INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2024, VOL. 70, NO. 2, PP.373-380
Manuscript received February 27, 2024; revised June, 2024. doi: 10.24425/ijet.2024.149555

# Reverberation divergence in VR applications

Patryk Rolkowski, Piotr Odya, and Bartłomiej Mróz

***Abstract*—This project aimed to investigate the correlation between virtual reality (VR) imagery and ambisonic sound. With the increasing popularity of VR applications, understanding how sound is perceived in virtual environments is crucial for enhancing the immersiveness of the experience.**

**In the experiment, participants were immersed in a virtual environment that replicated a concert hall. Their task was to assess the correspondence between sound scenes (which differed in reverberation times and their characteristics) and the observed invariant visual scene. The research was conducted using paired tests. Participants were asked to identify the sound scene they considered more closely matched the concert hall seen in the VR goggles for each pair. Each sound scene differed in the employed impulse response. All the impulse responses were recorded in real venues such as concert halls, auditoriums, churches, etc. To provide a realistic auditory experience, the sound scenes were processed using third-order ambisonics and decoded using binaural techniques with HRTFs. The virtual concert hall was generated using the Unreal Engine and was the same for all the tests.**

**One of the major conclusions drawn from the conducted research was confirming the role of spatial sound in creating immersive VR experiences. The study demonstrated that appropriately matching spatial sound to the VR visual scene is essential for achieving complete immersion. Additionally, expectations and preferences regarding reverberation characteristics in different types of spaces were discovered. These findings have significant implications for the design of virtual environments, and understanding these aspects can contribute to improving VR technology and creating more immersive and realistic virtual experiences for users.**

***Keywords*—Ambisonics; VR; 360-degree recordings**

## I. INTRODUCTION

VIRTUAL reality (VR) is becoming increasingly popular, and with it comes a growing demand for more immersive and realistic user experiences. Spatial sound is one of the tools to achieve this effect. The combination of VR technology and ambisonic sound allows the user to be seamlessly transported into a virtual world where sounds are perceived from various directions, mirroring the listening experience of the real environment.

However, despite the importance of spatial sound in VR, there is a need for comprehensive research to better understand the intricate relationship between visual scenes and ambisonic sound. This study aims to address this gap by investigating the correlation between VR imagery and ambisonic sound, focusing on how the interaction between vision and hearing influences the overall perception of the virtual environment. By gaining deeper insights into this dynamic interplay, we hope to contribute valuable knowledge that can help design more effective and immersive virtual experiences.

### A. *Virtual Reality*

Virtual reality is an immersive technology that allows users to experience a computer-generated environment as if it were real. The technology is mainly based on the use of special goggles that track the user's head movements and display a stereoscopic view of the virtual environment on screens placed close to the eyes.

The applications of virtual reality are extremely diverse and include many areas. Virtual reality offers a unique experience in video games and entertainment. With the ability to be completely immersed in a virtual world, players can experience a full sense of presence and participate in interactive adventures.

VR is used for training and simulation purposes in many industries. Employees can practice their skills in controlled virtual environments, which allows them to improve their techniques, prevent mistakes, and increase efficiency. Examples include pilot training, military training, or simulating emergency situations [1].

VR can be used to treat various phobias and mental conditions, such as anxiety disorders or post-traumatic stress disorder, by exposing patients to controlled virtual environments [2].

Virtual reality can be used in the field of design and architecture to visualize projects. Architects can create virtual models of buildings and spaces that allow clients and investors to have a thorough understanding of a project before it is realized.

### B. *Ambisonics*

To achieve high-quality spatial sound in VR environment, it was optimal to use ambisonic technology, which allows for the spatial experience of sound in a three-dimensional environment. Unlike traditional surround sound techniques, which are based on discrete channels, ambisonics uses a mathematical approach to represent the sound field. Michael Gerzon pioneered this field by proposing a technique for recording ambisonic sound using four cardioid microphone capsules in a tetrahedral arrangement [3].

Ambisonics is used widely in various applications that require realistic spatial sound. In music production, ambisonics is employed during music recording and mixing, enabling precise positioning of sounds in space and creating immersive recordings. In film and video game production, ambisonics is

This work was supported in part by the Statutory Funds of Electronics, Telecommunications and Informatics Faculty, Gdańsk University of Technology.

All the Authors are with Gdańsk University of Technology, Poland (e-mail: p.rolkowsk@gmail.com, pioodya@pg.edu.pl, bartlomiej.mroz@pg.edu.pl).

used for realistic sound experiences by reproducing sound effects from different directions and to better capture a scene's atmosphere. Virtual reality (VR) and augmented reality (AR) use ambisonics to provide sound with visual elements, resulting in a more immersive virtual experience.

There are different orders of ambisonics, which determine the difficulty of the sound encoding and decoding process. The order of ambisonics represents the level of detail in the mathematical representation of the sound field. Higher orders consist of more ambisonic channels and thus provide more details in the sound scene, but also require more computing power and disk space.

For the experiment described in this manuscript, it was important to localize sound sources in three-dimensional space accurately. With 16 channels in the third order of ambisonics, it is possible to achieve precise localization of sound in space, translating into a more realistic and immersive sound experience for the listener. At the same time, it will be possible to perform processing in a stand-alone VR application with low latency.

Listening to multichannel ambisonic sound is technically impossible without a special multichannel sound system, where speakers will be placed in the sphere around the listener. It is challenging to provide proper listening conditions, so it is necessary to decode the sound into formats with fewer channels to enable a natural and spatial effect. One of the most popular approaches is to use stereo headphones. In such a case, binauralization is necessary.

Spatial sound binauralization is the process of converting multichannel sound into stereo sound. Crucial components of binauralization are the Head-Related Transfer Functions (HRTFs); an HRTF is a transfer function related to the position and shape of the head and ears. HRTF is responsible for shaping the sound that reaches the ears according to the direction it comes from, considering the shape and size of the ear, as well as the interaction of sound with parts of the ear, such as the auricle and ear canal. This solution allows for evaluating that the sound is spatial and comes from a specific direction [4].

The Head-Related Transfer Function describes the changes that occur in sound as it passes through the head and outer ear of a listener. This happens because of differences in the distance between the sound source and the ear, as well as the shape and size of the head and ear, which introduce changes in the phase and amplitude of the sound [5].

### *C. Sound-Visual Correlation in 360-Degree Recordings*

Issues of audio-visual correlations, including the matching of visual and auditory cues, have been studied by many researchers.

In [6], the authors discuss the topic of virtual sound perception using headphones. It concludes that virtual sound reproduction is impaired if the acoustics of the synthesized room do not match those of the real listening environment. This effect has been well studied and is known as the room divergence effect (RDE). RDE is important for the perception of external virtual sound if the listeners are aware of the sound information associated with the room provided by the listening environment. In the case of VR applications, users receive a visual impression of a virtual room but may not be aware of the room's sound information. It is still unknown whether the acoustic correspondence between the synthesized room and only the visual listening environment is relevant to sound externality.

Psychoacoustic experiments using VR technology have found that the perception of external virtual sound depends on listeners' expectations of the acoustics of the virtual environment. Virtual sound scenes can be perceived as divergent despite the visual coherence with the listening environment. Sometimes, correct sound information about a room can lead to a worsening of the experience if the acoustic properties of the room do not match listeners' subjective expectations.

This research concludes that listeners' expectations of the acoustics of the visual environment affect the perception of external virtual sound. Acoustic compatibility between synthesized sound and the visual environment may be important for achieving optimal perception of virtual sound.

Zinah Al-Bayyar and Kivanc Kitapci also studied the relationship between sense of place and room acoustics indicators, among others [7]. A sense of place is achieved when all the senses are integrated, which creates certain expectations and behaviors. They have focused on the direct effects of two selected acoustic parameters: reverberation time and sound intensity. The main approach was to assess participants' subjective evaluations of the acoustic sense of place using questionnaires. They differed in reverberation time and sound pressure level. Correlations between the results of subjective assessments of the sense of place and data on room acoustics make it possible to determine the optimal conditions that are suitable for creating a sense of place depending on the nature of the environment. Due to the perception of space, different reverberation times were required to map the different sound characteristics of places, thereby providing spatial information. This solution is required to correlate with the participants' emotions, behaviors, sense of security, orientation, and cognitive experience in the context of place attachment. Similarly, expectations were related to different sound intensity conditions as they affect the creation of comfortable, satisfying spaces where cognitive experiences and behaviors are consistent with expectations. Conclusions from these studies provide information on optimal sound conditions that contribute to creating a positive sense of place. These findings may have important implications for the design and development of spaces, especially in the context of room acoustics and the impact on user experience.

In [8], the authors present the results of perceptual experiments related to the quality parameters of virtual acoustic environments. The study focused on the effects of divergence between the synthesized soundstage and the room where the listener is located and the adjustment for congruence or divergence between these rooms. Two experiments were conducted to investigate these effects. In the first experiment, the impact of room divergence on spatial sound perception was analyzed. The results indicated that when there was a discrepancy between the room in which the listener was located and the synthesized sound with the superimposed impulse response of another room, the sound externality was perceived as less distinct. In contrast, congruence between the two rooms led to an increase in localizability. This effect was most noticeable for sounds coming from front and rear directions. The second experiment analyzed the impact of adaptation on room divergence. Test subjects were trained in either congruent or divergent room conditions. The results suggested that training affected participants' perceptions, with the trained group showing a change in location expectations depending on the type of training. The study highlighted the negative impact of room divergence on localization capabilities. In addition, it was found that room adaptations have a role in determining the perception of sound. The results suggested that parameters such as room divergence and visual perception depended on each other. In conclusion, the research showed that room divergence and adaptation are important factors affecting the perception of sound externality in virtual acoustic environments. The study's findings contribute to improving synthesis systems, significantly improving sound spatial perception in virtual reality scenarios.

The aforementioned articles indicate expectations and preferences for sound characteristics in different types of rooms. For example, a long reverberation time is expected in a large concert hall, while in a conference room, the reverberation should be minimal. For instance, the recommended reverberation time for classical music in a concert hall is about 2-2.5 seconds, as described in [9].

## II. Objectives

Understanding how sound is perceived in virtual environments is critical to enhancing the immersive experience. This study attempted to investigate how the interaction between vision and hearing affects the perception of virtual environments by examining the relationship between visual and sound scenes.

One of the main objectives of the research was to confirm the role of spatial sound in creating immersive VR experiences. The expectation was that proper matching of spatial sound to the VR visual scene was necessary for complete immersion. Another goal was to prove that there are expectations and preferences for reverberation characteristics in different spaces. These findings have significant implications for the design of virtual environments, and understanding these aspects can help improve VR technology and create more immersive and realistic virtual experiences for users.

## III. Material and Methods

The experiment involved immersing users in a virtual concert hall and exposing them to different types of sound scenes while the visual scene remained constant in every trial. Each sound scene differed in the impulse response used, recorded in real locations such as concert halls, auditoriums, churches, etc. This variation allowed the exploration of the impact of different acoustic conditions on the perception of virtual sound regarding fixed vision conditions. To ensure a realistic listening experience, the sound scenes were processed using third-order ambisonics and decoded using binaural techniques with head-related transfer functions (HRTFs).

The research aimed to assess how variations in acoustic stimuli affect participants' preferences and expectations of the virtual environment, showing the relationship between visual and auditory stimuli in virtual reality.

Participants were asked to compare different sound scenes while viewing the same visual scene during pairwise comparison tests. They were requested to determine which sound scene was more appropriate for the room they were virtually in.

A pairwise comparisons test was used to simplify the participant's task to only the two situations presented and thus avoid the unwanted bias associated with scale selection. From the psychological point of view, pairwise comparisons can be described as the two-alternative forced choice (2AFC) method. It is a technique used in psychology and perceptual research to assess the ability to discriminate between two alternative signals or stimuli. It is also commonly used in sensory research and psychological testing [10]. The 2AFC method forces a choice between the two options; unfortunately, this involves situations where the participant's answer could be random. On the other hand, the pairwise comparisons method allows us to obtain listeners' subjective preferences and assess whether differences in the acoustic parameters of a virtual room affect the perception of external sound. Thus, we should be able to understand better what acoustic factors are important for creating sound spatialization in virtual environments and how to adjust acoustic parameters to improve the perception of virtual sound.

For the project, the Oculus Quest 2 standalone VR system was used. This type of VR headset is a standalone device that includes both a computing platform and display, as well as tracking sensors and controllers – it does not require connection to a computer or other external devices.

### *A. Participants*

In total, 25 volunteers participated in the experiment (18 males and 7 females). The age of the participants ranged from 18 to 55, which ensured a representative diversity in terms of age.

Before the study began, the participants were asked to complete a survey about their experience with audio, video, VR, and other relevant variables, along with self-rating their vision and hearing acuity. They were asked to indicate their familiarity with technologies using a scale from 1 to 10, where 1 means no familiarity and 10 means a high degree of familiarity.

The survey outcomes, including participants' self-reported familiarity with the technologies and their sensory acuity, are presented in Fig. 1.

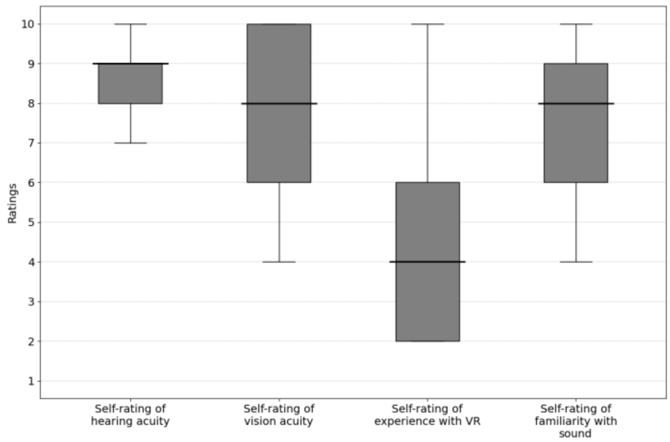


Fig. 1. Graphical representation of participants' evaluations.

Survey answers show varying degrees of experience with sound. Three participants described their abilities at level 4, suggesting a little experience. On the other hand, at levels 6 to 10, similar responses were registered, indicating more even levels of experience in this field. Hence, it can be concluded that the survey participants have varying degrees of sophistication/experience with sound, but it is at a satisfactorily high level.

It is worth noting that survey answers vary in the category of experience with virtual reality technology. Many surveyed described their experience as level 2, 3, or 4, suggesting limited familiarity with VR technology. In contrast, only a few individuals indicated a level 6, 7, or 8, indicating more advanced experience in virtual reality technology.

Regarding self-assessment of hearing abilities, all participants rated their hearing acuity as satisfactory, with their ratings at level 6 or above on the rating scale. Most respondents assessed their hearing at level 9, indicating nearly excellent hearing abilities.

Regarding visual acuity, many participants expressed satisfaction with their visual abilities, with their ratings at level 4 or higher on the rating scale. An intriguing aspect is that the highest number of respondents rated their visual abilities at level 10, suggesting the existence of practically perfect vision among many participants.

### *B. Virtual concert hall*

One of the key requirements was to create a realistic and convincing virtual concert hall. Unreal Engine, a game engine from Epic Games, was used. Among many features, it offered rendering optimizations for the Android platform, allowing it to run on the Oculus Quest 2 VR goggles. Real-looking objects and elements that characterize traditional concert halls were used to prepare the virtual concert hall. Designing such a virtual environment requires considering many details, such as the objects of the instruments, the musicians in appropriate costumes, the materials of the floor and walls, as well as the realistic shape. Accurate replicas of real instruments such as violins, violas, cellos, double basses, flutes, oboes, trumpets, horns, and others were made. Each instrument has been modeled with high attention to detail, reflecting its real appearance and features. Virtual musicians, who perform a concert in a virtual hall, have been depicted in appropriate costumes, typical of symphonic concerts. Details, such as elegant clothes and other elements, were included to complete the realism and authenticity of the concert stage.

The auditorium and balconies, characteristic elements of concert halls, were also included. The process of designing the auditorium involved meticulous considerations of multiple elements, such as the rows of chairs and the layout of the auditorium. The virtual concert hall was designed according to the norms and standards of typical concert halls. The hall's shape, dimensions, proportions, and acoustics have been thoroughly engineered to provide an authentic concert experience. Fig. 2 and Fig. 3 show visualizations of the designed concert hall.

### *C. Audio Material*

A set of recordings of a symphony orchestra performing Anton Bruckner's composition “Symphony No. 8,” the second movement (bars 1-61) [11], was used in the experiment. The decision to use existing recordings of a symphony orchestra was motivated by the need to obtain material of the highest sound quality. Conducting a complete recording of a symphony orchestra, especially when performing a full composition, would have required significant time and resources. In addition, an existing recording of Anton Bruckner's “Symphony No. 8” was available for academic purposes, making it available for familiarization and allowing for reproducibility of the experiment. Each orchestral voice in the set was recorded separately in a specially prepared anechoic room. The musicians played their parts while watching the conductor on a monitor and listening to the pianist performing the entire score. This allowed the musicians to adjust their playing style and tempo, and synchronization between different musicians was possible. The recording level settings were the same for all instruments. This means that in some passages, the signal-to-noise ratio is not ideal, as the gains were adjusted for the loudest instruments. All tracks contain “natural” dynamics played by the musicians. They provide excellent research material for a detailed analysis of the sound quality of an ambisonic system and a thorough examination of the effects of various sound processing techniques on its perception.

Fig. 2. View of the concert stage.

Fig. 3. View of the audience from the artist's perspective.

A Steam Audio plug-in was used to produce spatial sound, which allows the use of 3rd-order ambisonics and binaural decoding using HRTFs. The plugin provides full control over sound generation and processing, including the accurate creation of directionality and width of the sources, to obtain a realistic source beam. Each instrument was assigned an individual beam shape. It was based mainly on the cardioid shape as well as super/hyper-cardioid, for example, for the trumpet.

The binauralization plugin is based on the Finite Impulse Response (FIR) filter technique with phase shift. The binaural interpolation “Bilinear” was selected. It uses the HRTFs generated after interpolation from the four directions closest to the direction of the source for which HRTF data is available. Steam Audio uses HRTFs in connection with head-tracking technology (via headset sensors) to enable accurate localization of the sound source in virtual space. This allows the user to hear sounds that come from different directions and distances, enhancing the realism and immersiveness of the virtual world.

To achieve a natural sound, it was important to include sound reflections from various surfaces inside the room. This effect was achieved by using reverberation simulation techniques by convolution of the signal (preferably recorded in anechoic conditions) with the impulse response of the room. As a result, a sense of spaciousness is achieved through natural reverberation, which increases realism [12].

Impulse responses of five different rooms were used in the experiment:

A. Promenadikeskus concert hall in Pori, Finland, Helsinki University of Technology. T30 reverberation time ranges from 2 to 2.5 s [13].
B. 1st Baptist Church, Nashville. T30 reverberation time ranges from 2 to 5 s [14].
C. Auditorium Arthur Sykes Ryme, University of York. T30 reverberation time ranges from 0.5 to 1 s [14].
D. Elveden Hall, Elveden Estate, Elveden. T30 reverberation time ranges from 4 to 20 s [14].
E. Opera Hall Usina del Arte, Buenos Aires. T30 reverberation time ranges from 1.8 to 2.5 s [14].

Fig. 4 and 5 show the characteristics of reverberation time and Early Decay Time (EDT) of the selected rooms.

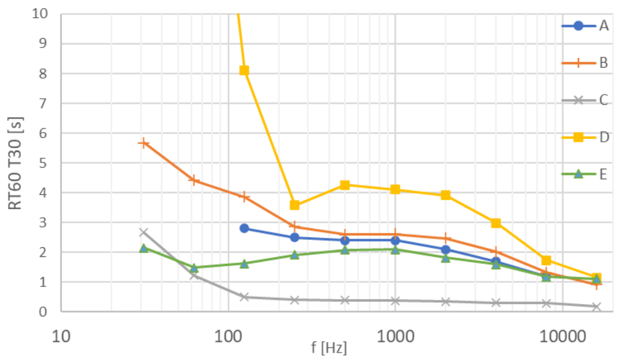


Fig. 4. Frequency characteristics of reverberation time (RT60 T30) of the selected rooms.

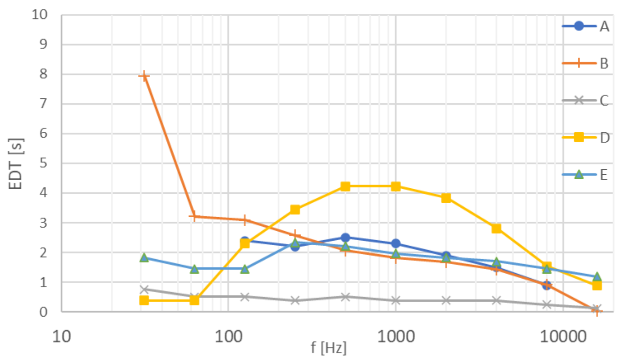


Fig. 5. Frequency characteristics of EDT of the selected rooms.

## D. Hardware

As mentioned before, the base platform for testing was the Oculus Quest 2 VR goggles, which provide high image quality and decent performance. As a result, the user encounters a very realistic virtual reality experience, which is particularly important in the project. The test setup also included Audio-Technica ATH-M50x headphones. The headphones have a wide frequency response range of 15 Hz to 28 kHz and can reproduce the full range of bandwidth (Fig. 6).

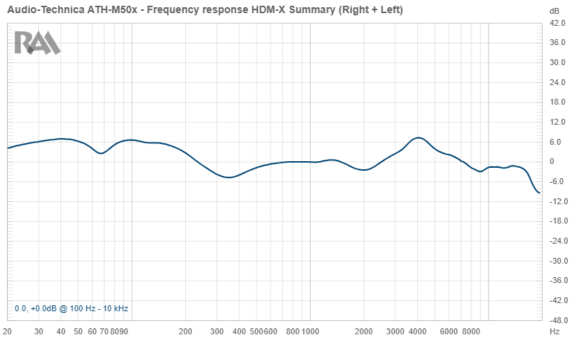


Fig. 6. Frequency response of Audio-Technica ATH-M50x headphones [15].

## E. Test scenarios

The study was conducted using 2AFC (Two Alternative Forced Choice) pairwise comparison tests, in which participants compared different sound scenes against each other while viewing the same visual scene. Their task was to assess which version of the sound scene more closely matched what they saw.

Five versions of the test application were prepared. Each version has the same visual layer. The user appears in the same place (in the VR scene). The versions of the application differ in the superimposed impulse response of real rooms such as concert halls, auditoriums, churches, etc. (Section III.C).

With 5 different variants available, 10 different sets were created, where each set contains ten pairs of sound scenes. The test sets differ in the combination of sound scene pairs, where each set is unique in terms of sequence. Due to technical difficulties associated with the presentation of the individual test pairs (including the need to load successive versions of the application each time), as well as time constraints on the test itself, we decided not to change the order of the sounds in the pair.

Each sound scene differs in the applied impulse response of the actual rooms. The test subject must indicate which version of the sound scene they would expect while in the projected concert hall. Each variant was played for about a minute to ensure that the test subject could listen accurately. Consequently, each task – a pair of scenes – takes about 2 minutes. It was assumed that each person would spend 25 to 30 minutes in the goggles, including the time for switching applications and collecting results. In addition, a survey on overall impressions to get feedback on the implemented solution was conducted at the end.

Example test set:

1. C vs. E
2. A vs. B
3. C vs. D
4. A vs. C
5. D vs. E
6. A vs. D
7. B vs. D
8. B vs. C
9. B vs. E
10. A vs. E

where the letters represent individual rooms – as in Section III.C.

A participant was positioned in the middle of a virtual listening hall and could move left and right and rotate around his/her axis. However, they could not move freely around the room. This was due to the inability to interpolate impulse responses for different points.

## IV. Results

### A. 2AFC analysis

For the 2AFC analysis, it was necessary to set the expected answer in each pair. This was determined based on the analysis of theoretical impulse response parameters and the subjective responses of the participants. In addition, for pairs where the choices included zero reverberation time and excessive reverberation time around 4 seconds, the participants leaned toward choosing a longer reverberation time. Participants' preferences were directed toward the version that they felt was more in line with their subjective expectations of the room's characteristics.

The 2AFC score, or Two-Alternatives Forced Choice score, which is also the ratio of selected sound spaces that were expected in each pair, is expressed by the following formula [10]:

$$2AFC\ SCORE(\%) = \frac{number\ of\ correct\ choices}{total\ numer\ of\ trials} \times 100 \qquad (1)$$

where:
*number of correct choices* – the number of trials in which the participant made the choice that was expected based on the analysis of theoretical impulse response parameters, i.e. chose the expected or reference option.
*total number of trials* – the sum of all trials carried out in the experiment.

Table I shows the results of the 2AFC analysis, along with a summary of each soundstage and the correct (expected) answers.

Table I
Results of the 2AFC test

| Pair | The sum of the selection of a given room | The sum of the selection of a given room | Correct answer | 2AFC score for each pair |
|---|---|---|---|---|
| C vs. E | E – 23 | C – 2 | E | 92% |
| A vs. B | A – 21 | B – 4 | A | 84% |
| C vs. D | C – 12 | D – 13 | D | 52% |
| A vs. C | A – 24 | C – 1 | A | 96% |
| D vs. E | D – 9 | E – 16 | E | 64% |
| A vs. D | A – 23 | D – 2 | A | 92% |
| B vs. D | B – 23 | D – 2 | B | 92% |
| B vs. C | B – 21 | C – 4 | B | 84% |
| B vs. E | B – 17 | E – 8 | E | 32% |
| A vs. E | A – 23 | E – 2 | A | 92% |

The following results are also shown in Fig. 7, which illustrates the percentage of correct answers in the pairwise comparison test. This way of presenting the data shows which test pairs were particularly troublesome for the participants. It clearly shows the combinations for which the participants had difficulties making a choice. In turn, the percentage chart also depicts the pairs for which the participants were most likely to choose the correct answer. This analysis provides a better understanding of which decisions were more intuitive for participants, and where they encountered more difficulty in choosing an answer. Also, some trends in participants' preferences can be observed. The strongest tendency to select expected answers was observed for those pairs where one of the options had a 2 seconds reverberation time, and the other was completely acoustically damped (reverberation time of 0.4 seconds).

However, due to the unbalanced nature of the test, the results may have been affected by order bias. Participants might be influenced by the order in which sound spaces were presented, leading to a systematic preference for the first or second item regardless of its content. This is the primary drawback of our study.

The results show that participants chose answer A most frequently, which represents the reverberation of the venue most similar to the virtually designed one. It is noteworthy that participants demonstrated the ability to accurately indicate this answer, as evidenced by an efficiency score of 90%. This high efficiency in indicating the expected reverberation time significantly exceeds the average, which proves that the participants were able to accurately define their expectations.

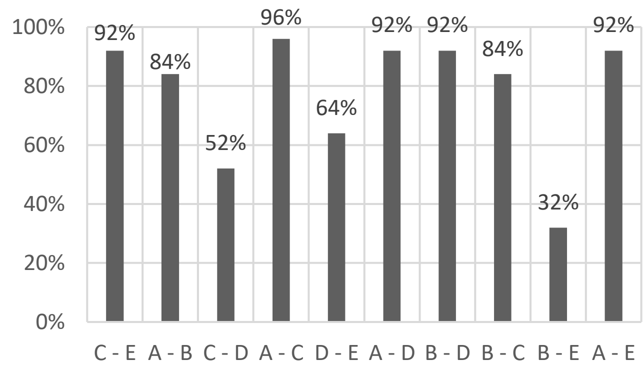


Fig. 7. Percentage of correct answers in the 2AFC test pair.

It is also worth noting that in most of the combinations, participants easily indicated option B, representing a hall with a slightly extended reverberation time (from 2 s to 5 s, depending on the frequency band), namely the 1st Baptist Church hall. This result suggests that the acoustic parameters of this hall also fit in with the participants' preferences for the designed virtual hall. The slightly longer reverberation time in option B can be interpreted in the context of the virtual hall's lack of an audience. The participants may have subjectively sensed that the absence of people in the room would affect the reverberation time and thus preferred a slightly longer time that was more in line with the conditions of no sound absorption by the absence of an audience. This shows that participants based their choices strictly on the relationship between vision and hearing.

An interesting point is the lack of consistency in the choices for the room labeled (E), which is represented by the professional opera hall Usina del Arte. This may suggest that other acoustical parameters, in addition to reverberation time, may have a significant impact on listeners' preferences. For example, the Initial Time Delay Gap, or the time between the arrival of the direct sound and the first reflection. This is a parameter largely responsible for the clarity of the sound. A high-level reflected sound arriving in a small time gap from the direct sound can adversely affect perception by creating an unpleasant echo. This area is a field for further research that may shed light on how different impulse response parameters affect sound perception and listeners' preferences.

However, it is worth noting that the results show the existence of situations in which participants' answers were much more varied. An example of such a situation is a pair of D vs. C soundstages. In this case, the first option was represented by Elveden Hall, where the reverberation time, depending on the frequency band, is up to 20 seconds. In contrast, the second option represented an auditorium without reverberation. This variation in choices may be due to a lack of clear preference among participants. Some participants may prefer the absence of reverberation or sound with a shorter duration, as opposed to the sound with a very long reverberation time that characterizes Elveden Hall. Differences in preferences may result from differing experiences and expectations of sound. Some participants may prefer a clearer and more focused sound, while others may be interested in more elaborate reverberant effects.

This variation in choices may be due to a lack of clear preferences among participants, as some participants prefer the absence of reverberation to its excessive duration. These differences may result from differing experiences and individual listening preferences among participants.

The low score of the 2AFC test can also be seen for pairs D vs. E and B vs. E, especially for the latter, where the score is around 30%. This may suggest that other acoustic parameters, besides reverberation time, may have a significant impact on listeners' preferences, or the impulse response was measured incorrectly so that the parameters do not represent the real state. It is also possible that, in these cases, additional factors affecting sound perception were not included in the study, which may impact the subjective preferences of the participants. Therefore, further research is recommended to understand further why there is such variation in choices in some cases and what additional factors may influence these results.

The analysis shows the diversity of listeners' preferences depending on the characteristics of each sound scene. The results of the study showed that there is no clear preference for all pairs of scenes, which may be due to individual expectations and participants' experiences in the context of sound in virtual spaces. The differences seem to be related not only to the reverberation time but also to other acoustic factors that affect the subjective perception of sound. However, some dominant tendencies in the choices can be discerned, such as a preference for sound scenes with more spatial reverberation times, indicating a general influence of acoustic parameters on participants' preferences.

The results showed that the participants chose soundstages that they thought more closely matched their expectations for room characteristics, especially related to reverberation time. The potential influence of other parameters, such as the latency of first reflection arrival, on sound perception and listener preference, was also indicated. The 2AFC's aggregate score, calculated from the percentage of expected answers, was 78%.

## B. *Survey results*

The survey results regarding participants' experiences with virtual reality technology are shown in Fig. 8. The plot provides participants' subjective assessments of image and sound quality, comfort regarding the VR platform, and willingness to attend concerts performed in virtual reality. The box plot serves as a concise overview, offering insight into responses and preferences expressed by the participants.

One of the other important aspects of the study was to assess the visual quality of the presented image and the comfort of using the virtual reality goggles. To this end, participants were asked to make a subjective evaluation of the image quality of the tested application based on their experiences.

Participants rated the degree of discomfort on a scale from 1 to 10, where 10 means “no discomfort felt” and 1 means “very high degree of discomfort.” Most participants expressed that comfort reached 9/10, indicating no discomfort. Nevertheless, individuals marked some degree of discomfort. In some cases, this may be due to problems associated with wearing glasses while using VR goggles, which may have affected overall satisfaction with the experience.

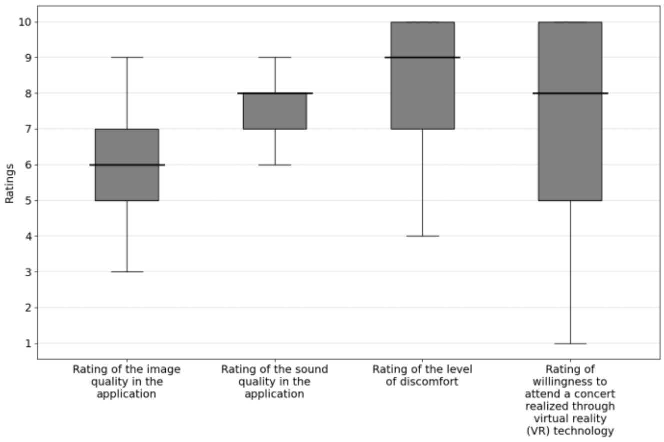


Fig. 8. Participants' subjective evaluations of VR experience.

Participants also rated image quality in terms of sharpness, color, and overall clarity. The results of this evaluation were included on a scale of 1 to 10, where 1 meant “inferior image quality” and 10 meant “very high image quality.” In their evaluations, the surveyed participants most often described image quality as 6/7, which may suggest some minor shortcomings. However, this level of ratings indicates an overall satisfactory visual quality that provided participants with an adequate representation of the virtual environment.

In the analysis of the participants' evaluation of sound quality, they indicated a level of 7, 8, and 9, suggesting a very satisfactory degree of satisfaction with sound quality. However, three people rated the sound quality at level 6. This deviation from the general trend may be important in the context of assessing the global sound quality presented in the app. It is worth investigating the reasons for these lower ratings further to understand whether there are specific aspects of sound that need improvement or adjustment to achieve even higher levels of satisfaction and quality of sound experience.

As part of the analysis of the image quality of the VR app studied, participants were asked to answer the question "What are your overall feelings about using VR?". Participants' responses vary. One person noted that VR technology still has a lot to offer and lacks its full potential. Another participant expressed the difficulty of using VR comfortably due to wearing glasses. Someone else was enthusiastic about the possibility of attending concerts through this technology. A person new to VR indicated that they was fascinated by the technology and would be happy to explore it. Among the positive feedback, there were also comments about imperfections, such as flickering acoustic panels and piercing light. Overall, the feedback suggests that while VR has potential, it still needs some improvement.

Another important issue of the survey was to ask participants about their willingness to attend a concert realized through virtual reality (VR) technology. Participants were asked to rate their interest in attending such a musical event to experience a virtual version of a concert. The results of this evaluation were included on a scale of 1 to 10, where 1 meant "Definitely not, I am not interested in concerts using VR technology." and 10 meant "I would be happy to participate in a concert realized using VR technology."

The results obtained suggest that some participants are strongly interested in attending virtual concerts using VR technology. This indicates the future potential of this technology and the existence of areas of significant interest for potential customers.

Participants were also asked what improvements should be introduced to the tested VR application. Some suggested that the movement of the artists on stage, for example, while playing instruments, would increase the impression of attending a concert. Another participant noted the importance of evaluating the room's acoustics, suggesting that more space or better reverberation would affect a more natural sound. In addition, another suggestion included adding an audience in the auditorium or making it clear whether to include the presence of an audience to emphasize the realism of the experience.

## V. Conclusions

One of the main conclusions of the study was to confirm the role of spatial sound in creating immersive VR experiences. The study showed that the proper matching of spatial sound to the visual scene is an important factor in achieving full immersion. Additionally, expectations and preferences for reverberation characteristics in different types of spaces were discovered. These findings have significant implications for the design of virtual environments, and understanding these aspects can help improve VR technology and create more immersive and realistic virtual experiences for users.

The results showed that the participants' preferences focused on sound scenes with a reference reverberation time of about 2.5 seconds. The potential influence of other parameters, such as the delay in the arrival of the first reflection, on sound perception and listeners' preferences was also indicated. Participants chose soundstages that they thought more closely matched their expectations of room characteristics. The 2AFC's aggregate score, calculated from the percentages of correct answers, was 78%. This confirms the thesis that there are expectations and preferences for listening experiences in different types of rooms.

The results suggest that VR application developers should attach great importance to adjusting sound to the designed environment. The research shows that user expectations are strictly defined and should be met to ensure the highest level of immersion. This is confirmed by the occurrence of the room divergence effect when the simulated room acoustics do not match the acoustics of the real environment.

When considering future research, there are many areas worth exploring in terms of the relationship between visual and auditory perception and virtual reality experience. It is worth investigating which factors influence interest in VR concerts and how to personalize concert offerings to meet the needs of different audiences. Also, acoustic parameters other than reverberation time, such as first-reflection delay, should be studied more thoroughly.

The study's results, despite some drawbacks, provide valuable guidance for the design of virtual spaces, allowing for better matching of audio-visual parameters with user expectations. This analysis is an valuable contribution to the development of virtual reality technology as it provides a better understanding of the critical aspects affecting the quality of immersion in a virtual environment.

## References


[1] L. Freina, and M. Ott, "A literature review on immersive virtual reality in education: state of the art and perspectives," in Proc. of The International Scientific Conference eLearning and Software for Education, Bucharest, pp. 133–141, 2015. https://doi.org/10.12753/2066-026X-15-020

[2] P. J. Costello, "Health and Safety Issues associated with Virtual Reality- A Review of Current Literature," pp. 1–23, 1997.

[3] M. A. Gerzon, „Periphony: With-height sound reproduction," Journal of the Audio Engineering Society, vol. 21, pp. 2–10, 1973.

[4] C. I. Cheng, and G. H. Wakefield, "Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space," Journal of the Audio Engineering Society, vol. 49, pp. 231–249, 2001.

[5] K. Iida, "Head-Related Transfer Function and Acoustic Virtual Reality," Springer, Singapore, 2019. https://doi.org/10.1007/978-981-13-9745-5

[6] S. Li, R. Schlieper, A. Tobbala and J. Peissig, "The Influence of Binaural Room Impulse Responses on Externalization in Virtual Reality Scenarios," Applied Sciences, vol. 11, 2021. https://doi.org/10.3390/app112110198

[7] K. Kitapci, and Z. Al-Bayyar, "A Room Acoustics Evaluation Framework Based on the Sense of Place Constructs," Forum Acusticum, pp. 1753–1761, 2021. https://dx.doi.org/10.48465/fa.2020.0788

[8] S. Werner, F. Klein, T. Mayenfels, and K. Brandenburg, "A summary on acoustic room divergence and its effect on externalization of auditory events," In Proc. of the 2016 Eighth International Conference on Quality of Multimedia Experience, Lisbon, pp. 1–6, 2016. https://doi.org/10.1109/QoMEX.2016.7498973

[9] Y. A. Zhang, "A method to predict reverberation time in concert hall preliminary design stage," A Dissertation, Georgia Institute of Technology, 2005.

[10] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen, "The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks," Psychological review, vol. 113, pp. 700–765, 2006. https://doi.org/10.1037/0033-295X.113.4.700

[11] J. Pätynen, V. Pulkki, and T. Lokki, "Anechoic recording system for symphony orchestra," Acta Acustica united with Acustica, vol. 94, pp. 856–865, 2008. https://doi.org/10.3813/AAA.918104

[12] Convolution reverb definition, https://science-of-sound.net/2016/09/does-convolution-reverb-work-part-1/

[13] J. Merimaa, T. Peltonen, and T. Lokki, "Concert Hall Impulse Responses – Pori, Finland: Reference," pp. 1–19, 2005.

[14] D. Murphy, and S. Shelley, "Openair—The open acoustic impulse response library," https://www.openairlib.net/

[15] Report of Audio-Technica ATH-M50x headphones, http://reference-audio-analyzer.pro/en/param6.php?&idhp0=529&idmain=529&ls=1&titplace=1&tit_head=1&tit_graph=6&tit_channel=2&tit_stand=2&addlistgroup=-2&cl0=%23094a78&stand0=HDM-X