Design and Evaluation of different Auditory Mappings for 3D Sound as Off-screen Technique for Head-mounted Augmented Reality Displays

Bachelor Thesis

submitted by

Johanna Bell (01/830337)

at the

Universität Konstanz

Human-Computer Interaction

Department of Computer and Information Science

1. Reviewer: Prof. Dr. Harald Reiterer

2. Reviewer: Jun.-Prof. Dr. Tiare Feuchtner

Konstanz, 1. August 2022

Abstract

Augmented reality head-mounted displays currently only offer a very small field of view, which means that objects that are located outside the field of view are not perceived by the user. Since humans can perceive their environment with different senses, the question here is whether auditory perception can also be used in augmented reality head-mounted displays to perceive information from objects outside the visual field of view of the device in a natural way, as we do in everyday activities.

This thesis describes the design and evaluation of three different auditory mappings used for 3D sound as off-screen technique for head-mounted augmented reality displays. First, research concerning related work was conducted. Thereby, a research gap in comparing different auditory mappings for 3D Sound as off-screen technique was identified. Therefore, a study prototype was designed in which important virtual objects are able to emit three different auditory mappings (nomic, metaphorical and symbolic) as 3D sound. Then, a user study was conducted with eighteen participants guided by research questions about search duration, accuracy in finding the correct object, object location memory and subjective preferences and perceptions. The results of the study are described and then discussed in relation to the research questions. The thesis ends with conclusions from the conducted evaluation. The main conclusions of the thesis are that overall, the use of the nomic auditory mapping for 3D sound as off-screen technique for head-mounted augmented reality displays is the most suitable for the tasks performed in the study, while the symbolic auditory mapping is the least suitable.

Contents

List of Figures

List of Tables

1 Introduction

Research has heavily studied the use of technologies such as virtual and augmented reality for individual and collaborative activities. During the last years, prices for consumer-grade products are steadily decreasing, allowing more and more people to purchase virtual and augmented reality head-mounted displays. But all these technologies have many challenges that need to be solved. As in real life, objects that are outside our visual field of view (FOV) or occluded cannot be perceived visually. In augmented reality, this is even more a problem, as augmented reality head-mounted displays, e.g., the HoloLens 2, currently only offer a relatively small field of view and objects are therefore more likely off-screen. Therefore, it is important to find a solution that allows users to also perceive these objects that are not visible. Furthermore, users should be able to orient themselves quickly and navigate confidently as well as efficiently through the environment, which makes it essential to provide navigational aids [1].

To solve these problems, various off-screen techniques have been developed. However, almost all of them are visual off-screen techniques. While these are already very helpful in perceiving objects outside the field of view, they still have some disadvantages. For example, if they are continuously visible, they not only waste space from the field of view but can also be distracting when they are not needed. One technique, which has been used only rarely or in combination with visual techniques, is the use of auditory cues as an off-screen technique. But especially the auditory techniques are the ones that could avoid some of the disadvantages of visual techniques. Auditory cues can be perceived easily at a given time, so there is no need to waste screen space on displaying a visual cue, and there is no need to constantly pay attention to the visual cues to avoid missing anything. In addition, auditory cues are natural, so people are already innately familiar with them [1]. Therefore, one might assume that auditory cues, in contrast to visual cues, allows users to perceive objects in an intuitive and non-obtrusive way.

Since there are currently only few studies on auditory off-screen techniques, it is important to find out how this technique could be used most efficiently. Among the aspects that this thesis addresses is the question of which auditory mapping (see Table 2.1) is best suited to help users locate invisible objects as quickly, easily, and accurately as possible, as well as to build up the best possible knowledge about the positions of the various objects. Therefore, this thesis designs, implements, and evaluates the three different auditory mappings (see Table 2.1) for 3D sound as off-screen technique for head-mounted augmented reality displays. The content of this work is divided into eight chapters.

This first chapter introduces the motivation and content of this work. The second chapter covers the theoretical backgrounds of the physiology of auditory perception and the different auditory mappings. The third chapter presents an overview of related papers about various studies that have used 3D sound as an auditory off-screen technique in augmented or virtual reality environments. In the fourth chapter, the prototype of the study is presented and an overview of its design and development is given. The fifth chapter explains the conduction and process of the user study and provides an overview of the used data collection techniques. The sixth chapter shows the results which could be obtained from the different data collection techniques. In the seventh chapter, the study results are discussed in relation to the research questions. The eighth and last chapter presents the conclusion of the thesis.

2 Theoretical Background

In the following, short introductions will be given on the topics physiology of auditory perception and auditory mappings. The section "Physiology of auditory perception" focuses on why and how humans can hear spatially. In the subsequent section on "Auditory Mappings," different types of auditory mappings and examples of them are given.

2.1 Physiology of auditory perception

Spatial hearing is made possible by three different effects, as Butz et al. [2] describe in their book. If the sound signal is not exactly at the same distance from both ears, i.e., neither above, below, in the middle in front of or behind us, it must take different paths to our ears [2].

The **Interaural intensity difference (IID)** is the first effect and can occur due to differences in loudness of the incoming sound at both ears [2]. For example, if the sound signal comes from the right side, the signal arrives much louder at the right ear than at the left ear, because the sound wave is shadowed by the head shadow. (see Figure 2.1).

The **Interaural time difference (ITD)** is the second effect and can be caused by the time difference between the arrivals of the sound at both ears [2]. For example, if the sound signal is closer to the right ear, it arrives significantly later at the left ear than at the right ear (see Figure 2.2).

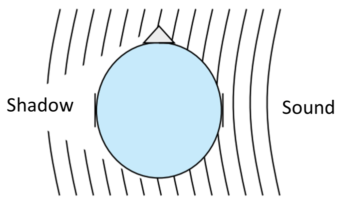


Figure 2.1: Spatial hearing with the help of IID. Sound arrives louder at the right ear than at the left ear. *Taken and adapted from [2]*

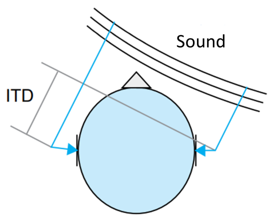


Figure 2.2: Spatial hearing with the help of ITD. Sound arrives earlier at the right ear than at the left ear. *Taken and adapted from [2]*

With IID and ITD, we are only able to distinguish between sound signals that come from the right or left side and only those that do not have a low frequency.

The third effect is the **Head-related transfer function (HRTF)** and enables us to determine the sound change depending on the direction and frequency of the sound. If, for example, the sound signal comes from behind the ear, it gets damped by the hair and the shape of the ear conch. However, each person has a unique HRTF because each person has their own geometric and material characteristics of the ears and head. Nevertheless, it is possible to determine an averaged HRTF that is approximately correct for many people and with which one can also distinguish sounds from every direction [2].

2.2 Auditory Mappings

Sounds can help us to distinguish between objects, activities and their properties. A sound has different parameters that can be changed. These include, for example, the frequency range, the volume range and the directionality.

There are simple sounds, which are tones that have only one frequency and are played for a certain time, or several tones that are played in a sequence. An example would be a "beep" sound, which can be used as a warning signal. Furthermore, there are designed sounds, which are used to convey information to the user. Designed sounds include auditory icons, earcons and compositions/music [3].

Gaver [4] defines auditory icons as informative sounds. Brewster et al. [5] add that Gaver's auditory icons, which have now been used in several systems, are environmental (real) sounds that have a semantic relationship to the represented object or action. Earcons, in contrast, are more general. Blattner et al. [3] define earcons as structured sounds that are not verbal but can consist of both synthetic and real sounds. Earcons can be used to assign a certain sound to a particular action or object, which then makes it easier to provide information to the user about these actions and objects.

In general, earcons are used for synthetic sounds and auditory icons for sounds that are recorded and played back, i.e. environmental (real) sound.

In order to better compare the use of natural sounds, i.e. auditory icons, versus synthesised sounds, i.e. earcons, Gaver [4] has divided sound into the kind of mapping between the data to be represented and the means used to represent it. As shown in Table 2.1, he has divided sound into symbolic, nomic and metaphorical sound. The given example describes what the user hears when searching for a trash can.

Mapping	Description	Example
Symbolic	The sound is arbitrary for the object or action it represents.	The user hears a "beep" sound.
Nomic	The sound is directly related to the object or action it represents.	The user hears the same sound that occurs when something is thrown into a trash can.
Metaphorical	The sound is a designed sound, which is inspired by real world experiences. Therefore, the sound is not wholly arbitrary like the symbolic sounds but also not wholly natural like the nomic sounds.	The user hears the crumpling of paper as a sound.

Table 2.1: Auditory representation system characterized by the kind of mapping between the data to be represented and the means used to represent it [4].

3 Related Work

Most off-screen techniques are visual off-screen techniques. Among them there are, for example, halos [6] or arrows [7], which help the user to perceive 2D or 3D objects in the environment that are outside of the user's field of view. With the Halo [6] technique, relevant objects and locations are circled by a ring, which is just large enough to reach inside the FOV of the used device. In this way, the user can estimate where the center of the circle, and thus the object, is located. With the Arrow [7] technique, an arrow is displayed in the FOV of the device, pointing in the direction where the relevant object is located. While these techniques are already very helpful in perceiving objects outside the field of view, they still have some disadvantages. For example, if they are continuously visible, they not only waste space from the field of view but can also be distracting when they are not needed. Therefore, it is necessary to find other solutions to provide the best possible user experience.

One solution could be the use of the auditory off-screen technique. Thereby, the user should be able to perceive invisible 2D or 3D objects with the help of a 3D sound. In the last chapter the basics of 3D sound were presented. In this chapter, three works will be presented, which have investigated the effect and benefit of 3D sound in virtual reality (VR) or augmented reality (AR).

3.1 The effects of 3D Sound on mixed reality remote collaboration

Yang et al. [8] describe in their paper *"The effects of spatial auditory and visual cues on mixed reality remote collaboration"* the conduction of a study in a remote collaboration system. The participant as the local worker had to search as quickly and as accurately as possible for a real Lego Brick inside a office. During the search, the remote worker gave the local worker instructions on where the target Lego Brick could be. The remote worker used a virtual reality head-mounted display (HMD) during the studies, which showed both the office and the position of the local worker moving around in it. He was able to teleport his virtual avatar in the virtual environment and thus navigate to Lego bricks and activate the sound on them. The local worker, on the other hand, used an augmented reality HMD. Besides the instructions, the participant additionally was able to see virtual visual cues for perceiving the positions of the objects under certain conditions, such as the head or hand of the remote worker's avatar and his FOV (see Figure 3.1). Furthermore, the local worker was able to hear one sound emitting from the target Lego Brick. The sound that came from the Lego Brick was a designed 2s long looping wide-band musical 3D sound. Regarding table 2.1, this sound could be classified as symbolic sound. During the study, the Lego Bricks were placed in such a way that they were either covered by another object or barely visible due to their colours.

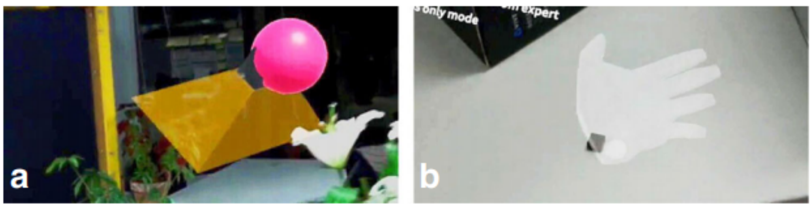


Figure 3.1: On the left picture, the remote workers head frustum, including his FOV can be seen, while the right picture shows the hand gestures of the remote worker. *Taken from [8]*

During their study Yang et al. [8] measured the task completion time, the social presence and the spatial presence of the participant. As conditions they used, that either the local worker heard just the instructions from the remote worker, or that he heard 3D sound or 2D sound from the Lego Bricks, or that the participant saw some of the visual cues in addition to the 3D sound.

The results for the task completion time show, that there was no difference between using 3D sound or 2D sound. However, the task completion time was significantly improved when the visual cues were included.

After the task completion Yang et al. [8] conducted a survey with the participants to indicate how strong the feeling of social presence and spatial presence was during the study. The results show, that the participants sense of social presence was generally strong with the auditory cues, but is not enhanced by the use of 3D sound as opposed to 2D sound. However, the results show that the social presence could be significantly increased by adding the remote workers head frustum.

In contrast to the social presence and task completion time, the use of 3D sound compared to 2D sound significantly increased the spatial presence of the participants. Furthermore, the use of visual cues, especially the use of the head frustum, could again clearly increase the spatial presence.

3.2 Usefulness of 3D Sound for searching and navigating through AR environments

Ruminski [9] describes in the paper *"Spatial Sound Localization in an Augmented Reality Environment"* the conduction of a study in an augmented reality environment which was displayed on a mobile phone. The participants had to locate certain virtual animals which were virtual objects and only visible when the participant held the mobile phone close enough to the animal. Thus, no visual cues were given. Instead, some of the participants received headphones that played the 3D sounds of the animals. These sounds were directly assigned to the sound of the animal, e.g. a pig makes "grunting" noises. Regarding Table 2.1, this sound could be classified as nomic sound. Depending on the position of the mobile phone, it was also possible to hear several animals at the same time.

In the study by Ruminski [9] were two groups of participants. The first group was the "3D sound" group, here the participants were able to hear the auditory cues while searching for the animals. The second group was the "control group", there the participants were not able to hear any auditory cues.

During the study, Ruminski [9] measured the task completion time of the participants and the results show, that the use of 3D sound compared to no sound can lead to a significant improvement in the task completion time.

Furthermore, most of the participants reported that they think that the use of 3D sound was a significant factor that contributed to the task completion time for the "3D sound" group.

3.3 3D Sound as a navigational aid in VR

In the third paper *"Using 3D sound as a navigational aid in virtual environments"* by Gunther et al. [1] the study was conducted inside a virtual reality environment. The participants wore virtual reality HMDs and were able to move around in a virtual environment consisting of twelve different rooms. There were markers and symbols on the walls outside the rooms to help with the orientation.

The rooms contained different objects, and one object in each room was able to emit a 3D sound, which the participant could hear if he was close enough to the room. For example, in one room there was a clock that made a clock ticking sound. Regarding Table 2.1, this sound could be classified as nomic sound. If the participant was between two rooms, he could hear the sounds of both objects from both rooms, so it was also possible to hear several objects at the same time. The aim was to find specific objects that could make a sound, starting from a given point. Visual cues were provided in so far that the virtual objects were visible when they were within the FOV of the device.

The study from Gunther at al. [1] consisted of two blocks of test rounds and three test groups. The first test group was the "FullSound" group. The participants in that group, were always able to hear the auditory cues. The second test group was the "PartialSound" group, in which participants only heard the auditory cues in the first round of testing. The last group was the "NoSound" group and there the participants were never able to hear the auditory cues.

Gunther et al. [1] measured the participants task completion time and spatial knowledge.

For results for the task completion time show that the "FullSound" group was always one of the groups that could find the objects the fastest in both test rounds. The "NoSound" group, on the other hand, took longer to find the objects in the first round of testing, but the participants in this group were able to achieve a certain learning effect, so that in the second round they were able to find the desired objects almost as quickly as the "FullSound" group. The "PartialSound" group was able to complete the tasks just as quickly as the "FullSound" group in the first round of testing, but their performance dropped significantly in the second round, in which they no longer received auditory cues anymore.

After the participants from the three test groups had found all the target objects in the virtual environment, they individually were asked which auditory or non-auditory objects they could remember. The participants did not know beforehand that they would be asked this question. In this way, a part of the spatial knowledge (the landmark knowledge) could be found out. The participants of the "FullSound" group were the ones who could recall the fewest objects.

Another part of the spatial knowledge (the survey knowledge) was found out, when the participants were asked to place the before recalled objects in the correct corresponding room. For this purpose they received a map with the 12 rooms of the virtual environment and then they were asked to indicate for each object in which room it was located. The participants correctly placed many more target objects that they had to search for beforehand than those they did not have to search for. In general, the "FullSound" group placed the fewest objects correctly.

One reason, mentioned by Gunther et al. [1], why the "NoSound" group had gained better spatial knowledge than the "FullSound" group could be that the participants from the "NoSound" group had to concentrate much more on the environment and where things were located than the "FullSound" group. The "FullSound" group could always rely on the auditory cues and thus had to concentrate less on the general environment. This could also be the reason, why the task completion time of the "NoSound" group has improved in the second test round, since they could already gain a certain basic understanding of the environment. The "PartialSound" group, on the other hand, was as lost in the virtual environment in the second test round as the "NoSound" group was before and had to get to know the environment first, as they could no longer rely on the auditory cues.

3.4 Conclusion

The three studies from the previous sections evaluated how suitable 3D sound is for searching and navigating to nonvisible objects inside a virtual or augmented environments. The study by Yang et al. [8] showed that 3D sound combined with visual cues produced the best results in searching and navigating. Furthermore, it was also found that the participants' sense of spatial presence could be clearly increased by the use of 3D sound. However, the participants' actual spatial knowledge was clearly reduced when using 3D sound compared to no sound, as Gunther et al. [1] showed in their study. Both Ruminski [9] and Gunther et al. [1] found out that 3D sound compared to no sound, provides a significant advantage in searching and navigating through virtual and augmented environments.

Table 3.1 shows the corresponding conditions used in the studies by Gunther et al. [1], Ruminski [9] and Yang et al. [8]. As can be seen, in each study 3D sound was tested either in comparison to no sound, only partial sound, 2D sound, or 3D sound plus visual cues. However, each time only one kind of auditory mapping was used for the objects with 3D sound. Furthermore, in each of these studies either only one object or more objects were heard at the same time, but both conditions were never given.

Conditions:		**Gunther et al. [1]**	**Ruminski [9]**	**Yang et al. [8]**
Auditory mapping	Symbolic			✓
	Nomic	✓	✓	
	Metaphorical			
Number of sounds at a time	One sound at a time			✓
	More sounds at a time	✓	✓	
Further Conditions	3D sound	✓	✓	✓
	2D sound			✓
	Partial 3D sound	✓		
	3D sound + visual cues			✓
	No sound	✓	✓	✓

Table 3.1: This table shows which conditions were used for each of the studies in the presented auditory off-screen technique papers.

Based on these findings, the goal would be to develop and evaluate a prototype which on the one hand can emit different auditory mappings, and on the other hand it should be possible to hear either only one sound or several sounds at a time. However, due to time constraints of this work, it has not been possible to realize both conditions. Therefore, it was decided that the goal of the prototype presented in the following chapter is to emit symbolic, metaphorical, and nomic sounds, but to hear only one sound at a time. In a subsequent study, it was evaluated which kind of auditory mapping would be best suited in an augmented reality environment to find invisible objects as quickly, easily, and intuitively as possible and to gain spatial knowledge about the position of the virtual objects.

4 Prototype

This chapter provides an overview of the study prototype. First, the requirements for the prototype are defined. Then, the technologies used are presented. The final section "Study Prototype" presents the prototype, its design and the development of the concepts.

4.1 Requirements

Since the prototype from the bachelor project still had to be modified for the study, the adapted requirements for the study were specified here, which means that they differ from those from the precedent bachelor project. Therefore, the following requirements are based on the findings from related work, the objective of finding out what kind of sounds are best suited for augmented reality applications and the study-relevant adaptions. In order to allow users to perceive objects based on the sound they emit, the prototype should be able to support 3D sound (R1). Unfortunately, there is a limitation that had to be solved. As already mentioned in chapter two, each person has their own HRTF, thus the used 3D sound might not be sufficient for all users. Therefore, it is important to find out how well the used 3D sound in the prototype can be perceived by each of the users (R5). In the conclusion of the related work chapter 3.4 it was pointed out that there is a research gap in comparing different auditory mappings for 3D Sound as off-screen technique for AR. Therefore, the objects in the prototype, that have to be found with the help of the used 3D sound, should have the ability to emit either the metaphorical, nomic, or symbolic [4] auditory mapping for 3D sound (R2). On the one hand, it should be found out which of the auditory mappings is best suited for finding objects as quickly, easily and intuitively as possible, without the help of visual indicators. Therefore, it is important that the user only has to rely on the 3D sound and that there are no visual cues about the positions of the objects (R3). For this prototype, the goal was to investigate the different auditory mappings when only one sound is heard at a time. Furthermore, it has the advantage that the users are able to concentrate better on finding an object at a time, since they only hear the 3D sound emitted by the object they are looking for. (R4). On the other hand, it should be found out which of the auditory mappings is best suited to gain the best possible spatial knowledge about the positions of the searched objects. Therefore it is important to include a spatial knowledge test in the prototype (R6). The resulting requirements for the prototype are listed again below. The requirements are classified according to their relevance for the auditory off-screen technique or for the study.

Auditory off-screen Technique:

R1 The prototype should support 3D sound.

Study-related:

R2 The prototype should provide target objects. Each target object should have the possibility to emit either the metaphorical, nomic or symbolic auditory mapping for 3D sound.

R3 The prototype should provide no visual cues for the position of the objects.

R4 The prototype should only emit the 3D sound of one target object at a time.

R5 The prototype should have the possibility to find out how well the used HRTF is adjusted to the user.

R6 The prototype should provide a spatial knowledge test to find out how well the users remember the positions of the target objects.

4.2 Technology used

The following section describes which hardware and which software were used for the prototype.

4.2.1 Hardware

The prototype was designed for augmented reality with head-mounted displays. The used device should be able to support spatial audio and should be as mobile as possible. Therefore, the Microsoft HoloLens 2 was used as hardware for the prototype.

4.2.2 Software

In this section the used development platform and other used software tools are listed.

Development Platform

The programming of the prototype was made with the *Unity3D* game engine [10], which is a platform for the development of 2D and 3D games and applications. In *Unity3D* there are scenes consisting of *Game Objects*. It is possible to assign *components* to the *Game Objects*. These *components* can be for example materials, physical properties or scripts, which are programmed with the scripting language *C#*. Another example of a component is an audio source, with which a certain audio source can be played from the corresponding *Game Object*, which was very important for the prototype.

Other software tools

By scanning an image target, with the help of the *Vuforia* [11] framework and the *Room Marker Package* [12] it was possible to give the virtual objects a fixed spatial position, regardless of where the prototype is started from. Furthermore, the *Mixed Reality Tool Kit (MRTK)* [13] from Microsoft was used, which provides a set of components and features to support the development of mixed reality apps in Unity.

To edit the *C#* scripts, *Visual Studio Code* [14] was used. In order to deploy the *Unity3D* application onto the HoloLens 2, *Visual Studio 2019* [15] was used.

The objects in the environment were either downloaded from the *Unity Asset Store* [16] or from the *Sketchfab* [17] website.

The nomic sounds, the symbolical sound and some of the metaphorical sounds were downloaded from the *freesoundslibrary* [18], *pixabay* [19], *freesfx* [20] or *freesound* [21] websites. The rest of the metaphorical sounds were created with the *audacity* [22] application. Using this, it was possible to edit existing audio files, such as changing the playback speed or pitch, or applying various effects to the audio files.

4.3 Study Prototype

The prototype was developed to meet the requirements of a planned study in which participants can test the different auditory mappings. The work flow of the prototype is as followed:

1. **Scanning an image**: To ensure that the objects have the same position in the room for each participant, they always had to scan an image at the beginning. After this happened, virtual objects are set to a position relative to the position of the image. Thus, the positions of the objects are not dependent on the position from which the application was started. Furthermore, it was possible to give the virtual objects a certain height. For each participant, the height of the virtual objects was set slightly below the height of the head (position of the camera). In this way it could be excluded that participants with different body heights would have an advantage or disadvantage in the processing of the tasks.

2. **Selecting an ID**: Since three different auditory mappings were used and each participant had to test each auditory mapping so that they could be compared between each other, there were three rounds within the prototype. For each round a different auditory mapping was used. Thus, there were six possible orders to assign the respective auditory mappings to a round. In order to test each sequence as often as the others, there should be a multitude of six persons participating in the study. In order to select the correct sequence of auditory mappings for each round, the participant had to select one of six IDs at the beginning, which was then used by the prototype to assign the corresponding auditory mapping to each round.

3. **HRTF test**: Before the tasks concerning the auditory mappings were performed with the prototype, an HRTF test was conducted to find out how well the used HRTF of the prototype is adapted to the participant's HRTF. The participant was asked to look in a direction so that he could press a button that activated a 3D sound. Afterwards the participant was supposed to look in the direction from which he expected the 3D sound and to press a button (see Figure 4.1), which always folled him in the FOV of the device, as soon as he thinks that he is looking in the right direction. Thus, it was possible to calculate an angle between the start direction and the end direction.

4. **Round one to three**: Each round started with a sequential search for the 12 virtual objects. Each object had the possibility to emit nomic, metaphorical and symbolic auditory mappings as 3D sound. At the start position, which was located in the middle of the room, it was possible to start a sound by pressing a button. Then this sound was emitted by the object to be searched for. All virtual objects in the environment were invisible until the user approached them within one meter (see Figure 4.3). Furthermore, they were distributed like a circle around the starting position, so that the participant hears once from each direction a sound when he is at the starting position (see Figure 4.2). Sometimes there were several objects close to each other, but only one of them had to be searched for at all (see Figure 4.4). If the user thought that he had found the object he was looking for, he was supposed to interact with that object. If it was the right object that was interacted with, the 3D sound that emanated from it automatically stopped. By pressing the button at the start position it was then possible to start the next sound. When all 12 objects were found, it was possible to start the object location memory test by pressing a button within the prototype. After the test was started, all objects were placed in the middle of the room (see Figure 4.5) and the participant was then able to place the objects back to the position he thinks they were before. During the test, the objects were always visible when they were within the FOV of the HoloLens 2, not just when the participant approached them within one meter. When the participant was finished placing the objects, he could end the test by pressing a button. After that, the next round could be started.To prevent that the participant already knows the position of the objects from the previous round in the second or third round, which could possibly make the search task easier for him, the objects were newly positioned in each round. This

was implemented by simply moving the objects for four positions. Thus, the distance between the respective objects could still be kept the same so that this could not have any effect on the task processing.

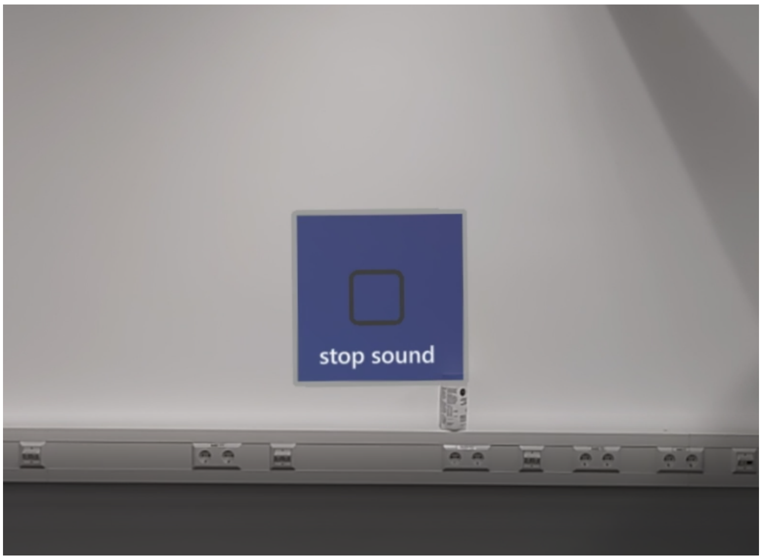


Figure 4.1: This button was used to stop the 3D sound during the HRTF test. All other buttons in the study prototype looked the same, just with different names and icons.

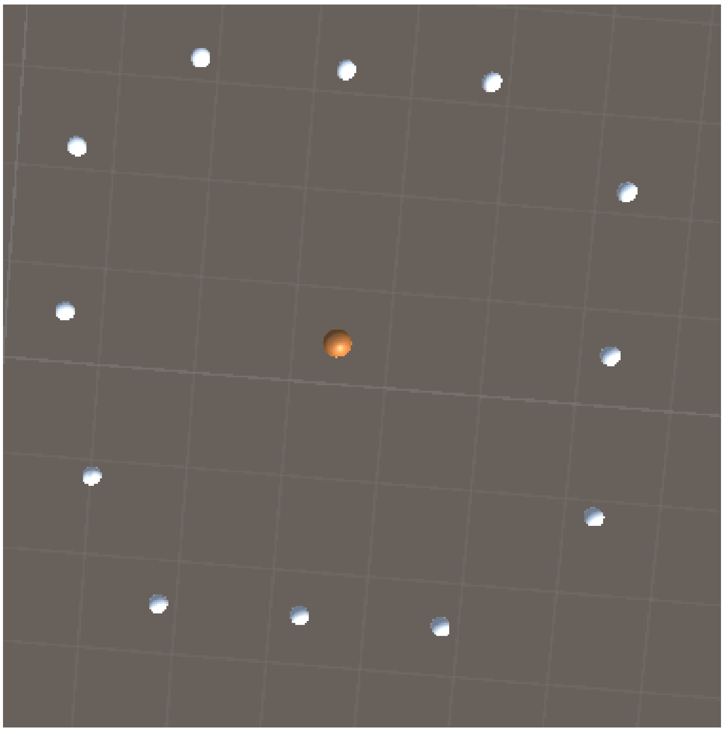


Figure 4.2: The orange circle signals the starting point, while each of the white circles represents a virtual target object.

Figure 4.3: On the left image it can be seen that the virtual object is not yet visible. However, as soon as the object was approached to within one meter, it became visible, as can be seen in the right image.

Figure 4.4: This is how it looked like when several objects were next to each other. However, only one object of them had to be searched for.

Figure 4.5: This is how the virtual objects were positioned at the beginning of the object memory task. From there, the participant was able to place the objects back to their previous position.

5 User Study

This chapter presents the user study that was conducted. At the beginning, the research questions are given, that specify what is supposed to be found out by conducting the study. Then, the study design is presented, such as the procedure and the tasks that had to be performed. In the last section, the data collection techniques that were used to answer the research questions are presented.

5.1 Research Questions

The overall research objective of the study is to compare three different auditory mappings in respect to the search duration, accuracy, object location memory and subjective preference. For this reason, the following research questions should be answered within the scope of the study:

RQ 1 How does the auditory mapping of 3D sound as off-screen technique for head-mounted AR displays affect the search duration?

RQ 2 How does the auditory mapping of 3D sound as off-screen technique for head-mounted AR displays affect the accuracy in finding the correct target object?

RQ 3 How does the auditory mapping of 3D sound as off-screen technique for head-mounted AR displays affect the object location memory?

RQ 4 How does the auditory mapping of 3D sound as off-screen technique for head-mounted AR displays affect the subjective preferences and perceptions?

5.2 Study Design

In this section the study design is presented. First, the used apparatus will be introduced. Afterwards, the procedure of the study will be described in detail, followed by a description of the different tasks and task objects used in the study. Finally, it will be discussed how the Covid-19 situation was addressed during the study.

5.2.1 Apparatus

The study was conducted in the Mixed Reality Lab at the University of Konstanz. The room is quadratic and was large enough for the use of the prototype. During the task executions, the Microsoft HoloLens 2 was used. However, the visualization of virtual objects can depend on lighting conditions. For example, virtual objects that are lighted by sun rays appear more transparent than others. By using blinds and lamps, it was tried to ensure uniform lighting conditions in the room independent of sunlight and daylight.

On the right side, a table and a seat were provided for the study participant. There, the study instructor welcomed the participants, explained the risk assessment for infection control and conducted the interview at the end of the study. Furthermore, the participants filled out the consent form, the demographic questionnaire and the questionnaires between the individual task rounds there. Audio recordings of the interview were made using a mobile phone. The following figure 5.1 shows a sketch of the room and where everything was located.

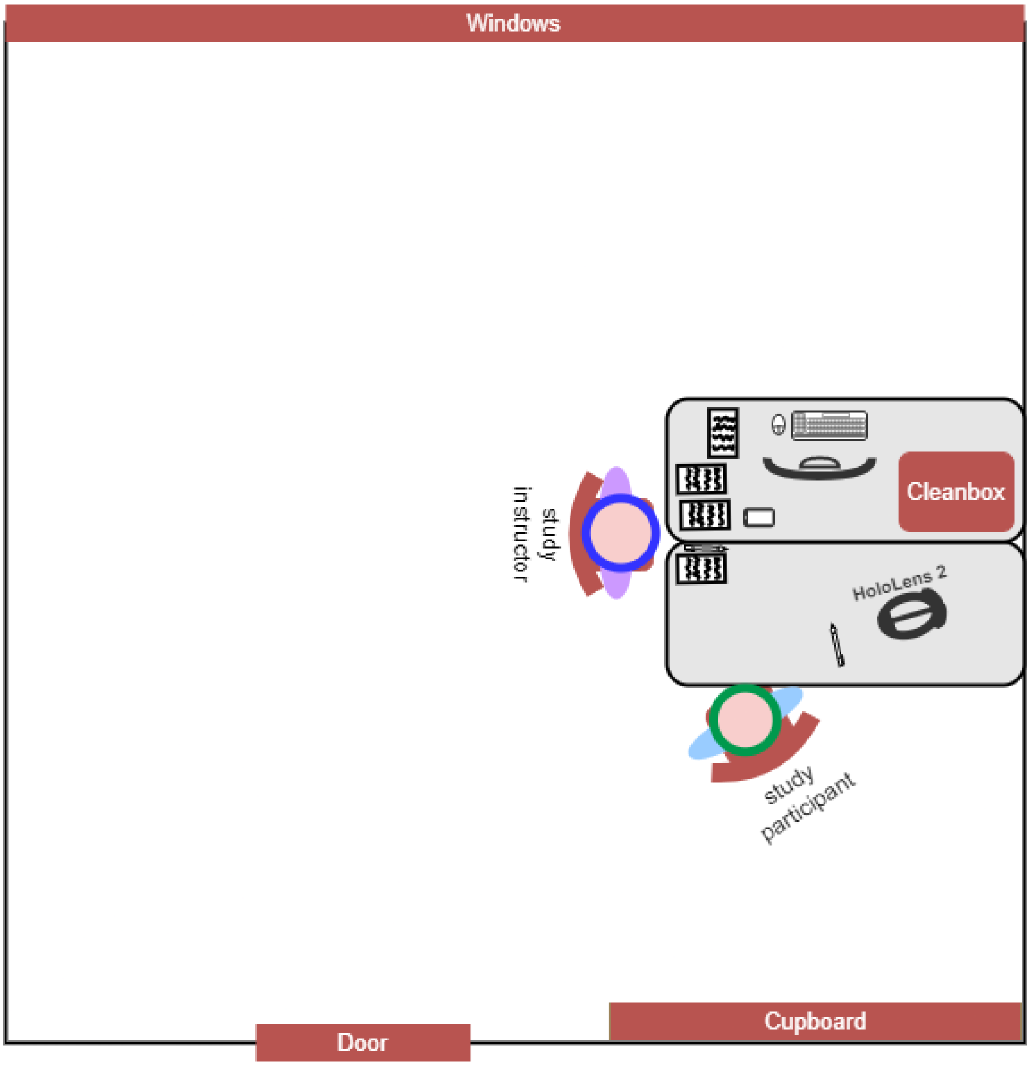


Figure 5.1: This was the distribution within the room at the beginning and at the end of the study.

During the tasks, the participants used the Microsoft HoloLens 2 and were able to move around the room freely. To ensure that the study instructor was not in the way, the instructor sat in the corner of the room during this time and made notes about the participants, their interactions, behavior, and comments. A sketch of the room and where everything was located can be seen on figure 5.2. In order to be able to clean the HoloLens 2 between the participants, the Cleanbox was used, which was located on one of the tables on the right side. The Cleanbox uses the proprietary and patented technology of UVC light to eliminate pathogens on the HoloLens 2 within one minute.

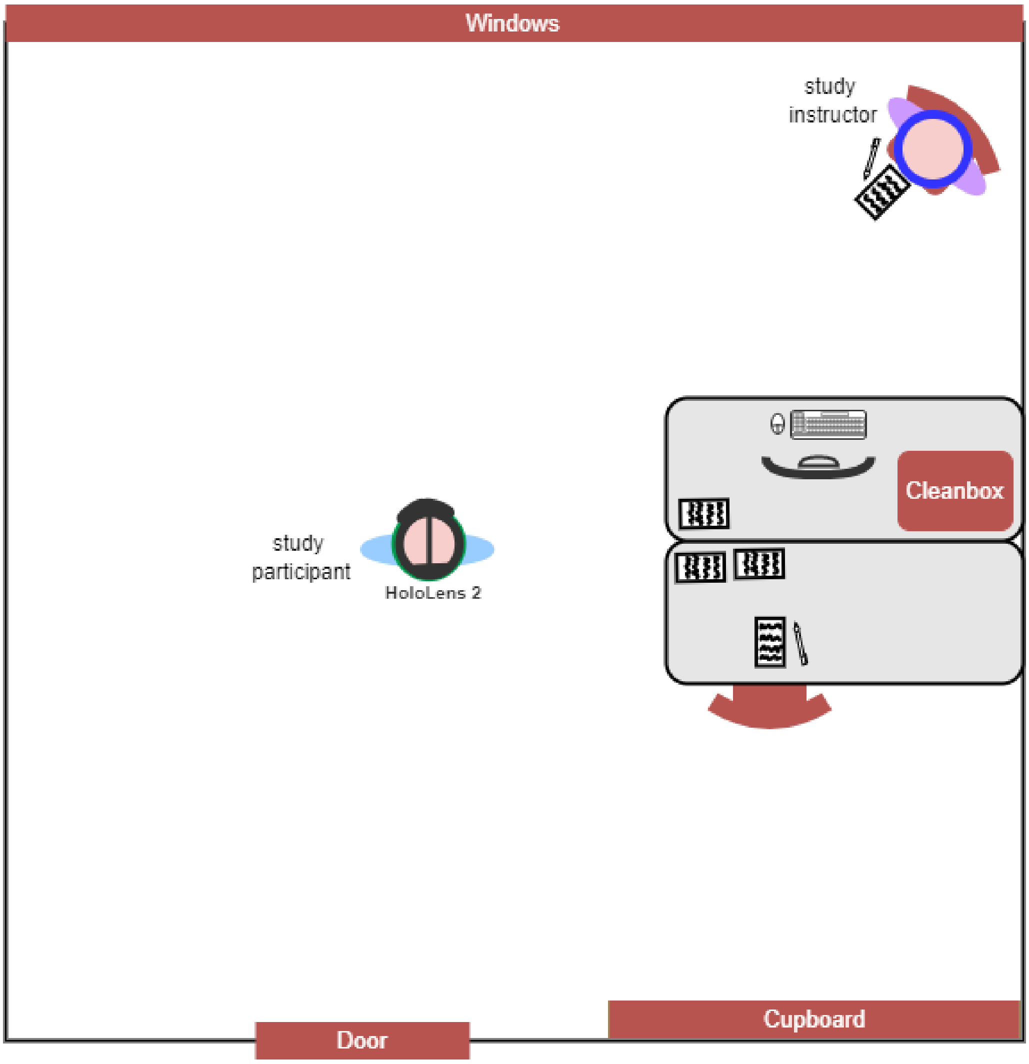


Figure 5.2: This was the distribution within the room during the task performance of the participants.

5.2.2 Procedure

The study consisted of three phases. An overview of these phases and their steps, can be seen in Table 5.1.

At the beginning, the study instructor welcomed the participant, introduced herself and guided the participant to a seat in the study setting. After that, the participant could read a welcome sheet, which contained information about what to expect during the study, the purpose of the study, the duration of the study and the compensation afterwards. Then, the study leader gave the participant a quick introduction about the behavior in case of fire and explained the covid risk assessment. Afterwards, the participant had to read and sign an informed consent form stating that he or she has been informed about the corona risks and that he or she agrees with the study procedure and the recording and logging of any data generated during the study. Then, the study participant filled out a demographic questionnaire, so that background information about all participants could be found out and compared. The welcome sheet, consent form, and demographic questionnaire can be found in the Appendix 8.

After the initial formal steps were completed, the study leader verbally explained in detail the process of the study and which tasks would have to be performed by the participant. At this point, the participant was also instructed to always speak aloud during the study so that subjective and qualitative data could be collected during the task performances (think-aloud technique). The study participant was then asked to stand up and put on the HoloLens 2. If the participant was not yet familiar with the use of the hardware, a small test scenario was performed. In that exercise the participant was asked to open an image from the gallery of the HoloLens 2 and to move the image around the room. In this way, all the hand gestures for the upcoming task processing could be practiced.

Afterwards, the participant started the prototype on the HoloLens 2 and simultaneously stood in front of the image target so that it could be scanned by the HoloLens 2 camera without any problems after starting the prototype.

Once the image target was scanned correctly so that the virtual objects were distributed to the correct positions in the room, the participant was able to start performing the HRTF test to check how well the used 3D sound was adapted to the participant's HRTF. Then the participant was able to begin the first of three rounds. Whereby the order of which auditory mapping was used in which round was counterbalanced (see section 4.3). In each round, the participant first had to search for 12 virtual objects inside the room. After that, he received a questionnaire in which he had to rate the auditory mapping that was used in that round. At the end of each round, the participant had to perform the object location memory test, in which he had to place all the previously searched objects back to where they had been before. The order in which the auditory mapping was used alternated between the participants. Each possible order in which the auditory mappings could be used was evaluated by three participants. The questionnaires between the rounds can be found in the Appendix 8.

After all three rounds were completed, there was a final semi-structured interview, which was audio recorded using a mobile phone. In this interview, the three used auditory mappings were compared by the participant. The basic structure of the semi-structured interview can be found in the Appendix 8.

After the semi-structured interview, the study instructor thanked the participant for his participation, handed over the compensation, asked him to sign the confirmation of compensation and said goodbye. The used confirmation form can be found in the Appendix 8.

Overall each run of the study took on average one hour. Depending on how easy or difficult it was for the participants to perceive the virtual objects with the help of the used 3D sound, deviations from this time were to be expected. The average time to complete the study tasks, starting from the HRTF test until the completion of the third spatial knowledge task, was 34 minutes and 8 seconds with a standard deviation of 7 minutes and 29 seconds.

Phase	Step	Documentation
Beginning	Welcome	
	SARS CoV-2 risk assessment	
	Informed consent	Signed consent form
	Demographic questionnaire	Demographic questionnaire
	Explain study process	
	Put on HoloLens 2	
	Test scenario	
Study	HRTF test	Logging
	First round:	
	- Searching for objects with the help of the first auditory mapping	Logging
	- 3D Sound questionnaire	Filled questionnaire
	- Object location memory test	Logging
	Second round:	
	- Searching for objects with the help of the second auditory mapping	Logging
	- 3D Sound questionnaire	Filled questionnaire
	- Object location memory test	Logging
	Third round:	
	- Searching for objects with the help of the third auditory mapping	Logging
	- 3D Sound questionnaire	Filled questionnaire
	- Object location memory test	Logging
Ending	Semi-structured Interview	Audio recording
	Thank you	
	Deliver Compensation	
	Confirmation of compensation	Signed confirmation of compensation
	Goodbye	

Table 5.1: Step by step procedure of the study.

5.2.3 Tasks and Task Objects

During the study, three different tasks could be performed with the prototype. These were conducting the HRTF test, searching for objects and placing objects back to their previous position. The objects and sounds associated with the tasks are presented below.

HRTF test

In the HRTF test, the direction from which a sound was heard had to be recognized, as described in section 4.3. For this purpose, nomic, metaphorical and symbolic sounds were used so that the participant could already get an example of the auditory mappings used later and that these could also be compared with each other. In total there were six different directions from which sounds were heard one after the other. Two times the symbolic sound was used, two times nomic sounds and two times metaphorical sounds. The positions of the different auditory mappings were fairly distributed. The nomic sounds were appropriate to the objects car and flute. The metaphorical sounds were appropriate to the objects horse and telephone.

Figure 5.3 shows the QR code with which one can access the audio files of the HRTF test objects.

Figure 5.3: With this QR code, one can access the audio files used in the HRTF test. [23]

Searching for objects

In total, the participant had to search for 12 different virtual objects inside the room in each round. The 12 objects were distributed in a circle around the starting position (see section 4.3). All objects have the same size (about 20cm), so that there is no advantage or disadvantage in finding the objects or remembering their positions. Since in each round a different auditory mapping was used for the 3D sound of the objects, it was necessary that each object which had to be searched for could emit a nomic, a metaphorical and a symbolic sound. Thus the decision was made to use three categories of objects. These included familiar sound objects from everyday life, instruments and animals, since these could all also emit a nomic sound. It was tried that the used audio files always have the same duration (about 3 seconds), so that there is no advantage or disadvantage when searching for the objects. In addition to the sound objects, six more objects, two from each of the three categories, were placed in the environment. These objects did not emit a 3D sound and therefore did not have to be searched for, but they were placed near the objects that had to be searched for. The aim was to find out how accurately the position of the 3D sound can be detected, even if there are several objects next to each other. In the following the used objects from the three categories are presented.

Everyday Objects:
Sound objects: trash can [24], alarm clock [25], basketball [26], and ping pong paddle [27] Non sound objects: telephone [28] and car [29]
Figure 5.4 shows pictures of the used everyday objects.

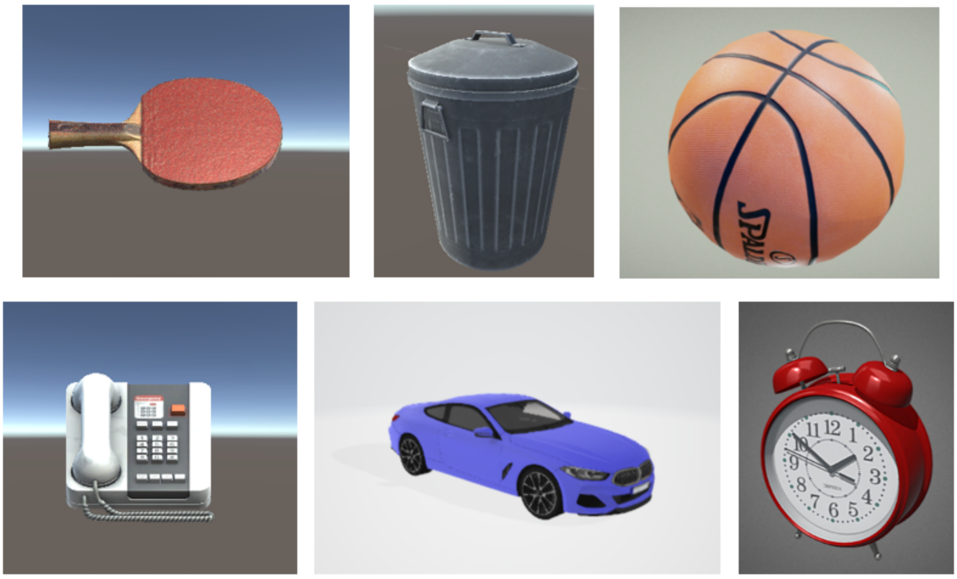


Figure 5.4: These are the everyday objects, which are used within the prototype.

Instruments: Sound objects: saxophone [30], guitar [31], violin [32] and trumpet [33] Non sound objects: flute [34] and piano [35] Figure 5.5 shows pictures of the used instrument objects.

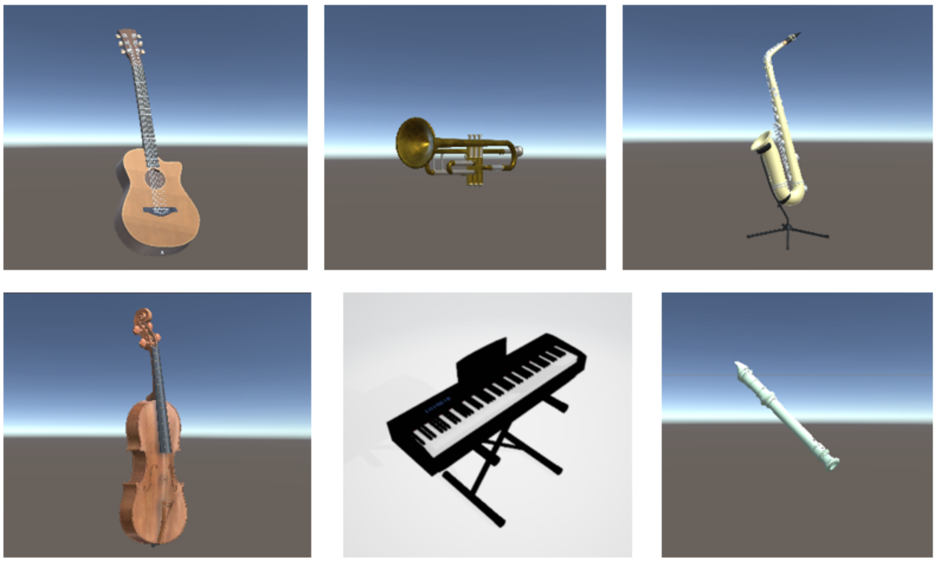


Figure 5.5: These are the instrument objects, which are used within the prototype.

Animals: Sound objects: cat [36], dog [37], chicken [38] and cow [39] Non sound objects: horse [40] and sheep [41] Figure 5.6 shows pictures of the used animal objects.

Figure 5.6: These are the animal objects, which are used within the prototype.

With the following QR code [42], one can access the audio files to the nomic, metaphorical and symbolic sounds of the 12 presented sound objects.

Figure 5.7: With this QR code, one can access the audio files used in the searching task. [42]

The following table 5.2 shows in which order the 12 sound objects had to be searched for in the respective rounds. For example, if the number four is assigned to an object in the first round, it means that this object is the fourth object that had to be searched for in this round.

Objects	First round	Second round	Third round
Trash can	12	9	10
Alarm clock	4	4	1
Ping-pong paddle	6	7	12
Basketball	9	5	4
Saxophone	2	6	7
Guitar	11	1	6
Violin	3	8	9
Trumpet	8	12	3
Cat	1	10	8
Cow	7	3	2
Chicken	5	11	5
Dog	10	2	11

Table 5.2: This is the order in which the objects had to be searched for in the respective rounds.

Object location memory test

The object location memory test was used to find out how well the participants could remember the positions of the previously searched objects in relation to the auditory mapping that was used. Therefore, this test was also conducted in each of the three rounds, always after the search task was completed and the questionnaire was filled out. For this purpose, the same 12 objects were used, which could emit a sound when searching for the objects.

5.2.4 Covid-19 Considerations

The Covid 19 pandemic was considered during the planning and conduction of the study. A requirement for participating in the study was the wearing of a FFP2 mask. During the study conduction, sufficient distance was always maintained between the study participants and the study instructor. Between the study participants, the room was adequately ventilated and the entire set-up, such as the tables, pens and the HoloLens was disinfected.

5.3 Data Collection

Various data collection methods were used to answer the research questions formulated in Section 5.1.

Demographic questionnaire
A demographic questionnaire was used to obtain some background information about the participants. Questions included age, gender, course of study, current occupation, whether one suffers from a hearing disorder and how familiar one is with using augmented reality glasses or applications. The used demographic questionnaire can be found in the Appendix 8.

Think aloud technique
The participants were asked to always talk aloud while performing the tasks. In this way it was possible for the study leader to always have an overview of the study participants progress in completing the tasks. In addition, problems and comments could be noted down at any time. Thus, it was also possible to gain a more precise insight into the participant's feelings while completing the tasks.

3D Sound questionnaire
In each round, after searching for the objects, the participants were asked to fill out a non-standardized questionnaire which used likert scale in which they evaluated the previously used auditory mapping. Among other things, it was asked how easy it was for the participant to recognize the direction from which the sound was emitted and how easy it was to identify the correct object. When all three completed questionnaires are compared with each other, it is for example possible to find out which of the auditory mappings allowed the participant to recognize the direction from which the sound came most easily. The used questionnaire can be found in the Appendix 8.

Semi-structured interview
After the three rounds were completed, a semi-structured interview was conducted. It contained e.g. questions about in which round (and thus with which auditory mapping) the participant found it easiest to complete the tasks. Furthermore, questions were asked about the subjective preference of the different auditory mappings and which advantages and disadvantages there were in the respective rounds due to the auditory mappings used. The participants were always asked to give reasons for their answers. The prepared questions for the semi-structured interview can be found in the Appendix 8.

Logging
The study prototype generates interaction logs during use. These logs include every interaction that occurs while the prototype is in use. Several variables can be derived from the data, which was stored in json files. Thus, it was possible to find out when it was interacted with which virtual object and thus the time it took to complete a task, e.g. the time it took to find each of the objects, or the total time it took to complete a round per participant. Furthermore, the positions of the different virtual objects could also be tracked, which allows to find out where the objects were placed in the object location memory task. In addition, for the HRTF test, it was possible to record in which direction the participant assumed the 3D sound to be heard by calculating the angle between the start direction (when the participant clicked on the "start sound" button) and the end direction (when the participant clicked on the "end sound" button) in which the participant was looking.

Table 5.3 shows which data collection is later on used to answer which research question.

	Data sources
RQ 1: Search duration	Logging, 3D Sound questionnaire, Semi-structured interview
RQ 2: Accuracy in locating objects	Logging, 3D Sound questionnaire, Semi-structured interview
RQ 3: Object location memory	Logging, Semi-structured interview
RQ 4: Subjective preferences and perceptions	3D Sound questionnaire, Semi-structured interview, Think-aloud technique

Table 5.3: These are the data collection methods used per research question.

6 Results

This chapter presents the data collected and the results of the study. The first section introduces the study participant demographics. The following sections show the results of the study in connection with the respective research question.

6.1 Study Participant Demographics

Eighteen participants took part in the study. Of these, nine were female and nine were male. The age range was between 20 and 46 years, with an average age of 24.88 years. Among all the participants, 15 were students and three were employees. Among the 15 students, 7 studied computer science, 3 life science, one chemistry, one mathematics, one business law, one education computer science and physics and one education chemistry and English.

Regarding the questions about previous experiences with augmented reality applications, seven persons stated that they had used augmented reality glasses before and 12 persons stated that they already had experiences with augmented reality applications.

Self-assessment of familiarity with the use of AR applications was done on a scale of 1-5. 1 was described as very familiar and 5 as not at all familiar. The median response to this question was 3.6. Overall, one person rated themselves as 1, one person as 2, eight persons as 3, two persons as 4, and six persons as 5.

Figure 6.1 shows gender, age, previous experience and familiarity as graphs.

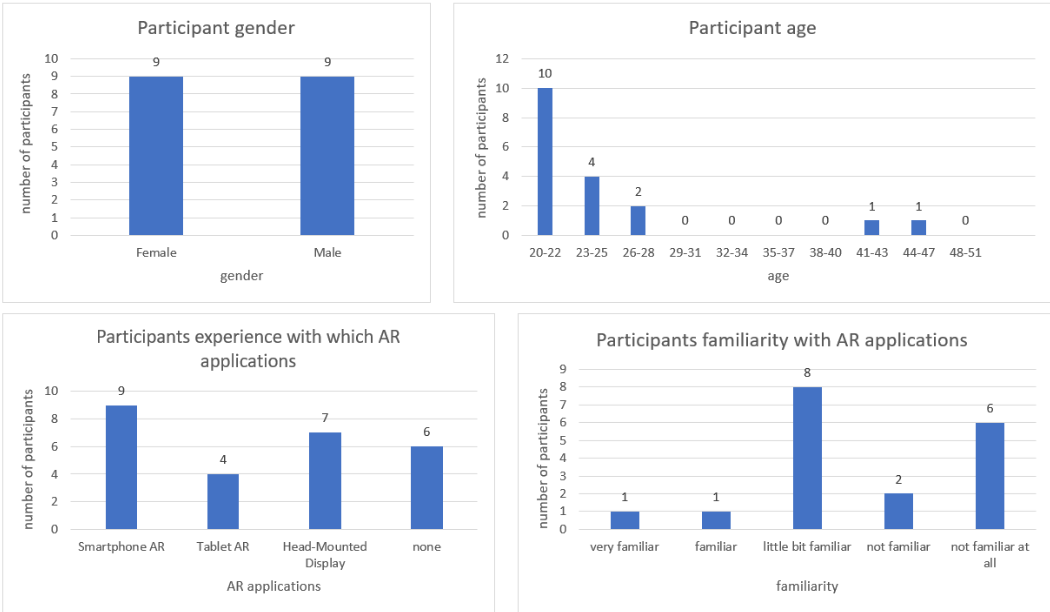


Figure 6.1: Gender, age, previous experiences with AR applications and familiarity with AR applications of the study participants.

None of the participants indicated that they suffer from a hearing disorder. During the study, it was tested for each participant how well the used 3D sound was adapted to their own HRTF. For this purpose, a test was conducted in which the participant was able to indicate from which direction he or she assumed the 3D sound to be heard. With the help of the prototype logging an angle was saved, which could be used to find out in which direction the participant assumed the sound to be saved. The implementation of this was described in Section 4.3. The evaluation was performed with two different intervals. The first time, a deviation of 10 degrees plus or minus was accepted for detecting the correct direction. In the second evaluation, the interval was reduced, so that only a deviation of plus or minus five degrees was accepted. In neither of the two used intervals there was a participant who could not correctly identify any direction. Figure 6.2 shows how many directions each of the 18 participants recognized correctly, respectively in an interval of [-10°, +10°] or [-5°, +5°].

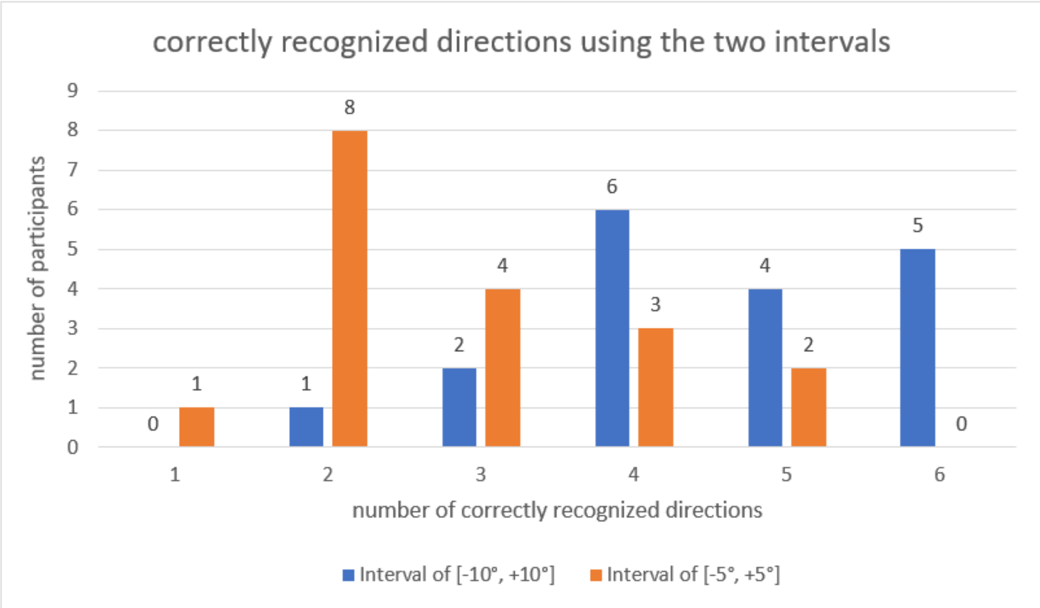


Figure 6.2: Number of correctly recognized directions during the HRTF test, using the two intervals

6.2 RQ 1: Search duration

The required search duration of the participants was recorded with the prototype logging and stored in json files. This made it possible to find out how long it took each participant to find an object. By selecting an ID at the beginning of each trial, it was then also possible to find out which search duration could be assigned to which auditory mapping. Thus, the search duration for each object could be compared together with the used auditory mapping and for each participant. With 12 objects and 18 participants, 216 different search durations could be compared for each auditory mapping. In the following the results of the investigation of the search duration for the respective auditory mappings are presented.

The mean time participants needed to find one of the 12 virtual objects during the evaluation of the symbolic auditory mapping was 14 seconds (SD=9.23). During the evaluation of the metaphorical auditory mapping, the mean time participants needed to find one object was 13.17 seconds (SD=7.61). A slightly smaller mean value was recorded for the nomic auditory mapping compared to the other two auditory mappings. Here the mean time the participants needed was 12.24 seconds (SD=8.25). Median (IQR) of the search duration for the nomic auditory mapping was 9 (7 to 13), for the metaphorical auditory mapping it was 11 (8 to 16) and for the symbolic auditory mapping, the median (IQR) was 12 (9 to 16).

Comparison of search duration between the auditory mappings

Since the search duration data is not normally distributed according to the *Shapiro-Wilk test* $(p > .05)$, a non-parametric procedure had to be performed to find out if there is a statistically significant difference between the three auditory mappings in respect to the search duration. As a non-parametric test, the *Friedman test* was chosen, which can be used to find out if there is an overall statistically significant difference between the mean ranks of the search duration for the different auditory mappings. The test revealed that there was a statistically significant difference in search duration depending on which auditory mapping was used, $\chi^2(2) = 16.565$, $p < .001$.

A post-hoc analysis using *Wilcoxon signed-rank tests* was performed with a *Bonferroni* correction, resulting in a significance level of $p < 0.017$ (significance level that was initially used divided by the number of tests that are running = 0.05 divided by 3 = 0.017). There was no significant difference of the search duration when comparing the use of symbolic auditory mapping with the use of metaphorical auditory mapping ($Z = -1.133$, $p = .257$). However, there was a statistically significant difference for the search duration when comparing the use of nomic auditory mapping with the use of metaphorical auditory mapping ($Z = -2.974$, $p = .003$). Furthermore, there was a statistically significant decrease in search duration when comparing the use of nomic auditory mapping with the use of symbolic auditory mapping ($Z = -3.678$, $p < .001$).

Since there are some outliers within the data series and they are not normally distributed, a box plot diagram is used to illustrate the data (see Figure 6.3).

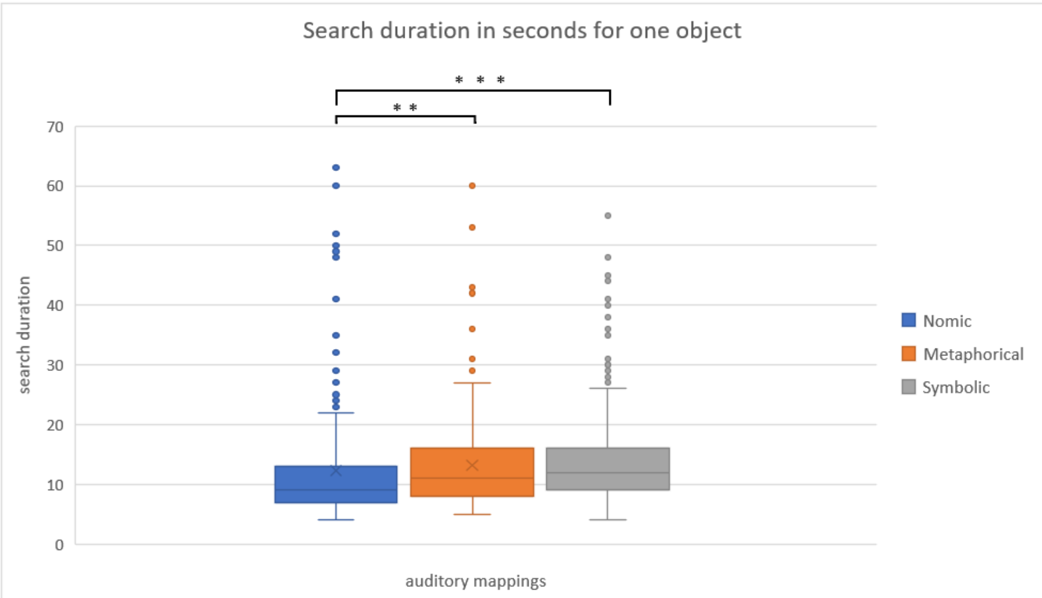


Figure 6.3: Median (IQR) and outliers in search duration for an object associated with the used auditory mapping. Bars connected by an asterisked bracket indicate statistically significant differences.

6.3 RQ 2: Accuracy in locating objects

In order to find out the accuracy in locating objects, it was logged how often a participant considered a wrong object to be the correct one when searching for an object. The data was stored in json files using the prototype logging. Thus, it was possible to find out how many times a participant made a mistake while searching for an object during the use of one of the three auditory mappings. In the following, the results of the investigation of the accuracy in locating objects for the respective auditory mappings are presented.

The participants mistook a mean of 0.35 objects (SD = 0.43) for the correct one during the use of the nomic auditory mapping. When the metaphorical auditory mapping was used, participants mistook a mean of 0.95

objects (SD = 1.06) for the correct one. When symbolic auditory mapping was used, a mean of 1.03 objects (SD = 1.2) were mistaken for the correct one. The median (IQR) of the number of objects that were incorrect but identified by the participants as the correct one while using the nomic auditory mapping was 0 (0 to 0.25), while using metaphorical auditory mapping it was 1 (0 to 2) and while using symbolic auditory mapping the median (IQR) 2 (0 to 3).

Comparison of accuracy in locating objects between the auditory mappings

As shown in Figure 6.4, there are outliers within the nomic auditory mapping data. Furthermore, a *Shapiro-Wilk test* ($p > .05$) revealed that none of the data sets were normally distributed. Thus, a non-parametric procedure had to be used to find out if there were statistically significant differences in the accuracy of locating objects. The *Friedman test* was chosen as the non-parametric test. The test revealed that there was a statistically significant difference of the accuracy in locating objects depending on which auditory mapping was used, $\chi^2(2) = 15.362$, $p < .001$.

A post-hoc analysis using *Wilcoxon signed-rank tests* was performed with a *Bonferroni* correction, resulting in a significance level of $p < .017$. There was no statistically significant difference of the accuracy in locating objects when comparing the use of symbolic auditory mapping with the use of metaphorical auditory mapping ($Z = -1.597$, $p = .110$). However, there was a statistically significant difference between the accuracy in locating objects when using the nomic auditory mapping and the accuracy in locating objects when using the metaphorical auditory mapping ($Z = -2.762$, $p = .006$). Also, there was a statistically significant difference of the accuracy in locating objects when comparing the use of nomic auditory mapping to symbolic auditory mapping ($Z = -3.100$, $p = .002$).

Figure 6.4 shows the box plot diagram for this evaluation. As can be seen, there are outliers when using the nomic auditory mapping.

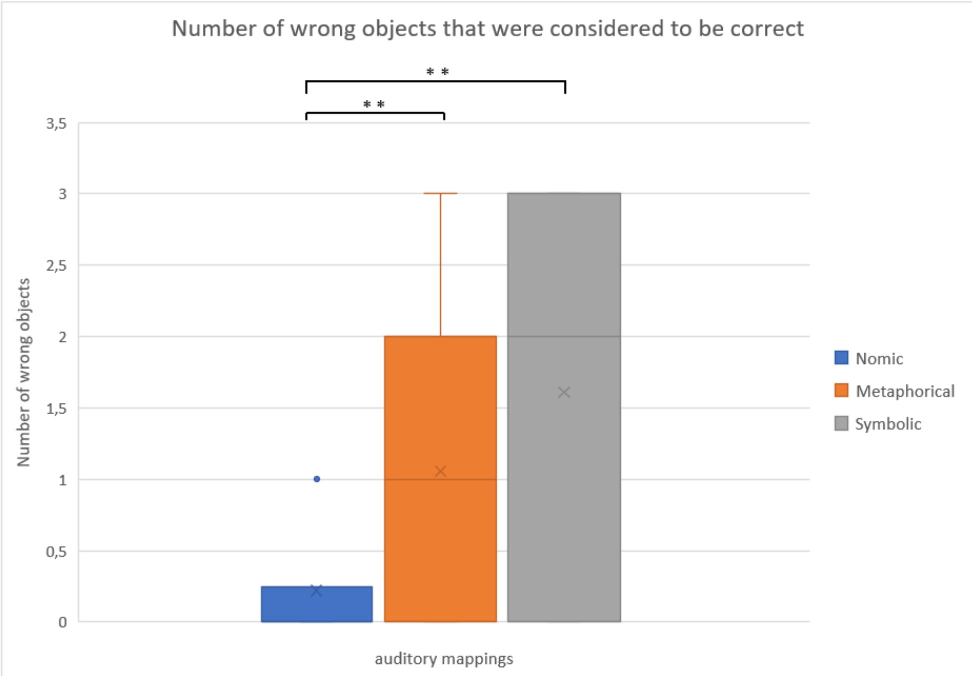


Figure 6.4: Median (IQR) and outliers of the number of wrong objects that were considered to be correct while searching for an object associated with the used auditory mapping. Bars connected by an asterisked bracket indicate statistically significant differences.

6.4 RQ 3: Object location memory

The object location memory of the participants was recorded with prototype logging and stored in json files as in RQ1 and RQ2. Thus, it was possible for each participant to find out where they placed the twelve virtual objects during the three object location memory tasks. For the evaluation of the object placement data, it was evaluated for each participant how many objects he was able to place correctly in each of the three object location memory tasks. An object was considered to be placed correctly if the placement made by the participant was closest to the true position of the object to be placed. However, if the object placement was closer to the true position of another object, the placement was considered to be incorrect. Thus, the number of correct placements per used auditory mapping could be determined for each participant. In the following, the results of the investigation of the object location memory for the respective auditory mappings are presented.

The mean number of correctly placed objects in the object location memory task after using the nomic auditory matching was 5.5 objects (SD=1.92). After using the metaphorical auditory mapping, the mean number of correctly placed objects in the object location memory task was 4.89 objects (SD=1.92). The mean number of correctly placed objects in the object location memory task after using the symbolic auditory matching was 3.78 objects (SD=2.13). Median (IQR) of correctly placed objects in the object location memory task for the nomic, metaphorical and symbolic auditory mappings were 5.5 (5 to 7), 5 (3 to 7) and 4 (2 to 5), respectively.

Figure 6.5 shows the box plot diagram for this evaluation. As can be seen, there is an outlier in the evaluation of the object location memory task after the nomic auditory mapping was used. This outlier shows that one participant could not place any of the 12 objects correctly after the nomic auditory mapping was used. Since at least three objects could otherwise always be placed correctly, this is an outlier.

Figure 6.5: Median (IQR) and outliers of the number of correctly placed objects during the object location memory task, associated with the used auditory mapping.

The following Figure 6.6 shows how many correct placements were performed by each participant in the object location memory task according to the used auditory mapping. It can be seen that five participants were able to place the most objects correctly after the nomic auditory mapping was used, compared to the other two auditory mappings. Four participants were able to place the same number of objects correctly after the use of the nomic and after use of the metaphorical auditory mapping and also more than with the symbolic auditory mapping. Furthermore, four participants were able to place the most objects correctly after using the metaphorical auditory mapping, compared to the other two auditory mappings. One participant was able to place the same number of objects correctly after the nomic and after the symbolic auditory mapping and more than with the metaphorical auditory mapping. Four participants were able to place the most objects correctly after the symbolic auditory mapping was used, compared to the other two auditory mappings.

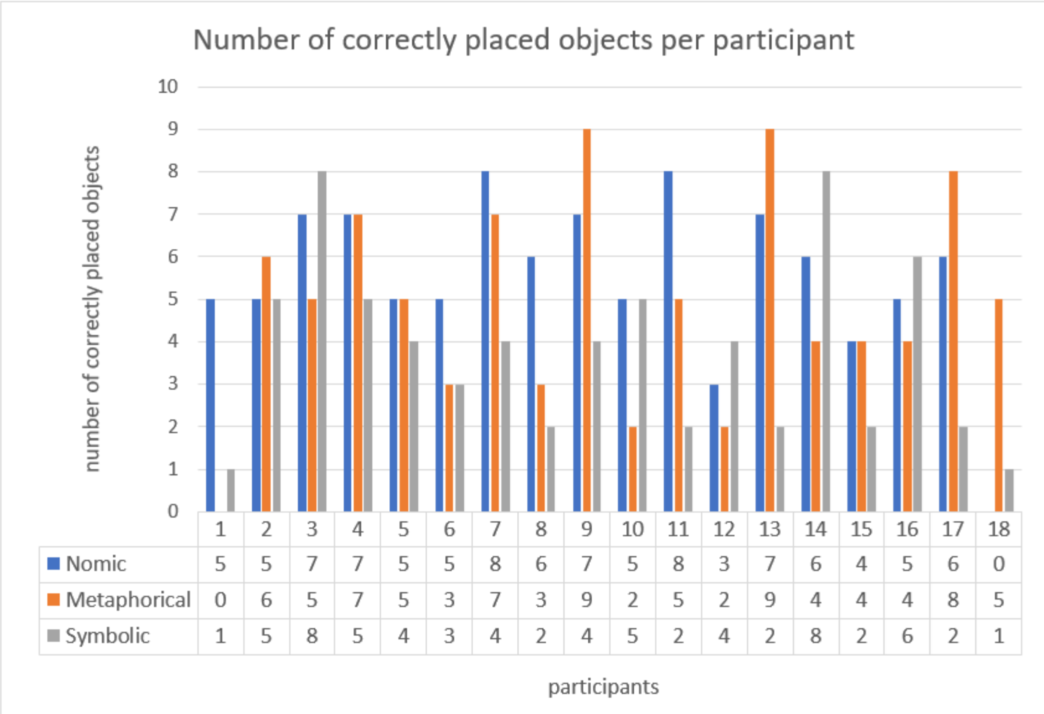


Figure 6.6: Number of correctly placed objects per participant during the object location memory task, associated with the used auditory mapping.

Comparison of object location memory between the auditory mappings

In order to find out whether there are statistically significant differences between the three auditory mappings regarding the placement of the objects in the object location memory task, an *ANOVA test* was performed. Since a within-subject design was used for the study, an *one-way ANOVA with repeated measures test* had to be performed. However, as previously shown, there is an outlier in the data, making this data set not normally distributed, according to the *Shapiro-Wilk test* ($p > .05$). Though, the *one-way ANOVA with repeated measures test* only needs approximately normal data because it is quite "robust" to violations of normality, meaning that the assumption can be a little violated and still provide valid results. After an *one-way ANOVA with repeated measures test* with the outlier and one with adjustment of the outlier was performed, it was confirmed that the results hardly differ from each other, respectively provide the same significant results. Therefore, the results of the *one-way ANOVA with repeated measures test* including the outlier will be presented here.

For the *one-way repeated measures ANOVA*, a *Greenhouse-Geisser* correction was applied ($\epsilon = .860$). The auditory mappings elicited statistically significant changes in the object location memory, $F(1.719, 29.229) = 3.726$, $p = 0.042$, partial $\eta^2 = 0.180$. Post hoc analysis with a *Bonferroni* adjustment revealed that the object location memory was statistically significant decreased from the nomic auditory mapping to the symbolic auditory mapping (1.722 (95% CI, .192 to 3.252) $p = .025$), but not from the nomic auditory mapping to the metaphorical auditory mapping (.611 (95%CI, $-.894$ to 2.116) $p = .888$) and from the metaphorical auditory mapping to the symbolic auditory mapping (1.111 (95%CI, $-.901$ to 3.123) $p = .483$). Figure 6.7 represents a bar chart, showing the means and the 95% CI of the correct placed objects associated with the used auditory mapping.

Figure 6.7: Means and 95% CI of the correctly placed objects during the object location memory task associated with the used auditory mapping. Bars connected by an asterisked bracket indicate statistically significant differences.

6.5 RQ 4: Subjective preferences and perceptions

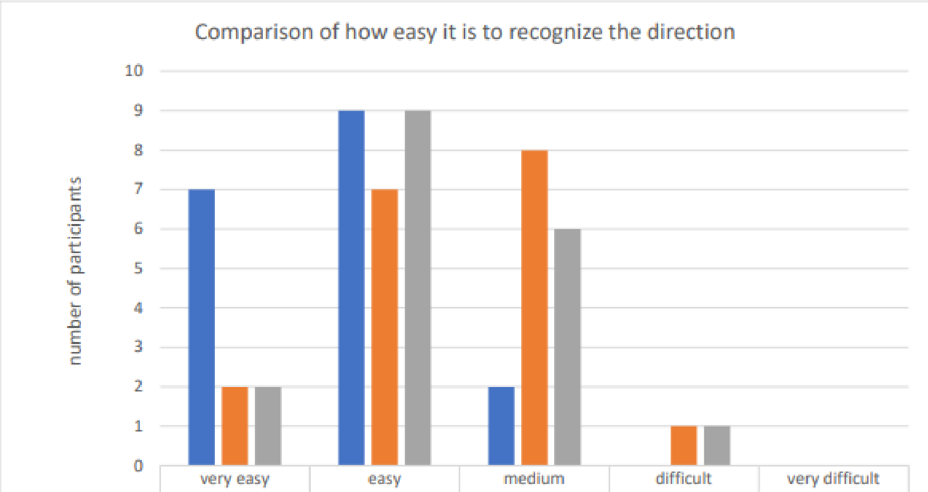
During the study, in addition to the demographic questionnaire, there were three further questionnaires, which the participants filled out in each round. In these, it was asked about the subjective perception about the previously used auditory mapping. First, there was the question "How easily could you recognize the direction from which the sound came?" And secondly, there was the question "If there were several objects next to each other, how easily could you identify the correct object from which the sound came?" After the three runs with the prototype, there was a final semi-structured interview in which participants were asked again about their subjective perception, but also about their subjective preference regarding the three auditory mappings. The participants were asked with which of the used auditory mappings they could recognize the direction from which the sound came best or worst and with which of the used auditory mappings they found it easiest or hardest to identify the correct object from which the sound was emitted. Furthermore, it was asked with which of the used auditory mappings it was most difficult or easiest for the participant to remember the positions of the objects afterwards. For the subjective preference, the participants were then asked which of the auditory mappings they would prefer to use in future augmented reality applications in order to perceive the object positions, to identify the correct object and to build up the best possible object location memory.

How easy could the direction from which the sound was emitted be perceived with the used auditory mapping?

The self-assessment for this could be answered on a scale of 1-5, where 1 was described as very easy and 5 as very difficult. As Figure 6.8 shows, most participants found it easiest to identify the direction with the help of nomic auditory mapping. The median (IQR) of how easy it was to identify the direction from where the sound was played using the nomic auditory mapping was 2 (1 to 2), using the metaphorical auditory mapping it was

2.5 (2 to 3) and when using the symbolic auditory mapping the median (IQR) was 2 (2 to 3).. To find out if there were statistically significant differences in the subjective perception of the direction recognition using one of the auditory mappings, a *Shapiro-Wilk test* ($p > .05$) was first used to test if the data were normally distributed. Since the test revealed that the data series were not normally distributed, the non-parametric *Friedman* test was performed to evaluate for significance. The test revealed that there was a statistically significant difference of subjective perception depending on which auditory mapping was used, $\chi^2(2) = 7.731$, $p = .021$.

A post-hoc analysis using *Wilcoxon signed-rank tests* was performed with a *Bonferroni* correction, resulting in a significance level of $p < .017$. There was no significant difference of subjective perception when comparing the use of symbolic auditory mapping with the use of metaphorical auditory mapping ($Z = -.632$, $p = .527$). Furthermore, there was no significant difference of the subjective perception when using the symbolic auditory mapping and the subjective perception when using the nomic auditory mapping ($Z = -2.221$, $p = .026$). However, there was a statistically significant difference between the subjective perception when using the nomic auditory mapping and the subjective perception when using the metaphorical auditory mapping ($Z = -2.543$, $p = .011$).



	very easy	easy	medium	difficult	very difficult
Nomic	7	9	2	0	0
Metaphorical	2	7	8	1	0
Symbolic	2	9	6	1	0

Figure 6.8: Subjective perception of how easily the direction from which a sound is emitted can be detected for each of the auditory mappings.

With which auditory mapping was it the easiest or most difficult to identify the direction from which a sound came?

In the semi-structured interview at the end, participants were asked with which auditory mapping they found it easiest or most difficult to identify the direction from which a sound was played. Here, 14 participants indicated that they found it easiest to recognize the direction with nomic auditory mapping, 2 participants indicated that they found it easiest with metaphorical auditory mapping and 2 participants indicated that they found it easiest to recognize the direction with symbolic auditory mapping. 11 participants found it most difficult when using sybolic auditory mapping, 6 participants when using metaphorical auditory mapping and one participant when using nomic auditory mapping (see Figure 6.9).

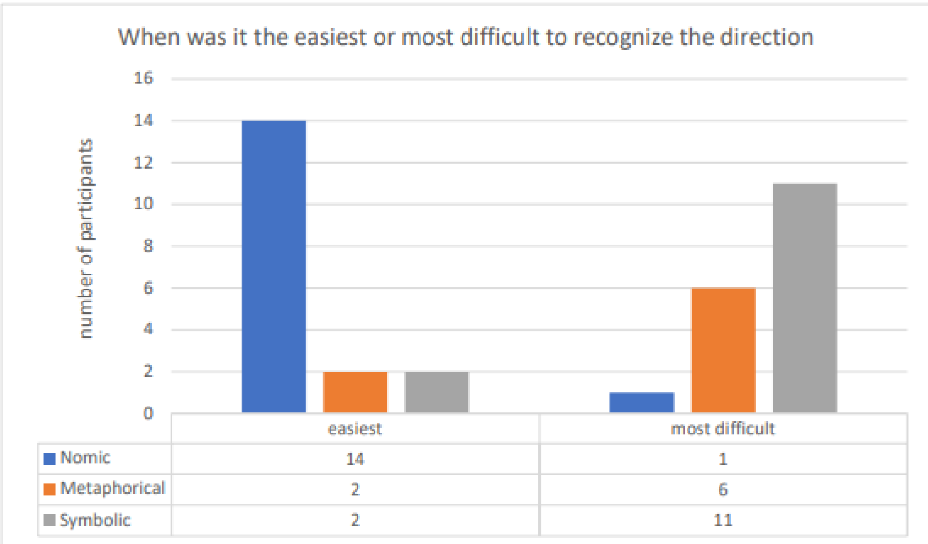


Figure 6.9: Subjective perception, with which auditory mapping it was the easiest or the most difficult to recognize the direction from which a sound was emitted.

How easy was it to identify the correct object from which the sound was emitted with the used auditory mapping?

The self-assessment for this could again be answered on a scale of 1-5, where 1 was described as very easy and 5 as very difficult. As Figure 6.10 shows, the participants found it easiest to identify the correct object with the help of nomic auditory mapping. The median (IQR) of how easy it was to identify the correct object using the nomic auditory mapping was 1 (1 to 1.25), using the metaphorical auditory mapping it was 2.5 (2 to 3) and when using the symbolic auditory mapping the median (IQR) was 3 (2 to 4). To find out whether there were statistically significant differences in the subjective perception regarding the identification of the correct object using one of the auditory mappings, a *Shapiro-Wilk test* ($p > .05$) was first used to test whether the data were normally distributed. Since the test revealed that the data series were not normally distributed, the non-parametric *Friedman test* was performed to evaluate the significance. The test revealed that there was a statistically significant difference of the subjective perception depending on which auditory mapping was used, $\chi^2(2) = 25.864, p < .001$.

A post-hoc analysis using *Wilcoxon signed-rank tests* was performed with a *Bonferroni* correction, resulting in a significance level of $p < .017$. There was no significant difference of subjective perception when comparing the use of symbolic auditory mapping with the use of metaphorical auditory mapping ($Z = -2.047$, $p = .041$). However, there was a statistically significant difference of the subjective perception when comparing the use of metaphorical auditory mapping with the use of nomic auditory mapping ($Z = -3.359$, $p < .001$). Also, there was a statistically significant difference of the subjective perception when comparing the use of nomic auditory mapping to the use of symbolic auditory mapping ($Z = -3.666$, $p < .001$).

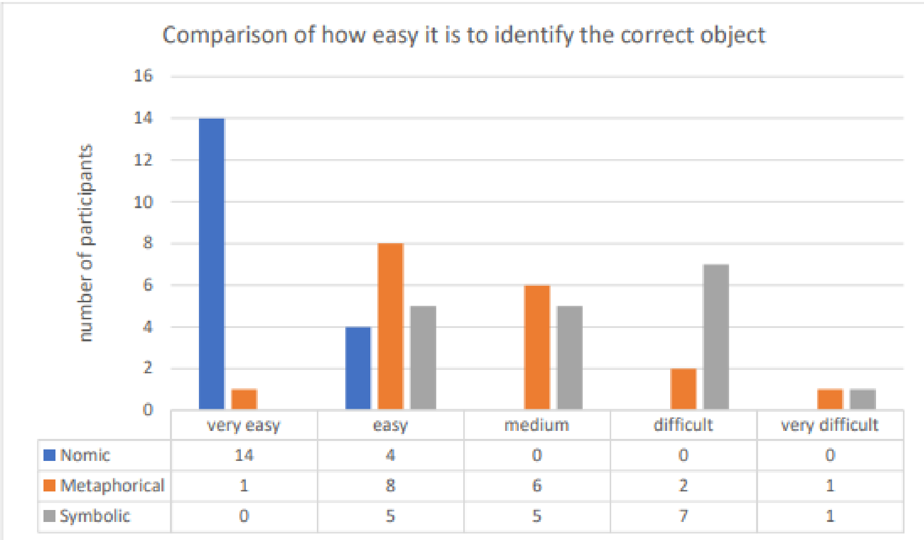


Figure 6.10: Subjective perception of how easily the correct object can be identified for each of the auditory mappings.

With which auditory mapping was it the easiest or most difficult to identify the correct object?

In the semi-structured interview at the end, participants were asked with which auditory mapping they found it easiest or most difficult to identify the correct object. Thereby, 16 participants indicated that they found it easiest to identify the correct object with the nomic auditory mapping and 2 participants indicated that they found it easiest to identify the correct object with the metaphorical auditory mapping. 16 participants found it most difficult when using symbolic auditory mapping and 2 participants when using metaphorical auditory mapping (see Figure 6.11).

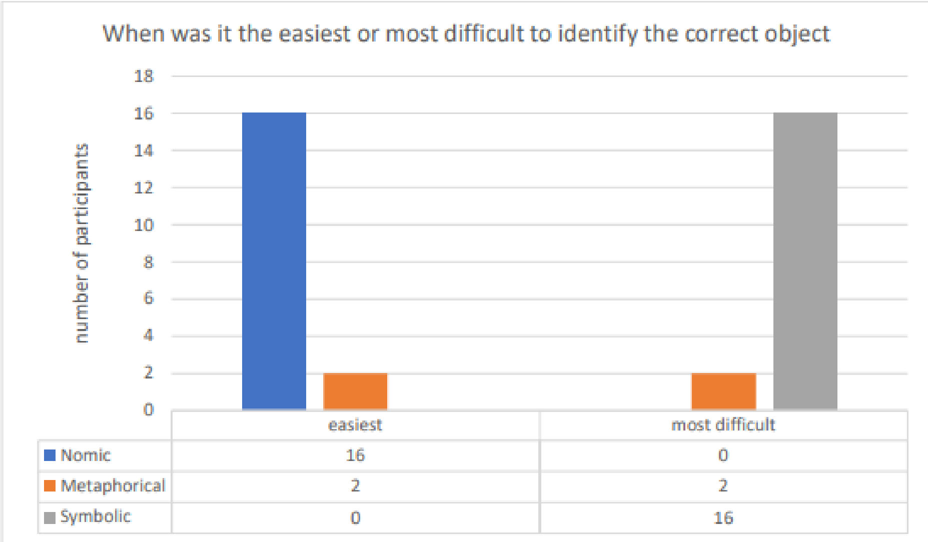


Figure 6.11: Subjective perception, with which auditory mapping it was the easiest or the most difficult to identify the correct object.

After using which auditory mapping, was it the easiest or most difficult to remember the positions of the objects afterwards?

In the semi-structured interview, participants were also asked whether they found it easiest or most difficult to remember the positions of the objects after using the nomic, metaphorical, or symbolic auditory mapping. Thereby, 10 participants indicated that they could remember the positions of the objects most easily after the nomic auditory mapping was used, 6 participants indicated that they could remember the positions of the objects most easily after the metaphorical auditory mapping, and 2 participants indicated that they could remember the positions of the objects most easily after the symbolic auditory mapping. Fifteen participants found it most difficult when using symbolic auditory mapping, 2 participants when using metaphorical auditory mapping and one participant when using nomic auditory mapping (see Figure 6.12).

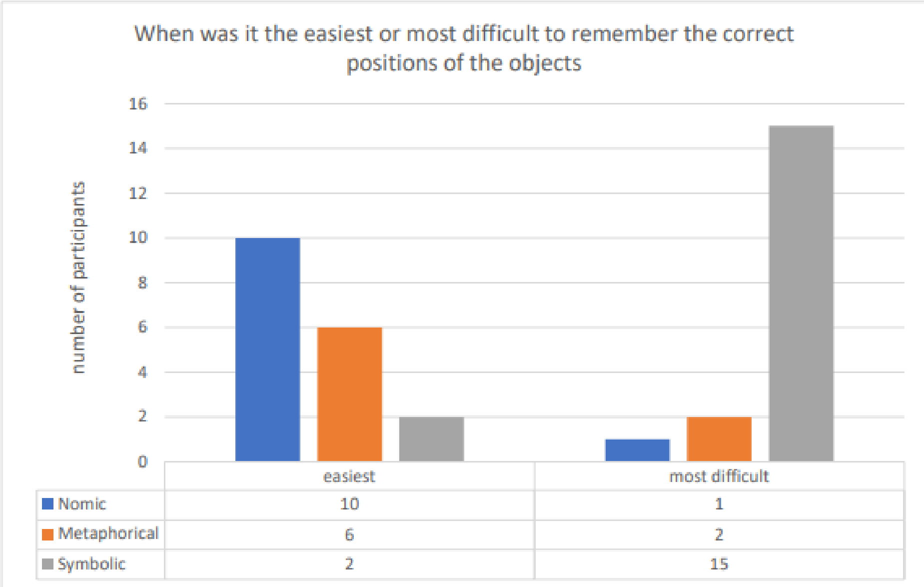


Figure 6.12: Subjective perception, after which auditory mapping it was the easiest or the most difficult to remember the correct positions of the objects.

Which auditory mapping would be most likely to be preferred in the future?

In addition, the participants were asked in the semi-structured interview which of the used auditory mappings they would prefer in future augmented reality applications. They were asked to consider the recognition of the direction as well as the perception of the actual position and the built-up object location knowledge in their assessment. 16 persons stated that they would prefer the nomic auditory mapping in the future and 2 persons stated that they would prefer the metaphorical auditory mapping in the future (see Figure 6.13).

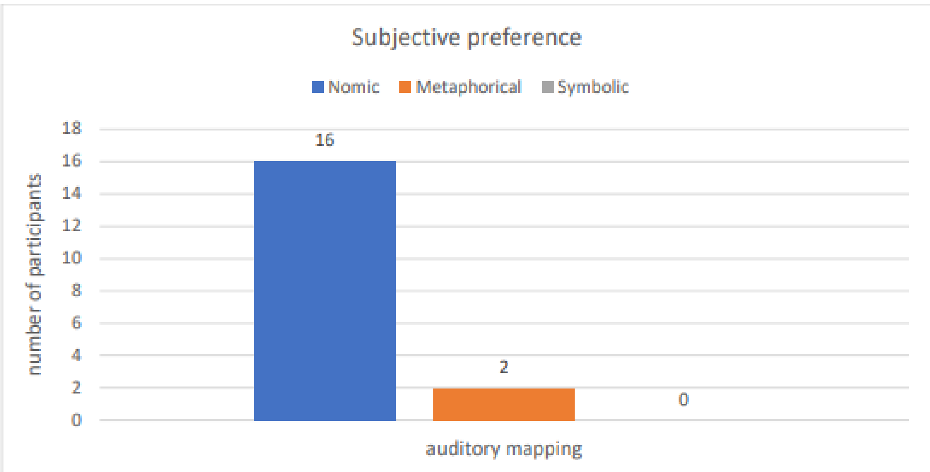


Figure 6.13: Subjective preference about auditory mapping would be most preferred in future augmented reality applications.

7 Discussion

In the following chapter, the study results are discussed in relation to the asked research questions. Section 7.1 discusses the first research question concerning the search duration, the participant needed for each auditory mapping. The next section 7.2 discusses the second research question concerning the accuracy in locating objects within each auditory mapping. Section 7.3 discusses the third research question about the object location memory after each of the auditory mappings. The last research question regarding participants' subjective preferences and perceptions is discussed in section 4. 7.4. The last section 7.5 discusses the limitations of this work and possible future works.

7.1 RQ 1: Search duration

RQ 1: How does the auditory mapping of 3D sound as off-screen technique for head-mounted AR displays affect the search duration?

This question can be answered by using the task completion time derived by the interaction logging. Also, the personal perception, with which auditory mapping, one can recognize the correct direction or identify the correct object most easily or with the greatest difficulty and interview responses were used to discuss the results from logging.

Significant differences were found in the evaluation of search duration. The use of nomic auditory mapping performed significantly better than the use of metaphorical auditory mapping. The use of nomic auditory mapping also performed significantly better than the use of symbolical auditory mapping. However, no significant differences were found between the use of symbolic auditory mapping and the use of metaphorical auditory mapping.

As shown in figures 6.8 and 6.9, participants indicated that they clearly found it easiest to recognize the correct direction and to identify the correct object when the nomic auditory mapping was used and clearly found it most difficult when the symbolic auditory mapping was used. When interviewing the participants, it also came out that most of the participants had the feeling that they found the objects with the nomic sound the fastest. The following describes what participants reported in the semi-structured interview regarding their perceptions and preferences. The participants stated that they already knew the used nomic sounds from everyday life, so they were already familiar with them, making it easier for them to decide faster for a direction. Furthermore they stated, that when several objects were located near to each other, the choice to identify the right object was much easier with the nomic sound, because they knew from the beginning after which object they have to look for, so that the decision in these cases could be made faster than when the metaphorical or symbolic auditory mapping was used.

Overall, the use of the nomic auditory mapping resulted in the significantly shortest search duration compared to the use of the metaphorical or symbolic auditory mapping, and this was also confirmed by the participants' perceptions. Furthermore, it is interesting that although the search duration results did not reveal a significant difference between the use of metaphorical and symbolic auditory mapping, participants' subjective perceptions revealed that they perceived the search time to be shorter when using the metaphorical auditory mapping than the search time when using the symbolic auditory mapping.

7.2 RQ 2: Accuracy in locating objects

RQ 2: How does the auditory mapping of 3D sound as off-screen technique for head-mounted AR displays affect the accuracy in finding the correct target object?

This question can be answered by using the number of wrong objects considered to be the correct one, when searching for one object which can be derived by the interaction logging. Also, the personal perception, regarding with which auditory mapping, one can identify the correct object most easily or with the greatest difficulty and interview responses were used to discuss the results from logging.

Significant differences were found in the evaluation of the accuracy in locating objects. The use of nomic auditory mapping performed significantly better than the use of metaphorical auditory mapping. The use of nomic auditory mapping also performed significantly better than the use of symbolic auditory mapping. However, no significant differences between the use of symbolic auditory mapping and the use of metaphorical auditory mapping could be found.

As shown in Figure 6.11, 16 out of 18 participants indicated that they found it easiest to identify the correct object when nomic auditory mapping was used, since with this auditory mapping they already knew which object they were searching for before they saw it and therefore it didn't make any difference to them, when multiple objects were next to each other. 16 out of 18 participants found it most difficult when the symbolic auditory mapping was used, since this auditory mapping did not convey any hint at all about the object to be searched for. In the semi-structured interview, some participants indicated that they had no chance to hear which object was supposed to be the correct one when several objects were next to each other, which led these participants to guess the correct object. However, some participants indicated that it was indeed possible to detect a difference in volume depending on how one was placed in front of the objects, which also made it possible to identify the correct object. Nevertheless, they complained that symbolical auditory mapping took them more effort and time than the other auditory mappings. The two participants who found it easiest to identify the correct object with the metaphorical auditory mapping stated that this auditory mapping was used in the last round, in which the participant was already used to the task and the 3D sound and could therefore make decisions more easily. The two participants who found it most difficult to identify the correct object with metaphorical auditory mapping had the metaphorical auditory mapping in the first round and stated that they were not as familiar with the task as they were in the subsequent rounds, which is why they found it most difficult then.

Overall, the use of nomic auditory mapping resulted in the significantly best accuracy in locating objects compared to the use of metaphorical or symbolic auditory mapping, which was also confirmed by the perceptions of the participants. Furthermore, it is interesting that although the results of accuracy in locating objects did not reveal a significant difference between using the metaphorical and symbolic auditory mappings, the majority of participants' perceptions revealed that they found it much more difficult to identify the correct object when using the symbolic auditory mapping.

7.3 RQ 3: Object location memory

RQ 3: How does the auditory mapping of 3D sound as off-screen technique for head-mounted AR displays affect the object location memory?

This question can be answered by using the the positions of the objects placed by the participants in the object location memory task and which can be derived by the interaction logging. Also, the personal perception, regarding with which auditory mapping, one can correctly place the objects to their correct position from before most easily or with the greatest difficulty and interview responses were used to discuss the results from logging.

Significant differences were found in the evaluation of object location memory. After using the nomic auditory mapping, participants were able to place significantly more objects correctly than after using the symbolic auditory mapping. However, there were no significant differences after using symbolic auditory mapping compared to after using metaphorical auditory mapping, as well as after using metaphorical auditory mapping compared to after using nomic auditory mapping.

As can be seen in Figure 6.12, 15 of 18 participants found it most difficult to complete the object location memory task after using the symbolic auditory mapping. In the interaction logging, the smallest number of correctly placed objects after using the symbolic auditory mapping was also recorded, but Figure 6.6 shows that only 10 participants could place the least number of objects correctly after using the symbolic auditory mapping compared to the other two auditory mappings. Furthermore, as Figure 6.12 shows, 10 out of 18 people indicated that they found it easiest to solve the task after using the nomic auditory mapping. However, as the interaction logging revealed and can be seen in Figure 6.6, only five participants were able to place the most objects correctly after using the nomic auditory mapping in contrast to the other two auditory mappings. In the semi-structured interview, participants indicated that they thought they remembered the most items correctly after the nomic auditory mapping was used because they were able to think about the object for a longer period of time during the previous search task. In nomic auditory mapping, they knew from the beginning which object they were looking for and were able to associate the path with the searched object while approaching the position of it. Thus, in the object location memory task, they felt more like they knew where the objects had been before. Other participants stated that they could remember the direction very well when the nomic sound was used, which made the object location memory task easier for them. This was because in nomic auditory mapping, participants could hear a different sound for each object, making it easier for them to associate the sound with the direction from whicht they heard sound before. In the symbolic auditory mapping, the participants only heard one type of sound for all objects, so they had no way to associate the direction with the sound. Furthermore, the participants stated that they had hardly concentrated on the object itself during the symbolic auditory mapping, but only on finding the object. Thereby they did not pay attention to the object itself, so that they could not remember afterwards which object was located where.

Overall, the participants were able to correctly place significantly more objects in the object location memory task after using the nomic auditory mapping than after using the symbolic auditory mapping. This result was also reflected in the participants' perceptions. However, interestingly, the participants perceived an even greater difference between the two auditory mappings than it was actually the case. The metaphorical auditory mapping did not lead to the most correctly placed objects in the interaction logging, nor did it lead to the fewest, which was also the perception of the participants.

7.4 RQ 4: Subjective preferences and perceptions

RQ 4: How does the auditory mapping of 3D sound as off-screen technique for head-mounted AR displays affect the subjective preferences and perceptions?

This question can be answered by using the subjective perceptions and preferences, which can be derived by the questionnaires and the responses in the semi-structured interview. Furthermore, the study observation will be taken into consideration.

The subjective perception to the object location memory task, the recognition of the correct direction and the identification of the correct object were already partly dealt with and compared in the previous research questions.

According to the subjective perception of the participants, it was significantly easier to determine the direction from which the 3D sound came when using the nomic auditory mapping than when using the metaphorical auditory mapping. This indicates that participants found it easiest to identify the direction when the nomic auditory mapping was used and hardest when the metaphorical auditory mapping was used. However, in the semi-structured interview it came out that the participants generally found it more difficult to recognize the direction with the symbolic auditory mapping than with the metaphorical auditory mapping. During the study, it was also observed that some participants found it easier to identify the direction from which a 3D sound was coming when it was either to their left or right side. However, when the 3D sound was in front of or behind them, they found it noticeably more difficult to identify the correct direction.

Furthermore, according to the subjective perception of the participants, it was significantly easier to identify the correct object when using the nomic auditory mapping compared to the metaphorical or symbolic auditory mapping. In the semi-sructured interview at the end, participants also indicated that they generally found it the easiest with the nomic auditory mapping and the most difficult with the symbolic auditory mapping to identify the correct object. Although no significant difference was found between the use of the metaphorical and the symbolic auditory mapping.

Regarding subjective preference, 16 participants indicated that they would prefer nomic auditory mapping in AR applications in the future and 2 participants indicated that they would prefer metaphorical auditory mapping. Here, participants indicated that nomic auditory mapping seemed the most realistic and familiar to them. Furthermore, they subjectively found the tasks with the nomic auditory mapping the easiest, which could also be recorded by the interaction loggings. Among the two participants who would prefer the metaphorical sound in the future, one stated that he could find the emitted sounds used in that round the easiest and would therefore prefer them and one stated that he had the feeling that the 3D sound was always played above his head during the nomic auditory mapping and would therefore prefer the metaphorical auditory mapping.

In general, during the study some participants indicated that they often had the feeling that they were hearing the 3D sound from above, i.e., as if the object was placed at a high height, although actually all objects in the prototype were placed slightly under the participants' head height, as described in section 4.3. However, participants further reported that the feeling that the 3D sound was above their heads diminished after a time. Likewise, participants reported that they became more used to the 3D sound over time. Thus, they felt that they were able to decide on a direction and an object more quickly over time. The interaction logging was able to confirm this. The mean time participants took to find an object in the first round was 15.91 seconds (SD=9.5), in the second round a mean time of 12.49 seconds (SD=8.23) was needed to find an object and a mean time of only 11 seconds (SD=6.48) was needed to find an object in the third round.

7.5 Limitations and Future Work

The findings of this study must be seen with some limitations.

This study was conducted with a relatively small sample for a quantitative study approach. Nevertheless, this was necessary due to the circumstances and time constraints of this work. However, this led to the resulting data having significant confidence intervals. A study with a larger sample could lead to greater confidence in the data and a more reliable evaluation of the three introduced auditory mappings.

Furthermore, the sample included, among others, some participants who were part of the study instructor's personal network. This may have had an additional impact on the outcome of the study, as participants may have provided positive ratings to appease the study instructor.

In this study, always the same objects were used for each of the three rounds. This resulted in the fact that from the second or third round on, the participants already knew to some extent which particular objects they had to search for, which made it easier to identify an object when there were several next to each other. Furthermore, this resulted in the fact that in the second or third round the participants still had the positions of the objects from the previous round in their memory, because the various positions for one object had been mixed in the memory. Thus, it was sometimes difficult for the participant to remember the current last position. Unfortunately, due to the time constraints of this work and the small participant sample, it was not possible to search and generate even more virtual objects and matching sounds and to examine them in different orders fairly and often enough. However, it would be an idea for a future work to have more virtual objects so that different ones can be used in each round.

In the conclusion of related work 3.4 it was pointed out that it would be interesting to test how well participants can perceive a certain 3D sound when several other 3D sounds are played at the same time. However, in this work the decision was made to leave out this investigation due to the time constraints. However, in future work it would be interesting to investigate the difference between several sounds at the same time and only one sound at a time.

Finally, the limitation that each person's HRTF is different should be addressed. Thus, the HRTF extension used in the prototype could have been insufficient for the study participant, which could have led to incorrect results. The solution chosen in this work was to perform a HRTF test at the beginning of the study, so that it could be found out how well the used HRTF was adapted to the participants. Fortunately, the used HRTF led to acceptable good results for all participants in the HRTF test. Otherwise, either it would have been necessary to invite another participant to replace the previous one, or the results of this participant would have had to be considered more critically. However, in future work, it would be worth for example considering implementing different HRTFs and finding out which one would be best suited to the participant.

8 Conclusion

Augmented reality devices (e.g., head-mounted display) are becoming more and more available and they offer great potential for spatial tasks that can support individual and collaborative activities. However, since this is still a relatively new technology, it still has many challenges that need to be addressed. A long-known challenge is to find a solution to make users as intuitive and non-obtrusive as possible aware of off-screen objects.

The initial literature research showed that there are still not many studies about the usage of 3D sound as off-screen technique for Head-mounted Augmented Reality Displays. Furthermore, it was found that there is a research gap in comparing different auditory mappings for 3D sound as off-screen technique.

Motivated by this finding, the aim was to find out whether using different auditory mappings for 3D sound as off-screen technique would bring different advantages or disadvantages. For this purpose, a prototype was developed which supports nomic, metaphorical and symbolic auditory mappings for different virtual objects. With three different auditory mappings available for evaluation, it was possible to individually evaluate the auditory mappings and create a comparative evaluation between them.

To evaluate the auditory mappings and to make a comparative evaluation, a study with a quantitative focus was conducted. Eighteen participants took part in the study and tested the different auditory mappings in three rounds. In every round a different auditory mapping was used, so that each participant was able to test all of them. In each round the participants had to find virtual off-screen objects with the help of the used auditory mapping as 3D sound and at the end of each round the virtual objects had to be placed back to their previous position in an object location memory task. The order in which the auditory mappings were tested was counter-balanced to avoid order effects. Several questionnaires and a semi-structured interview were used to document the participants' opinions.

The quantitative data generated during the use of the study prototype was used to answer research questions one to three. The qualitative data from the questionnaires, the semi-structured interview, and the study director's observations were used to answer research question four and to support and explain the results and findings from the quantitative data. The study results showed that the participants preferred the nomic auditory mapping the most, whereas the symbolic auditory mapping was the least preferred. Furthermore, it was found that most of the participants had the feeling that with the nomic auditory mapping they were able to solve the search task the fastest and to place the most objects correctly in the object location memory task. In contrast, when using symbolic auditory mapping, they felt that they needed the longest time and placed the fewest objects correctly. The quantitative results confirm the subjective perception of the participants, since the nomic auditory mapping led to the significantly shortest search time as well as the significantly best accuracy in locating the correct objects compared to the other two auditory mappings. In the object location memory task, it was found that significantly more objects could be correctly placed after using the nomic auditory mapping than after using the symbolic auditory mapping.

Overall, it can be stated that the use of the nomic auditory mapping for 3D sound as off -screen technique for head-mounted augmented reality displays is best suited for the tasks performed in the study, whereas the symbolic auditory mapping is the least suitable.

References

[1] Ryan Gunther, Rick Kazman, and Carolyn MacGregor. "Using 3D sound as a navigational aid in virtual environments". In: *Behaviour & Information Technology* 23.6 (2004), pp. 435–446. DOI: 10.1080/01449290410001723364 (cited on pages 1, 6, 7).

[2] Andreas Butz and Antonio Krüger. *Mensch-Maschine-Interaktion*. De Gruyter Oldenbourg, 2017. ISBN: 9783110476378. DOI: doi:10.1515/9783110476378. URL: https://doi.org/10.1515/9783110476378 (cited on pages 2, 3).

[3] Meera M. Blattner, Denise A. Sumikawa, and Robert M. Greenberg. "Earcons and Icons: Their Structure and Common Design Principles". In: *Human–Computer Interaction* 4.1 (1989), pp. 11–44. DOI: 10.1207/s15327051hci0401_1 (cited on page 3).

[4] William W. Gaver. "Auditory Icons: Using Sound in Computer Interfaces". In: *Human–Computer Interaction* 2.2 (1986), pp. 167–177. DOI: 10.1207/s15327051hci0202_3 (cited on pages 3, 8).

[5] Stephen A Brewster, Peter C Wright, and Alastair DN Edwards. "A detailed investigation into the effectiveness of earcons". In: *SANTA FE INSTITUTE STUDIES IN THE SCIENCES OF COMPLEXITY-PROCEEDINGS VOLUME-*. Vol. 18. ADDISON-WESLEY PUBLISHING CO. 1994, pp. 471–471 (cited on page 3).

[6] Patrick Perea, Denis Morand, and Laurence Nigay. "Halo3D: A Technique for Visualizing Off-Screen Points of Interest in Mobile Augmented Reality". In: *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. IEEE, 2017, pp. 170–175. DOI: 10.1109/ISMAR-Adjunct.2017.58 (cited on page 4).

[7] Stefano Burigat and Luca Chittaro. "Navigation in 3D virtual environments: Effects of user experience and location-pointing navigation aids". In: *International Journal of Human-Computer Studies* 65.11 (2007), pp. 945–958. ISSN: 1071-5819 (cited on page 4).

[8] Jing Yang et al. "The Effects of Spatial Auditory and Visual Cues on Mixed Reality Remote Collaboration". In: *Journal on Multimodal User Interfaces* 14.4 (2020), pp. 337–352. DOI: 10.1007/s12193-020-00331-1 (cited on pages 4, 5, 7).

[9] Dariusz Rumiński. "An experimental study of spatial sound usefulness in searching and navigating through AR environments". In: (2015), pp. 223–233. DOI: 10.1007/s10055-015-0274-4 (cited on pages 5, 7).

[10] *Link to Unity3D*: URL: https://unity3d.com/de/get-unity/download (visited on 01/15/2022) (cited on page 9).

[11] *Link to Vuforia Engine*: URL: https://developer.vuforia.com/ (visited on 01/15/2022) (cited on page 9).

[12] *Link to the Room Marker Package*: URL: https://gitlab.inf.uni-konstanz.de/ag-hci/research/room-marker (visited on 01/15/2022) (cited on page 9).

[13] *Link to the MRTK from Microsoft*: URL: https://www.microsoft.com/en-us/download/details.aspx?id=102778 (visited on 01/15/2022) (cited on page 9).

[14] *Link to visual studio code*: URL: https://code.visualstudio.com/s (visited on 01/15/2022) (cited on page 9).

[15] *Link to visual studio 2019*: URL: https://visualstudio.microsoft.com/de/downloads/ (visited on 01/15/2022) (cited on page 9).

[16] *Link to the Unity asset store:* URL: https://assetstore.unity.com/ (visited on 01/15/2022) (cited on page 9).

[17] *Link to the sketchfab website:* URL: https://sketchfab.com/feed (visited on 01/15/2022) (cited on page 9).

[18] *Link to the freesoundslibrary website:* URL: https://www.freesoundslibrary.com/ (visited on 01/15/2022) (cited on page 9).

[19] *Link to the pixabay website:* URL: https://pixabay.com/ (visited on 01/15/2022) (cited on page 9).

[20] *Link to the freesfx website:* URL: https://freesfx.co.uk/Default.aspx (visited on 01/15/2022) (cited on page 9).

[21] *Link to the freesound website:* URL: https://freesound.org/ (visited on 01/15/2022) (cited on page 9).

[22] *Link to the audacity application:* URL: https://www.audacity.de/ (visited on 01/15/2022) (cited on page 10).

[23] *Link to HRTF test audio files:* URL: https://cloud.uni-konstanz.de/index.php/s/HEyarKscoEbjnzq (visited on 05/26/2022) (cited on page 18).

[24] Memorie. *Trash Can.* URL: https://skfb.ly/6UPv7 (visited on 12/16/2021) (cited on page 18).

[25] andrey_r32167. *Alarm Clock.* URL: https://skfb.ly/6WQvS (visited on 12/16/2021) (cited on page 18).

[26] Scott VanArsdale. *Basketball.* URL: https://skfb.ly/6xxK7 (visited on 12/16/2021) (cited on page 18).

[27] Creaform 3D. *Ping pong Paddle scanned with ACADEMIA 50.* URL: https://skfb.ly/6XMoK (visited on 12/16/2021) (cited on page 18).

[28] maxdragonn. *Office Phone.* URL: https://skfb.ly/ooCp7 (visited on 12/16/2021) (cited on page 18).

[29] *Car.* URL: https://assetstore.unity.com/ (visited on 12/05/2021) (cited on page 18).

[30] Matt Caddie. *Saxophone.* URL: https://skfb.ly/oo97B (visited on 12/15/2021) (cited on page 19).

[31] Ivan Dnistrian. *Guitar.* URL: https://skfb.ly/6wG8w (visited on 12/15/2021) (cited on page 19).

[32] Ethan Savage. *Stradivari Violin.* URL: https://skfb.ly/6TwZU (visited on 12/15/2021) (cited on page 19).

[33] Charlie Tinley. *Trumpet.* URL: https://skfb.ly/UTHx (visited on 12/15/2021) (cited on page 19).

[34] alfa001100. *Flute.* URL: https://skfb.ly/6WH9I (visited on 12/15/2021) (cited on page 19).

[35] *Piano.* URL: https://assetstore.unity.com/packages/3d/props/furniture/digital-piano-203296 (visited on 12/15/2021) (cited on page 19).

[36] Tech Developer. *Cat 3d.* URL: https://skfb.ly/6Vo9Z (visited on 12/14/2021) (cited on page 19).

[37] Yury Misiyuk. *Dog.* URL: https://skfb.ly/6UpqR (visited on 12/14/2021) (cited on page 19).

[38] pooiloui2. *Chicken.* URL: https://skfb.ly/6SDEO (visited on 12/14/2021) (cited on page 19).

[39] Josué Boisvert. *Cow.* URL: https://skfb.ly/DTQH (visited on 12/14/2021) (cited on page 19).

[40] nikita. *Horse.* URL: https://skfb.ly/op9To (visited on 12/14/2021) (cited on page 19).

[41] ChemicalArgtist. *3d scan of a sheep toy.* URL: https://skfb.ly/6XtGy (visited on 12/14/2021) (cited on page 19).

[42] *Link to sound objects audio files:* URL: https://cloud.uni-konstanz.de/index.php/s/jWYzwWMYNk4zXaz (visited on 05/26/2022) (cited on page 20).

Appendix

Study documents

The following documents were used to conduct the user study.

Welcome letter

Herzlich willkommen!

Vielen Dank, dass du an meiner Studie teilnimmst. Bevor wir beginnen, möchte ich dir kurz erklären, worum es bei der Studie geht und welche Rolle du dabei spielst.

Studienziele und Ablauf

In dieser Studie möchte ich die Vor- und Nachteile von verschiedenen Soundarten in Augmented Reality untersuchen. Dabei wirst du eine HoloLens 2 tragen und unter anderem mit Hilfe von räumlichen Sounds nach verschiedensten virtuellen Objekten im Raum suchen. Du kannst während der Einweisung Fragen zum allgemeinen Ablauf stellen. Bitte habe jedoch Verständnis dafür, dass ich während der Aufgabenbearbeitung keine Fragen beantworten kann, um eine Datenverfälschung zu vermeiden.

Um möglichst umfassende Erkenntnisse zu erhalten, werde ich einige Daten speichern, welche während der Durchführung der Studie gesammelt werden. Dazu zählen alle Events, welche während der Studie getätigt werden, also zum Beispiel wann welches Objekt berührt oder ausgewählt wurde. Außerdem werde ich unser Interview am Ende mit Ton aufzeichnen. Dafür ist dein Einverständnis erforderlich. Im Gegenzug verpflichte ich mich, die Daten pseudonymisiert und ausschließlich zu Auswertungszwecken zu verwenden. In diesem Zusammenhang habe ich eine Einverständniserklärung vorbereitet. An dieser Stelle möchte ich darauf hinweisen, dass ich weder dich noch deine Leistung bewerte, sondern nur an der Eignung der verschiedenen Soundarten interessiert bin.

Dauer und Vergütung

Die Teilnahme an der Studie dauert etwa eine Stunde. Solltest du dich zu irgendeinem Zeitpunkt unwohl fühlen und deine Teilnahme beenden wollen, ist dies selbstverständlich ohne Angabe von Gründen möglich.

Nach Abschluss der Studie erhältst du für deine Hilfe 10 Euro. Vielen Dank für deine Unterstützung!

Informed consent form

Einverständniserklärung ID: ____

Informationen über die Studienleitung

Studienleiterin: Johanna Bell

Institution: Arbeitsgruppe Mensch-Computer Interaktion, Fachbereich Informatik und Informationswissenschaft, Universität Konstanz

Erklärung

Ich wurde über das Ziel, den Inhalt und die Dauer der Studie informiert. Außerdem erhielt ich eine Einführung zum Infektionsschutz (SARS CoV-2) und zum Verhalten im Brandfall und bei einer Evakuierung. Im Umfang dieser Studie werden personenbezogene Daten in einem Fragebogen erhoben. Darüber hinaus werden entstehende Bewegungsdaten und Events während der Studie aufgezeichnet. Dazu zählt jeweils, wann und wo mit welchem virtuellen Objekt interagiert wurde. Zusätzlich werden Audioaufnahmen bei der Beantwortung von Interviewfragen am Ende gemacht.

Hiermit wurde ich informiert, dass mit den personenbezogenen Daten vertraulich umgegangen wird und diese nicht an Dritte weitergeleitet werden. Nach der Studie werden die Daten von der Studienleiterin ausgewertet. Teil davon wird auch sein, Protokolle aus den ausgewerteten Daten und Audiodateien zu erstellen. Diese Protokolle sind pseudonymisiert, das heißt alle Verweise, die Rückschlüsse auf Sie als Person zuließen, werden entfernt. Die Publikation der Studienergebnisse wird ausschließlich pseudonymisiert stattfinden und wird zu keinem Punkt mit Ihnen als Person in Verbindung gebracht werden können.

Optionale Punkte (Bitte mit Kreuz markieren, falls du zustimmst)

[ ] Ich stimme zu, dass meine Audiodaten zu internen Präsentationszwecken genutzt werden dürfen

Hiermit stimme ich den unter "Erklärung" genannten Punkten *und den, mit Kreuz markierten, optionalen Punkten* zu:

____________________	________________________	_____________________
(Name)	(Ort, Datum)	(Unterschrift)

Hiermit stimmt die Studienleiterin zu, die Audioaufnahmen und jegliche andere gewonnenen Daten ausschließlich zu Auswertungszwecken im Kontext dieser Studie zu verwenden.

____________________	________________________	_____________________
(Name)	(Ort, Datum)	(Unterschrift)

Demographic questionnaire

Demographischer Fragebogen ID: _______

Vielen Dank, dass du an dieser Studie teilnimmst. Bevor wir beginnen, benötige ich einige weitere Informationen über dich und deine bisherigen Erfahrungen mit Augmented Reality. Alle gesammelten Informationen werden pseudonymisiert und nur zu Auswertungszwecken verwendet.

1. Persönliche Daten

Alter ______ Jahre

Geschlecht [ ] männlich [ ] weiblich [ ] divers

Bist du Student/in? [ ] ja [ ] nein

Falls ja, wie viele Semester und was studierst du? (z. B. 6. Semester - Mathematik BA usw.)

____

Falls nein, als was arbeitest du?

____

Leidest du an einer Hörstörung? [ ] Ja [ ] Nein

Falls ja, an welcher? (z. B. Tinnitus, Schallempfindungsschwerhörigkeit usw.)

____

Falls ja, wird die Hörstörung durch ein Hörgerät korrigiert? [ ] Ja [ ] Nein

2. Vorerfahrungen

Hast du schon mal eine Augmented Reality Brille verwendet? [ ] Ja [ ] Nein

Hast du bereits Erfahrungen mit Augmented Reality (AR) Anwendungen? [ ] Ja [ ] Nein

3D Sound questionnaire

Fragen nach erster Runde ID: ________

1. Wie leicht konntest du die Richtung erkennen, aus welcher der Sound kam?

[ ] sehr leicht [ ] leicht [ ] mittel [ ] schwierig [ ] sehr schwierig

2. Wenn mehrere Objekte nebeneinander waren, wie leicht konntest du das richtige Objekt identifizieren, von welchem der Sound kam?

[ ] sehr leicht [ ] leicht [ ] mittel [ ] schwierig [ ] sehr schwierig

3. Konntest du die Richtung, aus welcher der Sound kam, vom Ausgangspunkt aus immer direkt erkennen?

[ ] ja [ ] nein

4. Konntest du die gesuchten Objekte immer schon beim ersten Versuch finden? (Also, dass du nicht mit einem falschen Objekt interagiert hast, bevor du dann das richtige identifizieren konntest)

[ ] ja [ ] nein

5. Welche Objekte konntest du am einfachsten finden?

____

____

____

6. Bei welchen Objekten hast du dir schwer getan sie zu finden?

____

____

____

Fragen nach zweiter Runde ID: ________

1. Wie leicht konntest du die Richtung erkennen, aus welcher der Sound kam?

[ ] sehr leicht [ ] leicht [ ] mittel [ ] schwierig [ ] sehr schwierig

2. Wenn mehrere Objekte nebeneinander waren, wie leicht konntest du das richtige Objekt identifizieren, von welchem der Sound kam?

[ ] sehr leicht [ ] leicht [ ] mittel [ ] schwierig [ ] sehr schwierig

3. Konntest du die Richtung, aus welcher der Sound kam, vom Ausgangspunkt aus immer direkt erkennen?

[ ] ja [ ] nein

4. Konntest du die gesuchten Objekte immer schon beim ersten Versuch finden? (Also, dass du nicht mit einem falschen Objekt interagiert hast, bevor du dann das richtige identifizieren konntest)

[ ] ja [ ] nein

5. Welche Objekte konntest du am einfachsten finden?

____

____

____

6. Bei welchen Objekten hast du dir schwer getan sie zu finden?

____

____

____

Fragen nach dritter Runde ID: ________

1. Wie leicht konntest du die Richtung erkennen, aus welcher der Sound kam?

[ ] sehr leicht [ ] leicht [ ] mittel [ ] schwierig [ ] sehr schwierig

2. Wenn mehrere Objekte nebeneinander waren, wie leicht konntest du das richtige Objekt identifizieren, von welchem der Sound kam?

[ ] sehr leicht [ ] leicht [ ] mittel [ ] schwierig [ ] sehr schwierig

3. Konntest du die Richtung, aus welcher der Sound kam, vom Ausgangspunkt aus immer direkt erkennen?

[ ] ja [ ] nein

4. Konntest du die gesuchten Objekte immer schon beim ersten Versuch finden? (Also, dass du nicht mit einem falschen Objekt interagiert hast, bevor du dann das richtige identifizieren konntest)

[ ] ja [ ] nein

5. Welche Objekte konntest du am einfachsten finden?

____

____

____

6. Bei welchen Objekten hast du dir schwer getan sie zu finden?

____

____

____

Semi-structured interview questions

Abschließende Fragen ID: ________

1. In welcher Runde konntest du die Richtung, aus welcher der Sound kam am besten erkennen?

[ ] 1. Runde [ ] 2. Runde [ ] 3. Runde

2. In welcher Runde konntest du die Richtung, aus welcher der Sound kam am schlechtesten erkennen?

[ ] 1. Runde [ ] 2. Runde [ ] 3. Runde

3. In welcher Runde konntest du am leichtesten das richtige Objekt identifizieren, von welchem der Sound kam?

[ ] 1. Runde [ ] 2. Runde [ ] 3. Runde

4. In welcher Runde war es für dich am schwersten das richtige Objekt zu identifizieren, von welchem der Sound kam?

[ ] 1. Runde [ ] 2. Runde [ ] 3. Runde

5. Welche Sound Art würdest du in Zukunft bei Augmented Reality Anwendungen präferieren? Sound Art aus der:

[ ] 1. Runde [ ] 2. Runde [ ] 3. Runde

6. Welche Soundart hast du am angenehmsten empfunden? Soundart aus der:

[ ] 1. Runde [ ] 2. Runde [ ] 3. Runde

7. Welche Soundart hast du am unangenehmsten empfunden? Soundart aus der:

[ ] 1. Runde [ ] 2. Runde [ ] 3. Runde

8. In welcher Runde konntest du dich am einfachsten an die Position der Objekte erinnern?

[ ] 1. Runde [ ] 2. Runde [ ] 3. Runde

9. In welcher Runde konntest du dich am schwersten an die Position der Objekte erinnern?

[ ] 1. Runde [ ] 2. Runde [ ] 3. Runde

Was fandest du beim Sound aus der (1.Runde / 2. Runde / 3. Runde) positiv / was war daran hilfreich?

Was fandest du beim Sound aus der (1.Runde / 2. Runde / 3. Runde) negativ / was hat dich gestört?

Confirmation of compensation form

Bestätigung der Entschädigung

Hiermit bestätige ich, dass ich für die Teilnahme an der Studie „Verschiedene Sound Arten für 3D-Sound als Off-Screen-Technik in Augmented Reality“ im April 2022, eine Entschädigung in Höhe von 10 Euro erhalten habe.

____________________	________________________	_____________________
(Name)	(Ort, Datum)	(Unterschrift)