A Qualitative Investigation of Binaural Spatial Music in Virtual Reality

Sandra Mahlamäki
University of the Arts Helsinki
sandra.mahlamaki@uniarts.fi

ABSTRACT

Virtual reality (VR) games and applications strive to create as believable an illusion of "being there" in the virtual environment as possible. This implies fidelity in all aspects of technical reproduction, including the spatial qualities of audio content. "Spatial music" is a phenomenon that has incited growing interest among researchers, sound engineers and composers. Even though VR offers a perfect environment for spatial music experiences and experiments, most music soundtracks are typically presented to the player without situational spatial positioning or other game-scene-related spatial cues. This article studies how people experience spatial sound and music in a multisensory VR context, using multidisciplinary technical and artistic methodologies. It examines a VR experience with two music conditions: a normal stereo music soundtrack, and a spatially auralised music soundtrack. Qualitative interviews were conducted to gather information about the subjective experiences of music spatialisation in the VR experience, and the data gathered was evaluated using qualitative content analysis. The data reveals advantages and disadvantages of music spatialisation in VR, as well as the potential to increase immersion and to affect players' perceptions of virtual spaces. The findings indicate that spatial music could enhance the illusion of "being there" in the virtual environment.

1. INTRODUCTION

Overview and Conceptualisation

This paper investigates whether non-diegetic music spatialisation techniques offer new creative possibilities for the use of music in virtual reality (VR) games and applications, creating more immersive and interactive auditory VR experiences. Non-diegetic music can be defined as being situated in another place and another time than the events and objects currently represented in the game world – such as background music. Diegetic music can be defined as music with a visible source, or a source implied to be present in the narrative of the game – such as a musician playing in a corner of a street in the game. This research explores the spatial possibilities that exist between diegetic and non-diegetic music by utilising spatial attributes and techniques commonly used in games for audio object spatialisation to the music soundtrack.

In the majority of current VR games, the background music is presented as a stereo mix that is perceived as being localised inside the players head when listened to binaurally with headphones. Non-diegetic background music is usually presented to the player without situational spatial positioning or other game-scene-related spatial cues. By contrast, all diegetic sound design elements surrounding the player in the VR space, such as game object and character sounds, are spatialised using head-related transfer function (HRTF)-based techniques [1, 2, 3, 4]. In addition, other spatial cues such as reverberation can be used to mimic the given virtual space, allowing sounds to be audibly situated and localised in the virtual space around the player.

As VR offers visual and sometimes haptic stimuli that aim to convince our senses that we are actually in that virtual space, then why is background music still presented as if it was coming to us from some other space? In other words, why do we hear music playing in an unseen chamber music hall when we are exploring a virtual dungeon or a huge cave? Wouldn't the virtual dungeon or cave space be more convincing, and thus more immersive, if the background music were also to sound like it is playing in the dungeon or cave space?

Even though "immersion" is an academically ambiguous term [5], immersion and immersiveness are still highly appreciated features for gamers and game developers as well as music professionals working with spatial music. In their study, Peters et al. [6] asked 52 spatial music composers "what is the most important technical and compositional aspect regarding spatial music composition?" Immersion was rated as the most important aspect, even more crucial than localisation accuracy or distance perception, which were also highly rated features. Even if the formal definition of immersion is elusive, practitioners and gamers demonstrate an intuitive appreciation of the term. Therefore, the subjects in this study were asked to define immersion in their own words.

While there is an abundance of scientific research on object-based audio spatialisation and localisation using HRTFs [2, 3, 4], as well as on various audio quality assessments for spatial audio perception in VR [7, 8], *music* spatialisation is a much less researched topic [4, 6, 9], and there is little to no scientific literature on non-diegetic music spatialisation in VR games and applications. Award-winning composer and author

Winifred Phillips [10] has addressed some of the questions that arise from background music spatialisation in VR. The most important of these include: “Where should video game music be in a VR game?” and “Should it feel like it exists inside the VR world, weaving itself into the immersive 3D atmosphere surrounding the player? Or should it feel like it’s somehow outside of the VR environment and is instead coasting on top of the experience, being conveyed directly to the player?” [11] In her Game Developers Conference presentation [12], Phillips explores methods that blur the distinction between diegetic and non-diegetic music.

This research presented here explores the creative possibilities that exist between diegetic and non-diegetic music in VR by introducing situational spatial positioning and other game-scene-related spatial cues to the background music soundtrack. The novelty of this research as a preliminary study and an orientation on the possibilities in the new field of spatial music in VR should be emphasized.

2. THEORETICAL FRAMEWORK: AURALISATION

2.1 Geometry-Based Auralisation

Most research in the field of music spatialisation focuses on room modelling with computational room acoustics, one aim being “to implement an auralisation system that renders audible a 3D model of an acoustic environment” [13].

The term “auralisation” has been defined by Kleiner et al. [1] as follows: “Auralization is the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space.” Geometry-based auralisation is an important research topic for diegetic music. It requires first recording the music in an anechoic chamber and then convolving these echo-free signals with binaural room impulse responses (BRIRs). The BRIRs represent the acoustic transmission response of a sound source in a measured space to the eardrums of a listener.

2.2 Perception-Based Auralisation

The approach in this experiment differs from the geometry-based computational room acoustics in that it does not aim to model the geometry of the virtual spaces faithfully. One objective is to minimise the computational costs in order to make music auralisation a viable option for (VR) game developers. Geometry-based auralisation requires heavy processing, making it an unrealistic option for non-diegetic music in VR games in the near future, given the limited computational resources allocated for music in such games. Moreover, the geometry-based approach requires any music to be recorded in an anechoic chamber, which most composers and game music studios do not have access to.

Instead, this study uses a perception-based approach to achieve auralisation, in other words designing the spatial rendering from a perceptual point of view. Perception-based auralisation can be produced without applying the precise geometry of a room [13]. The exact response of a given room, or the physical propagation of the sound in the given room, is not necessary for synthesising the auralisation [14]. “The perceptual approach seeks to reproduce only the perceptually salient characteristics of reverberation” [15].

This paper investigates the perceptual possibilities for non-diegetic background music auralisation that are least intensive in terms of processing power and other resources. This study suggests that it is possible to achieve perception-based music auralisation for VR by recording musical instruments in a basic studio environment and then binaurally rendering the instruments with BRIRs of a similar (but not geometrically identical) space. The hypothesis in this study is that the background music can be spatialised by introducing perceptual cues that are salient to the acoustic characteristics of the space in VR – thereby externalising the music to this surrounding virtual space. This in turn can increase immersion and/or affect a player’s perceptions of the virtual space.

The aim of this study is to present creative auralisation ideas and strategies that are easy to use and implement, are processor-friendly (i.e. do not use too much computational resources), are production-pipeline efficient and are more artistically flexible than geometry-based auralisation methods. Adding spatial attributes to non-diegetic game music could be seen as an opportunity to expand musical language, as well as a creative possibility for composers to broaden their musical expression and to add new layers of meaning to the music. These perception-based music auralisation strategies can offer new creative opportunities for game composers and music content creators, as well as for game audio directors and audio leads when designing music for VR. These strategies can also be applied to augmented reality (AR) and extended reality (XR) music development.

3. RESEARCH QUESTIONS AND METHODOLOGY

In this study, the spatial music phenomenon is subjectively evaluated by human listeners, based on a complex virtual scene with multiple spatial attributes. The spatial attributes are described in more detail in Section 4. A pre-defined set of semi-structured [16a] questions was used to conduct a qualitative interview (see [Appendix 1]). These questions were designed to elicit information that might help us to understand and answer the following research questions:

- What are the advantages and disadvantages of music spatialisation in VR/AR/XR games and applications?

- Can the spatialisation of background music affect players perception of the virtual space?

The qualitative research methodology [16, 17] used in this paper relies on expert interviews [16b] based on a purposive sample [16c]. The primary selection [16b] of experts in the field of music technology consists of doctoral students, alumni and faculty members at the music technology department at the University of the Arts

Helsinki. The experts interviewed were chosen for their knowledge and experience necessary for critical evaluation of the phenomenon of spatial audio and music. Besides addressing the main research questions, the goal of the interviews was to identify concepts and generate conceptual frameworks to explore the new field of spatial music in VR. The interview data was analysed using inductive in vivo coding, descriptive coding and axial coding to identify relationships between categories [17]. This type of qualitative content analysis is a method mainly used to “analyse subjective viewpoints, collected with interviews” [16d]. The primary selection of experts in this study was limited to seven participants due to a shortage of suitable available experts. In order to reach “saturation”, the point at which further interview data provides little or no new information [17, 18], further interviews will be required in later stages of the research. As this is a preliminary study aimed at identifying concepts in the emerging field of spatial music in VR, even a limited amount of data is valuable in providing an orientation to the field and determining whether the concepts identified invite further research.

4. EXPERIMENT PROCEDURE

Seven expert subjects (four male and three female) played the two experiment versions (normal stereo and spatial music version) consecutively and were interviewed afterwards. The experimental conditions were not randomised because the aim was to make the experts aware of the differences between the two experimental versions. As normal stereo is the standard way of presenting music in VR and in general, it was presented first to make it easier for the experts to compare the less conventional spatialised music version with the standard stereo version. In future phases of this research, randomisation of the experimental conditions (stereo and spatialised) will be considered.

The subjects were first introduced to the VR system (HTC Vive with Sennheiser HD600 headphones) with a tutorial scene, where they learned to move around within the walking simulation used for the experiment. When they felt comfortable with this set-up, they were asked to proceed to the experiment proper. The subjects were asked to purposefully pay attention and critically evaluate the music and audio in the experiment scenes.

First, they experienced the VR experience with normal stereo background music. They were allowed to play for a maximum of three minutes, then the experience was stopped and started again with the same music spatialised so that it should sound as if the music was playing and reverberating in the surrounding virtual environment – in this case, a large cave. The participants were allowed to play for a maximum of three minutes with each of the two experiment versions and were interviewed immediately after playing.

The VR experience developed for the experiment takes place in a virtual cave environment, where the subjects start from a very large dark cave and are lured with a light source towards a tunnel (a small space), from which they enter another large red cave space. These three virtual cave spaces can be seen in Figures 1, 2 and 3. (Note: it is not possible to properly represent a 3D space in a 2D image, for a better understanding of the scale of the spaces see video excerpts from the experiment [23] and [24]).

Figure 1. The starting position of the VR experience, which takes place in a dark, very large cave, where the entrance of a tunnel (the second space, see Figure 2) can be seen as a light source in the distance.

Figure 2. The second space of the VR experience is a small tunnel, from which the player enters the final large cave space (see Figure 3).

Figure 3. The third and final space of the VR experience, which is a very large cave with water and a waterfall.

4.1 The Two Experiment Versions Explained – Differences in Music Spatialisation

The music composed and recorded for the experiment is for a string ensemble (violins, violas, cellos and double basses) and a grand piano.

The first version of the experiment presents the music in basic stereo with no spatial BRIR rendering. The musical instruments are recorded and mixed down to stereo with typical classical music production standards, to sound as though the music is playing in a small to moderately sized chamber music hall. A basic stereo mix is perceived as being localised inside the subjects’ head when listened to binaurally with headphones.

The second version of the experiment presents the music as spatialised. The same musical instruments are spatialised with a hybrid of two auralisation techniques, one static and the other object-based. The auralisation techniques used are described in detail in the next sections.

4.2 Spatial Music Static Auralisation

Most of the score (everything except the lead violin) in the second experiment version was auralised with *static auralisation* by convolving the instrument sound files (recorded in a music studio environment) with binaural room impulse responses (BRIRs) measured in a cave (a large cave in Helsinki) with a high-quality dummy-head stereo microphone. See [Appendix 2] for a detailed description of the BRIRs measurement procedure employed. In other words, the main part of the score did not use object-based/head-tracked localisation, but was spatialised with a static binaural mixing [9, 19] technique, where the instruments were convolved using BRIRs of a large cave and rendered as a static (not head-tracked) binaural stereo mix. Listened to with headphones, the aim of this static auralisation was to externalise the music from the inside of the subject's head, to sound as if the music was playing in the surrounding large cave.

4.3 Spatial Music Object-Based Auralisation

Object-based auralisation was used *for one instrument* in the second experiment version, *the lead violin*. The lead violin sound file was first convolved with a mono impulse response from a tunnel (a Space Designer stock IR [20]), and then added as an audio source to a game object in the VR scene. This was done using the DearVR [21] auralisation plugin that renders real-time head-tracked positional HRTFs to the violin signal. The DearVR plugin was also programmed to simulate volume and frequency drop-off over distance, meaning that when the subject approaches the (invisible) violin source, the volume of the violin sound increases, and high frequencies become more prominent.

The lead violin source was attached to a light source leading to a tunnel in the cave scene as an aurally localisable but invisible game object. Pairing the directional light cues with directional spatial music cues could possibly be used as a subtle or subliminal guidance tool in VR, luring the player towards a desired direction in the virtual environment.

4.4 Adaptive Music Auralisation in the Second Experiment Version

Both the static and the object-based music auralisation techniques used in the second experiment version adaptively follow the size of the different spaces represented in the scene. For example, when the player enters a tunnel (a smaller space), the music auralisation changes from a large cave BRIRs to a smaller tunnel BRIRs, and vice versa when exiting from the tunnel to a second large cave. In other words, the music auralisation adaptively follows the player's movements between the different-sized spaces in the VR scene. This method of adaptive music auralisation has its origins in game sound design practices, where adaptive spatial design is widely used for everything from player footsteps to other object and environmental sounds. This paper does not discuss in detail the techniques and methods commonly used in sound design to implement spatial interpolation, as this is beyond the scope of this paper. The adaptive music auralisation in this experiment was achieved by interpolating between two spatialised versions of the music, one rendered with a BRIR from a large cave, the other with a BRIR from a tunnel.

In both versions of the experiment (stereo and spatialised), the footstep sounds adaptively follow the player's movements between the different-sized spaces in the VR scene. This means that when a player walks in a large cave, the footstep sounds have a high degree of reverberation as well as a slap-back echo from a distant wall. By contrast, when the player walks into a tunnel, the reverberation of their footsteps is shortened accordingly. It is important to note that all the sound design elements (object sounds such as footsteps) stay the same through both versions of the experiment. The only difference between the two versions is the music spatialisation.

Excerpts from the normal stereo and spatial music versions of the VR experiences used in the research experiment can be found via the following reference links:

[23] A video excerpt from the first experiment version with basic stereo music.

[24] A video excerpt from the second experiment version with spatial music.

It should be noted that videos from VR experiences never correspond to the actual experience of being in a virtual space, as the head tracking for object-based auralisation is not properly conveyed in a video form. Therefore, only rough conclusions and impressions can be drawn from watching and listening to the excerpts provided.

4.5 Hybrid of Auralisation Techniques

The hybrid of two auralisation techniques used in the second spatialised experiment version are merely one choice from numerous auralisation possibilities [4, 9, 19]. The combination of static and object-based auralisation used was selected primarily for its simplicity and processor friendliness compared with more complex sound-field-based or purely object-based auralisation techniques, as "processing and transmission cost increase linearly with the number of sound events to be synthesised simultaneously" [14].

Everything except for the lead violin was spatialised with the static auralisation technique, resulting in one stereo music track with BRIR rendering. During the development and iteration of the experiment, the hybrid of static rendering combined with one object-based head-tracked instrument source (the lead violin) was perceived to externalise the music to the surrounding virtual space.

A clear indication that the hybrid techniques used were successful is that the expert subjects in the experiment were able to detect and evaluate the music auralisation. Six out of seven subjects were also able to detect the object-based auralisation – the lead violin source at the entrance of the tunnel. The hybrid auralisation approach was perceived to externalise the musical content to the surrounding virtual cave. It can thus be interpreted that the hybrid of static and object-based music elements, in combination with the variation and mixture of ambient sounds and other object sound effects and reflections, resulted in constant changes in the amplitude and spectra of the overall perceived soundscape.

5. RESULTS OF THE QUALITATIVE INQUIRY

All expert subjects reported a difference between the two experiment versions (stereo and spatialised), giving answers in the affirmative to the first question asking whether they perceived a difference in the music between the two VR scenes (see [Appendix 1]).

In response to the second question, which asked what happened in the music in the second scene compared with the first, the subjects described their experience in a variety of ways, using the words "spatialised" and "diegetic" when analysing the music. All subjects used descriptions that indicated their recognition that the music was spatialised, such as describing the music as playing in the virtual space around them, as being "more integrated" with the space, "revealing" the space, or having "spatial cues".

Six out of seven respondents recognised *adaptive music auralisation*, describing that the spatialised music gave them information about the sizes of the virtual spaces around them, and that the music reverb was modulated to follow the changes between the different sizes of the virtual spaces. Some subjects paid attention to the tunnel, where they described feeling the walls of the tunnel close to them because "the music reflected off the walls" or had "fast reflections", in contrast to when the music reverberation "opened up" when entering the large cave space, where the music had a longer reverberation.

Six of the seven participants detected *object-based music auralisation*, reporting that the music was "head-tracking" and that they localised either the music or the violin as coming "from the direction of the tunnel". Some subjects noted that the music was "directing" or "leading" them towards the tunnel.

The chosen music spatialisation strategy also elicited some negative comments, with participants stating that "the space didn't complement the instruments, they had this metallic and hard tone". Some participants described the spatialised version of the music as being more "blurry" and "unclear" and that there was less "clarity" and "rhythm" in the instrumental aspects.

The subjects with the most prior gaming experience, who are used to hearing background music as a non-diegetic element in games, were less accepting of the new way of presenting background music as a part of the virtual space. One participant disliked some aspects of the spatialisation and adaptive music auralisation but liked another, describing how they are used to having the music on a "different level" but agreeing that the spatialisation "tends to put everything in the same space". Another respondent with gaming experience also wondered about the function of music in the virtual environment, saying that in a first-person shooter or FPS game they don't usually pay attention to the music: "I usually just try to listen to where the enemy is". They found that "the stereo music stayed nicely in the background and didn't raise questions like the spatialised music did". They added that "if the point was to follow the music, the spatialisation would work really well". This subject was also surprised by their own reaction: "I'm so used to background music coming from somewhere else than the virtual space I'm in, so it surprised me how out of old habit I automatically accepted the stereo music more readily than the spatialised music".

When considering the appropriate use of spatial music, many subjects stated that there are game situations where music should not be spatialised. They stated that spatial music might be appropriate for situations where game designers want to "draw attention to sound and space" or to offer the player a possibility for "lingering in and exploration of the virtual space". Several respondents felt that spatial music would not be appropriate for fight-and-flight situations, as "if you want the player to move quickly from one place to another, it's probably not the best choice to try this [spatial music] for that situation".

The subjects were then asked to consider the concept of immersion: "How do you understand the term 'immersive'? What does 'immersive' mean to you?" After the subjects gave definitions of the term in their own words, they were asked to rate which experiment version was in their opinion more immersive. Their definitions of immersion (Question 6, listed in [Appendix 1]) include: "being completely in the medium"; "immersed to a different medium such as water"; "surrendering completely to the experience"; "submerged in something, completely surrounded"; "being totally inside the ambiance or the second environment"; "all-encompassing, blurring the line of reality and virtual reality"; and "you can forget reality and immerse yourself in an artificial world".

All seven participants rated the second version of the experiment (with spatial music) as more immersive (question 7, see [Appendix 1]). Reasons given included the observations that spatial music makes the experience or virtual space more "holistic", "interesting" or "stimulating". They felt more "involved and present" or more "related" to the virtual space. One participant stated that when you spatialise music to sound like "it was in the cave", it makes the experience more immersive. One participant found that the sound effects and spatial music blended better in the spatial version: "the mix is much smoother". Two participants stated that a more immersive experience was not necessarily a better one, as "the metallic sound of the violin" or "the blurriness" caused by the auralisation disturbed their experience.

When asked about their perception of the space (question 8), the subjects reported that the spatialised music gave "information" about the virtual spaces, "revealed" the space and "emphasised" the differences in the sizes of the caves. They found that the spatialised music gave "clues" and "cues" about the spaces and made them "pay more attention to the virtual space" they were experiencing. The question about perception also elicited several comments about being in an "explorative mode".

Six of the seven participants felt that the spatialised music version "involved" them in being more "curious" about their surroundings, inviting them to "explore" and "linger" in the virtual space.

Many interesting observations emerged from the final few questions, which focused on how the experiment had affected the views of the participants, or invited further discussion of the issues raised. One subject even invented new concepts – calling the normal stereo music the "detached soundtrack" and spatial music the "integrated soundtrack". Subjects variously described stereo music as "wallpaper" and "more static" than spatial music. Two subjects reported having a "bodily experience" from spatial music. Spatial music seems also to draw attention to other sound effect-based spatial cues, as "it [the music spatialisation] helps me to pay more attention to other spatial sonic cues".

6. DISCUSSION

As the qualitative aspects of music spatialisation are multifaceted and subjective, subjects are affected by various contextual factors and stimuli when experiencing a music soundtrack in VR. Due to the variety of stimuli, participants are easily distracted from their task of listening and critically evaluating the music, and it is often unclear what aspects of the experience they are assessing factually.

Half of the participants in this experiment had minimal experience (fewer than three times) with virtual reality systems. Despite having little prior experience with VR, subjects were able to assess the musical qualities in question, although some were so overwhelmed with the experience of being in VR that they sometimes forgot to pay attention to the music. It should be considered therefore that a lack of prior VR experience can affect such music assessments. For future research, similar research and experiments could target VR gamers or VR developers instead of music technology experts.

The subjective nature of qualitative research predisposes the study for bias both from the researchers and from the subjects. As the sample here solely includes experts in music, there might be a tendency for respondents to favour the spatial music version, as it might evoke their professional interests more. Even when the interviewer tried to stay completely neutral, as all aspects of the phenomena are equally interesting for the purposes of study, some subliminal compliance issues might be conveyed to the subjects in the interview process.

In terms of the experimental set-up, the unwanted timbral coloration caused by the auralisation process cannot be overlooked. For example, the blurring effect of a long reverberation from the cave BRIRs, used to auralise the music to the big cave, hampered the instrumental expression and clarity and inflicted a metallic and hard tone on the string instruments. Such unwanted colorations need to be taken into consideration at the composition stage, as in many situations they might override the possible benefits of the auralisation.

Since certain instruments prefer warmer, concert hall-like acoustics, it would be useful to study the possibilities of spatialising the music in a particular surrounding environment without including the undesirable sonic properties of that space. Instead of modelling the space as faithfully as possible, as in this experiment, future experiments could exploit perceptual possibilities by considering the acoustic needs of the instruments and better adapting the perceptual spatialisation to these requirements. Different VR spaces and game situations require different approaches for perception-based music auralisation. Subsequent research on other auralisation strategies could be studied and tested on this experimental space and a variety of other types of VR games and virtual spaces.

As described in chapter 5, many of the expert subjects expressed the opinion that spatial music is not suited for every VR game situation, but that there might be game situations where narrative variation could be achieved or enhanced with spatial music. Many of the participants described how spatial music evokes the sense of exploration and lingering in the VR experience. This indicates that spatial music could, for example, be used to signal a change in a game state to the player, for example, using spatial music – the integrated soundtrack – for an exploration setting, then changing the music back to normal stereo – the detached soundtrack – for a fight-or-flight scenario.

When analysing the data with axial coding [16e], it became evident that when the music was auralised, the subjects paid more attention to other sound effect-based spatial cues such as footstep reverberation. This suggests that spatial music could also be used to draw a player's attention to their virtual surroundings. Both experiment versions had the same footstep auralisation (as sound effect-based spatial cues remained constant in both experiment versions), but in the spatial music version the subjects seemed to pay more attention and pick up more information from the spatial cues of footsteps and other environmental sounds. This indicates that the spatial cues based on sound effects might be emphasised by spatial music. Further research is needed to justify this argument.

Concerning perception of the space by emphasising virtual spaces with adaptive music auralisation: most subjects acknowledged the adaptive auralisation, many saying that they enjoyed the adaptivity, meaning that they enjoyed the music spatialisation that followed the player's movements between the different sized spaces in the VR scene. Almost all subjects reported getting information about the surrounding virtual space from the auralised music. This leads to a conclusion that spatial music could be used to emphasise virtual spaces in VR games, for example, evoking or emphasising awe and astonishment by spatialising music for a particularly impressive VR scene or space, where the game developers want to draw players' attention to the surrounding virtual space.

The claim that pairing the directional light cues with directional spatial music cues could potentially be used as a subtle or subliminal guidance tool in VR, luring or guiding the player to a desired direction in the virtual environment, was also proven to work as claimed. Six of the seven participants detected this object-based music auralisation, represented by the invisible violin source at the entrance of the tunnel. Many of them reported that the violin sound guided or invited them to approach the tunnel.

The hypothesis that spatialising the background music with perceptually salient acoustical characteristics could increase immersion was also supported by the experiment. All expert subjects rated the spatial music version as more immersive. This indicates that when used appropriately, spatial music holds the potential to increase immersion in VR game situations. What the appropriate auralisation approaches might be for different game situations and virtual spaces, invites further experimenting, testing and inquiry.

This sample size allows the identification of concepts and similarities in the experiences of the expert subjects, as described in Section 5 of the paper. The identified concepts and similarities in the reported experiences encourage further research into the phenomenon of spatial music in VR.

Different approaches and perspectives on music spatialisation require further research. This research should explore different methods of spatialisation, for example, testing the spatialisation of music without aiming for the lowest possible computational cost. In addition, different research methods should be used, such as statistical analysis to assess the perceived quality of spatial music compared to regular stereo music in VR using established quantitative methods such as MUSHRA (Multiple Stimulus with Hidden Reference and Anchors) [7, 8], or more generalisable qualitative research methods such as DELPHI [25]. These extensions would naturally lead to triangulation [16f] of both data and research methods, resulting in a better understanding of the phenomena of spatial music in VR. The sample size needs to be increased and extended to other expert groups, such as gamers and VR game developers, and eventually to the general public.

7. CONCLUSIONS

This study serves as a pilot study for exploring the possibilities of spatial music in virtual reality; the purpose of the study was not to generalise the phenomena found, but to identify possible new avenues for further research into the use of spatial music in virtual reality, and to invite discussion among game music composers and music content creators about the creative possibilities of using spatial music in VR.

This research indicates that spatial music can contribute to the perception of a given virtual space and, if used appropriately, has the potential to increase immersion and enhance the illusion of “being there” in the virtual environment. On the basis of this research, the phrases “detached soundtrack” for ordinary stereo music, and “integrated soundtrack” for spatial music, are proposed as standard terms when talking about the perceptual characteristics of music soundtracks in the context of VR games and applications.

Acknowledgments

This research has been supported by the Finnish Cultural Foundation. The author would like to thank the Centre for Music & Technology, Sibelius Academy, University of the Arts Helsinki.

8. REFERENCES

[1] M. Kleiner, B.-I. Dalenbäck and P. Svensson, “Auralization – An Overview,” *Journal of The Audio Engineering Society,* p. 861, 1993.

[2] J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localization, Massachusetts Institute of Technology: MIT Press, 1999.

[3] F. Wightman and D. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *The Journal of the Acoustical Society of America,* vol. 85, 1989.

[4] C. Pike, Evaluating the Perceived Quality of Binaural Technology, University of York, 2019.

[5] A. Sarvesh, A. Simon, S. Bech, K. Bærentsen and S. Forchhammer, "Defining Immersion: Literature Review and Implications for Research on Immersive Audiovisual Experiences," in *Audio Engineering Society Convention Paper,* New York, 2019.

[6] N. Peters, G. Marentakis and S. McAdams, "Current Technologies and Compositional Practices for Spatialization: A Qualitative and Quantitative Analysis," *Computer Music Journal,* MIT Press, 2011.

[7] F. Brinkmann and S. Weinzierl, "Audio Quality Assessment for Virtual Reality," in *Sonic Interactions in Virtual Environments*, M. Geronazzo and S. Serafin, Eds., Human-Computer Interaction Series, 2023, pp. 145-174.

[8] O. Rummukainen, T. Robotham, S. J. Schlecht, A. Plinge, J. Herre and E. A. P. Habets, "Audio Quality Evaluation in Virtual Reality: Multiple Stimulus Ranking with Behavior Tracking," in *AES Conference Audio for Virtual and Augmented Reality*, Redmond, WA, 2018.

[9] T. Parnell and C. Pike, "An efficient method for producing binaural mixes of classical music from a primary stereo mix," in *Audio Engineering Society Convention Paper*, New York, 2018.

[10] W. Phillips, A Composer's Guide to Game Music, MIT Press, 2014.

[11] W. Phillips, "Virtual Reality Developers Conference takeaways: VR music for the game composer," 2017. [Online].Available: https://www.gamedeveloper.co m [Accessed February 2023].

[12] W. Phillips, "Game Developers Conference Vault," 2018.[Online].Available: https://www.gdcvault.com [Accessed February 2023].

[13] T. Lokki, Physically-based Auralization – Design, Implementation, and Evaluation, Espoo: Helsinki University of Technology, Publications in Telecommunications Software and Multimedia, 2002, p. 1.

[14] J.-M. Jot, "Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces," *Multimedia Systems, Special Issue on Audio and Multimedia*, vol. 7, p. 55–69, 1996.

[15] W. Gardner, "Reverberation Algorithms," in *Applications of Digital Signal Processing to Audio and Acoustics*, Cambridge, Kluwer Academic Publishers, 2002, p. 87

[16] U. Flick, An Introduction to Qualitative Research, 6th edition, SAGE Publications Ltd, 2018, pages [16a] p. 207 [16b] p. 237 [16c] p. 182 [16d] p. 487 [16e] p. 458 [16f] p. 191

[17] M. Miles, M. Huberman and J. Saldana, Qualitative Data Analysis A Methods Sourcebook, Arizona State University: SAGE Publications, 2014, p. 74 and 81

[18] G. Guest, A. Bunce and L. Johnson, "How Many Interviews Are Enough? An Experiment with Data Saturation and Variability," *Field Methods*, vol. 18, no. 1, p. 59–82, 2006

[19] Sennheiser electronic GmbH, "Approaching Static Binaural Mixing with AMBEO Orbit," 2018. [Online].Available: https://assets.sennheiser.com [Accessed February 2023].

[20] Apple, "Space designer plugin Logic Pro X," [Online]. Available: https://support.apple.com [Accessed February 2023].

[21] DearVR, "DearVR Unity plugin," [Online]. Available: https://www.dear-reality.com [Accessed February 2023].

[22] Apple, "Impulse Response Utility manual," [Online]. Available: https://support.apple.com/ [Accessed February 2023].

[23] S. Mahlamäki, “A video excerpt from the first experiment version with basic stereo music,” 2023. [Online].Available: https://youtu.be/dnADjJpwQFI

[24] S. Mahlamäki, “A video excerpt from the second experiment version with spatial music,” 2023. [Online].Available: https://youtu.be/0EaVQgdwH4g

[25] A. Silzle, "Quality Taxonomies for Auditory Virtual Environments," in *AES 122nd Convention*, Vienna, 2007.

[Appendix 1]

Qualitative interview questions:

1. Was there, in your opinion, a difference in the music between the two VR scenes you were in?
2. Tell me about what happened in the music in the second scene compared with the first?
3. What was it like?
4. What did you think about it?
5. In your opinion, which musical version fits the virtual scene better?
6. How do you understand the term ‘immersive’? What does ‘immersive’ mean to you?
7. Which musical version was, in your opinion, more immersive? Could you describe why you think that/so?
8. Do you feel that either one of the music versions affected somehow your perception of the virtual space you were in? If so, how?
9. Tell me about how your views (about background music in VR or in general) may have changed since you experienced these VR environments?
10. Is there anything that you might have not thought about before that occurred to you during this interview/these VR experiences?

[Appendix 2]

The binaural room impulse responses (BRIRs) of the large cave were measured at azimuth angles of -90°, -60°, -30°, 0°, 30°, 60°, and 90°. At a distance of approximately 20 m from the dummy head with an elevation angle $\theta \approx 20$, a sine sweep that covered the entire audible frequency range of 20 to 25 kHz was played through a high-quality speaker. The sine sweep produced by a loudspeaker in the measurement space (a large cave) was binaurally recorded with a Neumann KU 100 Dummy Head Binaural Stereo Microphone, though any real- or dummy-head binaural stereo microphone could have been used. This recording resulted in a sine sweep sound file containing the binaural room response of the cave space including the directional information and filtering of the pinna and head (HRTF). This recorded sound file, namely the sine sweep with the BRIR attached, was then deconvolved with the sine sweep. Deconvolution is a mathematical process used to reverse the effects of convolution on recorded data. This means that the sine sweep was removed from the signal, leaving only the BRIR information. This information can be used as a static auralisation filter with any convolution software available (e.g., Space Designer, Altiverb, etc). The whole BRIR measurement procedure described above was conducted using the MacBook Pro software Impulse Response Utility [22].