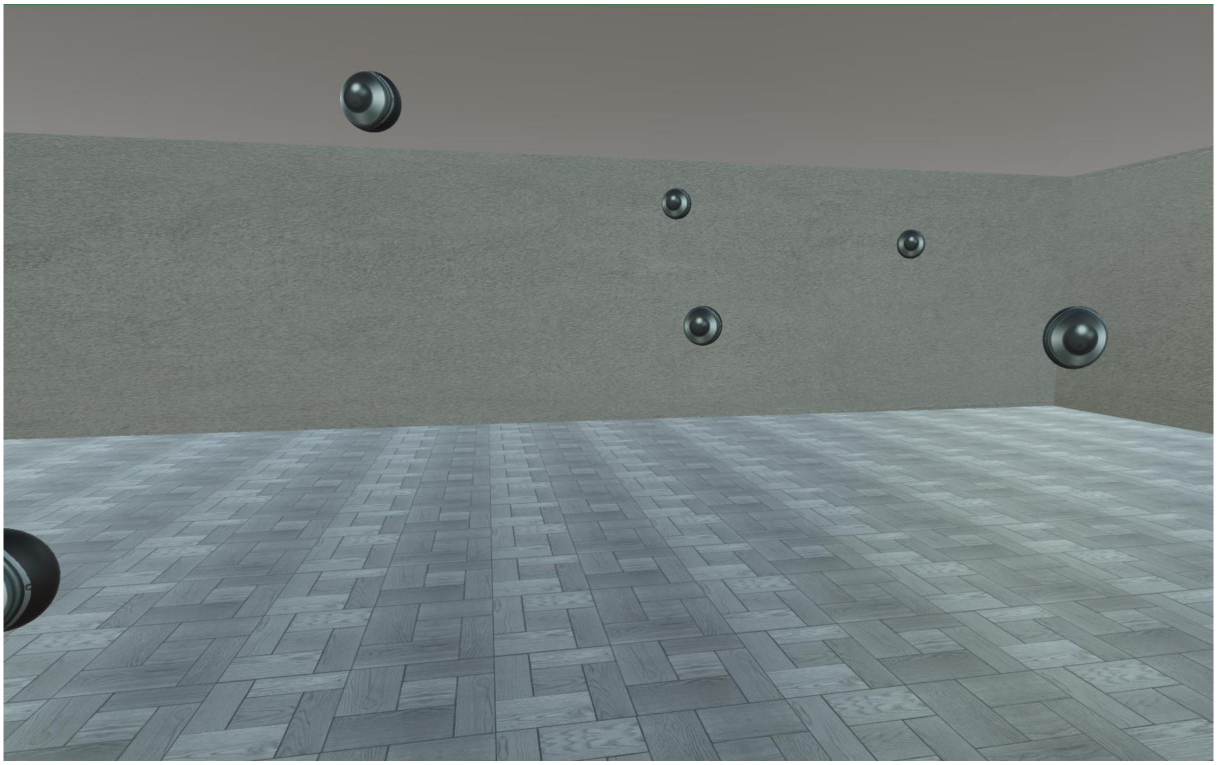


Figure 9: Screenshot of what the experience experiment can look like.

Prior to the experiment, the participants are only asked to focus on their overall experience, and think about previously discussed aspects of the simulation. This way, participants will only describe their own experience and will only write down what they think is important. Afterwards, they are asked to elaborate their experience as thoroughly as possible. This would provide diverse qualitative feedback and may show different views on the installation and what kind of experiences are perceived. A screenshot of the information and question given to the participants is included as Figure 20 in Appendix F.

3.3.3. Results

All results will first be written down in the chronological order of which the experiment has been conducted. The direct results will be visualised and analysed in 3.3.3.1. After that, the processed results will be analysed for potential correlations between different experiment scenarios and the theoretical variables analysis conducted in 3.1.3. These results will be shown in 3.3.3.2.

3.3.3.1. Practical research

The practical research is conducted with nine participants in total. All participants study Creative Technology at the University of Twente. This implies that the participants are already known and comfortable with virtual simulations, and are able to distinguish reality from simulation. As the study contains courses aimed at using state-of-the-art technology combined with art to create new forms of (interactive) art, it can be accepted that participants have a sense of artistic value. They could be considered as a subgroup of the potential user groups of the actual installation.

As the experiments are conducted with a total of nine participants, it cannot be statistically proven that the sample (with n = 9) is a representative for the total population. Although quantitative feedback will be evaluated in the results, most emphasis will be on the qualitative feedback. The qualitative feedback will be processed and may be used for (new) insights concerning the design suggestions of *Sound Swarm*. The quantitative data will mainly be used to verify qualitative feedback. It is highly recommended that the results of this research are valued as explorative design recommendations only.

Validation

All participants have encountered problems trying to locate the sound sources. The relevant problems that participants have encountered are shown in Table 7, combined with a count of how many participants have addressed the problem in their feedback. The table is shown in descending order, based on the amount of times participants have addressed the problem. The collected data concerning the actual localisation has unfortunately been lost.

Problem	**Count**
Difficulty estimating the height of the sound source.	5
Difficulty estimating the distance of the sound source.	3
Static design: head movement did not work.	2
Difficulty differentiating whether sound source was in the front or the back.	2
There was no example given for reference.	2

Table 7: Most relevant problems that occurred during the localisation of sound sources.

Whereas all participants seem to have encountered problems localising the sound sources, every participant does think that the room acoustics sound realistic. Two participants state that the reverb may be less than usual, one of them describing the phenomenon as if "the room walls were soft, like they absorbed some of the sound". The other participant states to not have noticed any reverb, but also to not have noticed any absence of reverb, concluding that the room might be realistic enough that the acoustics feel very natural. A screenshot of the answers regarding these questions is included in Appendix G as Figure 22.

Grouping

The results of the amount of sound sources perceived by the participants are shown in Figure 10. The figure seems to show a lot of similarities in terms of the perceived amount of sources per scenario. However, the first scenario contains twenty sound sources, which is not visible from the figure.

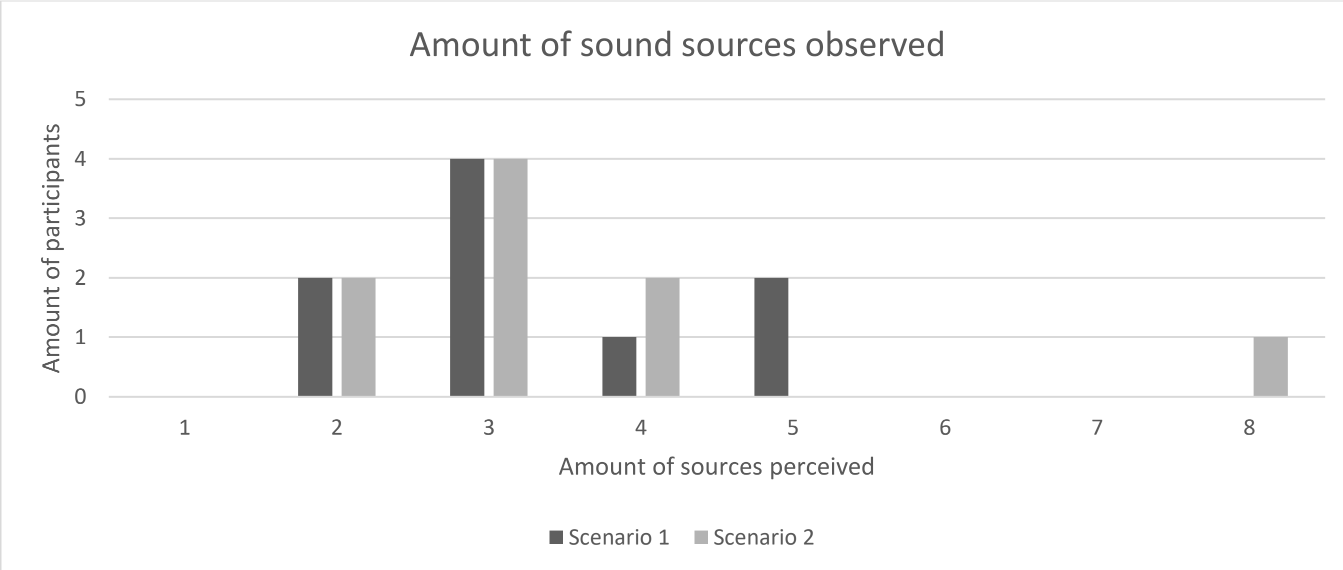


Figure 10: Amount of sound sources perceived per scenario concerning the grouping experiment.

The original qualitative results are included in Appendix G as Figure 23. These results are analysed for keywords and each keyword is counted, and checked whether it was a remark to the first or second scenario. Participants who have provided more qualitative feedback have most likely discussed more keywords. This means that these participants may have more influence on the results than participants who discussed less in their feedback. The keywords are counted per scenario in Table 8.

	Scenario 1	Scenario 2
More movement	1	3
More depth	0	2
More fluctuation	0	1
More speakers	1	1
More dynamic	1	0

Table 8: Amount of counted keywords to describe the differences in the grouping experiment.

Division of sound

Whether participants think a change in the amount of speakers would positively increase the experience is visualized in Figure 11. The graph covers the results of both scenarios. The graph shows a divided opinion concerning the amount of speakers, for both scenarios. For scenario 2, the amount of participants is equally divided over whether to increase, decrease, or keep the amount of speakers. Participants who voted in favour of changing the amount of speakers, are equally divided over whether the amount should increase or decrease. The latter holds for both scenarios.

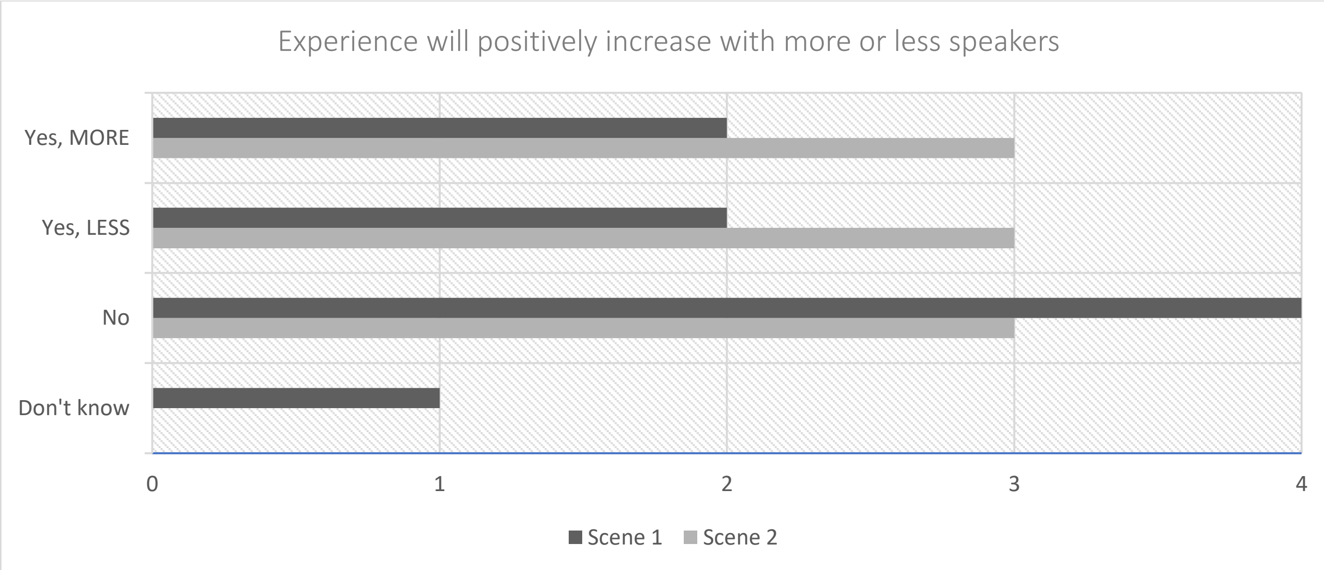


Figure 11: Results concerning whether a change in the amount of speakers would positively increase the experience per scene.

The experience description for scenario 1 seems to describe the overall impression and perception participants have experienced concerning the whole simulation scenario. For scenario 2, participants have mainly described the similarities and differences they have experienced between the two scenarios. Screenshots of the original experience descriptions are included in Appendix G as Figure 24 (scenario 1) and Figure 25 (scenario 2).

The descriptions for both scenarios have been analysed for general experience feedback. The feedback can be applied to both scenarios. It is split into two categories (positive and negative). Specific remarks are listed into one of the categories. Remarks without specific explanation are listed as 'general'. The remarks and how many times participants have mentioned them are listed in Table 9. Some participants have described their experience more thoroughly, which means that it is possible that these participants are contributing multiple remarks. In terms of participants, eight are positive whereas one participant experienced the scenarios negatively.

Category	Remark	Amount
Positive remark	*Total*	**9**
	Moving speakers (visual)	2
	Different intensities (auditory)	2
	General	5
Negative remark	*Total*	**2**
	Claustrophobic	1
	Too much input (chaotic)	1
	General	0

Table 9: Positive and negative remarks and how many times it has been mentioned concerning the division of sound experiment.

Analysing the descriptions of scenario 2 provides the results listed in Table 10. Participants have mainly described the differences they noticed or seemed to notice between the first and second scenario. The remarks are split up into two categories: sound perception and general perception. The answers have been analysed using the same method as the results in Table 9. However, in Table 10 the remarks are describing how the second scenario differs to the first scenario.

Category	Remark	Amount
Sound perception	*Total*	**7**
	Sounded off	2
	More lower tones	2
	High notes more prominent	1
	Messed up treble and mid	1
	Bigger sound difference	1
General perception	*Total*	**5**
	More chaotic	2
	Less speakers	1
	Faster speakers	1
	More distributed speakers	1
	Better speaker density	1

Table 10: Participants' remarks about the differences of scenario 2 compared to scenario 1, and how many times the remarks have been mentioned.

Of the nine participants, five prefer the first scenario and two prefer the second scenario. One participant does not provide a clear preference, but states that the sound was nicer in scenario one, and the second scenario had a better speaker density. Another participant clearly states “both were not pleasant to listen to”. A screenshot of the participants’ answers concerning their preferences has been included in Appendix G as Figure 26.

As for the speed of the speakers, three participants suggest a lower speed, two suggest a higher speed and one thinks the speed does not have to be changed. One participant states that the speed in scenario 1 was about right, but would prefer a lower speed in scenario 2. Another participant suggests to apply different speeds to the individual speakers. One of the three participants that suggest a lower speed, describes that a lower speed would be more suitable for the composition that is being played, thus making a link between speed and composition.

A screenshot of the answers provided by all participants is included in Appendix G as Figure 27. Unfortunately, one of the participants compares the two scenarios, but addresses both with “first scenario” and “first one”, making it unclear which scenario is meant. This feedback will therefore not be included in the results.

Experience

The experience descriptions have been analysed similar to the previous experiment results. The remarks are categorised in several relevant subjects. It has been counted how many times remarks have been mentioned by different participants. Since some participants have been more thorough with describing their experience, these participants most likely have addressed more remarks. The counted remarks and relevant categories are listed in Table 11. A screenshot of the original feedback is included in Appendix G as Figure 28.

Category	Remark	Amount
Experience	Nice experience	6
	Perfect combination of both division of sound scenarios	1
Acoustics	Much reverb	2
	High realism	2
	Localisation easier with visuals	3
	Hard to hear difference front and back	1
	Hard to hear difference in height	1
Amount	Nice amount of speakers	3
	More sources (than previous scenarios)	2
	Too many speakers to distinguish	2
	Only hear difference when speakers are close	2
	Same amount as scenario 1 of division of sound experiment	1
	Too many speakers to hear reverb	1
Speed	Speed too high	3
Division	Good sound quality	2
	Higher tones sound like razor	2

Table 11: Remarks about the experience experiment, categorised and counted.

One participant mentions a possible relation between the amount of speakers, the speed of the speakers and the composition that is being played. The participant mentions that there are too many speakers to distinguish and that the speed of the speakers does not seem to fit the style of the composition. The participant then suggests that the amount of speakers might make the scene too chaotic, but the speed of the speakers may also influence the perception that the scene is too chaotic for the composition that is being played.

Another participant uses an extraordinary metaphor to describe the perceived experience in one sentence. The simulation experience is compared to a visit to a discothèque while being under the influence of a certain drug. The way this metaphor is written down suggests that the comparison is positive, thus the feedback has been marked as 'nice experience'.

3.3.3.2. Combined results

To be able to better understand the value of previous results, this subsection will look at potential correlations between the conducted experiments, while also addressing the relevant theoretical analyses done in 3.1.3.1 and 3.1.3.2. Most of the results are qualitative feedback, which means that feedback given at one experiment might also hold true for another experiment where this feedback has not directly been given.

Acoustics

In both the validation and experience experiments, participants have pointed out that they found it hard to hear differences in the vertical plane. It is also mentioned by participants that estimating the distance of a sound source is hard. In 2.2, several scientific articles have been addressed which state that distance and height localisation is less precise than horizontal localisation. This means that the auditory localisation perception might be in line with its underlying theory.

Amount and movement

All experiment scenarios, except for the validation experiment, contain setups with a different amount of speakers. The theoretical analyses imply that the amount of sources has a certain optimum if all other variables are constant. This optimum derives from the movement freedom per speaker, according to the correlation that the room volume divided over the amount of speakers leads to the movement freedom per speaker.

The movement freedom of the speakers in the grouping scenarios has been very limited. In the second scenario, the simulation contains five times less speakers than in the first scenario, but the total movement freedom remains the same. This means that an individual speaker in the second scenario has five times more distance to cover. However, the movement of the speakers has been scripted and is not random, meaning that the movement freedom the speaker would achieve using the theoretical formula is not applicable. The results of the grouping experiment imply that more participants have noticed more movement, depth or fluctuation of the speakers in the second scenario, where there are only four speakers instead of twenty.

The division of sound experiment consists of two scenarios both containing eight speakers. The room remains the same for all experiments, therefore the movement freedom does not change either. However, participants have stated that the second scenario would be more chaotic and contains a better speaker density with less speakers that are distributed more. With the experience scenario, which contains twenty speakers, one participant states that the amount of speakers is the same as the first scenario of the division of sound experiment. Three remarks in Table 11 on page 46 of the experience experiment describe difficulties with sound localisation because of too many sound sources. This has been remarked five times in total. Three participants, however, are explicitly positive about the amount of sound sources. No participant has explicitly mentioned a negative remark about the amount of sources.

Speed

In the variables analysis, it is suggested that the speed should not go higher than 5.4 m/s. In both scenarios of the division of sound experiment and the experience experiment, all speakers move with a speed of 4 m/s. Three participants suggest a lower speed for both experiments. One participant makes a connection between the speed, amount and music composition. The participant states that both speed and amount can make the experience seem less or more chaotic, and that the style of the composition also influences the perceived chaos.

Division and grouping

The second scenario of the division of sound experiment has received seven remarks about the sound perception. The experience experiment and the first scenario of the grouping experiment both use the same sound sources. The only difference is that in the experience experiment the sources move randomly and do not flock together. All the sound sources of the experience experiment and the second scenario of the division of sound experiment contain only a certain frequency range. In the experience experiment, there is one participant who has a remark about the higher tones sounding like a razor.

3.4. Evaluation

The practical research that has been conducted needs to be evaluated as well. This is to examine how much value the results may contain and what could be improved to gain more value for potential future experiments. In 3.4.1, the research methods will be evaluated. This will be done with regard to the methods used to design the experiment and questions, and the design of the experiment scenarios. The prototype will then be evaluated and examined in 3.4.2.

3.4.1. Research methods

With a total of nine participants, it is not possible to test a hypothesis with statistical significance. The critical questions from Table 3 on page 26 require a specific and numerical answer. Although it has been the aim of this research to provide answers to these questions regarding the optimal experience, the conducted research is more valuable as explorative design suggestions. The critical reason for this is the fact that 'the optimal experience of *Sound Swarm*' is a theoretical phenomenon and does not exist yet. As there is no existing definition of this potential phenomenon, it has not been possible to test certain design setups and let participants decide which one has their preference. Therefore, the experiments had to be designed such that participants are free to describe their own experience and describe what parts are perceived positively or negatively. Based on the qualitative feedback results, certain design suggestions can be created.

The qualitative data has been analysed by the researcher. For some given feedback, there is the possibility that the interpretation has been biased, due to language usage. It is possible that the use of some words is interpreted as positive or negative, while this is not meant by the participant. All results have been analysed as objectively as possible. However, if more certainty is desired, the experiments should have been designed such that participants would have to provide more quantitative data.

The theoretical analyses in 3.1.3 have provided the most important variables and critical questions. That chapter suggests that numerical answers to each question would provide the optimal experience, in theory. Ideally, the practical research would have provided these answers. This has not completely been the case; the practical research did not directly satisfy the questions raised during the theoretical analyses. However, both theoretical and practical researches have tackled the same question of what should be the perfect experience. Results of both researches can be combined to complement each other. The practical research should not be considered as attempt to verify the theory. Instead, the theoretical and practical researches are different approaches to find and define the perfect experience.

3.4.1.1. Experiment methods evaluation

The validation experiment does not have a reference point in the beginning. In Table 7 on page 42, two participants have remarked this as well. A reference point or example would most likely have made the localisation process more precise, because participants can use the reference point to estimate distance better. However, the validation experiment is meant to validate the simulation environment. To better accomplish this and to prevent an early learning-curve, the experiment is designed such that participants use their ears and possibly previous experiences in real-life to estimate the source location.

Participants have not been allowed to move the virtual camera around during the validation experiment. This might be unnatural compared to real-life situations, where it is more common to freely move around. However, the participants could also not obtain visual stimulation, as the experiment is only about auditory localisation. If participants were to be able to change the camera angles, the answer sheet of the localisation experiment would no longer be representative of the virtual environment. A solution could be to let the researcher move the camera back to the original position before letting the participants localise the sound source.

In the grouping experiment, participants have also not been allowed to change the virtual camera angles, or move the virtual head. However, as this would be more natural and would enable a better auditory overview of the sources that are present, having allowed camera movement might have increased the validation of the results. In contrast to the validation experiment, there are no substantial disadvantages of having allowed participants to move the virtual head.

In the division of sound experiment, all participants first experienced the scenario with the composition separated by instrument and then the scenario separated by frequency ranges. All participants have written down feedback about the whole experience when asked after the first scenario. The feedback regarding the second scenario only consists of how it differs compared to the first scenario. This is not necessarily affecting the results in a negative manner, but the first overall impression will be assigned to the first scenario.

3.4.1.2. Experiment scenarios evaluation

As stated in 3.3.3.1, the collected results of the actual localisation assignment has been lost. This means that it cannot be validated how adequately the source locations have been estimated. However, the goal of the experiment is to let participants truly focus on the auditory perception, without an already critical mindset concerning the realism of the audio. The actual localisation is not relevant for the results of the experiment. The sample size is too small for significance, meaning there also is no margin of error. Results can also depend on the localisation ability of the participants themselves, and do not necessarily rely on the audio output of the simulation. Although the localisation results would probably have provided an overview of how well the participants could localise the sources, due to the previously mentioned points, this data could not have been analysed or used in any further way.

The movement behaviour of the sound sources in the grouping experiment most likely does not represent actual speaker movement of the final art installation. This might affect the perception of the grouped sound sources. However, the experiment has been designed such that the effect of grouped sources should theoretically be more prominent than when present in random movement behaviour. To atone for this, the same sound sources are being used in the experience experiment, with the only change that the movement behaviour is random, and grouped sound sources do not flock.

In both the division of sound experiment and the experience experiment, the sound sources move randomly. While this is according to Maas' concept idea, it is unlikely that the speakers of the actual installation will show this movement behaviour. In both experiment results, some participants have stated that they noticed either more lower or higher tones than usual. This might be the result of the randomisation of the speaker movement. Speakers with either the lower or higher frequency ranges might have passed the participant more often during the experiment, leading the participant to think that this is what the installation should sound like. However, due to time constraints it was not possible to generate proper movement designs to test separately. With random movement behaviour and qualitative feedback, there is a likelihood that one or more substantial movement design suggestions would appear.

3.4.2. Prototype

An important aspect of the prototype is the auditory spatial environment. As described in 3.2.3, the simulation uses the Google VR SDK plugin, which provides a more elaborate algorithm to calculate the three-dimensional auditory space. However, after the experiments it has been found that the HRTF (head-related transfer functions) do not work when the audio output is two-channelled. The most notable change is the inability of hearing differences between the front and back, which has been addressed by some participants as well. This means that, for example, a sound source on the horizontal plane at 45° relative to the ear sounds the same as a source that is placed at -45°. When the participant is able to look around, this becomes less of a problem, as this changes the relative angle between the source and the ear. If the source would be in front of the left ear, the angle would decrease if the head is turned to the right. If the source would be behind the left ear, the relative angle would increase.

The prototype also does not seem to provide for volume experiments. This is because all audio output regulators work proportionally and relatively to the original digital sound loudness. Therefore, it is hard to determine the actual output in decibel sound level pressure. Next to that, the virtual soundwaves created by the speakers are not objected by objects such as other speakers or the audience. The room itself does reflect the sound as reverb.

Although the auditory experience has been said to be realistic, the prototype does not provide an overall realistic experience. Firstly, the visual experience is not as optimised as the auditory experience. This is because the 3D-models are of low detail and contain basic textures. The visuals provide enough information to let the observer feel like being inside a space with four walls, a floor and a roof, but it does not show any details of an actual room. Secondly, where the audio can be perceived as three-dimensional, the visuals cannot, since the screen is two-dimensional. Looking around is done by moving a computer mouse while looking at a screen. In real-life, looking around would be done by moving the head. Lastly, the speakers have no physical constraints. This means that the speakers are able to move through each other and the audience.

4. Conclusion

The conclusion can be generated from both theoretical research and analyses and the practical research. Since Maas explicitly requested to conduct this research under the condition that all sound sources are able to float in every direction, the conclusion will also cover this. However, this research is also initiated to support Maas in realising an actual installation. This means that realistic solutions or suggestions will also be provided where possible.

This research has been aimed to answer the research question: *What is the best setup design of Sound Swarm in order to provide the most exquisite experience of being inside a music composition?* Results of both the theoretical and practical research have provided more knowledge of *the most exquisite experience*, but have been unable to give a direct answer to the whole research question. The gained knowledge concerning the experience can be used for further research to examine actual design setups. As fundamental design suggestions, theoretical and practical findings have been applied to the concept idea of *Sound Swarm*, which provide a part of the answer to the research question in 4.1. It has also been found that there are two kinds of experiences concerning *Sound Swarm*. These experiences will be explained in 4.2.

Based on the qualitative results in 3.3.3 concerning the validation of the virtual environment in which the practical research has been conducted, it can be concluded that the auditory experience is reliably realistic. This means that any conclusions based on the auditory perception during experiments of the practical research can be considered reliable as well.

4.1. General theory

All statements supporting the conclusive design suggestions in this subsection are based on the theoretical research conducted in 2.2 on page 15. Conclusive statements that use a different source of information for support, such as practical results, will exclusively specify where the information can be found. If no source is given, the fundamental knowledge has been presented by the scientific articles from 2.2.

Frequencies below 200 Hz are inaccurately localisable by the human brain. The brain uses the interaural phase differences to localise low frequencies. However, when the duration of one period is too long, the relative phase difference perceived by both ears is not big enough for the brain to accurately calculate the delay, leading to the misperception that the soundwave is coming from both sides. This occurs to frequencies below 200 Hz, due to the size of the human head. Since low frequencies are produced by subwoofers, which are relatively big, it would be highly inefficient to let these speakers float around as well. Therefore, it is suggested to separate the low frequencies and use stationary subwoofers near the audience. If the subwoofers are placed close to the audience, the volume can be lower and there is less chance of interference.

Humans are more accurate in localising sound on the horizontal plane in terms of angular differences than estimating the height of a sound source. The least accurate auditory localisation parameter is the estimation of distance between the observer and the source. However, by providing relatively big movements on either the vertical plane or in terms of distance, the chances of these movements being noticed will increase. It is thus suggested to focus the speaker movement on rotational motion for a fundamental perception of being surrounded by sound, but include big changes in height and distance relative to the audience to create more depth and dynamics.

Sound sources containing higher frequencies could be placed physically higher as well. This way, the illusion of tones occupying their correlated height in a spatial location can be used to strengthen the localisation abilities on the vertical plane. However, it might also be possible to use this illusion and turn it completely around; suddenly physically lowering sound sources containing higher frequencies to break the illusion, possibly leaving the audience more focussed on differences on the vertical plane.

The speed of the speakers can mainly be used to create a level of chaos. Participants have been very divided concerning the speed of 4 m/s used in the relevant experiments. None of the participants have had any remarks about the Doppler effect. It can be concluded that 4 m/s is a safe speed to test future designs with. It is recommended to keep the speed lower than 5.4 m/s, to prevent people with absolute hearing noticing the Doppler effect, possibly leading to a lesser experience. Also, since the Doppler effect is relative, it is suggested to keep all speakers at the same speed.

The amount of speakers and the corresponding movement freedom seems to have less influence on the experience of *Sound Swarm* than expected. This can be concluded using the results of 3.3.3.2 on page 47, where results of all relevant scenarios are combined to gain more knowledge on the amount and movement. Other variables, such as speed and the division of sound could influence the perception of the level of chaos, catering an illusion of an increased or decreased amount of speakers. In the correlation analysis in 3.1.3.2, the amount of speakers and division of sound are expected to be equally important. However, results have provided reason to believe that the exact amount of speakers is less important than the method of dividing the original composition. Therefore, it is recommended to first focus on the division of sound.

4.2. Experience development

From the practical research, it has been discovered that there can be two different ways of experiencing *Sound Swarm*. This conclusion is based on the results provided by the division of sound experiment as well as the experience experiment. By dividing the sound over frequency ranges, most of the participants have stated at least one remark about the sound perception. This means that the sound was different than expected, creating a new sort of experience.

The experience of *Sound Swarm* can be focused on either the 'sound' or the 'swarm' aspect, as discovered during the practical research. The sound aspect will focus more on a new experience of music, where the swarm aspect will focus more on the chaotic experience of a swarm. The two possible experiences will be explained in 4.2.1 and 4.2.2. Suggestions concerning certain variables such as the amount or speed of speakers will be in terms of relativity, as there has not been a definite conclusion for specific values.

4.2.1. Swarm-like

The main focus of the swarm-like experience will be on auditory localisation, moving sources and chaos. This design will most likely be less satisfying to listen to, but the experience would be more unique. To optimise the auditory localisation, it is advised to use a relatively small amount of speakers. Dividing the sound over frequency ranges will create a greater sense of diversity and lower density, also allowing for easier distinguishing the sound sources. The movement behaviour of the speakers can contain quick and sudden movements, whereas the speed can also be relatively higher.

4.2.2. Music-like

The music-like experience should contain enough speakers such that the audience will always perceive a fundament of the composition. This decreases the distinguishability of sound sources, but would make the installation more pleasant to listen to. The density of the speakers is focused in a circle around the audience, with sometimes having random speakers pass close by. This will move the attention of the audience towards the speakers that are close, while the rest of the composition is still provided in the background. The speed of the speakers would be relatively lower, and the movement behaviour should be more fluent and elegant.

5. Future work

This chapter will provide further suggestions that are more specific than the design recommendations in the conclusion. However, where the previously mentioned recommendations have also been researched, all suggestions provided in this chapter have not. In 5.1, possible further development will be examined. This includes enhancing the virtual simulation used in this research, as well as recommended further experiments. Important aspects concerning the actual built of the installation will also be reflected on. The section in 5.2 will discuss the purpose of this research and the value it should add.

5.1. Further development

This research has provided information about possible experiences, based on subjective feedback and theoretical analyses and literature. Therefore, the first step towards the realisation of *Sound Swarm* has been made, but further research and development is needed to provide more concrete results. Based on current results and conclusions in combination with the theory, certain development suggestions can be made. These suggestions will include consultation in improvement of the prototype, as well as experiment proposals.

In 5.1.3, all relevant acquired knowledge will be analysed again in an attempt to provide design suggestions for the actual art installation. The main difference is that this subsection will not assume the ability of sound sources to float in any direction. As with all suggestions presented in this chapter, every given advice is supported by scientific literature or relevant results of the practical research. However, the suggestions themselves have not been tested during this research and are not guaranteed to work. Therefore, it is recommended not to implement any features without first conducting practical research.

5.1.1. Prototype enhancement

Several remarks about the prototype used in the practical research have already been mentioned in 3.4.2. These points are mainly with regards to improving the quality of the research results. Nonetheless, in order to enhance the prototype, these issues have to be improved. The suggested prototype enhancements are to improve the realism of the virtual experience of *Sound Swarm*. Besides the auditory experience, this also includes more focus on the visuals and interaction.

The first improvement would be to fully implement virtual reality. This means that actual head movements control digital head movements, and that the visual stimuli are perceived in 3D. To enhance the auditory experience, an audio output that enables surround sound is suggested. The most suitable option would be to go for a 5.1 surround sound audio system, which includes five channels for full bandwidth and one for low frequencies. It is possible to use a headphone or complete speaker set for this audio realisation. A headphone has the advantages of being portable and relatively cheaper than a full speaker set. It also prohibits any undesired auditory influences from the real world. However, using a headphone will make the auditory experience rely more on the accuracy of the algorithm that calculates the auditory spatial awareness.

Besides the experience improvements, the prototype can also be enhanced in enabling the user to produce different designs more easily. This can be done by building a user-interface on top of the basic functions, such as changing the amount of speakers or their movement behaviour. With an intuitive user-interface, the prototype will become more accessible to use.

5.1.2. Future experiments

The conclusion in chapter 4 has provided more knowledge about different kind of experiences that *Sound Swarm* could provide. In order to fully comprehend the actual setup design, further research concerning the important variables is highly recommended. This subsection will cover suggested experiments that can be conducted to gain more concrete information about these variables. Most of these experiments could not have been conducted due to time constraints. For all experiments, it is recommended to have a sample size large enough to provide statistical significance. For more reliable results, it is also advised to conduct the suggested experiments in the improved virtual environment of *Sound Swarm*, according to the prototype enhancements provided in the previous subsection.

5.1.2.1. Suggested experiments

As evaluated in 3.4.1.2, the grouping experiment will most likely provide more reliable results when participants are able to move the virtual head, enabling them to 'look around'. The speakers may show random movement behaviour, but with a relatively low speed. This is so participants can more easily focus on the amount of sound sources.

The original composition is concluded to be an important component of the experience, although this research has barely covered its influence. The type of composition is considered to be the undertone of how the experience will be perceived. A slow and melodic composition will likely cater a more intense experience if the speed of the speakers is not too high. However, such an assumption should be tested. Therefore, it is suggested to conduct experiments concerning the correlation between the type of composition and the speed of the speakers.

The movement freedom or behaviour of the speakers is considered to be one of the most important variables in terms of contributing to the overall experience. It is suggested to research the movement behaviour of different animals that tend to form swarms, in order to copy a more realistic 'sound swarm' experience. This might add a more natural perception of movement by the audience.

Once a few movement designs have been selected, and it has been decided what kind of experience the installation should provide, research regarding the amount of speakers can be conducted. A suggestion for testing the amount would be to begin the experiment with the amount that is preferred by Maas, and create a few alternative designs with more and less speakers. An important note is that changing the amount of speakers will also change the division of sound.

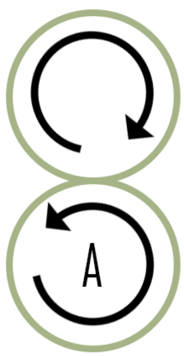
5.1.2.2. Design possibilities

Based on the design suggestions provided in the conclusion, two design concepts have been generated as possibilities for further research. These concepts potentially enable a more music-like experience, as the research question is more focused on the experience of *Sound Swarm* regarding a music composition. The designs are not completely thought-out, but may show potential. Therefore, these designs are explicitly mentioned in this subsection.

Flower-design

This concept design enables a combination of rotational movement and difference distances, as well as the option of having a denser circle around the audience. The speakers will follow a track that is shaped as an '8', with one of the circles surrounding the audience, as shown in Figure 12 (left). A single track could contain multiple speakers. It could be decided to have speakers follow the circle that is not surrounding the audience most of the time, and only sometimes make the switch towards the audience. This will create the higher density as suggested in 4.2.2.

To decrease the possibility of a noticeable Doppler effect, it is suggested that the rotational movement of neighbouring eight-shapes is countered. If the speakers on the outside circle of one of the tracks move clockwise, then the speakers on the neighbouring tracks should move counter-clockwise, as shown in Figure 12 (right). As the level of perception of the Doppler effect is relative to the velocity between the source and the observer, with this design the speakers from both tracks will be perceived with the same level of the Doppler effect, decreasing the possibility of being noticed. If the rotational movement on all tracks would be in the same direction, one speaker would have a positive velocity relative to the observer, and the other would have a negative velocity. This would make the change in the perceived pitch differ per speaker, leading to the possibility that the combined sources would sound off key.



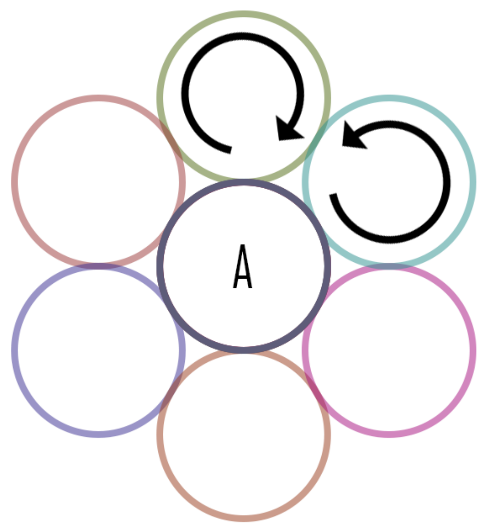


Figure 12: Left: Movement of speakers shown on a single track concerning the flower-design. Right: Movement of speakers shown on six tracks concerning the flower-design.

Speakers that move clockwise on the outside circle will move counter-clockwise on the circle around the audience, and vice-versa. However, this will not matter concerning the Doppler effect, as the distance between the observer and the sources will not change. The Doppler effect will not occur, regardless of the rotational velocity. The speakers should be pointed inwards for the best effect. The number of tracks should be even. This concept design does not have proper suggestions concerning the vertical plane.

Gyroscope-design

This concept design is based on the dynamics of a gyroscope combined with the design of the *Multi Axis Trainer* (MAT) by NASA, as shown in Figure 13. However, concerning *Sound Swarm*, the installation would consist of two circles that are cut in half, encapsulating the audience. This way, movement in both the horizontal and vertical plane is rotation-wise. Both semicircles contain a row of speakers. The outside semicircle rotates on the horizontal plane, where the inside semicircle rotates on the vertical plane. Both use the audience as pivot point. However, this design does not show movement behaviour that would be associated with swarms.

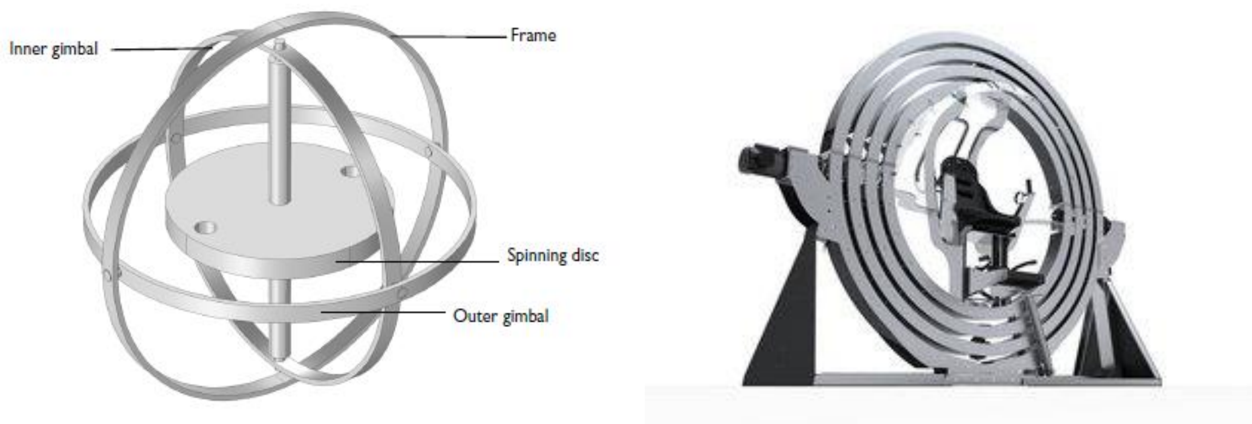


Figure 13: Left: Dynamics of a gyroscope. Right: Example of a Multi Axis Trainer.

5.1.3. Physical installation

The aim of this research has been to provide Maas a definite answer to the optimal design of *Sound Swarm*, which would bring the most exquisite experience. Using the conclusion of this research, Maas could then add constraints of the physical world and build an installation that would be based on this research. However, most of the results have led to suggestions rather than concrete methods. While designing certain experiment scenarios or formulating design suggestions, some aspects concerning the real installation have emerged as well. These aspects are formulated here, including possibilities which are recommended to be researched further.

Considerably the biggest difficulty with the design of the actual art installation of *Sound Swarm* is the method to allow sound sources to seemingly move in all directions. An option would be to design a grid skeleton which can be placed inside the room. Speakers will be suspended from this grid towards the floor. Regarding the movement freedom, speakers are restricted by the grid itself, but may be able to move up and down by decreasing or increasing the chords that hold the speakers. For more movement freedom concerning the horizontal plane, speakers can also be attached to poles that are placed on wheels. The higher the pole would be, the more unstable the speakers would stand and the higher the risk that it will lose its balance. Another possibility is using surround sound systems, using the Ambisonics or ultrasonic sound that can create virtual sound sources. Concerning the movement freedom of the speakers, this enables the same possibilities as in a virtual environment. However, there will be loss of the visual experience.

For real implementation, the gyroscope-design in the previous subsection shows potential. For added sense of depth, both horizontal and vertical plane could contain layers of multiple semicircles with speakers. The installation would likely require an outside framework. The pivot mechanism concerning the horizontal plane could be above the audience, suspended from the framework, or using a rail on the floor, surrounding the audience. The pivot mechanism concerning the vertical plane could be placed at floor-height. Distance is estimated using reference points and relative volume changes, which could be used to create the illusion of distance changes by fluctuating the volume of certain sound sources.

It is recommended to aim the sound sources towards the audience, especially in rotational movements. Unlike the virtual simulation, where the sound sources have been arranged as omnidirectional, default speakers only focus on one direction. However, it is also possible to conduct certain experiments containing directional speakers aiming in different directions.

The final general design suggestion is concerning the placement of the audience. Based on previous theoretical findings, it is suggested to put the audience in chairs on a podium. This way, there is more space below the audience, enabling sound sources to go relatively lower as well. The subwoofer(s) could be placed underneath the podium, such that there are no static speakers visible. Assumingly, the low frequencies would also make the podium vibrate, adding a strong somatic sensation to the experience. The chairs of the audience can be placed in a circle, facing outwards.

5.2. Discussion

The goal of this research has been to design a setup that would provide the optimal experience of an art installation like *Sound Swarm*. However, from Maas' concept idea, analysed in 3.1.2, it becomes clear that this research is initiated, because the experience itself needs to be examined. Such an experience is subjective and there are no indications of exactly what the experience should sound like. The concept idea also does not clarify any variables.

5.2.1. Optimal experience

As there is no specific goal, nor predefined variables, it would be close to impossible to find the actual optimal experience, especially when including the time constraints for this research. The theoretical analyses hypothetically claim that the optimal experience would be created with a design that would allow all sound sources to be distinguishable and able to localise. However, as the sound sources will move (randomly) across the room, it differs per time whether sound sources can be distinguished or not. The theory does allow for better understanding important variables concerning the design of *Sound Swarm*, but it cannot function as a design fundament.

The definition of the optimal experience can also differ based on the indication of the definition. Besides the theoretical definition, there are three possibilities for optimising the experience. However, these possibilities require existing design setups that already have the potential of being realised into the actual art installation of *Sound Swarm*. These designs can be experienced by Christine Maas, as she is the main person of interest, to determine which design will be considered to provide the optimal experience. This would define the optimal experience as the experience Maas wants to present to the audience. The other two definitions are based on changing Maas' decision to the majority of potential visitors. The difference between these two definitions of the optimal experience, is based on whether it is more important that most visitors can enjoy the installation at least a bit, or that the design should provide the most intense experience, but to a more select group of visitors.

It would be highly insufficient to randomly design setups and test which one is most suitable for the actual installation. There would be no supporting theory as to why a specific design is considered better than the others. It would also cost an unnecessary amount of time to design random setups, and there is a high likelihood that the best possible setup has not been realised. Therefore, a more explorative research to examine possible experiences and develop an underlying theory has been more efficient.

5.2.2. Research value

This research has explored and defined what kind of experiences *Sound Swarm* could offer. It has also included a focused literature review that has been conducted to provide relevant scientific knowledge to optimise the design suggestions and narrow the research field. Although there are no definite design results, the research has provided research recommendations to enable more concrete results concerning a more defined experience. This research has also provided consultation the design of a virtual simulation which could be used to create a realistic experience of *Sound Swarm*, and further advice regarding important aspect concerning the built of the actual art installation, including design suggestions.

6. References

[1] V. Tabry, R. J. Zatorre and P. Voss, "The influence of vision on sound localization abilities in both the horizontal and vertical planes," *Frontiers in Psychology,* vol. 4, p. 932, 2013.

[2] C. J. Plack, "Chapter 9: Spatial Hearing," in *The Sense of Hearing: Second Edition*, New York, Taylor & Francis Group (Psychology Press), 2014, pp. 173-192.

[3] G. Cook and T. Lesoing, "A simple derivation of the Doppler effect for sound," *American Journal of Physics,* vol. 59, p. 218, 1991.

[4] J. C. Middlebrooks and D. M. Green, "Sound Localization by Human Listeners," *Annual Review of Psychology,* vol. 42, pp. 135-159, 1991.

[5] W. S. K. Westerdijk, *Auditory Localization in Art: Exploring the Human Ability of Localizing Sound Sources as Design Fundament,* Enschede: University of Twente, 2016.

[6] S. R. Oldfield and S. P. Parker, "Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions.," *Perception,* vol. 13, no. 5, pp. 581-600, 1984.

[7] J. C. Makous and J. C. Middlebrooks, "Two-dimensional sound localization by human listeners," *The Journal of the Acoustical Society of America,* vol. 87, no. 5, pp. 2188-2200, 1990.

[8] J. C. Middlebrooks and D. M. Green, "Sound Localization by Human Listeners," *Annual Review of Psychology,* vol. 42, pp. 135-159, 1991.

[9] A. J. Hudspeth and M. Konishi, "Auditory neuroscience: Development, transduction, and integration," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 97, no. 22, pp. 11690-11691, 2000.

[10] H. Wallach, E. B. Newman and M. R. Rosenzweig, "The Precedence Effect in Sound Localization," *The American Journal of Psychology,* vol. 62, no. 3, pp. 315-336, 1949.

[11] L. Rayleigh, "On Our Perception of Sound Direction," *Philosophical Magazine Series 6,* vol. 13, no. 74, pp. 214-232, 1907.

[12] T. Sandel and D. Teas, "Localization of Sound from Single and Paired Sources," *The Journal of the Acoustical Society of America,* vol. 27, pp. 842-852, 1955.

[13] B. C. J. Moore, An Introduction to the Psychology of Hearing: Fifth Edition, Academic Press, 2003.

[14] D. W. Grantham, "Interaural intensity discrimination: insensitivity at 1000 Hz," *The Journal of the Acoustical Society of America,* vol. 75, no. 4, pp. 1191-1194, 1984.

[15] D. W. Batteau, "The Role of the Pinna in Human Localization," *Proceedings of the Royal Society of London B,* vol. 168, no. 1011, pp. 158-180, 1967.

[16] S. K. Roffler and R. A. Butler, "Factors That Influence the Localization of Sound in the Vertical Plane," *The Journal of the Acoustical Society of America,* vol. 43, pp. 1255-1259, 1968.

[17] C. C. Pratt, "The spatial character of high and low tones," *Journal of Experimental Psychology,* vol. 13, no. 3, pp. 278-285, 1930.

[18] P. Zahorik, "Direct-to-reverberant energy ratio sensitivity," *The Journal of the Acoustical Society of America,* vol. 112, no. 5Pt1, pp. 2110-2117, 2002b.

[19] V. Tabry, R. J. Zatorre and P. Voss, "The influence of vision on sound localization abilities in both the horizontal and vertical planes," *Frontiers in Psychology,* vol. 4, p. 932, 2013.

[20] B. R. Shelton and C. L. Searle, "The influence of vision on the absolute identification of sound-source position," *Perception & Psychophysics,* vol. 28, pp. 589-596, 1980.

[21] B. E. Stein, M. A. Meredith, W. S. Honeycutt and L. McDade, "Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli," *Journal of Cognitive Neuroscience,* vol. 1, pp. 12-24, 1989.

[22] A. J. King, "Visual influences on auditory spatial learning," *Philosophical Transactions of the Royal Society B,* vol. 364, no. 1515, pp. 331-339, 2009.

[23] D. Alais and D. Burr, "The Ventriloquist Effect Results from Near-Optimal Bimodal Integration," *Current Biology,* vol. 14, no. 3, pp. 257-262, 2004.

[24] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature,* vol. 415, pp. 429-433, 2002.

[25] P. W. Battaglia, R. A. Jacobs and R. N. Aslin, "Bayesian integration of visual and auditory signals for spatial localization," *The Journal of the Optical Society of America A,* vol. 20, no. 7, pp. 1391-1397, 2003.

[26] J. Heron, D. Whitaker and P. V. McGraw, "Sensory uncertainty governs the extent of audio-visual interaction," *Vision Research,* vol. 44, no. 25, pp. 2875-2884, 2004.

[27] N. Cain, "Florian Hecker's Chimerization: Surround sounds," The Wire, January 2013. [Online]. Available: http://www.thewire.co.uk/in-writing/columns/nick-cain-on-florian-hecker_s-chimerization. [Accessed 4 11 2016].

[28] J. Cardiff, "The Forty Part Motet," Cardiff Miller, 2001. [Online]. Available: http://www.cardiffmiller.com/artworks/inst/motet.html#. [Accessed 4 11 2016].

[29] N. Parry, “The Exploded Sound,” Nye Parry, 12 3 2011. [Online]. Available: http://www.nyeparry.com/exploded/. [Accessed 4 11 2016].

[30] N. Parry, “Exploded sounds: spatialised partials in two recent multi-channel installations,” *Divergence Press*, no. 3, 2014.

[31] L. Capece, “Project # 2 - EPOCHÉ- Flying Speakers, Analog Synthesiser, Sine Waves and PA,” 31 12 2015. [Online]. Available: http://luciocapece.blogspot.nl/2012/11/conditional-music.html. [Accessed 30 1 2017].

[32] A. Mader and W. Eggink, “A design process for creative technology,” *16th international conference on engineering & product design education "Design education & human technology relations"*, pp. 568-573, 2014.

[33] Nieuw Dakota, “Organisatie,” Nieuw Dakota, [Online]. Available: http://www.nieuwdakota.com/colofon.php?pool_page_id=7. [Accessed 21 10 2016].

[34] I. I. o. B. Analysis, “MoSCoW Analysis (6.2.5.1.),” in *A Guide to the Business Analysis Body of Knowledge (2 ed.)*, Toronto, International Institute of Business Analysis, 2009, p. 165.

[35] S. W. Smith, “Chapter 22 - Audio Processing / Human Hearing,” in *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Publishing, 1997.

[36] P. Papacosta and N. Linscheid, “The Confirmation of the Inverse Square Law Using Diffraction Gratings,” *The Physics Teacher*, vol. 52, no. 4, p. 243, 2014.

[37] C. Doppler, Über das farbige Licht der Doppelsterne und einiger anderer Gestirne des Himmels, Prague: Borrosch und André, 1842.

[38] F. I. Petrescu, A New Doppler Effect, Norderstedt: Books on Demand GmbH, 2012.

[39] S. Lauesen, Software Requirements: Styles and Techniques, Glasgow: Pearson Education Limited, 2002.

[40] T. Nolinske, “Minimizing Error When Developing Questionnaires,” *To Improve the Academy*, vol. 17, pp. 291-310, 1998.

[41] B. Razavi, W. E. O'Neill and G. D. Paige, “Auditory Spatial Perception Dynamically Realigns with Changing Eye Position,” *Journal of Neuroscience*, vol. 27, no. 38, pp. 10249-10258, 2007.

[42] J. Blauert, in *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, 1997, pp. 122-123.

[43] M. B. Gardner, "Distance Estimation of 0° or Apparent 0°-Oriented Speech Signals in Anechoic Space," *The Journal of the Acoustical Society of America,* vol. 45, no. 1, pp. 47-53, 1969.

Appendices

Appendix A

The following Table 12 is referred to in 3.1.3.1 on page 27. The table is used as reference for the maximum volume per speaker.

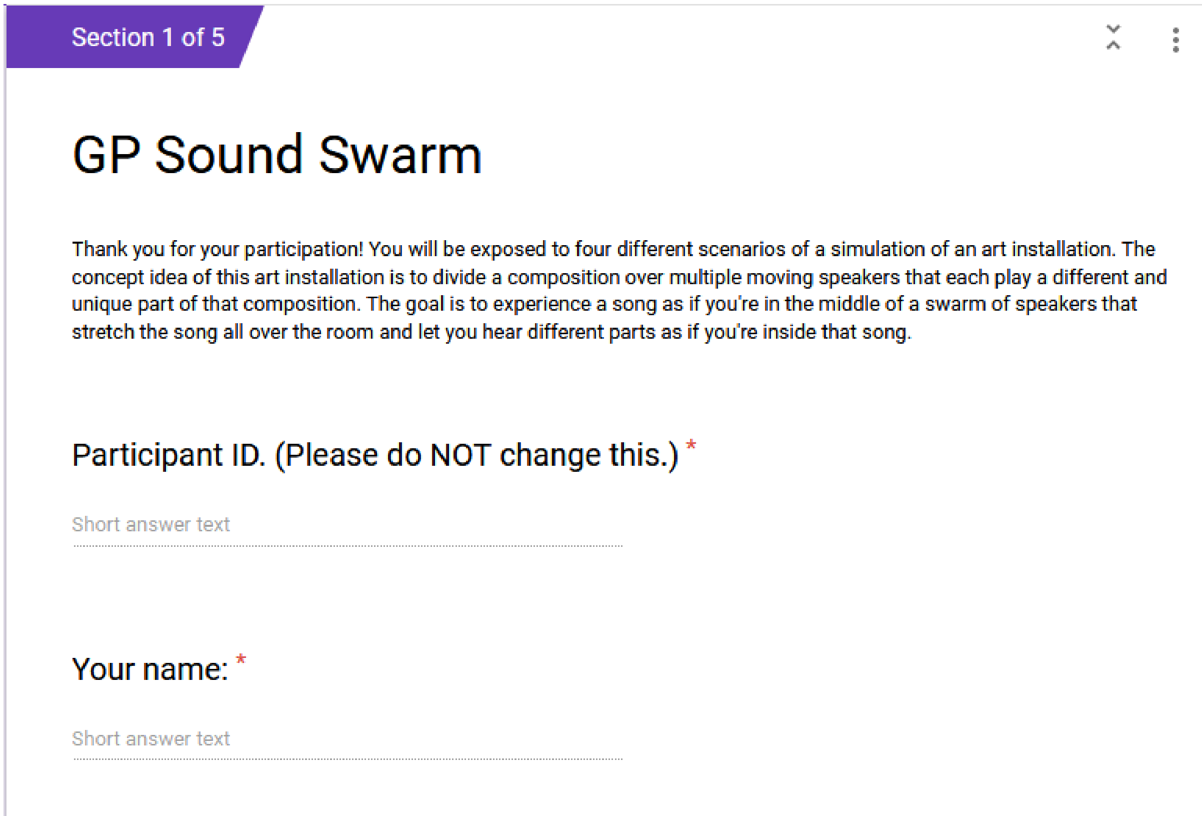
TABLE 22-1
Units of sound intensity. Sound intensity is expressed as power per unit area (such as watts/cm^2), or more commonly on a logarithmic scale called *decibels SPL*. As this table shows, human hearing is the most sensitive between 1 kHz and 4 kHz.

Softer Louder

Watts/cm^2	Decibels SPL	Example sound
10^{-2}	140 dB	Pain
10^{-3}	130 dB	
10^{-4}	120 dB	Discomfort
10^{-5}	110 dB	Jack hammers and rock concerts
10^{-6}	100 dB	
10^{-7}	90 dB	OSHA limit for industrial noise
10^{-8}	80 dB	
10^{-9}	70 dB	
10^{-10}	60 dB	Normal conversation
10^{-11}	50 dB	
10^{-12}	40 dB	Weakest audible at 100 hertz
10^{-13}	30 dB	
10^{-14}	20 dB	Weakest audible at 10kHz
10^{-15}	10 dB	
10^{-16}	0 dB	Weakest audible at 3 kHz
10^{-17}	-10 dB	
10^{-18}	-20 dB	

Table 12: Units of sound intensity and corresponding reference points. Taken from **[35]**.

Appendix B



Section 1 of 5

GP Sound Swarm

Thank you for your participation! You will be exposed to four different scenarios of a simulation of an art installation. The concept idea of this art installation is to divide a composition over multiple moving speakers that each play a different and unique part of that composition. The goal is to experience a song as if you're in the middle of a swarm of speakers that stretch the song all over the room and let you hear different parts as if you're inside that song.

Participant ID. (Please do NOT change this.) *

Short answer text

Your name: *

Short answer text

Figure 14: Screenshot of the introduction section of the questionnaire.

Appendix C

The appendices included in this subsection, are related to the validation experiments, see 3.3.2.1 on page 38.

Appendix C.i

This is the code that is being used in the scenario for validating the simulation.

```
using UnityEngine;
using System.Collections;

public class ValidationMoveSpeaker : MonoBehaviour {
    public Vector3 q1;
    public Vector3 q2;
    public Vector3 q3;
    public Vector3 q4;
    public Vector3 q5;
    public Vector3 q6;
    public Vector3 q7;
    public Vector3 q8;
    public Vector3 q9;
    public Vector3 q0;

    // Use this for initialization
    void Start () {

       }

       // Update is called once per frame
       void Update () {
        if (Input.GetKeyDown(KeyCode.Alpha1)) transform.position = q1;
        else if (Input.GetKeyDown(KeyCode.Alpha2)) transform.position = q2;
        else if (Input.GetKeyDown(KeyCode.Alpha3)) transform.position = q3;
        else if (Input.GetKeyDown(KeyCode.Alpha4)) transform.position = q4;
        else if (Input.GetKeyDown(KeyCode.Alpha5)) transform.position = q5;
        else if (Input.GetKeyDown(KeyCode.Alpha6)) transform.position = q6;
        else if (Input.GetKeyDown(KeyCode.Alpha7)) transform.position = q7;
        else if (Input.GetKeyDown(KeyCode.Alpha8)) transform.position = q8;
        else if (Input.GetKeyDown(KeyCode.Alpha9)) transform.position = q9;
        else if (Input.GetKeyDown(KeyCode.Alpha0)) transform.position = q0;
    }
}
```

Appendix C.ii

The following code is made in Processing, which creates a sketch that is used to enable participants of the validation experiment to mark the sound source.

```
/*
  Original Processing version: 3.2.3

  The code will create a grid, which represents the topview of a room, and a single grid with
four big squares, representing the height.
  The program will show a green circle where the user clicked, and calculates its coordinates
and saves this. It is important to first click the
```

```
  topview and then the height, as the program calculates these differently and changes the cal-
culations based on click-counts.

  Programmed by Wouter Westerdijk
*/

String participant = "Wouter";

PFont font;
int gridAmount = 20;
int gridSize;
int roomSize = 800;
int heightSize = 200;
int clickCount = 0;
int countID;
int saveX = 0, saveY = 0, saveHeight = 0;
boolean click = false;

void setup()
{
    size(1000, 800);
    background(255);
    gridSize = roomSize / gridAmount;
    font = createFont("Ariel", 32);
    textFont(font);
    textAlign(CENTER, CENTER);

    // draw the left grid (horizontal)
    for (int i = 0; i <= roomSize; i += gridSize)
    {
        if (i == roomSize) strokeWeight(4);
        line(0, i, roomSize, i);
        line(i, 0, i, height);
    }

    // draw the circles in the grid
    noStroke();
    fill(0, 50);
    for (int i = 0; i <= roomSize; i += gridSize)
    {
        for (int j = 0; j <= roomSize; j += gridSize)
        {
            ellipse(i, j, gridSize / 1.5, gridSize / 1.5);
        }
    }

    // draw the right grid (vertical)
    stroke(1);
    strokeWeight(1);
    for (int j = 0; j < roomSize; j += (roomSize / 4))
    {
        line(roomSize, j, width, j);
    }

    // draw middle circle
    noStroke();
    fill(220, 0, 10, 190);
    ellipse(width / 2 - heightSize / 2, height / 2, gridSize, gridSize);
```

```
    fill(0, 100);
    text("U", width / 2 - heightSize / 2, height / 2);

    countID = 0;
    noLoop();
}

void draw()
{
    if (click)
    {  //prevents anything happening without clicking (when starting up)

        // draw circle where you clicked
        fill(50, 110, 10, 190);
        ellipse(mouseX, mouseY, gridSize, gridSize);
        fill(0, 100);
        text(countID + 1, mouseX, mouseY);

        // calculate the position in terms of grid
        if (clickCount < 1)
        {
            //convert location points in pixels to same as Unity (10,10 top right, 0,0 being in
the middle, -10,-10 bottom left)
            float flgS = gridSize;  //save gridSize integer as float for correct calculations
            float calcX = round(mouseX / flgS) - 10;
            float calcY = 10 - round(mouseY / flgS);
            saveX = int(calcX);
            saveY = int(calcY);

            clickCount++;
        }
        else if (clickCount == 1)
        {
            float flrS = roomSize;  //save roomSize integer as float for correct calculations
            float calcH = round(4 - (mouseY / (flrS / 4)));
            saveHeight = int(calcH);

            countID++;
            println(participant, "\t", countID, "\t", saveX, "\t", saveY, "\t", saveHeight);
            clickCount = 0;
        }
        click = false;
    }
}

void mousePressed()
{
    click = true;
    redraw();
}
```

Appendix C.iii

Figure 15 shows a screenshot of the code in the previous appendix in use. The lower part shows the console of Processing itself, which outputs the coordinates of the green circles with identical numbers with tab-separated-values, enabling easy copy-paste into a spreadsheet. The columns are: *participant ID*; *speaker location ID*; *X-axis coordinates*; *Z-axis coordinates*; *Y-axis coordinates*.

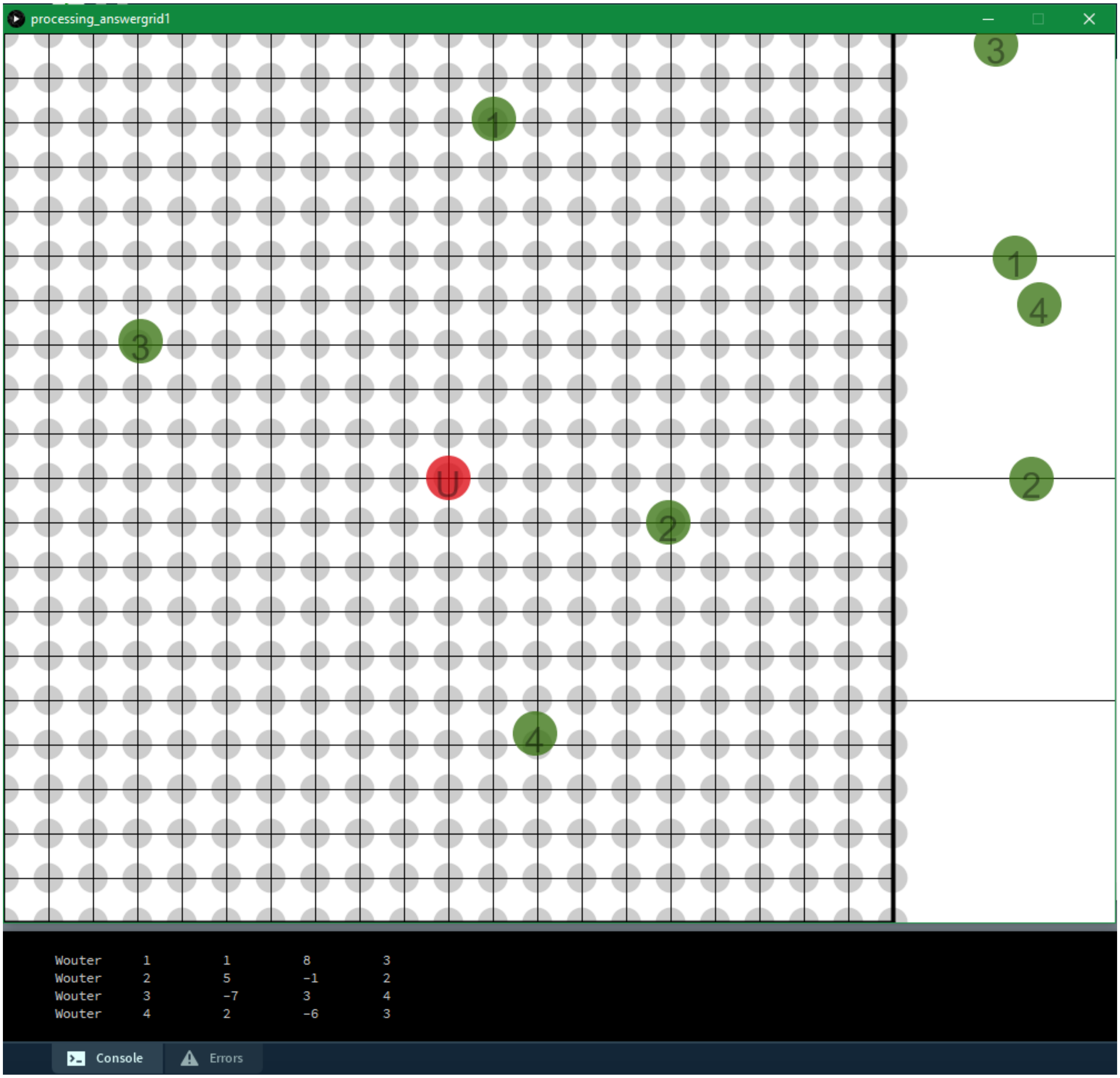


Figure 15: Screenshot of the Processing sketch used in the validation experiment. The inputs are an example and are not related to actual results.

Appendix C.iv

Figure 16 shows a screenshot of the information and questions given with the experiments concerning validating the virtual simulation.

Section 2 of 5

Scenario 1: Validation

You will hear one sound source at a time. Your goal is to use your auditory localization to locate the sound sources. In front of you is a Processing sketch with on the left side a top-down overview of the room with yourself in the middle. This room is 20x20m with every grid 1m. The grey circles are possible locations. In terms of this 'map', you are facing upwards. On the right side is the elevation, every block is 1m (maximum height of 4m). Please always first fill in the left part and then the right. After this, please inform me so I can activate the next position. This will be repeated 10 times. After trying to localize the sources, please fill in the next questions:

Did you encounter any problems trying to locate the sound? Please elaborate. *

Long answer text

Given the fact that this is an empty room of 20x4x20 meters, did the acoustics (such as reverb) come across as realistic? Please elaborate. *

Long answer text

Do you have any other remarks?

Long answer text

Figure 16: Screenshot of the questions regarding the validation tests.

Appendix D

All appendices included in this subsection are related to the grouping experiment in 3.3.2.2 on page 38.

Appendix D.i

This is the code being used in the grouping experiment, scenario one. All four speakers move according to a plus-shaped pattern. The speaker first moves on the X-axis until it has reached its final coordinates (movfree = 2.5, meaning the speakers can move 2.5 meters up, down, left and right relative to the center). Meanwhile, the speaker moves up and down on the Y-axis. Once the speaker, starting at the center, has moved all the way up and down and back to the center, it will move on the Z-axis to repeat the same pattern.

```
using UnityEngine;
using System.Collections;

public class GroupingSolo : MonoBehaviour {

    private Vector3 startPos;
    private bool movx = false;
    private bool movz = true;
    private int count = 0;
    public float movfree = 2.5f; //moving freedom
    public float speedxz = 1f;
    public float speedy = 0.8f;
    public float upperLim = 4f;
    public float lowerLim = 0f;
    public float radius = 0.5f;

       // Use this for initialization
       void Start () {
        startPos = transform.position;
       }

       // Update is called once per frame
       void Update () {
        if(count < 2)
        {
            MoveOnY();
            if (movx) { MoveOnX(); }
            else if (movz) { MoveOnZ(); }
        }
        else if(count == 2)
        {
            transform.position = Vector3.MoveTowards(transform.position, startPos, speedxz *
Time.deltaTime);
            if (transform.position == startPos)
            {
                count = 0;
            }
        }
    }

    void MoveOnX()
    {
        transform.Translate(speedxz * Time.deltaTime, speedy * Time.deltaTime, 0);
```

```
        if(transform.position.x >= startPos.x + movfree)
        {
            transform.position = new Vector3(startPos.x + movfree, transform.position.y, trans-
form.position.z);
            speedxz *= -1;
            count++;
        } else if(transform.position.x <= startPos.x - movfree)
        {
            transform.position = new Vector3(startPos.x - movfree, transform.position.y, trans-
form.position.z);
            speedxz *= -1;
            count++;
        }

        if (count == 2)
        {
            movz = true;
            movx = false;
        }
    }

    void MoveOnZ()
    {
        transform.Translate(0, speedy * Time.deltaTime, speedxz * Time.deltaTime);

        if (transform.position.z >= startPos.z + movfree)
        {
            transform.position = new Vector3(transform.position.x, transform.position.y,
startPos.z + movfree);
            speedxz *= -1;
            count++;
        }
        else if (transform.position.z <= startPos.z - movfree)
        {
            transform.position = new Vector3(transform.position.x, transform.position.y,
startPos.z - movfree);
            speedxz *= -1;
            count++;
        }

        if(count == 2)
        {
            movz = false;
            movx = true;
        }
    }

    void MoveOnY()
    {
        if (transform.position.y >= upperLim - radius)
        {
            transform.position = new Vector3(transform.position.x, upperLim - radius, trans-
form.position.z);
            speedy *= -1;

        }
        else if (transform.position.y <= lowerLim + radius)
        {
```

```
            transform.position = new Vector3(transform.position.x, lowerLim + radius, trans-
form.position.z);
            speedy *= -1;
        }
    }
}
```

Appendix D.ii

This code is being used by the middle speaker in every corner, which is the target speaker for the four speakers surrounding it. This speaker will only move up and down.

```
using UnityEngine;
using System.Collections;

public class GroupingMasterMov : MonoBehaviour {
    public float upperLimit = 4;
    public float lowerLimit = 0;
    public float radius = 0.5f;
    public float speed = 1f;

       // Use this for initialization
       void Start () {

       }

       // Update is called once per frame
       void Update () {
        transform.Translate(0, speed * Time.deltaTime, 0);

        if(transform.position.y >= upperLimit - radius/2)
        {
            transform.position = new Vector3(transform.position.x, upperLimit - radius/2, trans-
form.position.z);
            speed *= -1;
        } else if(transform.position.y <= lowerLimit + radius/2)
        {
            transform.position = new Vector3(transform.position.x, lowerLimit + radius/2, trans-
form.position.z);
            speed *= -1;
        }
       }
}
```

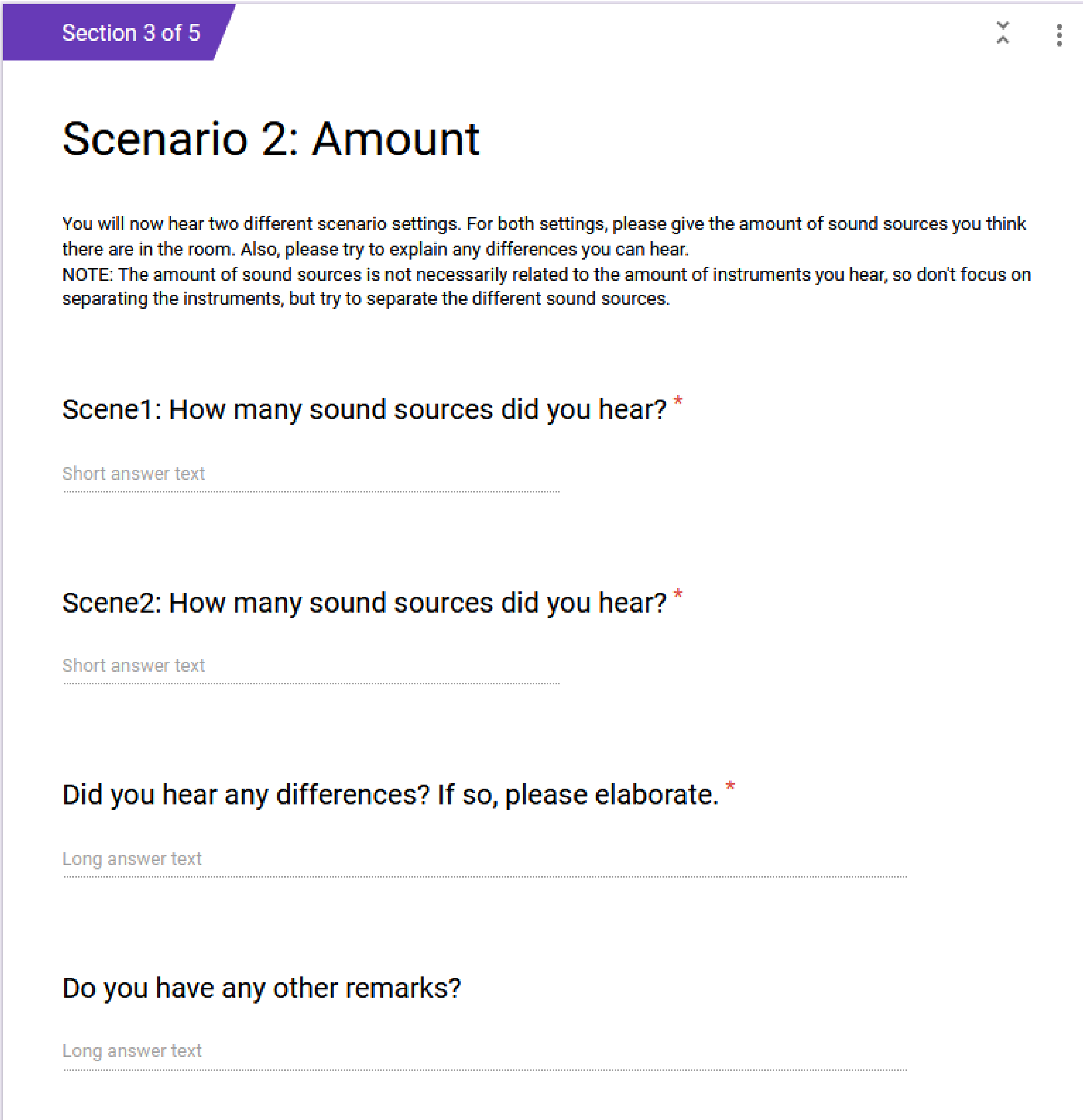
Appendix D.iii

This code is being used for the four speakers surrounding the middle speaker. The speakers move towards the middle speaker and back to their original coordinates.

```
using UnityEngine;
using System.Collections;

public class GroupingSlaveMov : MonoBehaviour
{

    public Transform target;
```

```
    private Vector3 startingPoint;
    private bool returning = false;
    public float speed = 0.9f;
    public float radius = 0.5f;

    // Use this for initialization
    void Start()
    {
        startingPoint = transform.position;
    }

    // Update is called once per frame
    void Update()
    {
        if(!returning)
        {
            TowardsTarget();
        } else if (returning)
        {
            FromTarget();
        }
    }

    void TowardsTarget()
    {
        transform.position = Vector3.MoveTowards(transform.position, target.position, speed *
Time.deltaTime);
        if (Mathf.Abs(Vector3.Distance(target.position, transform.position)) <= radius)
        {
            returning = true;
        }
    }

    void FromTarget()
    {
        transform.position = Vector3.MoveTowards(transform.position, startingPoint, speed *
Time.deltaTime);

        if (transform.position == startingPoint)
        {
            returning = false;
        }
    }
}
```

Appendix D.iv



Section 3 of 5

Scenario 2: Amount

You will now hear two different scenario settings. For both settings, please give the amount of sound sources you think there are in the room. Also, please try to explain any differences you can hear.
NOTE: The amount of sound sources is not necessarily related to the amount of instruments you hear, so don't focus on separating the instruments, but try to separate the different sound sources.

Scene1: How many sound sources did you hear? *

Short answer text

Scene2: How many sound sources did you hear? *

Short answer text

Did you hear any differences? If so, please elaborate. *

Long answer text

Do you have any other remarks?

Long answer text

Figure 17: Screenshot of the questions regarding the grouping experiment.

Appendix E

All appendices included in this subsection are related to the experiment on the division of sound, which is described in 3.3.2.3 on page 39.

Appendix E.i

The following code is used to make the speaker move to a random location inside the room. It moves from its starting point to a target point, and once it gets there, the target point is set as starting point and a new target point is randomly created.

```
using UnityEngine;
using System.Collections;

public class RandomMovement : MonoBehaviour {

    private Vector3 targetPoint;
    public Transform target;
    public float speed = 2f;
    public int sizeX;
    public int sizeY;
    public int sizeZ;
    public float radius = 1f;

       // Use this for initialization
       void Start () {
        setNewTargetPoint();
    }

       // Update is called once per frame
       void Update () {

        transform.position = Vector3.MoveTowards(transform.position, target.position, speed *
Time.deltaTime);

        if (transform.position == target.position)
        {
            setNewTargetPoint();
        }
       }

    void setNewTargetPoint()
    {
        float new_x = Random.Range(-sizeX / 2 + radius, sizeX / 2 - radius);
        float new_y = Random.Range(0 + radius, sizeY - radius);
        float new_z = Random.Range(-sizeZ / 2 + radius, sizeZ / 2 - radius);
        targetPoint.Set(new_x, new_y, new_z);
        target.position = targetPoint;
        //print(targetPoint);
    }
}
```

Appendix E.ii

Section 4 of 5

Scenario 3: How

You are now able to also see the simulation and you can use the mouse to look around. You will hear two different scenarios. Can you notice and explain the difference(s), and explain with one you'd prefer?
Next to that, please pay attention to the speed of the speakers.

Scene1: How would you describe your experience? *

Long answer text

Scene1: Do you think that adding or removing speakers would increase your experience? *

- Yes, MORE speakers would increase the intensity of the experience.
- Yes, LESS speakers would increase the intensity of the experience.
- No, it's just right.
- I have no clue.

Scene2: How would you describe your experience? *

Long answer text

Figure 18: Screenshot of the questions regarding the division of sound experiment (1/2).

Scene2: Do you think that adding or removing speakers would increase your experience? *

- Yes, MORE speakers would increase the intensity of the experience.
- Yes, LESS speakers would increase the intensity of the experience.
- No, it's just right.
- I have no clue.

What did you think of the speed of the speakers? Would you prefer them going faster, slower or was it about right? Also, please mention if this applies to both scenes, or not. *

Long answer text

Compare: which one has your preference and why? *

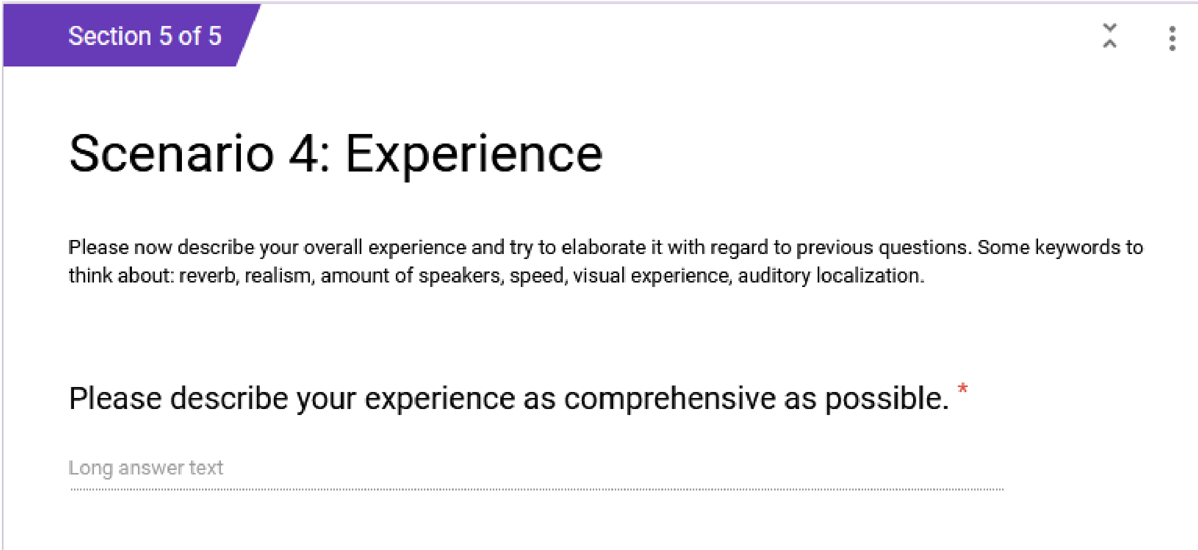
Long answer text

Do you have any other remarks?

Long answer text

Figure 19: Screenshot of the questions regarding the division of sound experiment (2/2).

Appendix F


Section 5 of 5

Scenario 4: Experience

Please now describe your overall experience and try to elaborate it with regard to previous questions. Some keywords to think about: reverb, realism, amount of speakers, speed, visual experience, auditory localization.

Please describe your experience as comprehensive as possible. *

Long answer text

Figure 20: Screenshot of the question regarding the experience experiment.

Appendix G

All subsections included in this appendix are related to the results, discussed in 3.3.3 starting on page 41.

Appendix G.i

Did you encounter any problems trying to locate the sound? Please elaborate.

9 responses

yes, sometimes I found it difficult. mainly because turning my head did not help

Yes it was difficult to estimate how far the sound was

Yes, the interface with the selecting the height is unclear
Would make more sense to select it in a 3d space
Overall choosing a direction wasn't hard to figure out but the height was

I found height very hard to differentiate between. I also thought everything was from behind me.

The height was difficult to figure out

It was especially hard to set a baseline of how far the first few were since you don't know how loud it gets, afterwards it became easier to compare to compare what you heard to sounds you had heard before.

I was trying to locate the sound by turning my head, this however was not working. Furthermore I found it really hard to determine the height of the sound and whether the sound was coming from the front or the back.

Of course but I think it is because you are trying to imagine an environment while you are trying to locate the sound. Then you are trying to guess the where abouts of the speaker. I think the harder part for me was to determine the sound level rather than the direction because I believe as the direction changes the volume is affected as well

It was difficult to reference how far the sound away was, because you did not know how loud the music would be from 20 metres away. Left and right went fine, but height was also difficult to determine.

Figure 21: Screenshot of the answers regarding problems encountered during the validation experiment.

Given the fact that this is an empty room of 20x4x20 meters, did the acoustics (such as reverb) come across as realistic? Please elaborate.

9 responses

one time, it sounded like it music came from another room. but usually it was realistic

yes it sounded like i was i a normal room

Yes

Yes, there was an unexpected sense of closeness or far away

Yes, very very

Yes, it was subtle but you could definitely hear some reverb on the other side when you knew where it was coming from. It did sound like the room walls were soft, like they absorbed some of the sound.

I did not really notice any reverb. However this might also mean that it was realistic because I did not notice anything missing

Yes it was realistic, in fact it sounded a little bit like, the background noise that you hear when you are leaving from a concert area

Yes, sometimes the sounds sounded as if they were right next to you and sometimes really far away. It was easy to notice the difference and sounded pretty realistic.

Figure 22: Screenshot of the answers regarding the room acoustics of the validation experiment.

Appendix G.ii

Did you hear any differences? If so, please elaborate.

9 responses

the speakers moved in the second scene
Yes the bass was less loud and there was less depth in the music
Yes, the second scene seemed to have more fluctuation in the tss woeing sound
The sounds felt further away from each other the second time
First one moved around more.
The second scenario sounds less dynamic so less speakers, but it seems to move around you.
During the second scene the sound was moving, therefor it was really hard to determine how many sources there were.
In the first scene I felt like the standard sounds moved through the different sound sources. I think there was a sound source right behind me that had the lower base sound. However in the second one compared to the first scene there were more sounds included I tried to add to the number of sources from scene 1. This makes me think that there were definitely more sound sources in the second scene.
The music sounded the same, but in scene 2 the sources were probably further away from each other so it was easier to differentiate.

Figure 23: Screenshot of the qualitative answers regarding the differences between the grouping scenarios.

Appendix G.iii

Scene1: How would you describe your experience?

9 responses

the fact that you are not able to move freely feels a bit claustrophobic, but it was nice to see the speakers moving around
it was very interesting to see the sound boxes move and hear the differences.
Painfull, and confusing mostly because the sound sources give to much input
It's a lot of fun and really cool
Super cool and I'm speachless. My ears are confused
It was remarkably pleasant compared to usual audio experiences, the different intensities of the instruments while maintaining one song sounded nice.
It's like I am standing in the middle of a swarm of bees. However it does sound really interesting and different.
I think this is what vr systems need. Sound vr system on point!
Really nice, the sounds really sound close by when they move past you and when you turn the soundsource also changes in direction so to say.

Figure 24: Screenshot of the experience description concerning the division of sound experiment, scenario 1.

Scene2: How would you describe your experience?

9 responses

same, only the speakers sounded "off"
there was a bigger difference in soudn when the boxes moved. It was still cool tho
See input for Scene 1, although this scene seemed to be even more chaotic
Very much like the first time, but the high notes seem more prominent this time
Les pleasing, don't mess with the Treble and Mid
It felt like there were less speakers and the quality of the individual speakers seemed to have gone down as well. Overall still a pleasant experience.
I found this one more pleasing. It felt like only the lower tones were crossing me and therefore it did not feel like a swarm of bees. This was way more relaxing
More distributed and messy and faster sound vr
Quite the same as scene 1 but the sounds seemed as if they move underwater if they move behind you or further away.

Figure 25: Screenshot of the experience description concerning the division of sound experiment, scenario 2.

Appendix G.iv

Compare: which one has your preference and why?

9 responses

the second one has a better speaker density, but the sound was nicer in scene 1
the first one because it had more difference
None, because both were not pleasant to listen to
The first one, the sound seems 'fuller'.
First one, sick experience without weird added effects. Just pure fun
The first scenario
The second, since this one had a lower tone and therefore was more relaxing
Scene 2 was more unorganized, as in the sound level and the speed of speakers were much better, scene 2 almost makes you a little bit dizzy
The first one, because I don't really like the dampened sound of scene 2.

Figure 26: Screenshot of the answers to which scenario had the participants' preference concerning the division of sound experiment.

Appendix G.v

What did you think of the speed of the speakers? Would you prefer them going faster, slower or was it about right? Also, please mention if this applies to both scenes, or not.

9 responses

a bit slower would be nice, it would be nice i you could follow the speakers with your head, now they are too fast in both scenes

it think it was right for both the setting, if you have more speaker you can change the speed . I think you can make it faster

Maybe faster, but I should be able to compare it before I can give certain answer

If you want to make music I'd let them go slower but if you make a soundscape/game background sounds it really depends on what you want to convey.

Maybe have different speakers at different speeds.

In the first scene, they seemed to go faster and group up more, so there was only one big variance in loudness instead of like in the first one, where it seemed like different instruments had different intensities depending on their distance.

I would like to try with slower speakers, see whether that enhances the relaxing vibe. When looking at the speakers I feel as though they are moving way to fast for the tune they are playing. However when listening I feel it being just fine.

Scene 1 was about right, scene 2 however was fast.

A little bit too fast for my preference, but not perceived as annoying. This applies for both scenes.

Figure 27: Screenshot of the answers concerning the speed of speakers during the division of sound experiment.

Appendix G.vi

Please describe your experience as comprehensive as possible.

9 responses

it was a nice experience overall, but the reverb was too big to have a clean experience. the amount of speakers was nice, but the tempo was of the speakers was still a bit fast. visually, it looks like a prototype (which it probably is), but the sounds are of good quality. the localization was very nice because you could see the speakers for yourself and follow them.

It looked llike there were more sources. The realism was pretty high. I only think that they speakers cant move like this normally because they go through you. The localization is easier with visuals

More sound sources actually made it easier to listen to, because the position of sound came from is more often filled with another sound unit

The amount of speakers is larger than you can easily distinguish, especially since all the sounds sound quite similar. You can only really hear the difference when a speaker passes close by. I think this system would be really good in video games. It felt very realistic, although the difference between what is in front and behind you I find hard to hear. The visual experience mattered a lot in this regard, only when combining the two you get a real sense of spatial awareness.

As if your stoned as fuck and in a disco high on laughing gas

This final scenario seemed to have the same amount of speakers as the first scene from scenario 3, there seemed to be more reverb as well. The audio quality seemed higher, the speakers didn't group up as much as the second scene from scenario 3. Personally the best experience so far, closely followed by the first scene from 3. Room did seem darker than scenario 3.

It sounds really good. However it looks way too messy, the amount of speakers and the speed at which they are going doesn't match the style at all. I do not know whether the amount of speakers also effects this, I think that if they are going a lot slower it might not look as messy and then the amount would be good. I really feel like the speakers are passing right by me, although the visuals are also helping at lot. The higher tones sometimes sound like a razor cutting of the hair off the back of my head. You only can really locate the speakers when they are getting close to you, otherwise it sounds like the sound is just all around you (which it also basically is). I cannot really hear any reverb, since the sound is coming from all around and there are so many sound sources.

It is the perfect combination of scenes 1 and 2 from the previous experience. It is almost like when you close your eyes you can keep track and move with the speakers which I think is really cool.

A really nice immersive experience, coupled with VR this would be really nice because while the sound is immersive, there are no nice visuals associated with it. The speed is still a little bit too fast for my taste, but the amount of speakers seems right. The auditory localization is easy in left-right and front-back direction, but I think the difference in height is (at least for me) too difficult too distinguish.

Figure 28: Screenshot of the experience descriptions concerning the experience experiments.