Fakultät für Mathematik und Informatik
Institut für Informatik
Lehrstuhl für Informatik IX
Games Engineering

Bachelorarbeit

zur Erlangung des Grades
Bachelor of Science (B.Sc.)
im Studiengang Games Engineering
an der Julius-Maximilians-Universität Würzburg

Comparative Study of Sound Propagation in Game Environments for Effective Auditory Navigation and Localization

vorgelegt von
Tim Joshua Grün
Matrikelnummer: 2626954

am 24.10.2024

Prüfer: Prof. Dr. Sebastian von Mammen
Betreuer: Yaozhong Zhang, M.Sc.
Lehrstuhl für Informatik IX
Julius-Maximilians-Universität Würzburg

Zusammenfassung

Audio is ein essenzieller Teil von Videospielen, der immersivere Spielerlebnisse ermöglicht und Spielern Hinweise darüber gibt, was um sie herum geschieht. Akustische Lokalisation and auditive Navigation sind dabei nützliche Werkzeuge, welche die räumliche Information der akustischen Signale nutzen. Die Fähigkeit der Spieler, Klänge räumlich einzuordnen, ist jedoch stark von der Implementierung des Audios abhängig. Eine unterschiedliche Modelierung der Schallausbreitung könnte dabei einen signifikanten Einfluss auf die Navigations- und Lokalisationsfähigkeit der Spieler haben.

Zu diesem Zweck haben wir ein Experiment entwickelt, das unterschiedliche Modellierungsarten von Schallwegen vergleicht. Hierbei haben wir drei verschiedene Audiokonfigurationen in sowohl simplen, als auch komplexen, mehrräummigen Innenumgebungen untersucht. Die Ergebnisse zeigen, dass eine realistische Modellierung der Schallwege die Navigaitonsfähigkeit verbessern, die Lokalisierungsfähigkeit jedoch verschlechtern könnte.

Abstract

Audio is an essential part of video games, enabling more immersive experiences and providing players with cues about what is happening around them. Sound localization and auditory navigation are thereby invaluable tools for players, making use of the spatial information these cues provide. However, the ability for players to spatially reason using sound cues is highly dependent on how audio is implemented. Modeling sound propagation in different ways could thereby have a significant impact on navigation and localization performance.

For this purpose, we have designed an experiment to compare different ways of modeling sound paths. We hereby compared three different audio configurations in both simple and complex indoor, multiroom environments. The results show that accurately modeling sound paths may improve navigation performance but worsen localization performance.

Contents

1 Introduction

Audio is an essential part of games for both immersion and atmosphere, but also for gameplay (Larsson & Kleiner, 2002). It hereby provides an additional layer of information, beyond visual information, which does not require direct line-of-sight of the player.

We want to look at two common tasks, which are essential when reacting to sound in games.

1. *Sound localization* is what we define as determining the position of a sound source in an environment. This process is vital in games and useful for players to quickly locate potential threats like enemies.
2. *Auditory navigation* is what we define as navigating through an environment using auditory cues as orientation. In our context, the cues are provided by a single sound source that has to be reached by the player. Games can use auditory navigation to, e.g., guide players to points of interest. It is thereby instrumental in navigating complex environments that are hard to navigate using visual information alone.

For simplicity, this paper will from now on refer to these tasks as *localization* and *navigation*.

The problem with these tasks is that they might not always be easy to perform, especially in complex multiroom environments. It can be especially frustrating when a threat like an enemy is localized incorrectly, or navigating to an audible point of interest leads to the player getting lost. A possible goal could therefore be to maximize localization and navigation ability when designing the audio implementation for a game. Different implementations could hereby lead to different localization and navigation behavior and performance (Lokki et al., 2000; Mehra et al., 2015).

Despite this, the way audio is implemented in games is often very limited (Broderick et al., 2018). Especially the way in which sound waves propagate in the real world is

rarely accounted for in games. Games commonly assume that the sound travels in a straight line to the listener, even when the audio source is behind a wall. In the real world, however, the sound would reach the listener by propagating, e.g., through an open door or around a corner to reach that room. This is because sound waves can bend around corners (diffraction), bounce off walls (reflection), etc. (Kuttruff, 2016). Depending on what effects are modeled, the volume, frequency content, and directionality of the sound can be affected. These effects would be especially pronounced when the sound has to travel around objects or through corridors to reach the listener.
Mehra et al. (2015) already hint at an improvement in navigation performance when wave propagation effects are simulated more accurately. The first-person shooter Overwatch (Blizzard Entertainment, 2016b) also models certain propagation effects in their audio implementation. They focus on better modeling of sound occlusion and attenuation, with the goal of improving localization performance (Blizzard Entertainment, 2016a).

There are only a few similar research papers that compare navigation and localization performance across different audio implementations (Kuppanda et al., 2015; Lokki et al., 2000; Mehra et al., 2015). This is especially true for complex multiroom environments which are not covered by the existing research. Research on this topic is assumed to be mostly internal to game companies and rarely published (as with Blizzard Entertainment (2016a)). In a scientific context, there are no publications from game companies to be found. The aim of this thesis therefore is to fill that research gap.

For this purpose, we have designed and conducted an experiment, testing navigation and localization performance in indoor multiroom environments. We hereby compare three different audio configurations (*Basic*, *Pathing*, and *Mixed*), which were implemented using Steam Audio (Valve Corporation, 2024).
These configurations have different sound propagation approaches, which differ in the way the sound path is modeled. This most notably affects the attenuation and perceived sound direction change along the sound paths.

The following Chapter 2 will cover the current state of audio systems, used in games. It will also provide an overview of prior studies examining navigation and localization performance. The design of our experiment is described in Chapter 3. Next, Chapter 4 will cover the implementation of this experiment, along with providing detailed information on the different audio configurations. Chapter 5 will lay out the results of the experiment. These results are then discussed in Chapter 6, including notable observations about participant behavior. Finally, Chapter 7 will summarize the findings of this thesis, state its limitations, and elaborate on the possibility for future work.

2 Related Work

This chapter will cover the current state-of-the-art in game audio systems and existing research on the topic of navigation and localization.

There is a wide range of possibilities regarding implementation of audio. Despite this, popular game engines like Unreal Engine (Epic Games, 2024) and Unity (Unity Technologies, 2024), offer only basic audio system features out of the box. Often audio middleware like Wwise (Audiokinetic, 2023) is used to allow for more control over the sound and more realistic effects (Fırat et al., 2022). However, many games stick with the capabilities of their respective game engine, and do not allocate additional resources to achieve more realistic spatial audio.

In recent years, AAA games have started to invest more resources into implementing spatial audio effects into their games. Spatially accurate audio is especially important in competitive first-person shooters. Therefore, games like Overwatch (Blizzard Entertainment, 2016b) and Counter-Strike: Global Offensive (Valve Corporation, 2017) use sound path calculation to more accurately model sound occlusion effects. In these examples, sounds are smoothly attenuated and filtered, depending on how much the sound has to travel around corners and obstructing objects to reach the listener. Rainbow Six Siege (Ubisoft, 2015) goes a step further and additionally models how the perceived direction of the sound changes. This is done based on what door, window, or other opening the sound has to travel through to reach the listener (Game Developer, 2017).

Although these implementations are quite interesting, they are often proprietary and not available to the public. However, there are some spatial audio plugins that can be integrated into game engines like Unreal Engine and Unity. The main commercially used plugins that support sound propagation effects are Wwise Spatial Audio and Steam Audio. These plugins only approximate sound propagation effects, while the now discontinued Project Acoustics (Microsoft Corporation, 2022) even supported pre-calculated wave simulations.

Steam Audio supports many different features relevant to this paper. It supports 3D positional audio rendering using Head-Related Transfer Functions (HRTFs). This is a significant improvement over traditional panning methods. It enables users to perceive the direction of the sound, including elevation and differentiation between front and back (Cheng & Wakefield, 2001). Sound occlusion can also be modeled using Steam Audio. It hereby applies filters to the sound whenever the sound is occluded by solid objects like walls. Transmission can also be adjusted to model sound penetrating through objects. Finally, it supports sound propagation modeling by calculating multiple paths through the environment to reach the listener. It can then model how the perceived direction changes, how the sound is attenuated along the path, and how the sound is distorted after diffracting around corners. This is the most important feature and the reason why Steam Audio was chosen.

After establishing the background on game audio systems, we will now look at previous research on auditory navigation and sound localization.

Lokki et al. (2000) conducted a navigation experiment, where subjects needed to find an audio source in a simple (single room or open field) environment. They showed that participants could successfully navigate to the audio sources using the auditory cues. Additionally, three variables were tested. The results show that 1. Noise performed better than the other two sounds (flute or guitar), 2. Reverberation had a negative effect on navigation time, and 3. Simple panning methods were sufficient for the navigation tasks. In a follow-up study, Gröhn et al. (2005) performed another navigation experiment. This time, the experiment used a Virtual Reality (VR) environment, utilizing a specialized input device to allow for full 3d movement in all directions. They compared audio, visual, and audiovisual cues for navigation and found that audiovisual cues performed by far the best, while sole audio cues performed the worst.

Research on auditory navigation is also done in the field of auditory displays for the visually impaired. Walker and Lindsay (2006) hereby conducted a successful navigation experiment in VR, with the only relevant issue being participants overshooting the waypoints.

Navigation was also used to evaluate the wave-based sound propagation system developed by Mehra et al. (2015). Their system produced multiple sound propagation effects like diffraction and reflection. In their evaluation, they observed better navigation performance for their wave-based sound propagation system compared to a previous geometry-based system.

Based on the overwhelming success of the previous navigation experiments, we expect navigation to still be possible, even in our more complex multiroom environments. We hereby hypothesize each audio configuration to perform differently, based on the differences observed by Mehra et al. (2015) and Lokki et al. (2000).

Regarding localization, the ability of humans to localize direct sound has been extensively studied (Fastl & Zwicker, 2006; Middlebrooks, 2015). More recently, Kuppanda et al. (2015) also developed a VR framework to test direct sound localization performance to aid in the development of auditory displays. However, there seems to not be any research regarding sound localization across multiple rooms.

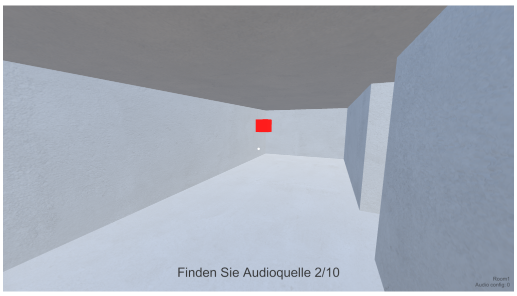
Summarizing the existing research, all studies examine only simple, mostly single room environments. The only exception being Mehra et al. (2015) which used outdoor game scenes containing a small number of obstacles. Environments of large complexity or consisting of multiple rooms, on the other hand, are not covered by the existing research.
This paper thereby opens a new area of research, examining navigation and localization performance in complex, multiroom environments.

3 Experiment Design

We designed an experiment to test the navigation and localization performance with different audio configurations and environments. The following sections describe the experiment in detail, its procedure, and how it was conducted.

3.1 Procedure

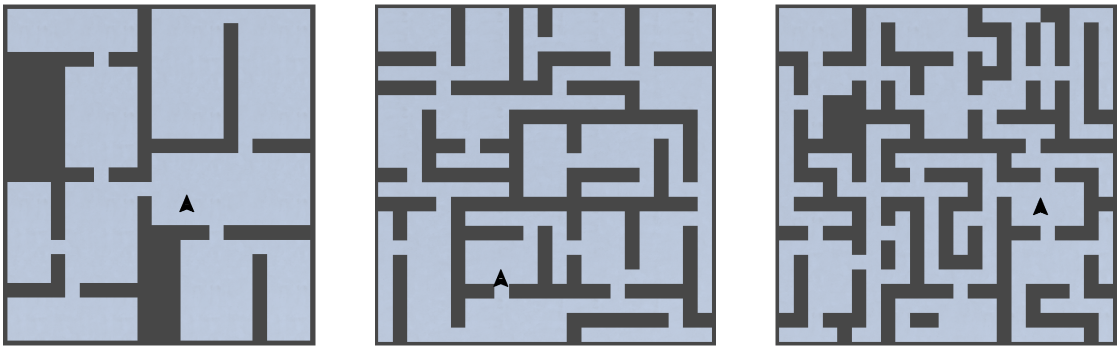
The entire procedure takes around 20 minutes to complete and consists of 1. Reading the participant information document (see Appendix B) and listening to additional instructions 2. Playing the *study game*, and 3. Answering questions about the participant's demographic.
Most of the trials were conducted remotely on the participants' PCs, so they were instructed to download the study game beforehand. They should also verify that they have working stereo headphones and wear them while playing the study game. It is important to note that the experiment was conducted in german. Some instructions in the following sections may be translated for consistency.



(a) Navigation (the red box is the audio source)

(b) Localization

Figure 3.1. Screenshots taken during the study



(a) Scene 1 **(b)** Scene 2 **(c)** Scene 3

Figure 3.2. Maps of the different environments that were used

3.2 Study Game

In the study game (shown in Figure 3.1), participants control a first-person character using a standard keyboard control scheme. The WASD or arrow keys are used for movement, while the mouse is used to turn and look around. The game takes place in three different environments (also called "scenes") of increasing complexity (see Figure 3.2). In these scenes, multiple navigation and localization tasks have to be completed. The tasks for a specific scene are hereby always the same across all trials. When all tasks in a scene are completed, the scene and audio configuration change. The participants are notified of this change, but not informed in what way the audio configuration is now different.

The scenes are always played in the same order, while the order of audio configuration is different across the trials (more on this in Section 3.2.3). The pairing of an audio configuration and a scene is what we call a "scenario". In addition to the three main scenarios, there is a brief tutorial scenario, which contains fewer navigation and localization tasks. The goal of the tutorial scenario is to teach the participants how each task works. Therefore, the performance in this scenario is not measured toward the final results.

The following sections explain the previously mentioned specifics, like the different environments and audio configurations in more detail.

3.2.1 Environments

As mentioned before, there are three different environments with different levels of complexity. They have a plain look with a maze-like structure, consisting of multiple rooms and corridors.
The three environments are:

1. Simple (see Figure 3.2a): Small number of connected rooms. There are no loops, which means that there is only a single way to get from one room to another.
2. Complex (see Figure 3.2b): Larger number of rooms. There are many corridors with dead ends. There are also multiple loops and ways to get from one room to another.
3. Very complex (see Figure 3.2c): Mainly consists of erratic corridors and a few small rooms. Corridors are often looping or have dead ends, which are only visually discernible after entering them.

3.2.2 Audio Configurations

The three different audio configurations each differ in the path, the sound takes to the listener.

1. *Basic*: The sound always travels in a direct line to the listener, even if there are obstructing objects.
2. *Pathing*: The sound always travels around obstructing objects and through corridors to reach the listener.
3. *Mixed*: The sound partly travels with *Basic* and partly with *Pathing* .

More details on the differences between these audio configurations are provided in Section 4.2.

3.2.3 Order of Audio Configurations

The three scenes are always played in the same order, while the order of audio configurations is different across trials. Each trial is assigned one of the six possible orders at the start of the experiment. This assignment is mostly random but manually adjusted to ensure that the orders are equally distributed across all trials. The manual assignment is also used to ensure that the reported experience level of the participants is not disproportionately larger for trails of one order than another.

3.2.4 Navigation Tasks

For the navigation tasks, participants need to navigate between ten audio positions at predefined locations. At the start of each task, the audio source starts playing. The player then needs to find the audio source by navigating to it as quickly as possible. The source is hereby visualized as a small red box (see Figure 3.1a). As soon as the player finds it and comes into contact with the red box, the audio source counts as found. After a short pause of 1.25s, the audio plays at the next location, and these steps are repeated until all 10 audio sources are found.

During this task, data on the task execution is collected. This data includes the time needed to find each audio source and the path the player takes.

3.2.5 Localization Tasks

For each localization task, participants need to guess the position of an audio source at a predefined location. The goal hereby is to be as accurate as possible, while taking as much time as needed for the guess. During these tasks, the player character is positioned at a specific location and cannot be moved unlike in the navigation tasks. The head of the player character can still be turned, which may help in discerning the direction of the sound (Wightman & Kistler, 1999). At the start of each task, audio is played at six random example positions while being shown on a map of the environment. These positions can be used as references to be able to discern sound coming from different locations in the environment. Participants are hereby instructed to pay close attention to the direction and perceived distance of the sound. Each example is played for 1.6s with a pause of 0.15s between them. The example positions are especially important, as participants need to adapt to each new audio configuration after each scenario change.
After the example positions, audio is played at the position, the participants need to make a guess on. Participants can listen to the audio as long as the need to make

a guess. When they are ready, they can open the map again, and click on it to place a marker (see Figure 3.1b). When the participant is confident about their guess, they can confirm it or listen further to revise it.
These steps, including the example audio positions, are performed for each of the seven different audio positions.

The data collected during this task includes the guessed position and the time the participants take to make their guess.

3.2.6 Sound Choice

The base of sound, participants hear in the navigation and localization tasks is white noise. White noise is used, as it is established to perform quite well in direct sound localization tests (Kuppanda et al., 2015; Lokki et al., 2000). It also covers the full spectrum of audible frequencies, so effects like audio filters are more noticeable. This noise is additionally ring modulated with a 33hz sine wave, which causes a fast, flutter-like effect. The sound can most closely be described as sounding like the sped-up audio of a helicopter. The ring modulator was added to make the sound less harsh, and more pleasant, and interesting to listen to over a long period of time.

3.3 Participants

Trials were conducted with a total of 13 participants, consisting of mostly personal contacts and colleagues. Twelve participants were able to complete the experiment without issues. The exception being one trial, in which the participant experienced some issues during the navigation tasks. They reported that the sound was too loud for them, which made them stop several times to adjust the volume. Therefore, it was decided to discard their results and exclude them from any of the following data.

All participants were required to have normal hearing with no known hearing problems, although this was not additionally verified. To collect data on the demographic of the participants, they were asked to answer questions about their age and gaming experience: 1. “How old are you?” 2. “How often have you played video games?”

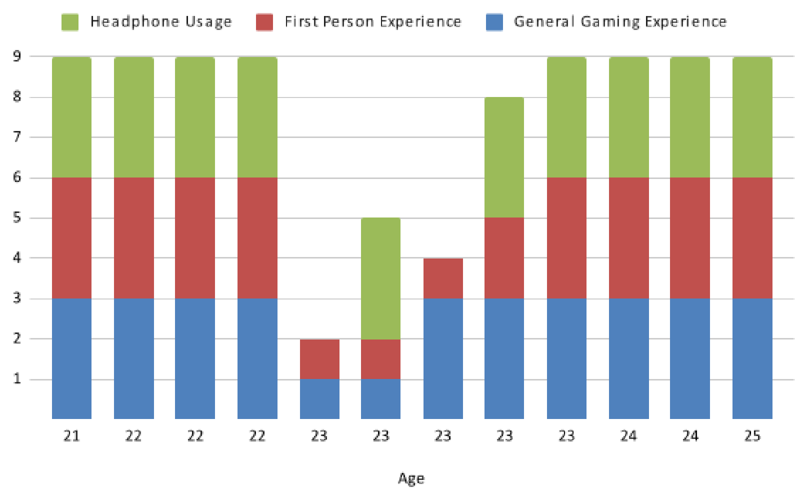


Figure 3.3. Age and gaming experience of participants, likert scale answers are converted to a numerical scale of 0–3

3. "How often have you played first-person games?" 4. "Did you wear (stereo) headphones while doing so?" The answers about gaming experience and headphone usage were constrained to a four-point likert scale with the answers *Never*, *Rarely*, *Occasionally*, and *Often*.

The age of the participants ranged from 21 to 25. All of them had at least some general and first-person gaming experience, while most of them had great experience. Most of the participants also answered *Often* for headphone usage, while only 2 answered *Never*. An overview of the age and gaming experience for each participant is shown in Figure 3.3.

4 Implementation

The study game was developed in Unity (2022.3.34f1) using the Steam Audio Unity Plugin (v4.5.3). The next sections will cover the implementation of the study flow and detailed information on the audio implementation. Furthermore, the generation of the environments and the utility window, used to analyze and visualize the data, will be described.

4.1 Study Flow

The study was implemented in C# and makes heavy use of unity coroutines. Coroutines are functions that can pause their execution at any point, and continue on the next frame or after a certain condition has been fulfilled. Coroutines still run on the main unity thread though, which enables them to call unity functions that are restricted to this thread. In the study game, coroutines are commonly used to wait for user input, for a goal to be achieved or for brief pauses between tasks.
To allow for this functionality at all stages, the entire study flow is made up of a hierarchy of coroutines that each wait for the completion of their subroutines. A simplified version of this hierarchy is shown in 4.1.

4.2 Audio

As mentioned before, Steam Audio was used to spatialize the sounds and simulate the sound propagation effects. The general Steam Audio settings enabled sound occlusion and disabled reflections and reverb. Reflections were explicitly disabled to not introduce an additional variable. Additionally, reflections could have had a negative effect on localization performance, as shown by Lokki et al. (2000). Apart from that, we also enabled HRTFs to allow for more accurate spatialization of sounds. The full details of the Steam Audio settings can be found in Appendix C.

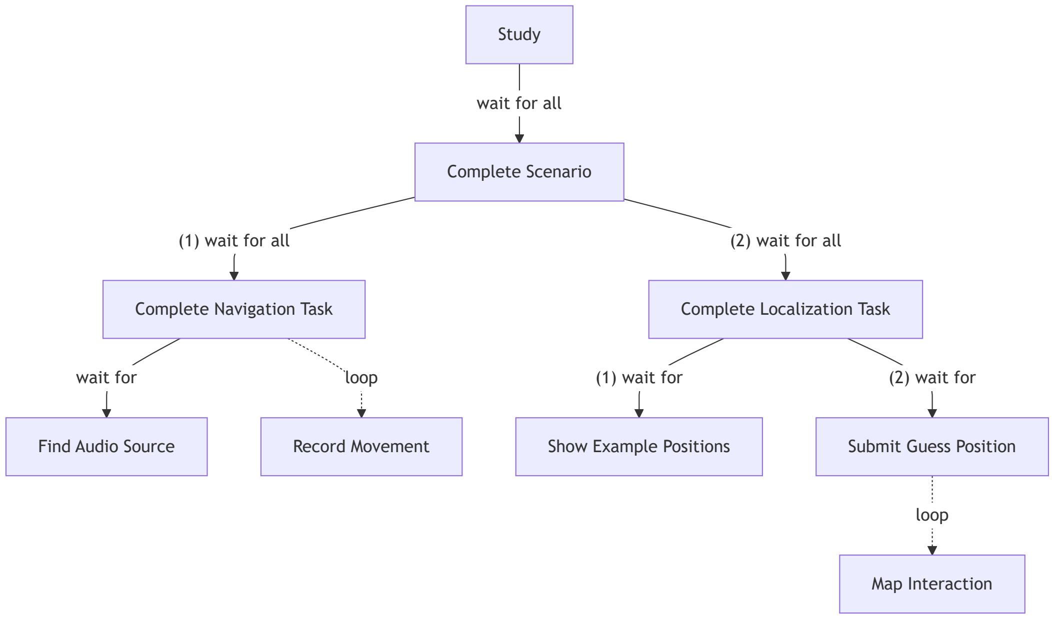


Figure 4.1. Execution Hierarchy of Coroutines (Simplified)

The different audio configurations in the experiment additionally change the pathing and transmission mix in the Steam Audio settings. Note that the transmission mix for the high and mid frequency band are always set to 0%, and only the mix for the low frequency band is adjusted. The following sections show the settings for each audio configuration. To put the results in more context, we also want to provide a brief explanation on how the perceived sound changes, when the position of the audio source changes.

4.2.1 *Basic*

Settings: Pathing mix = 0%, low transmission mix = 100%.

The perceived sound direction is the direct vector between the listener and the audio source. Additionally, the sound becomes quieter as the direct distance between the listener and the audio source increases.
It follows that every position on the map has a unique combination of volume and direction. In practice, two audio positions should be distinguishable whenever the distance between them is large enough for the change in volume and direction to be perceivable.

Unlike with *Pathing*, the perceived sound only carries information about the position of the source and not how to reach the source. This could be disadvantageous for navigation tasks if the listener does not know how to reach the assumed location. Therefore, this configuration was hypothesized to perform well at localization tasks, while performing badly at navigation tasks, especially in the most complex environments.

4.2.2 *Pathing*

Settings: Pathing mix = 100%, low transmission = 0%.

The perceived sound direction is the last segment of the shortest path between the listener and the audio source. As an example, if the sound is coming from a connected, adjacent room, the sound is perceived as coming from a door to that room. This also means that the perceived direction does not change when the source is moved inside that room. For volume, the sound becomes quieter whenever the length of the shortest path between the listener and the audio source increases. Additionally, Steam Audio applies filters, depending on how the sound should be distorted after diffracting around corners. In the tests, this was not noticeable enough to help in localization. This effect might also be masked by the filter that is applied when the sound is occluded.
With this configuration, not every position on the map can be uniquely identified. It is likely not possible to distinguish two audio positions when the shortest path to the listener is similar (i.e. they share the last segment), while the length of the paths is the same. A small advantage over *Basic*, however, is that when two source locations have a short direct distance[1] but a long path distance[2], the difference in volume is more pronounced (e.g., when they are on other sides of a wall). In specific cases, this could be useful when differentiating, which room a sound is coming from. With respect to navigation, this audio configuration does carry information about how to reach the source. The listener can hereby follow the path, the sound takes using the perceived sound direction. With these impressions, this audio configuration was hypothesized to perform the best at navigation tasks, while performing the worst at localization tasks.

[1]Direct distance: The distance between two points along a direct line
[2]Path distance: The distance between two points along a traversible path through the environment

4.2.3 *Mixed*

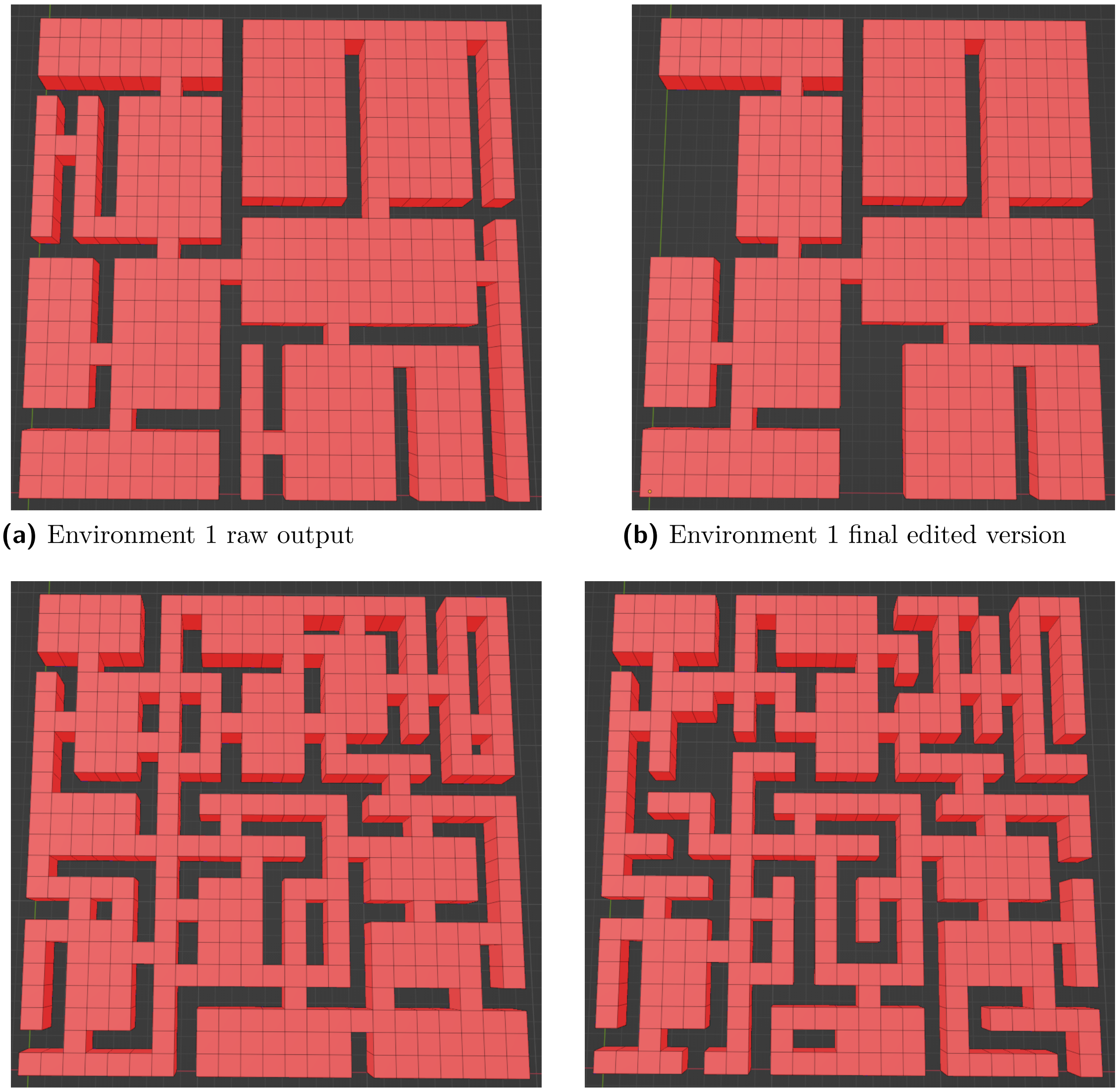
Settings: Pathing mix = 100%, low transmission mix = 40%.
The transmission mix value was hereby fine-tuned to provide a good balance between pathing and transmission information.

As this audio configuration mixes the sound of *Basic* and *Pathing*, the perceived sound direction is somewhere in the middle of where *Basic* and *Pathing* would be perceived. It follows that the sound is attenuated over both the direct and path distance.
Mixing two configurations, the sound balances information about the path and the position. The differences and also strengths of the two configurations are likely less pronounced though. Therefore, *Mixed* was hypothesized to perform average on both the navigation and the localization tasks.

4.3 Environments

The environments were generated procedurally using a recursive subdivision algorithm. This was implemented to test different kinds of environments with different levels of complexity. The algorithm is implemented as a python script in blender and is based on an existing implementation (Olson, 2019, February 9/2024), that was slightly adapted for this paper.

The environments were all generated with the same size of 23x23 blocks, each block being 1.25m wide and 2.25m tall. After generating a large number of environments with different complexity, the best results were selected. Additionally, testing feedback was collected to check whether the environments were suitable for the study. Over this iterative process, some initially simple environments were replaced or modified to increase navigation complexity. Modification was also often used to remove unwanted artifacts, as seen in Figure 4.2a and 4.2b.
The process of creating more complex environments involved two modification steps on top of the generated output. Firstly, additional connections between rooms and corridors were added to increase the number of possible paths, as seen in Figure 4.2c. Then some connections were removed to make finding the path to a particular room harder. Environment 3 was specifically edited in such a way to produce corridors



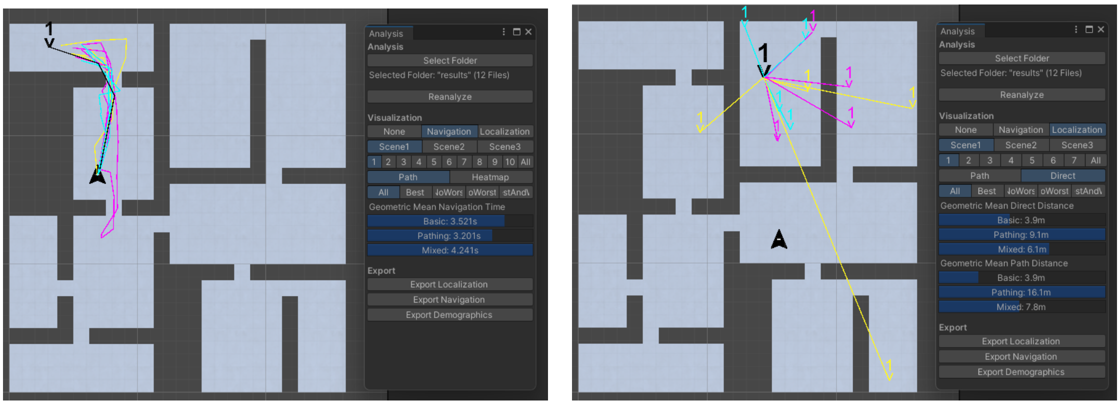
(a) Environment 1 raw output

(b) Environment 1 final edited version

(c) Environment 3 after some additional connections were added

(d) Environment 3 final edited version

Figure 4.2. Environments before and after editing



(a) Navigation analysis example **(b)** Localization analysis example

Figure 4.3. Analysis Window

with dead ends, that have at least one additional turn. This makes them not visually identifiable as dead ends without entering them. The final result can be seen in Figure 4.2d

To quickly set up new environments, a unity editor utility was created, that replaces the meshes, exports the scene for Steam Audio, and generates the sound probes used for sound path calculations.

4.4 Data Analysis

The data collected during the study is exported as a JSON file, after the study game is completed. To be easily able to analyze, group, and transform multiple data objects in code, they were then re-imported into unity. We designed a unity editor window specifically for this purpose, so we could also display different metrics and visualize the data.

For the navigation task, the paths that players took can be either visualized as lines on the game map or aggregated as a heatmap. Additionally, the average navigation time for each audio configuration can be displayed as a bar graph (see Figure 4.3a).

For the localization task, each guess can be visualized as a pin on the map. The average direct and path distance between guess and audio position is also calculated for each audio configuration and displayed as a bar graph (see Figure 4.3b).

The path distance is hereby calculated using the pathfinding implementation of Steam Audio. It is important to note though, that this distance is only approximate, as the paths calculated by this implementation are simplified and therefore slightly shorter.

The window additionally allows for exporting data as tables, so more elaborate graphs can be created in spreadsheet programs. The tables are exported in the xlsx format using the ClosedXml library (ClosedXML, 2016, August 27/2024).

5 Results

This section will lay out the results of the navigation and localization tasks based on the data that we collected.

5.1 Navigation Results

Evaluating the navigation times, we calculated the average navigation time for each scenario (see Figure 5.1).
To explain how this value is calculated, each scenario was done a total of four times. The four navigation times are then averaged per task using the geometric mean. This value is then averaged across all tasks in a scenario to create the final average navigation time. We chose the geometric mean to not skew the results as much, as there are some outliers in the data, that take an unusually long time to find the audio source.
During the study, it was also observed that participants sometimes took time to start moving after a task had already begun. These breaks were very inconsistent across participants and tasks. Due to the inconsistency and small sample size, we decided to count the time before the player has started moving with a highly reduced factor. Even though some participants may have unknowingly exploited this to pre-listen and thus improve their navigation time, this behavior may also have been unintentional. Overall, the navigation times become more consistent after applying this adjustment.

Looking at the navigation times in Figure 5.1, generally all participants were able to navigate to the audio source in a reasonable amount of time. This was true across all audio configurations, but with some having consistently faster navigation times in certain scenes.
For scene 1, all audio configurations have similar navigation times. For scene 2, there are major differences between the navigation times with *Basic* performing the

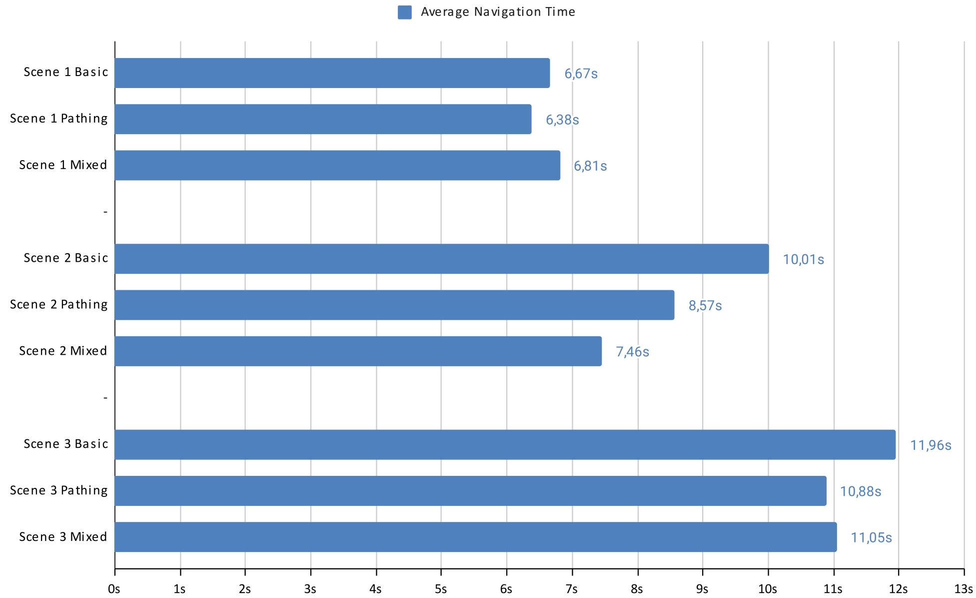
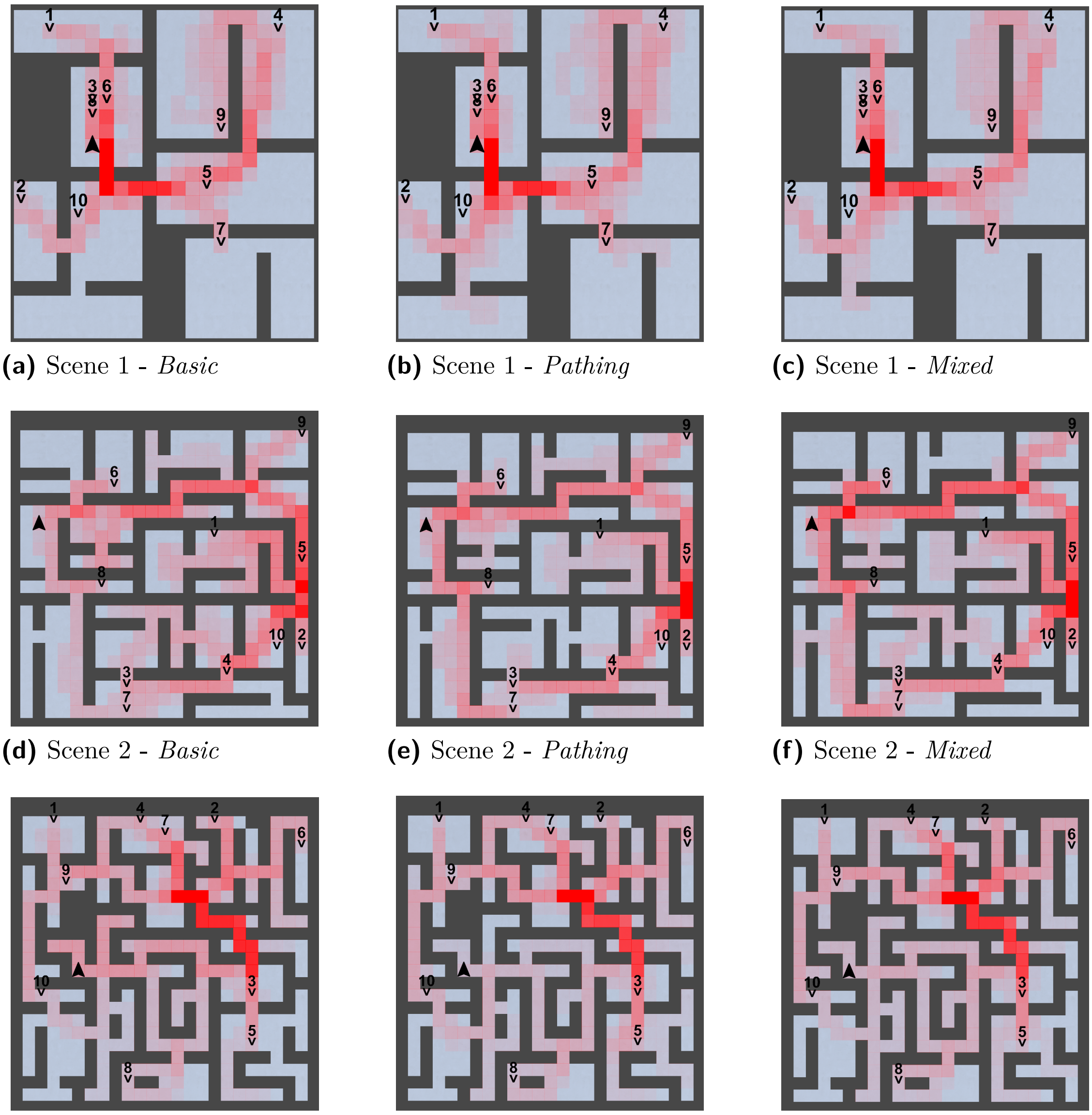


Figure 5.1. Average navigation time for each scenario

worst at 10.01s. *Pathing* was 1.44s faster on average, while *Mixed* was the fastest, being an additional 1.11s faster than *Pathing*. For scene 3, *Basic* is roughly a second behind *Pathing* and *Mixed*, which both have similar navigation times.

The collected heatmaps of player positions across all navigation tasks (Figure 5.2) show only slight differences.
The heatmaps for scene 1 look almost identical, while the heatmaps for scene 3 also look quite similar. The only immediately visible feature in the heatmap for scene 3 is, that players using *Basic* were more likely to enter the dead ends near the start point. For scene 2, players with the *Pathing* configuration can be seen to have visited dead ends or alternative paths less often then players with the *Basic* configuration. *Mixed* seems quite similar to *Pathing*, but players seem to have stayed on the main path a little more consistently.
To more closely analyze differences in player behavior, individual tasks need to be analyzed. This is done in the following Chapter 6.



(a) Scene 1 - *Basic* **(b)** Scene 1 - *Pathing* **(c)** Scene 1 - *Mixed*

(d) Scene 2 - *Basic* **(e)** Scene 2 - *Pathing* **(f)** Scene 2 - *Mixed*

(g) Scene 3 - *Basic* **(h)** Scene 3 - *Pathing* **(i)** Scene 3 - *Mixed*

Figure 5.2. Heatmap of player positions during navigation tasks

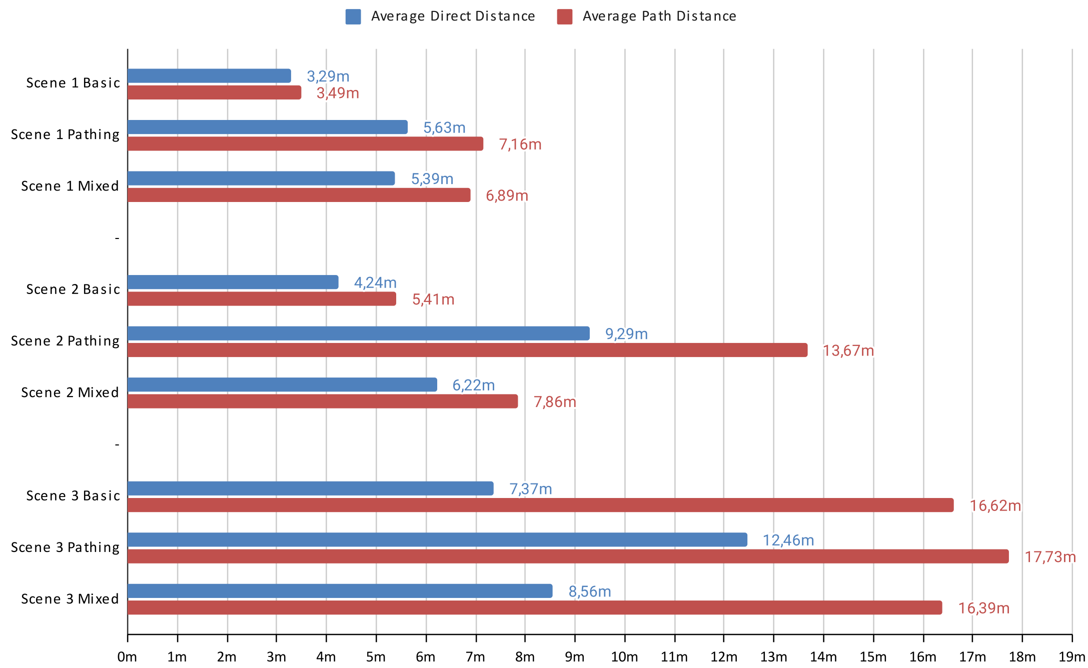
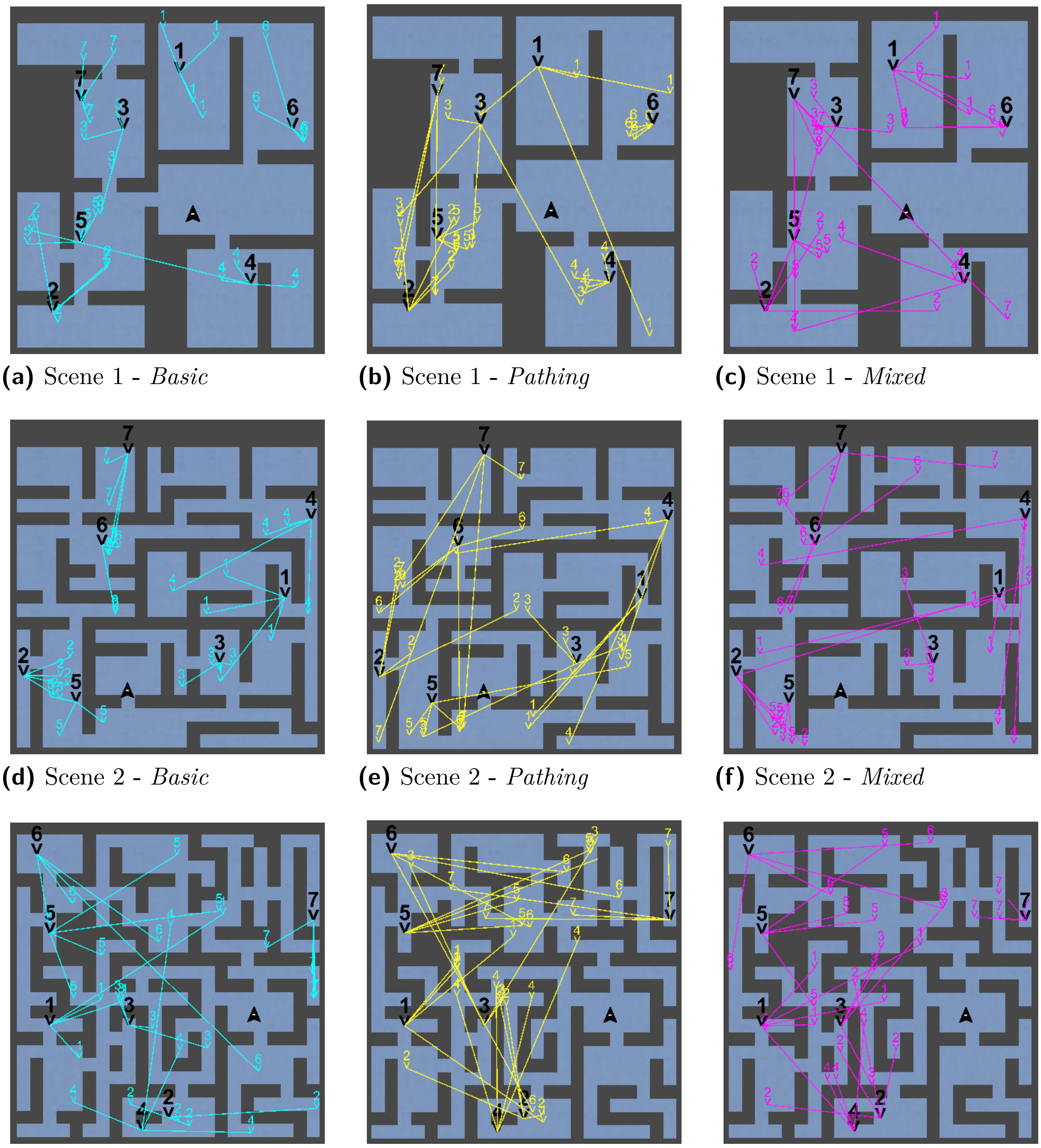


Figure 5.3. Average distance between guess and audio source for each scenario

5.2 Localization Results

The results for the localization tasks are visualized in two figures. Figure 5.3 shows the average direct and path distance across all tasks of a particular scenario. This average is calculated over the geometric mean of the four guesses per task in a particular scenario. For context, this is the same way the data was averaged for the navigation tasks. The second Figure 5.4 visualizes all guessed positions for each scenario on their respective maps. The following paragraph will describe the guess positions and distances shown in the two figures.

In scene 1, the guesses of *Basic* are mostly in the correct room and always quite close to the actual audio position. For *Pathing* the guesses in adjacent rooms (4, 5, 6) are extremely precise, generally even closer than the other audio configurations. The other guesses are way more inaccurate, however. The audio positions (2,3,7) in the rooms on the left partition of the map are generally identified to come from the left side, but the correct room could not be identified consistently. *Mixed* often has two guesses, that are very close to the actual audio position but also some outliers, which make the distance on average similar to *Pathing*.
In scene 2, the guesses of *Basic* become only slightly less accurate and are still quite



(a) Scene 1 - *Basic* **(b)** Scene 1 - *Pathing* **(c)** Scene 1 - *Mixed*

(d) Scene 2 - *Basic* **(e)** Scene 2 - *Pathing* **(f)** Scene 2 - *Mixed*

(g) Scene 3 - *Basic* **(h)** Scene 3 - *Pathing* **(i)** Scene 3 - *Mixed*

Figure 5.4. Map of guessed positions during the localization task

consistent. *Pathing* on the other hand, already becomes very inconsistent in this environment. Only the close audio positions (3, 5) can be identified with some level of accuracy, while the others are extremely inaccurate. The path distance is also disproportionately large for this configuration. The guesses of *Mixed* are still quite accurate with less difference to *Basic* in this scene. There are still one or two outliers per task though, which mainly contribute to this difference.
In scene 3, the inaccuracy of all guesses increases significantly again, especially for *Basic*. After this increase *Basic* and *Mixed* are quite similar, while *Pathing* is even more inaccurate. Interestingly, *Mixed* has more consistent guesses per task with no consistent number of outliers. Even though the direct distance is still much better for *Basic* and *Mixed* than *Pathing*, the path distance is similar across all configurations.

6 Discussion

Like we hypothesized, navigation was successful across all audio configurations. We also confirmed the difference in performance between the audio configurations, but only for the more complex environments. *Pathing* performed very well, achieving consistently better navigation times than *Basic*. *Mixed* exceeded our expectations by performing similarly to *Pathing* in scene 1 and 2 and even outperforming it by a significant margin in scene 2. Possible explanations for this are covered in Section 6.1.3.

Localization performance significantly decreased as the environments got more complex. However, we could still confirm our initial hypothesis about which audio configurations would perform the best. *Basic* produced consistently accurate guesses in the first two scenes, while *Mixed* was a little more inaccurate on average, commonly having outliers. *Pathing* performed by far the worst, only producing good guesses when the audio source was very close or located in an adjacent room. For scene 3, localization was very inaccurate across all audio configurations. Despite the still significant differences in direct distance, the average path distance was similar and also very high. These guesses would therefore not be that useful, thinking of an actual game context.

With the following sections, we want to share some notable observations about the navigation and localization behavior. Our analysis also aims to provide possible explanations for the observed results.

6.1 Navigation Strategy

All participants seemed to follow a similar navigation strategy. To reach the sound source, participants overwhelmingly moved in accordance with the perceived sound direction. Details on the perceived sound direction were previously explained in Section 4.2.

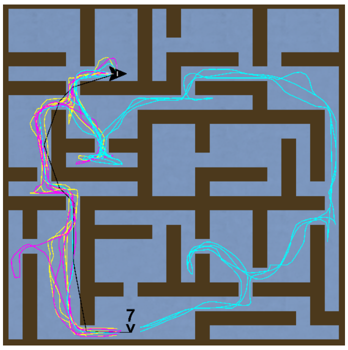
The described strategy seems quite intuitive and reasonable, as the environments were both unknown and too complex to memorize in such a short time. The following sections will analyze the observed navigation behavior and the effects of using this strategy across the different audio configurations.

6.1.1 Common problems

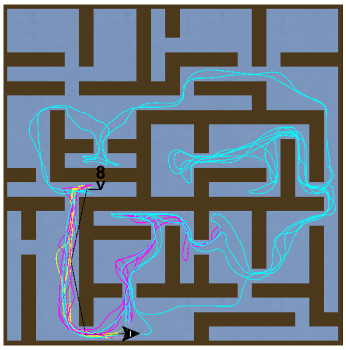
Following the perceived sound direction is a very effective strategy, when it is continuously possible to move in that direction. Within the tested environments, this should always be the case for *Pathing*, as the sound can only travel through the traversable rooms and corridors. For *Basic* on the other hand, it is often not possible to move in this direction, e.g. when there is a wall blocking the way. This is also the case for *Mixed*, although the mix of *Pathing* provides at least some bias towards a traversable direction. To illustrate this point, we can look at two tasks where this issue is very noticeable. Firstly, we will look at task 7 of scene 2 with the navigation paths illustrated in Figure 6.1a. After turning the first corner, players are faced with multiple different options. The players using *Pathing* and *Mixed* all make the correct initial choice with only two of them initially entering the dead end below. On the other hand, the players using *Basic* all enter the dead end. Unlike the other configurations, *Basic* can not use the sound direction to infer which way to go after exiting the dead end. Three out of the four times, they end up making the wrong decision, which results in a way longer path to the audio source.
In the following task in Figure 6.1b, we can see that all participants using *Pathing* make the correct initial decision to reach the source on an optimal path. This is again in accordance with the perceived sound direction. It may play a factor that they all traveled on that path in the previous task. However, everyone using *Mixed* also took that path and most still end up making the wrong initial choice. For *Basic*, the shortest path is also not the intuitive choice. Therefore, three out of the four players using *Basic* end up making the wrong initial choice, which again results in a way longer path.

6.1.2 Optimal Paths

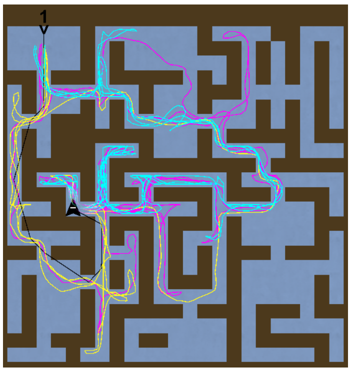
The major upside of using an audio configuration like *Pathing* is that the user not only gets information on how to navigate to the sound source. The sound also always



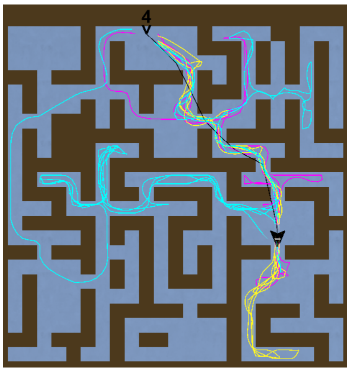
(a) Scene 2, task 7 - *Basic* sound direction is block by wall



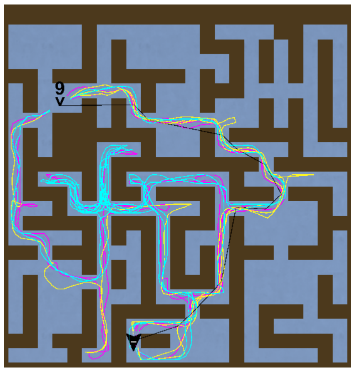
(b) Scene 2, task 8 - *Basic* sound direction is block by wall



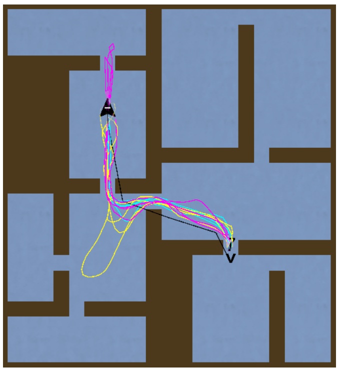
(c) Scene 3, task 1 - *Pathing* is able to find the shortest path



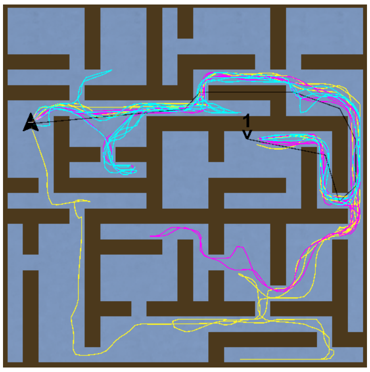
(d) Scene 3, task 4 - *Pathing* is able to find the shortest path after initial confusion



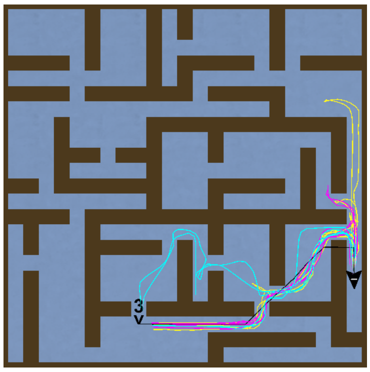
(e) Scene 3, task 9 - *Pathing* is able to find the shortest path



(f) Scene 1, task 7 - *Pathing* overshoots at corner



(g) Scene 2, task 1 - *Pathing* overshoots at multiple corners



(h) Scene 2, task 3 - *Pathing* overshoots at corner

Figure 6.1. Navigation paths for different audio configurations
(cyan: *Basic*, yellow: *Pathing*, magenta: *Mixed*, black: simplified optimal path)

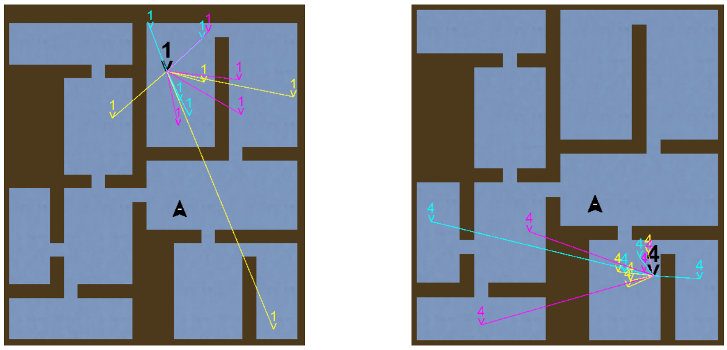
leads them on the shortest, likely optimal path. This is most easily shown with task 1 of scene 3 (see Figure 6.1c). To reach the audio source on the shortest path here, the player must first move away from the source. This goes against all intuition for *Basic*, but as *Pathing* models the sound direction as coming from behind, players are intuitively able to find the shortest path. To provide a few additional examples, Figure 6.1d and 6.1e also show players using *Pathing* to most consistently find the shortest paths.

6.1.3 Problems with *Pathing*

The problem with *Pathing* is that the perceived sound direction often changes. Players are not always that quick to notice this change, which commonly leads to them overshooting past the corridor, that the sound is coming from. This is observable on the previously analyzed maps, but especially noticeable in Figures 6.1f, 6.1g, and 6.1h. Players using *Pathing* also seem to become confused sometimes at the start of a task. This leads to them initially walking into the wrong direction and then correcting. This issue, combined with overshooting, likely resulted in *Mixed* performing better than *Pathing* in scene 2 (see Figure 5.1). It is likely that this is an issue linked to *Pathing* and not special participants, as it can also be observed in scene 3 with different participants (see Figure 6.1d)

6.2 Localization Problems

The localization task has generally proven to be quite hard. Despite this fact, participants often tried to make guesses to the best of their ability. However, there are some guesses that are very far off from the other guesses. For example, in Figure 6.2a there is one odd guess, and in Figure 6.2b there are three odd guesses to be seen. These guesses are hard to intuitively explain, as they don't correspond to the perceived sound direction at all. They also do not always come from the same participants. In the mentioned examples, all four guesses come from different participants, two of them even having great experience with first-person games and using headphones. We do not know yet if these guesses are just random or if there is a reason to why they were made.



(a) Scene 1, task 1 - one odd guess in the lower right corner

(b) Scene 1, task 4 - three odd guesses in the lower left corner

Figure 6.2. Examples of odd guesses during the localization task (cyan: *Basic*, yellow: *Pathing*, magenta: *Mixed*)

One issue we found was that some participants did not realize at first that they were able to turn their heads. This may have decreased their accuracy for some of the earlier tasks. Also, some participants found the example positions at the start of each localization task to be quite tedious to sit through. This might have caused some participants to pay less attention, which may have also affected localization performance negatively.

The final interesting observation we made about the localization results is the difference in performance for *Basic* across the different scenes. In theory, *Basic* should be expected to perform about the same across all scenes. The reason for this is that the sound of *Basic* only depends on the position relative to the player. The specific arrangement of walls does not affect the sound in any way. Steam Audio also does not apply a stronger occlusion effect when the sound is behind multiple walls. As seen in Figure 5.3 this is partly the case between scene 1 and scene 2. The direct distance only slightly increases despite the significant increase in environment complexity. Between scene 2 and scene 3, however, the direct distance increases by a great margin, despite having essentially the same sound perception. This could be caused by many different factors. For example, participants may just be overwhelmed by the environment or less attentive at this point in the experiment. We hypothesize that the changes in audio configuration also play a large role here. Participants may have not fully adapted to the new audio configuration yet or are confused by the previous localization tasks with different audio configurations. This has likely diminished the localization performance for all audio configurations.

7 Conclusion

The goal of this thesis was to test navigation and localization performance for different environments and audio configurations. For this purpose, we designed an experiment and implemented it in Unity using Steam Audio.
The results show that navigation was possible across all audio configurations and environments. We thereby saw significant differences in navigation time when the environments were sufficiently complex. The results suggest that navigation performance may be improved by modeling certain sound propagation effects. We also observed the perceived sound direction to be one of the most important cues during navigation. The audio configurations *Pathing* and *Mixed*, which based this direction on a realistic sound path, were able to lead participants much more efficiently to the audio sources compared to *Basic*. In our experiment, a *Mixed* approach resulted in the best navigation times.
For localization, however, these audio configurations were observed to generally produce more ambiguous sound cues. To make accurate guesses, *Basic* seems to be the most intuitive for participants, followed by *Mixed*. *Pathing* only allowed for consistently accurate guesses when the audio source was located in an adjacent room. Otherwise, it produced quite inaccurate guesses. With increasing complexity of the environments, the localization error also increased. While it was possible to make fairly accurate guesses in the simpler environments, participants were not able to accurately localize sounds in the most complex environment.

While we think the study was an overall successful experiment, there are some limitations. While the observed results were fairly consistent, it is hard to make generalized statements because of the limited sample size. There are also factors that were not uniform across all trials, like the different experience levels of participants. Furthermore, the testing environments and hardware differed across trials as the study was mostly conducted remotely. Another factor that may have had a greater impact than initially expected was the switching of audio configurations. We assume that this did not influence the navigation results as much, as this task was much more intuitive for participants. For localization, however, this may have noticeably decreased the accuracy in the later scenes.

Looking at the transferability of the results, this thesis focuses on the comparison of very specific audio configurations. Hereby, we intentionally excluded factors such as reverberation and only relied on a fairly simple sound occlusion model. Our results might not be directly transferable to other audio implementations in games, as they could differ greatly. Games may also want to include additional effects for realism or artistic purposes. Finally, the geometries of our environments lack in detail compared to most games. Open windows and other scene elements could hereby significantly affect the sound paths.

More research may be needed to prove the statistical significance of the results, using a study with a larger sample size. We would also like to see how transferable our results are to actual game contexts.
It would also be useful to research how the results change after participants have significant practice with the environment or one of the audio configurations. Further research might also investigate additional audio configurations and effects. This could lead to valuable insights to designing an audio system optimized for navigation or localization performance.

Bibliography

Audiokinetic. (2023). *Wwise* (Version 2023.1). https://www.audiokinetic.com/en/wwise/overview/ (cit. on p. 5).

Blizzard Entertainment. (2016a, March). *Overwatch - The Elusive Goal: Play by Sound (GDC 2016)*. Retrieved October 20, 2024, from https://www.gdcvault.com/play/1023317/Overwatch-The-Elusive-Goal-Play (cit. on p. 2).

Blizzard Entertainment. (2016b, May 24). *Overwatch*. (Cit. on pp. 2, 5).

Broderick, J., Duggan, J., & Redfern, S. (2018). The Importance of Spatial Audio in Modern Games and Virtual Environments. *2018 IEEE Games, Entertainment, Media Conference (GEM)*, 1–9 (cit. on p. 1).

Cheng, C., & Wakefield, G. (2001). Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space. *AES: Journal of the Audio Engineering Society*, *49*, 231–249 (cit. on p. 6).

ClosedXML. (2024, October 23). *ClosedXML/ClosedXML*. Retrieved October 23, 2024, from https://github.com/ClosedXML/ClosedXML (cit. on p. 21).

Epic Games. (2024). *Unreal Engine 5* (Version 5.4). https://www.unrealengine.com/ (cit. on p. 5).

Fastl, H., & Zwicker, E. (2006, December 14). *Psychoacoustics: Facts and Models*. Springer Science & Business Media. (Cit. on p. 7).

Fırat, H. B., Maffei, L., & Masullo, M. (2022). 3D sound spatialization with game engines: The virtual acoustics performance of a game engine and a middleware for interactive audio design. *Virtual Reality*, *26*(2), 539–558 (cit. on p. 5).

Game Developer. (2017, January 18). *Game Design Deep Dive: Dynamic audio in destructible levels in Rainbow Six: Siege*. Retrieved October 20, 2024, from https://www.gamedeveloper.com/design/game-design-deep-dive-dynamic-audio-in-destructible-levels-in-i-rainbow-six-siege-i- (cit. on p. 5).

Gröhn, M., Lokki, T., & Takala, T. (2005). Comparison of Auditory, Visual, and Audiovisual Navigation in a 3D Space. *TAP*, *2*, 564–570 (cit. on p. 6).

Kuppanda, T., Degara Quintela, N., Worrall, D., Thoshkahna, B., & Müller, M. (2015). Virtual Reality Platform for Sonification Evaluation (cit. on pp. 2, 7, 13).

Kuttruff, H. (2016, September 30). *Room Acoustics* (6th ed.). CRC Press. (Cit. on p. 2).

Larsson, P., & Kleiner, M. (2002). Better presence and performance in virtual environments by improved binaural sound rendering (cit. on p. 1).

Lokki, T., Ohn, M., Savioja, L., & Takala, T. (2000). A Case Study of Auditory Navigation in Virtual Acoustic Environments (cit. on pp. 1 sq., 6 sq., 13, 15).

Mehra, R., Rungta, A., Golas, A., Lin, M., & Manocha, D. (2015). WAVE: Interactive Wave-based Sound Propagation for Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics*, *21*(4), 434–442 (cit. on pp. 1 sq., 6 sq.).

Microsoft Corporation. (2022). *Project Acoustics*. https://github.com/microsoft/ProjectAcoustics (cit. on p. 5).

Middlebrooks, J. (2015). Chapter 6. Sound localization. *Handbook of clinical neurology*, *129C*, 99–116 (cit. on p. 7).

Olson, A. (2024, September 24). *Aaronjolson/Blender-Python-Procedural-Level-Generation*. Retrieved October 22, 2024, from https://github.com/aaronjolson/Blender-Python-Procedural-Level-Generation (cit. on p. 18).

Ubisoft. (2015, December 1). *Tom Clancy's Rainbow Six Siege*. (Cit. on p. 5).

Unity Technologies. (2024). *Unity* (Version 2022.3.34f1). https://unity.com/ (cit. on p. 5).

Valve Corporation. (2017, December). *Counter-Strike: Global Offensive* (Version 1.36.1.7). (Cit. on p. 5).

Valve Corporation. (2024, March 27). *Steam Audio* (Version v4.5.3). https://valvesoftware.github.io/steam-audio/ (cit. on p. 2).

Walker, B. N., & Lindsay, J. (2006). Navigation Performance With a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice. *Human Factors*, *48*(2), 265–278 (cit. on p. 6).

Wightman, F. L., & Kistler, D. J. (1999). Resolution of front–back ambiguity in spatial hearing by listener and source movement. *Journal of the Acoustical Society of America*, *105*(5), 2841–2853 (cit. on p. 12).

Appendix A. Source Code & Results

The source code of the unity project and a release of the study game can be found on GitHub: https://github.com/Tim1Blau/audio-study The repository also contains all data collected during the trials.

Appendix B. Participant Information Document

Tim Grün
Julius-Maximilians-Universität Würzburg
tim.gruen@stud-mail.uni-wuerzburg.de

Allgemeine Informationen für Teilnehmende

Titel der Studie: Comparative Study of Sound Propagation in Game Environments for Effective Auditory Navigation and Localization

Herzlich willkommen bei unserer Studie! Wir danken Ihnen für Ihr Interesse.

Wir untersuchen in dieser Studie, wie effektiv Spieler in virtuellen Umgebungen Audioquellen lokalisieren und zu diesen navigieren können. Hierbei werden drei verschiedene Audiokonfigurationen verglichen.

Ablauf der Studie

Die Teilnahme wird insgesamt etwa 15-20 Minuten in Anspruch nehmen.

Vor Beginn des Experiments werden Sie ein paar Fragen zu ihrem Alter und Erfahrung mit Videospielen beantworten. Das Experiment findet in einer virtuellen Umgebung statt, sie werden dazu aufgefordert während des Experiments Stereokopfhörer zu tragen. Das Experiment wird in drei verschiedenen virtuellen Umgebungen mit jeweils unterschiedlichen Audiokonfigurationen durchgeführt. Die Umgebungen bestehen hierbei aus mehreren Räumen und Korridoren. Hierin befindet sich eine Audioquelle an einer zufälligen Position, die kontinuierlich einen Ton abspielt.

In diesen Szenarien gibt es zwei Aufgaben, die jeweils mehrmals durchgeführt werden:

1. Navigation: Finden Sie die Audioquelle, indem Sie den Spielercharakter zur Position der Audioquelle bewegen.
2. Lokalisierung: Schätzen Sie die Position der Audioquelle, ohne den Spielercharakter zu bewegen

In diese Aufgaben werden Sie in einem kurzen Übungsszenario eingeführt.

Während des Experiments wird die Dauer für die Erfüllung der Aufgaben und die jeweilige Effizienz und Genauigkeit erfasst. Hierfür wird außerdem die Bewegung des Spielercharakters aufgezeichnet.

Sollten Sie noch Fragen haben, wenden Sie sich damit bitte an die Versuchsleitung.

Freiwilligkeit und Anonymität

Die Teilnahme an der Studie ist freiwillig. Sie können jederzeit und ohne Angabe von Gründen die Teilnahme an dieser Studie beenden, ohne dass Ihnen daraus Nachteile entstehen. Falls Sie die Studie abbrechen, werden die erhobenen Daten gelöscht.

Die im Rahmen dieser Studie erhobenen, oben beschriebenen Daten und persönlichen Mitteilungen werden vertraulich behandelt. Des Weiteren wird die Veröffentlichung der Ergebnisse der Studie in anonymisierter Form erfolgen, d. h. ohne dass Ihre Daten Ihrer Person zugeordnet werden können.

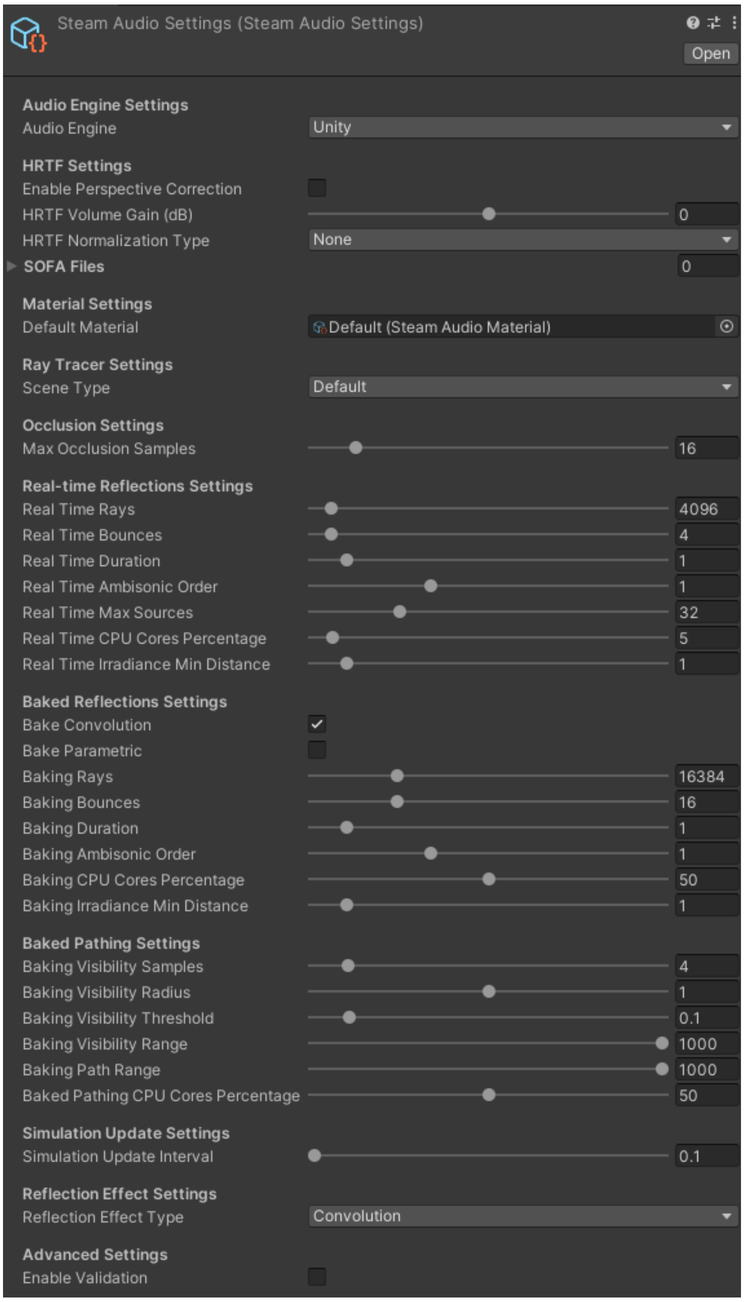
Datenschutz

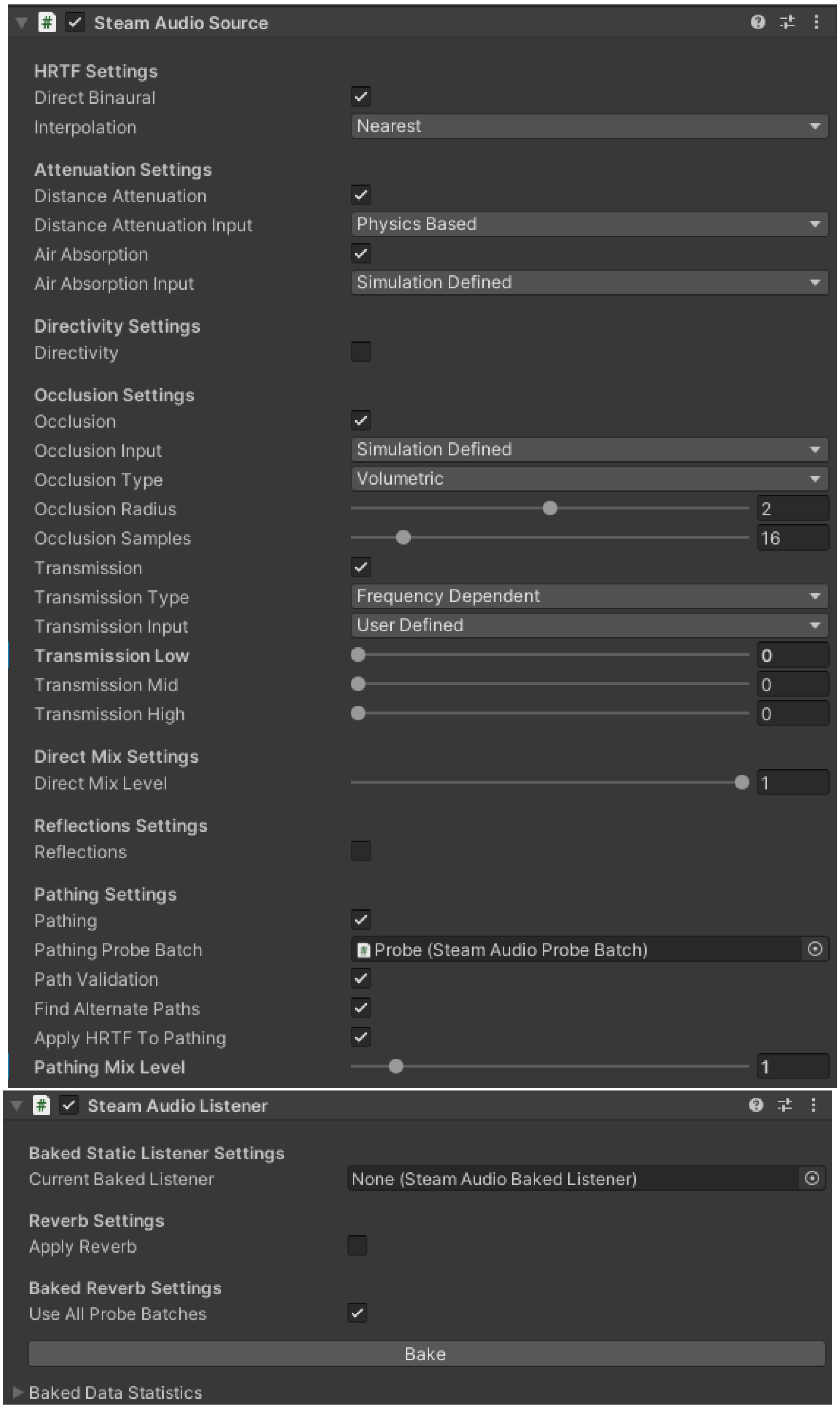
Die Erhebung und Verarbeitung Ihrer oben beschriebenen persönlichen Daten erfolgt anonym. Damit ist es niemandem möglich, die erhobenen Daten mit Ihrem Namen in Verbindung zu bringen. Eine Löschung der Daten ist daher nicht möglich. Die anonymisierten Daten werden mindestens 10 Jahre gespeichert.

Ausschlusskriterien

Um an der Studie teilnehmen zu können wird vorausgesetzt, dass Sie über ein normales, gesundes Gehör verfügen.

Appendix C. Steam Audio Settings





Steam Audio Source
HRTF Settings
Direct Binaural
Interpolation
Nearest
Attenuation Settings
Distance Attenuation
Distance Attenuation Input
Physics Based
Air Absorption
Air Absorption Input
Simulation Defined
Directivity Settings
Directivity
Occlusion Settings
Occlusion
Occlusion Input
Simulation Defined
Occlusion Type
Volumetric
Occlusion Radius
2
Occlusion Samples
16
Transmission
Transmission Type
Frequency Dependent
Transmission Input
User Defined
Transmission Low
0
Transmission Mid
0
Transmission High
0
Direct Mix Settings
Direct Mix Level
1
Reflections Settings
Reflections
Pathing Settings
Pathing
Pathing Probe Batch
Probe (Steam Audio Probe Batch)
Path Validation
Find Alternate Paths
Apply HRTF To Pathing
Pathing Mix Level
1
Steam Audio Listener
Baked Static Listener Settings
Current Baked Listener
None (Steam Audio Baked Listener)
Reverb Settings
Apply Reverb
Baked Reverb Settings
Use All Probe Batches
Bake
Baked Data Statistics

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Würzburg, October 23, 2024

Tim Joshua Grün