Spatial Audio in VR - A Comparison of Ambisonic Sound with Regular Sound

Freek De Sagher
dept. Computer Science
University of Antwerp
Antwerp, Belgium
freek.desagher@student.uantwerpen.be

***Abstract*—Sound is a key aspect in Virtual Reality (VR) immersion. In order to create immersive sound, methods such as the Head Related Transfer Function or Room Impulse Response are used to add the feeling of directionality to sounds. These functions are implemented in a variety of free to use tools and frameworks such as Steam Audio. Furthermore, the format of the sound file has an impact on how immersive the sound of an application feels. Ambisonic sound, often referred to as 3D sound, is a sound format that includes directionality in its format. This paper evaluates differences in user behavior between ambisonic sounds and regular sound formats. To this end, we performed a user experiment using a simple VR application containing both sound formats with an assisting survey. The survey asked for the experience of the participants with respect to the sounds in the VR environment. The results of the experiment and the survey are analyzed in order to give an insight in how users react to the two different sound formats.**

***Index Terms*—VR, immersion, spatial audio, ambisonics, auditory cues**

I. Introduction

Virtual reality (VR) became more popular during the last decade [1], [2]. Using a Head Mounted Display (HMD), users can experience realistic virtual environments. A key aspect to make the experience better is immersion. The more immersed users are in a VR application, the better their experience will be [1].

A few important cues are required to achieve immersion. The first and most important one is the visual cue. The resolution of the HMD display is a key aspect to achieve good visual immersion. Furthermore, the look of the environment and the way everything is combined visually helps to immerse the user.

Another, important but slightly less prominent cue for immersion is sound [3]–[5]. To achieve immersive sound, a few tricks have to be used. First, the way in which the sound is propagated to the ear is a key aspect. This is achieved by the Head Related Transfer Function (HRTF). The HRTF takes the HMD rotation into account and propagates the sound based on the user's head rotation. Besides user's rotation, environment modeling improves sound immersion as well. Only using a HRTF makes sounds sound flat and it does not sound well externalized [6]. A sound that is well externalized, sounds as if it is not playing from right besides your head. Functions such as the room impulse response (RIR) help to upgrade externalization. The RIR takes the environment into account and applies the effect of reflections, reverberation and absorption of the environment to the sounds.

An important difference between the visual and auditory cues is that the former is not always perceived while the latter is constantly perceived: a user can shut their eyes or look away, but cannot stop listening [7]. The most important aspect of the auditory cue is that it can affect the user even when it is looking away from the sound source. Furthermore, sound is a more global cue than the visual cue as it is omni-directional [7].

A handful of tools exist to implement an HRTF and/or RIR. The quality of the sound created by these tools differs based on the functions used, and on the anatomy of the ear of the user [6], [7]. A couple of industry-adopted frameworks are Steam Audio [8], Oculus Spatializer [9] or Microsoft HRTF (MS HRTF) [10].

Besides techniques to adapt the sound for immersive VR applications, the format of the sound might have an impact as well. Three-dimensional sound, also called ambisonic sound, is a special sound format recorded in a specific way [11]: uniquely designed microphones are required that house not one, but four sub-cardioid microphones pointed in different directions [13]. Ambisonic sound can then be represented as a sphere around the listener in 3D. Based on the rotation and location of the HMD in relation to the sound source, the directionality of the sound is delivered partially by the sound format itself and partially by the HRTF.

Not much research is performed in comparing ambisonic sound with regular sound in the context of VR. An overview of the ambisonic sound format specification is given in other research [11], but no direct comparison is made with other sound formats with respect to user responsiveness. The practical difference and advantages of using ambisonic sound over other formats is not clear yet.

That is why this research experimented with ambisonic sounds in comparison to normal MP3 files. We experimented with a couple spatialization methods provided by spatialization frameworks. We assess whether users are affected more by spatialized ambisonic sound in comparison to normal spatialized sounds.

Our research focuses to answer the following Research Questions (RQ).

- *RQ1: Is there a difference in reaction of users to ambisonic sounds in comparison to normal sounds (MP3)?*
- *RQ2: Which sound format is better for spatialization?*
- *RQ3: How noticeable is spatialized sound for VR users?*

The remainder of the paper is structured as follows. First some additional background information is given (Section II). Section III describes the methodology of the experiment. A subsection is given where the performed survey is explained. This section is followed by Section IV, where the experiment results are elucidated. Finally, the results are discussed in Section V. The work is concluded in Section VI with some pointers to future work.

II. Background

Before the methodology and experiment results are clarified, some additional background information is presented in the following sections. Some extra details are given on topics such as the HRTF, RIR, sound types and ambisonics.

A. Head Related Transfer Function

Humans are capable of localizing sound in a 3D space, despite only having two ears [3]. To do that, the brain interprets sound that reaches both ears and compares them in time and difference while taking the reflection and absorption of the environment and the body into account. The HRTF tries to mimic these differences to trick our brain to think that a sound comes from a specific direction. The HRTF is considered a scattering based cue. It is a function that takes the head rotation as input and outputs localization cues that should be applied to the sound.

The HRTF contains a dataset of measurements for certain head rotations. These measurements are used to propagate the sound to the user in such a way that directionality can be perceived. When a user's rotation matches one of the measurements, the characteristics of the measurement can be used to change the way the sound is propagated. If a rotation does not match one of the measurements, an interpolation method is used to determine which characteristics to apply to the sound. Different interpolation methods achieve different performances as mentioned by [6]. Some industry-adopted frameworks allow to specify the interpolation method. For example for Steam Audio can switch between nearest interpolation and bilinear interpolation.

Note that the HRTF behaves differently for the left and right ear. This means that separate measurements should be made for each ear. The dataset used by a HRTF is measured by applying sound stimuli to a listener that is equipped with microphones at the ears. Important to note is that each ear is different: for each human, a different HRTF should be used for the optimal performance. Creating personalized HRTFs could improve the performance for individual users, but is still an open research topic [6]. Three-dimensional audio is adopted in the industry as well. For example the PlayStation 5 uses an HRTF for 3D sound [12].

B. Room Impulse Response

Note that the human brain takes the reflections and absorption of the environment into account when localizing sounds. With an HRTF alone, sounds are not well externalized. When not using some form of reflection and absorption model, sounds appear to be played right besides the listener's head rather than at the sound source.

The Room Impulse Response (RIR) is another scattering based cue. The RIR includes effects due to reflection at the boundaries, sound absorption, diffraction around edges and obstacles, and low frequency resonance effects [6]. For that, the location of both listener and sound source should be taken into account. A model of the environment should be created to calculate the effects of reflections and absorption. Image models can be used to model the environment [14]. Based on these models reflections can be calculated with ease. For absorption, different image model strengths can be multiplied with an absorption constant (β) for each material type [6].

C. Sound Types

In general, sounds can be categorized into four categories based on their function [16]:

- Speech/dialogue: aids the player to follow the game/application's story.
- Music: creates an ambience, feeling or tension. Enhances player experience.
- Voice over: increases engagement with the player.
- Sound effects: give feedback to the player about their actions or about things that happen in the environment

Furthermore, sound can be categorized into two specific types: diegetic sounds and non-diegetic sounds [16], [17].

- Diegetic sounds: originate in the virtual environment. This means that all in-game characters (other entities or Non-Playable Characters (NPCs) would hear those sounds if they were real people. This includes speech/dialogue and sound effects.
- Non-diegetic sounds: sounds that do not originate from the virtual environment. The in-game characters and NPCs would not be able to hear this sound if they were real people. Music and voice-over usually fall into this category.

D. Ambisonics

The ambisonic sound format is capable of providing three-dimensional sound. This is useful in VR applications as it encodes sound direction into its format. Sound directionality on its turn helps to achieve higher player immersion.

The encoding of ambisonic sounds is based on spherical harmonic functions, which are weighted orthogonal functions. A soundfield can then be described using the sum of these functions [11], [19].

The sound wave equation can be encoded in the spherical coordinate system. Directionality of the sound can be encoded

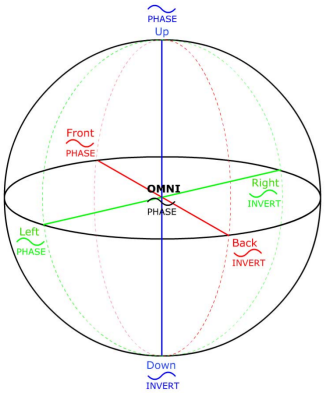

Fig. 1. Graphical representation of the sound origin and the three figure-of-eight microphones for first order ambisonics [15]

Fig. 2. Screenshot of the 3D VR environment in Unity

in four different channels (W, X, Y, Z) of the B-format. The B-format is the basis for first order ambisonics [19]. Channel W contains information about the pressure field at the origin and is recorded by an omnidirectional microphone (see the origin in Figure 1). Channels X, Y and Z contain information about the acoustic velocity at the origin [18]. These channels are recorded using three figure-of-eight microphones at the origin, one for each axis (see the blue, red and green lines in Figure 1). A figure-of-eight microphone is sensitive to sound coming from the front and from the back, but not to sounds coming from the sides. Each microphone measurement is encoded in its own channel.

With all these measurements combined, a sphere can be created around a sound source. A listener can then be placed inside the sphere and by using an ambisonic decoder the directionality of the sound can be extracted from the sound format to create immersive sound. Steam Audio has an ambisonic decoder built-in for VR applications.

III. Methodology Overview

This section explains the methodology of the experiment. The first subsection describes the VR environment, while the second explains the performed survey.

A. Environment

In order to search for an answer for the research questions, a simple VR experiment related to sound was created.

The environment was created in Unity[1] version 2020.3.33f1. Figure 2 shows a screenshot of a part of the environment. For the VR sound library, the Unity package version of the Steam Audio API [8] was used. The Microsoft HRTF [10] and Oculus Spatializer [9] were considered as well, but the Steam Audio API has more options to tweak the behavior, and it has an ambisonic sound decoder built-in. Furthermore, a simple test scene was created to test the sound localization performance of the three toolkits. Based on this short informal experiment, we found that by using the Steam Audio API, it was easier to localize the origin of spatialized sounds, both in a horizontal and vertical direction. For the MS HRTF toolkit and Oculus Spatializer, it was harder to determine.

The HMD used for the experiments was an Meta Quest 2[2] (which was called Oculus Quest 2 before they rebranded), developed by the Meta company. The controls for the player were created using the Oculus Quest Integration package for Unity [9]. For the experiment, a continuous move system was used rather than a teleportation system. This may impact the performance of the participants of the experiment because systems with continuous movement are known to cause motion sickness easier. To reduce influence of motion sickness on the participants, the experiment was kept rather short (between 5 and 10 minutes per participant).

Using features provided by Unity and free low poly assets from various asset packs, a simple 3D VR environment was assembled in which VR users could move freely. Sound effects were placed in the environment.

Since we wanted to measure the responsiveness of players to in-game sound types, all sounds used fall in the category of a diegetic sound effect (see Section II-C). The different sounds placed around the world would start playing based on player proximity. Important to indicate is that there was no visual cue attached to the sound. The experiment tried to look at the impactfulness of the sound alone. For example, when a sound with visual cue would start playing, the player would almost certainly look at the sound source, since the visual cue is right there. This eliminated the possibility that players look at the visual cues alone rather than in the direction of the auditory cues.

One of the experiment's goals was to create a comparison between spatialized ambisonic sounds, normal spatialized sounds and their non-spatialized variant. Table I summarizes the different sound classes that were scattered around in the world. For each of these sound types, two different sounds were used, resulting in a total of 10 sounds separated across the environment. The sounds were placed in a particular way, such that no two sounds would play at the same time.

In order to measure user response to a particular sound, it is checked whether they look in the direction of the sound source. A raycast is constantly being emitted where the player looks

[1]https://unity.com/

[2]https://store.facebook.com/be/quest/products/quest-2

TABLE I
SOUND TYPE ABBREVATIONS

Abbreviation	Sound Type
NUS	Non-ambisonic, Unity spatialized
NSS	Non-ambisonic, Steam spatialized
NNS	Non-ambisonic, non-spatialized
AS	Ambisonic, Steam spatialized
AN	Ambisonic, non-spatialized

Fig. 3. First person view: this is how the players perceive the environment

at. Three different accuracy levels are used: far, medium and close, implemented using a collider around the sound source. This can be seen on Figure 2. The sphere with the longest radius acts as proximity sensor for a player: when a player character enters the sphere, the sound starts playing. The other three spheres are the accuracy range colliders. Figure 3 shows a first person view of how players perceive the environment. If the raycast hits the far-collider, the user looks at approximately 10 meters from the sound source. A hit with the medium collider results in a direction of approximately 5 meters and a hit with the close-collider indicates that the user looks at the sound source with a 1 meter accuracy.

In regular games or applications, it is rarely the case that sound is the main source of interest to the player. In games for example, an objective is created, such as collecting artifacts, while sounds are there to enhance the overall experience, without being required to complete the objective.

A similar approach is used in the experiment. In order to see whether users reacts to a specific sound unknowingly, a separate objective was given to them. Their main objective was to collect 8 coins scattered around the map. Each coin was placed inside the detection radius of a sound. This ensured that the sounds were played each time a player collected a coin. This increased the amount of data that could be collected during one experiment.

In order to gather data, a number of participants (15) was asked to run through the entire experiment. Information on when a sound starts playing, when a player looks at a certain sound and for how long and the accuracy level were constantly monitored. Furthermore, the position and rotation of the HMD was monitored constantly as well. The Unity project containing the experiment is accessible via GitHub[3].

After the experiment, the participants were asked to fill in a small survey to gauge their experience related to the sound in the experiment.

B. Survey

Before the experiment started, the participants were explained the controls of the experiment, and what their goal was in the environment. The conductor of the experiment explained that the goal was to collect 8 coins scattered around the 3D environment. It was mentioned that sound would play at certain locations. Important is that no instructions were given relating to the sounds; only the main objective of collecting the coins was given.

After a participant finished the experiment, a short survey was filled in. It asked for some demographic information such as gender, age, occupation or study and for their experience with VR. Furthermore, some questions related to sound were asked. The participants were asked how well they could position the sound, both horizontally and vertically, how well they could estimate the distance to the sound source and how accurate the nature of the sound was. The participants answered using a Likert scale where 1 means 'totally disagree', and 5 means 'strongly agree' [4].

Finally two general questions were asked. First was whether or not the controls were intuitive and second it was asked whether the user experienced motion sickness or not. The former is important to know whether the user was not distracted figuring out the controls during the experiment. The latter was asked since motion sickness would diminish the performance of the participant as well. Both questions had to be answered using the same Likert scale.

To conclude, an open feedback box was provided.

IV. RESULTS

The following sections partition the result in different categories. First, the demographics of the participants are discussed (Section IV-A), then we present the survey results (Section IV-B). Finally, Section IV-C describes the results of the gaze direction of the players.

A. Participant Demographics

In total, 15 people participated in the experiment. The age of the participants ranged from 16 to 77 years old. Mostly male participants participated (73.3%) (see Figure 4).

In general, the VR experience of most people was limited to only having tried VR once.

[3]https://github.com/FreekDS/SpatialAudio-Experiment
[4]https://forms.gle/J3ug3sNZs254A8oH7

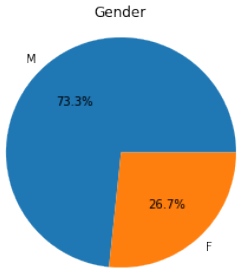

Fig. 4. Gender of participants

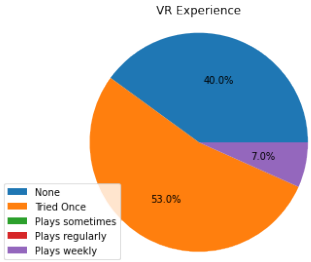

Fig. 5. VR experience of participants

One person is an experienced user and uses VR almost on a weekly basis (see Figure 5). As for occupation, mostly students with a study in IT were asked to perform the experiment.

B. Survey Results

The result of the survey is summarized in the bar plots shown in Figure 6. Generally speaking, most participants found localizing the location of the sound relatively easy. The horizontal direction is especially well localized. Only three out of 15 participants gave the horizontal localization a score of 3 or less. Positioning the sound vertically was feasible for most participants, but it was experienced harder than with the horizontal direction. For the ease of positioning vertically, 8 persons gave a score of 3 or less. Note that there is no distinction made between the different accuracy ranges in the survey.

In fact, 46.67% of the participants (7/15) found positioning the sound horizontally easier than vertically. 13.33% (2/15) found vertical easier than horizontal and 40% (6/15) found it of equal difficulty for both directions.

Most participants found it not trivial to estimate the distance to the sound source: Only 7 people gave a score of 3 for this question. On the other hand, they did not find it infeasible since no one gave it a score less than 3.

The quality of the sound nature was experienced as realistic for most participants. Quality mainly is related to the way the volume of the sound changed as a player moved farther or closer to a sound source. One participant noticed the sudden sound drop when a sound stopped playing and mentioned it in the open feedback: *"There is a certain distance at which the sound drops drastically. This was worse with some sounds."* Another participant mentioned that the sounds were heard, but since the main goal was to collect the coins, they were partially ignored: *"The main goal is to collect the coins, I heard the sounds, but was not bothered to look closer into it."*

TABLE II
THE AMOUNT OF TIMES LOOKED AT A CERTAIN SOUND CLASS IN PERCENTAGE

Type	Times looked at (%)
AS	27.32 %
NUS	26.78 %
AN	20.21 %
NNS	19.67 %
NSS	6.01 %

Even though most participants had absolutely no or little experience with VR, the controls were experienced as intuitive and easy to use. Most participants did not experience motion sickness at all. A few experienced some motion sickness. One person even gave it a score of 4 on the Likert scale.

Overall, the participants felt completely immersed in the scene and found it an enjoyable experience: *"I was completely involved.", "Easy to play, fun, easy controls."*

C. Gaze Direction Data Results

Data is analyzed using a Python Notebook[5]. A couple of metrics were analyzed based on the look direction data. For each participant, a data file is generated that contains information on the look direction of the player combined with the moments at which certain sounds start and stop playing. The start time, end time, sound name, sound type abbreviation and sound position are logged. The abbreviation of the sound types is defined in Table I.

When a sound is played, it is checked whether or not a player looks at the direction of a sound. If that is the case (the raycast hits one of the three accuracy categories) a log entry is created that shows the current time and the accuracy type (close, medium or far). Finally, the time at which a sound stops playing is monitored as well.

Based on these data files generated by all participants, a few things can be analyzed.

1) Percentage per type: First, it can be checked how many times the participants looked at a specific sound type for all accuracy classes combined. Combining these accuracy classes is important as a log entry of medium accuracy is not automatically flagged as far accuracy, and a close accuracy hit is not automatically recorded as a medium accuracy hit. Only the most accurate gaze accuracy is logged.

Table II shows the results for each sound class. Note that the sum of the column results in 100%. In 27.32% of the times where a player looked at the sound, it was at an ambisonic

[5]https://github.com/FreekDS/Analysis-Spatial-Audio-In-VR

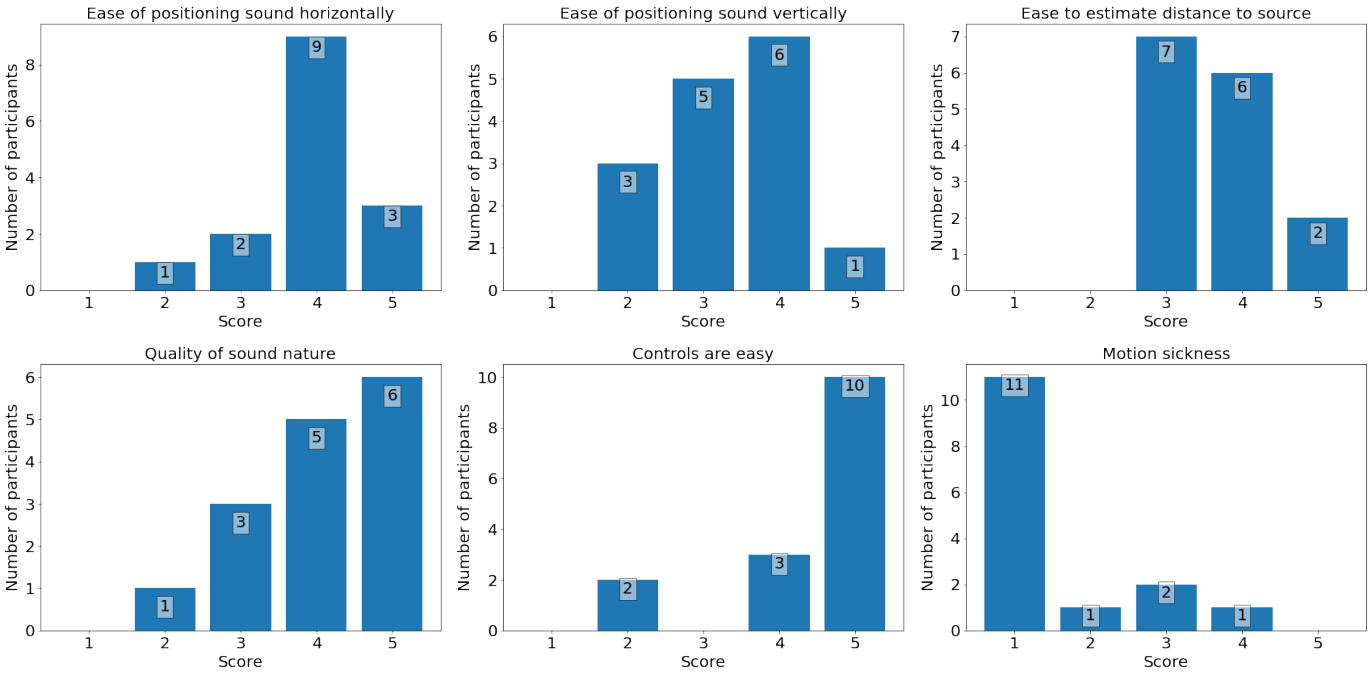

Fig. 6. Survey results summary

TABLE III
THE AMOUNT OF TIMES LOOKED AT A SOUND TYPE PER ACCURACY CLASS IN PERCENTAGE

	Times looked at (%)		
Type	Far	Medium	Close
AS	24.18 %	37.50 %	66.67 %
NUS	27.45 %	25.00 %	16.67 %
AN	18.95 %	33.33 %	0.00 %
NNS	22.22 %	4.16 %	16.67 %
NSS	7.19 %	0.00 %	0.00 %

TABLE IV
PERCENTAGE OF PEOPLE LOOKED AT SOUND TYPE

Type	**Avg amount of participants looked at (%)**
AS	63.33 %
NUS	70.00 %
AN	50.00 %
NNS	73.33 %
NSS	23.33 %

spatialized sound. In comparison to the other sound classes, this class is looked at the most. The second most looked at sound type was the non-ambisonic Unity spatialized class. In general 47.54% of the times there was looked at the sound, it was at an ambisonic sound. It is not a great difference with the other sound formats.

This first analysis of the data looked at all accuracy types combined. If the three different accuracy ranges are separated, and the same analysis is performed, the data shows the results as shown in Table III. Here, the sum of each column is 100% as well.

This data indicates an interesting trend. Although overall, there does not seem to be a difference between ambisonic sounds and non-ambisonic sounds, a different trend pops up if the different accuracy classes are separated from each other. In a whopping 70.83% of the times of which a player looked at a sound source with a medium accuracy, the sound was originating from an ambisonic source. It clearly outperforms the other data classes in this accuracy category. The same trend is continued when looking at close accuracy, even though there is not much data on this accuracy class.

2) Percentage of participants that looked at sound type: Rather than looking at all the times that participants looked at a sound, it is checked how many participants looked at a certain sound type. To do that, the average percent of participants that looked at the sound is taken from the two sounds of each sound type. The results of this metric are displayed in Table IV. Note that in this case, the results of the column are not cumulative and do not add up to 100% due to the nature of the metric. It can be seen from the data that on average, more than 70% of the participants looked at the Unity spatialized audio sources. This is a higher average rate than with the ambisonic variants.

Again, the data can be separated into the accuracy classes (see Table V). In doing so, similar results can be seen: the Unity spatialized audio sources are looked at by more different participants than the ambisonic counterparts, except for the

TABLE V
PERCENTAGE OF PEOPLE LOOKED AT SOUND TYPE PER ACCURACY CLASS

	Avg amount of participants looked at (%)		
Type	Far	Medium	Close
AS	60.00 %	16.67 %	10.00 %
NUS	70.00 %	33.33 %	6.67 %
AN	50.00 %	23.33 %	0.00 %
NNS	73.33 %	6.67 %	6.67 %
NSS	23.33 %	0.00 %	0.00 %

TABLE VI
AVERAGE DURATION LOOKED AT

Type	Avg duration looked at (s)
AS	1.004861
NUS	0.698804
AN	0.755357
NNS	0.742727
NSS	1.272222

close accuracy class. Combining this information with the results of the previous metric provides the following conclusion: Non-ambisonic sounds are less looked at by different people, but when people noticed the particular sound, they were triggered to look at it multiple times.

3) Look duration: Finally, it can be checked how long users looked at a certain sound type. As it turns out, people looked longest to the non-ambisonic Steam spatialized sound class on average, followed by the ambisonic sound classes (see Table VI).

V. DISCUSSION

This section discusses the results presented in Section IV. An interpretation to them is given in V-A and some validity threats are given in Section V-B.

A. Results Interpretation

Based on the results obtained from the experiments, an answer to the research questions posed in Section I can be formulated.

1) RQ1: Is there a difference in reaction of users to ambisonic sounds in comparison to normal sounds (MP3)?: Based on the analyzed data, there is no clear distinction between the behavior of users towards ambisonic sounds or regular sounds.

The only difference observed, is that when users found an ambisonic sound, they looked at it multiple times and at greater accuracy.

2) RQ2: Which sound format is better for spatialization?: When analyzing the results of the experiments, only a feint difference can be observed between the ambisonic sound format and the regular MP3 format.

Most participants give equal notice towards ambisonic sounds and regular sounds. When people look at the ambisonic sounds though, the accuracy of looking at the sound source is significantly higher than with the regular sound type. Based on that observation, the conclusion for this research question is that ambisonic sound is better for spatialization than regular sound.

Do note that there is a significant trade-off in memory consumption, besides the cost of spatialization [20]. Ambisonic sound files are usually way larger than their regular counterpart. Developers of sound based VR applications should determine whether they want to pay the price of more memory consumption for better spatialization performance or not.

3) RQ3: How noticeable is spatialized sound for VR users?: Spatialized sound is very noticeable for VR users. Both users with no VR experience, or users with a lot of VR experience looked at the most sounds with some accuracy. All sounds that were present in the experiment were at least looked at once by the participants. This observation is confirmed in other research [21].

B. Threats to Validity

1) Sample size: The amount of people that participated in the experiment is rather small (15). It is possible that with more people the trends described earlier are slightly different.

2) Participants variation: Most people that participated in the experiment were active, either by study or by work, in the IT sector. People working in IT look at different aspects when participating in a technical experiment compared to others. It is possible that with other people, other trends arise.

3) Accidental log entries: it is possible that participants gazed in the direction of the sounds merely by coincidence. To combat this, sounds were placed smartly such that accidentally looking at them is minimal. For example, sounds are placed above the eye level of the user or on a location relatively far away from the coins.

4) Sound nature: Whether or not a user looks at a certain sound is influenced by the sound itself: if a sound is appealing to a person, it is more likely that it is looked at by that person. Think of horror games for example, when a scary sound plays, players are more likely to look away from the source that might scare them. To combat this, only sounds were used that sounded naturally.

5) Ambisonic sounds: While there exists a huge library of regular sounds, the amount of ambisonic sounds that are freely available is rather sparse. Because of this, some ambisonic sounds might sound more strange in comparison to others.

VI. CONCLUSION

As a conclusion we can say that compared to normal sounds, ambisonic sound exhibits a slightly better spatialization performance. Users were able to pinpoint the source of the sound more accurately with the ambisonic sound format. Furthermore, we discussed the obtained results with respect to memory consumption. Trade-offs should be made between more computation and memory usage and better spatialization performance. In that light, we would suggest developers to use ambisonics for the more important sound cues where

accurate sound localization is required. We also mentioned that special hardware is required to capture ambisonic sound which is an important factor when making the balance of using normal sound in comparison to ambisonic sound.

Based on the initial observations made in this work in regards to the difference of user noticeability between the two sound formats, additional research can be performed.

It can be verified whether the same trends occur with a larger set of people and with a different kind of people. Furthermore, an experiment can be performed where the same sounds are used, but in the two different formats to compare them directly. This would combat the *sound nature* threat discussed earlier in Section V-B.

Furthermore, since people have the tendency to look at spatialized sounds, it is possible to analyze the rotation and location of the HMD in comparison with the sound location. This data is collected using this experiment, but is not used.

REFERENCES

[1] Boas, Y. A. G. V. *"Overview of virtual reality technologies."* Interactive Multimedia Conference. Vol. 2013. 2013

[2] Bozgeyikli, Evren, et al. *"Point & teleport locomotion technique for virtual reality."* Proceedings of the 2016 annual symposium on computer-human interaction in play. 2016.

[3] Broderick, James, Jim Duggan, and Sam Redfern. *"The importance of spatial audio in modern games and virtual environments."* 2018 IEEE Games, Entertainment, Media Conference (GEM). IEEE, 2018.

[4] Ohn, M.G.R., Lokki, T., and Takala, T. (2005). *Comparison of Auditory, Visual, and Audiovisual Navigation in a 3D Space*, 2(4), 564–570.

[5] Walker, B. N., and Lindsay, J. (2006). *"Navigation performance with a virtual auditory display: effects of beacon sound, capture radius, and practice"*. Human Factors, 48(2), 265–278.

[6] Zotkin, Dmitry N., Ramani Duraiswami, and Larry S. Davis. *"Rendering localized spatial audio in a virtual auditory space."* IEEE Transactions on multimedia 6.4 (2004): 553-564.

[7] Serafin, Stefania, et al. *"Sonic interactions in virtual reality: state of the art, current challenges, and future directions."* IEEE computer graphics and applications 38.2 (2018): 31-43..

[8] "Steam Audio", *Valve Software*. Accessed: Jun. 14, 2022. [Online]. Available: https://valvesoftware.github.io/steam-audio/

[9] "Oculus Native Spatializer for Unity", *Oculus*. Accessed: Jun. 14, 2022. [Online]. Available: https://developer.oculus.com/documentation/unity/audio-osp-unity

[10] "spatialaudio-unity", *Microsoft*. Accessed: Jun. 14, 2022. [Online]. Available: https://github.com/microsoft/spatialaudio-unity

[11] Yagunova, Yulia, Mark A. Poletti, and Paul D. Teal. *"Ambisonics and Sonic Simulation in Virtual Reality."* TENCON 2021-2021 IEEE Region 10 Conference (TENCON). IEEE, 2021.

[12] "PS5 3D audio: what is it? How do you get it?", *What Hi-Fi?*. Accessed: Jun. 20, 2022. [Online]. Available: https://www.whathifi.com/features/ps5-3d-audio-what-is-it-how-do-you-get-it

[13] Paul Virostek, "An Introduction to Ambisonics", *Creative Field Recording*. Accessed Jun. 17, 2022. [Online]. Available: https://www.creativefieldrecording.com/2017/03/01/explorers-of-ambisonics-introduction/

[14] J. B. Allen and D. A. Berkeley, *"Image method for efficiently simulating small-room acoustics"*, J. Acoust. Soc. Amer., vol. 65, no. 5, pp. 943–950, 1979.

[15] A. Sorensen, "Ambisonic Sound - the Basic Way", *Arnfinn*. Accessed Jun. 19, 2022. [Online]. Available: https://arnfinn.blog/191024_ambisonics/191024_ambisonics.html

[16] Rogers, Katja, et al. *"Vanishing importance: studying immersive effects of game audio perception on player experiences in virtual reality."* Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018.

[17] Beck, Thomas, and Sylvia Rothe. "*Applying diegetic cues to an interactive virtual reality experience.*" 2021 IEEE Conference on Games (CoG). IEEE, 2021.

[18] Arteaga, Daniel. "Introduction to ambisonics." Escola Superior Politècnica Universitat Pompeu Fabra, Barcelona, Spain (2015): 21.

[19] A. Rana, C. Ozcinar and A. Smolic, "Towards Generating Ambisonics Using Audio-visual Cue for Virtual Reality," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 2012-2016, doi: 10.1109/ICASSP.2019.8683318.

[20] Innami, Satoshi, and Hiroyuki Kasai. *"On-demand soundscape generation using spatial audio mixing."* 2011 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2011.

[21] Jordan, Dennis, et al. *"Spatial audio engineering in a virtual reality environment."* Mensch und Computer 2016-Tagungsband (2016).