DIGITAL CREATIVITY
https://doi.org/10.1080/14626268.2024.2386353

Handheld controller-based locomotion in Virtual Reality as an approach to interactive music composition: insights from composers' preferences

Matteo Tomasetti and Luca Turchet

Department of Information Engineering and Computer Science, University of Trento, Trento, Italy

ABSTRACT
In this paper, we propose a novel interactive composition approach in which locomotion within a virtual environment is used as a means to trigger predefined sound events, attenuations of sound parameters, immersive sound spatializations, and reverberations. To implement such an approach, we developed a virtual environment consisting of a large auditorium shaped like a nautilus, where users move via handheld controllers. We tested the proposed virtual environment and composition approach by conducting a user study with 15 expert composers, where we compared the 3 locomotion techniques most used in VR to move in a 3D virtual space using a handheld controller: *continuous* (i.e. walking), *discrete* (i.e. teleporting), and *mixed* movement (i.e. walking and teleporting). Our results showed a preference for the *mixed* locomotion technique. We provide a critical reflection on such a result as well as on various insights that arose from the composers' comments.

ARTICLE HISTORY
Received 26 February 2023
Accepted 22 July 2024

KEYWORDS
Musical XR; virtual reality; interactive composition; computer music; spatial audio

Introduction

The last twenty years have witnessed a considerable expansion in artistic and scientific research using the Extended Reality (XR) medium in the music scenario. This expansion has led to the development of a new research area referred to as Musical XR, which includes all interactive music processes in virtual and augmented environments (Turchet, Hamilton, and Çamci 2021). Moreover, recent developments in Virtual Reality (VR) technology have brought back interest in techniques related to spatial audio, such as Ambisonics and binaural audio, which offer significant advantages for immersive audio experiences (Çamci 2019).

In general, audio design in VR experiences should be dynamically generated, contain forms of interaction, and be immersive (Atherton and Wang 2020). Typically, the 3D interaction techniques that the user in VR can experience are divided into locomotion, selection, manipulation, and application control (Berthaut 2020). During the last decade, several systems related to musical composition in virtual environments (VEs) have emerged that would consider these interactions. Nevertheless, to the best of the authors' knowledge, no work in the literature on VR has considered locomotion as an interaction tool for generating and managing music composition by composers as well as for allowing users (intended as listeners) to explore the composition at their leisure, becoming a sort of active performer of the composition itself. With the term

CONTACT Matteo Tomasetti matteo.tomasetti@unitn.it Department of Information Engineering and Computer Science, University of Trento, Via Sommarive 9, Povo, Trento 38123, Italy

locomotion, we refer to the fact that the user can move freely in any direction in the VE.

The objectives of this study are threefold. First, we aim to propose a novel, replicable approach for sound designers and composers to create interactive compositions for VR based on the user locomotion technique. Second, we aim to assess which is the most preferred locomotion technique for both composing and exploring the open-form musical piece we created. Third, we aim to understand the composers' user experience interacting with the developed system.

A video showcasing our system in action is available online[1]. Our approach leverages the user's locomotion to trigger different composition layers in real-time and manage specific sound parameter attenuations, which involve amplitude, spread, focus, filtering, reverberation zones, and immersive sound spatialization techniques dependent also on the user's head movements. We refer to our proposed approach with the term 'composition' because it relates to sound material organization, interactive creation, and open-form structure. By the term 'open-form', we refer to the fact that the overall structure of the piece we have composed with our approach is an open-form, in the sense that the music and sounds employed change each time the piece is played because they depend on the user's position in the VE. This type of art format (i.e. open-form) is closely associated with the concepts of nonlinear interactive music and sound design inspired by game audio methodologies (Zdanowicz and Bambrick 2019).

The user, by moving around, triggers the predefined set of sounds and also manages the predefined attenuation curves of various parameters, which we explain in detail in this paper. By the term 'predefined', we denote that the sounds and the parameter settings of the attenuation curves are determined by the authors of this paper and are triggered by user interactions. Consequently, users do not generate or compose the sounds themselves.

By means of locomotion using handheld controllers, the user becomes an active performer, as the whole composition approach is directly linked to the user's behaviour. User locomotion brings different structural functions that change according to the individual experience because each user makes different movements and routes. We provide an evaluation of our approach to composition by gathering insights from the experience of 15 expert acousmatic music composers interacting with the developed system. In addition, we also propose the evaluation of the role of locomotion techniques in VR music composition, studying and investigating which locomotion technique is preferred for composing and exploring a musical piece. Specifically, we compared 3 different locomotion techniques that can be carried out in VR using a handheld controller: the first is the classic walk, which we refer to as *continuous* locomotion movement; the second is the possibility of teleporting from one point to another of the VE, which we refer to as *discrete* locomotion movement; the third, referred to as *mixed* locomotion movement, is a mix between the two, where users are given the freedom to choose between walking, teleporting or using them simultaneously. Our investigation was driven by two main research questions:

1) How can we design a system to generate music by user locomotion in VR?
2) What are the most preferred locomotion techniques to explore an interactive composition in a VR environment?

To address these questions, we conducted the user study described above, where we investigated further aspects such as the design we proposed, the advantages/disadvantages of this design, the agency in controlling the different parameters proposed, and more general aspects regarding sound organization and localization.

Additionally, we investigated cybersickness, often referred to as visually induced motion sickness in VR environments, characterized by symptoms such as nausea and headache (Kwok, Ng, and Lau 2018). This dimension was investigated because previous studies have found that such locomotion techniques we used can cause this phenomenon (Mayor, Raya, and Sanchez 2019).

2. Related works

This Section describes the previous related works from which we have drawn inspiration. We have divided it into two parts: the first describes the locomotion techniques most commonly used in VR; the second part focuses on previous works related to creating musical experiences using locomotion and similar techniques.

2.1. Locomotion techniques in VR

Several locomotion techniques are currently available in VR, and the one that delivers a significant immersive experience is 'real' or 'natural' walking (Slater, Usoh, and Steed 1995; Steinicke et al. 2013), which means that a user physically walks around the space without using additional controllers (see e.g. Turchet, Burelli, and Serafin 2012). However, natural walking is not widely used because the size of the positional tracking area bounds it, or it entails the usage of complex, cumbersome, and costly systems such as omnidirectional treadmills (Lohman and Turchet 2022). In addition, in several situations VR users do not have as much physical space to move freely, and because of these space constraints, artificial locomotion techniques delivered via controllers have been increasingly used (Liu et al. 2018).

Nevertheless, various locomotion techniques have been developed in the last decades, such as *Walk-in-Place* (Steinicke et al. 2013), flying (Robinett and Holloway 1992), redirected walking (Freitag, Rausch, and Kuhlen 2014). The reader is referred to the works presented by Bozgeyikli et al. (2019) and Zayer, Paul, and Eelke (2018) for in-depth surveys and comparisons of the most common locomotion techniques utilized for room-scaled tracked areas. As these surveys show, the most common locomotion techniques currently utilized in VR are *continuous* (i.e. using a handheld controller for moving in a walking fashion) and *discrete* (i.e. teleporting). The work presented by Riecke and Zielasko (2021) discussed in detail the benefits and disadvantages of using such techniques for different VR scenarios. However, to the best of the author's knowledge, no study has thus far tested the preferences of locomotion techniques concerning exploring the open-form type acousmatic composition in VR.

2.2. Locomotion techniques in musical contexts

A substantial work that we were inspired by to create this study is *Versum* (Barri 2009b), where the author presented an audiovisual sequencer for creating different audiovisual works. For example, in the audiovisual work created with *Versum* entitled *Lought* (Barri 2009a), the user could fly and, therefore, move via a joystick within an abstract virtual world, and every *entity* (sphere or line) present in this abstract virtual world produced a single sound. The closer the user got to each *entity*, the louder the sound. However, our work is very different from that proposed by Barri, as the author did not use VR as a means of fruition. Versum was designed to create sound and visual effects in a 3D environment as a new form of artistic practice. Our work instead aims to investigate the best locomotion technique from those most used in VR to explore a musical piece we created with our approach. It also aims to offer a detailed technical description to allow for the replication of our approach from composers and sound designers.

Furthermore, our work was tested by composers during the evaluation phase, aiming to bring insights regarding the use of locomotion techniques for the exploration or creation of interactive musical pieces in VR environments.

Another example is *SongVerse* (Costa et al. 2019). The authors of this system have created a VE where it is possible to compose loop-based music by selecting and manipulating virtual objects represented in the form of stars, satellites, and planets, which enable the creation of a complex musical composition. Another example is the *Wedge* (Moore et al. 2015). The authors developed a VR interface for composing and playing custom music compositions where the user must use a 3D virtual keyboard formed by cubes representing the different octaves of the musical system to compose. Another composition work is *Maps and Legend* (Hamilton 2008), which tracks users' virtual positions and activity data within a VE. By streaming these data over a network, the actions and paths performed in the VE generate a music content output in a real physical room. In this case, locomotion and movement interactions within the VE occur via keyboard and mouse.

However, except for Barri's work, these two systems use interaction forms related to selection, manipulation, and application control. Only one uses locomotion performed through classical input, such as a mouse and keyboard, to generate musical materials in a real environment.

Another interesting work is the one developed by Wozniewski, Settel, and Cooperstock (2006), where authors developed several more natural controls (e.g. head rotation and movements) that allow one to manage different spatial sound parameters. Such a work has some similarities with the sound mapping strategies used in the present study. However, it does not deal with locomotion as a technique for creating a musical composition and exploring a musical piece.

Another essential example is the VR work presented by Naef and Collicot (2006). Such work allowed the user to position the sound sources at will in the 3D space and change their relative sound spatialization. However, our work follows another paradigm since it does not give users the freedom to position the sounds as they wish. Still, our work aims to assess the preferred locomotion technique by composers to explore our musical piece actively and whether our proposed approach to the interactive composition may be of interest to composers. Furthermore, during the user study, we investigated whether the precise localization of the spatialized sound sources arranged in the VR environment changes perceptually with the variation of the locomotion techniques tested.

The last important example we want to cite is the one developed in the *PHASE* project proposed by Cahen, Rodet, and Lambert (2005), where authors used gestural control with haptic, visual, and sound feedback to achieve the goal of playing music. However, this work differs from ours for the purpose since we were more interested in studying how and what the best form of locomotion that the user can perform in our proposed VR auditorium to explore the composition at will and what it elicits from the composers' perspective during the evaluation about this form of composition's exploration and approach.

3. Design

To address the two research questions outlined in Section 1, we first created a 3D model of a large auditorium. Externally, the auditorium resembles the shape of a nautilus, while internally it is designed to function as a large venue for music listening or organizing Musical XR concerts. We selected a virtual auditorium because we wanted to simulate a conventional mode of musical fruition, which, in reality, usually takes place in theaters and auditoriums. In addition, an auditorium represents an ideal virtual place to install a spatially distributed virtual loudspeaker system, as detailed in Section 3.2.

3.1. The nautilus-shaped auditorium

We divided the whole nautilus-shaped auditorium into two parts: the first part is that of the foyer, and the second part is that of the central area of the auditorium, where we installed the different virtual loudspeakers. To get to the foyer, the user has to enter through an entrance tunnel to the auditorium (see Figure 1). After the foyer, the user descends via virtual steps into the auditorium's central part to experience our composition approach.

3.2. The spatially distributed virtual loudspeaker system

We developed a network of 32 virtual loudspeakers within our auditorium, each with distinguishable shapes and sizes. The design of these virtual loudspeakers was inspired by the loudspeakers commonly used in a real acousmonium, which is a sound system consisting of various groups of loudspeakers spatially distributed and differing in construction sizes and shapes (Kermit-Canfield 2016). After designing the virtual loudspeakers, we positioned them with different orientations in the auditorium's main central area. We located each loudspeaker at a distance ranging from 20 to 60 meters from the other to cover the main central area (see Figure 2).

In detail, we positioned 16 virtual loudspeakers on the auditorium floor, an octophony of loudspeakers positioned at the height of 3 meters, and another octophony of loudspeakers positioned on the floor at the end of the auditorium, where there are some arcades. This area represents the end of the auditorium, where we have implemented 8 arcades that are not widely spaced with each other and can be crossed by the user. For this reason, this is

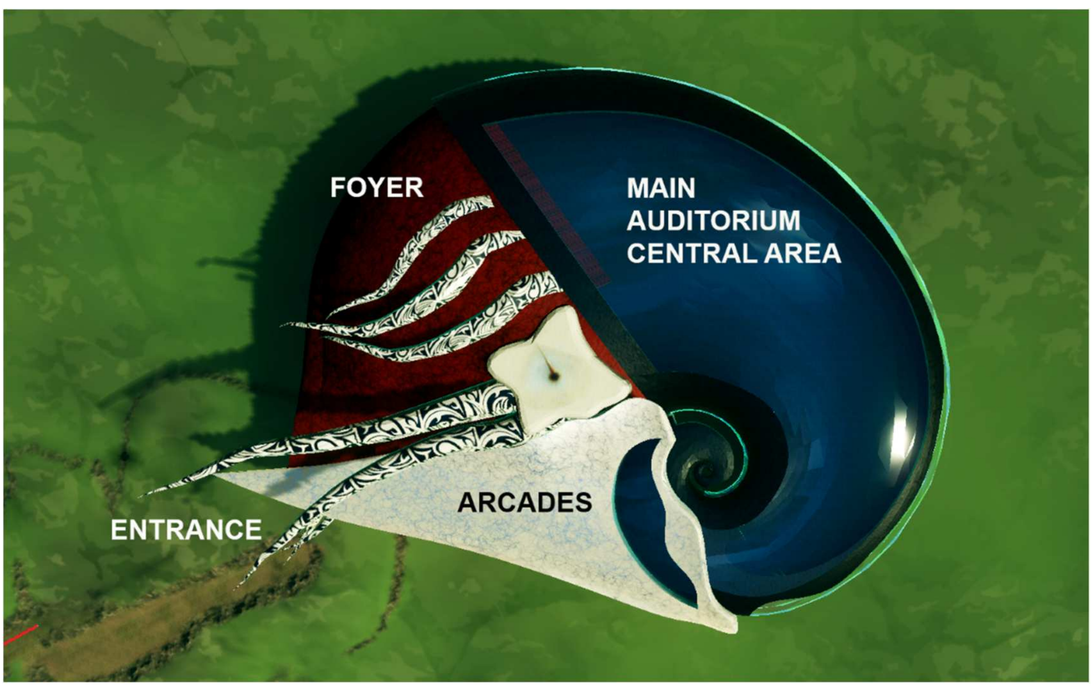

Figure 1. Screenshot of the 3D model of the entire nautilus-shaped auditorium, with the indication of the different areas corresponding to the entrance tunnel, foyer, auditorium, and the auditorium's final part formed by the arcades.

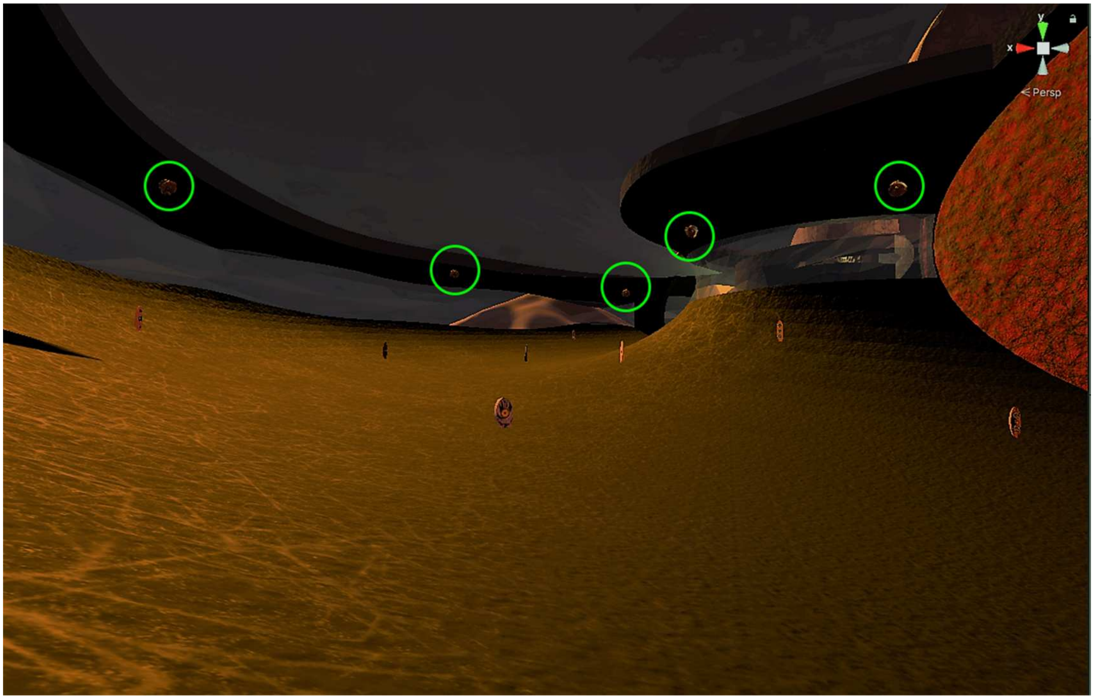

Figure 2. A screenshot of the main auditorium central area where it is possible to see the different loudspeakers that were located on the floor at a distance ranging from 20 to 60 meters. Moreover, it is possible to see some loudspeakers (those circled in green) of the octophony positioned at 3 meters high.

the only space where we have deliberately placed the 8 loudspeakers closer together, one for each arcade. Our vision was that inside the auditorium, our virtual sound diffusion system had the role of increasing the sense of immersiveness and playing a fundamental role in the composition processes that we detail in Sections 4 and 5.

3.3. Interaction design and pilot study regarding two parameters of the XR Interaction Toolkit

We aimed to develop a single-user VR system with first-person viewing. We used the Meta Quest 2 head-mounted display (HMD), connected to the computer via the Meta Quest link cable. Each user managed the locomotion via the Meta Quest 2 handheld controller. In the *continuous* technique, we used the controller thumbstick to manage the directionality and walking in the VE. In the *discrete* technique, however, the user chooses the point to teleport by simply moving her/his head, looking for the point with the controller, and then using the trigger button to carry out the actual teleportation toward the point s/he has chosen. In the *mixed* technique, the user utilizes the controller thumbstick to walk in the VE as in the *continuous* technique, but s/he simultaneously could teleport as in the *discrete* technique. In the *discrete* and *mixed* techniques, the user always had an active raycast showing the arrival point of the teleportation.

To analyze the walking speed and length of the raycast, we performed a pilot study with 3 expert composers. They were left the freedom to choose the walking speed in the range of 1–10 of the XR Interaction Toolkit Unity's component used for developing the locomotion techniques in our system. The composers chose the value 7 in the 'Speed' parameter of the walk, which represents the speed, in units per second, to move forward. They reiterated that, in their opinion, it was the best walking speed to experience the composition. In Unity, the unit of measurement is the meter, and a value of 7 denotes a speed of 7 meters per second. While this value might initially appear exaggerated, it is appropriate given the dimensions of the nautilus-shaped auditorium we designed.

Regarding the length of the raycast, which represents the teleportation arrival point, we used the same Unity toolkit, and we left the pilot study's composers the freedom to choose the raycast length in the range of 5–30 meters. They chose the 15-meter raycast length because, in their opinion, this was the correct value to carry out the experience accurately. In the *mixed* technique, we therefore used both values. Participants in the pilot study did not take part in the evaluation phase.

4. Composition approach

In this section, we present our composition approach. Nevertheless, we do not describe in detail how we created the sounds used for our open-form acousmatic composition because our primary intent is to make clear our general approach, which can be extended especially to electronic musical genres where the beat is not a predominant element (e.g. ambient, experimental, and acousmatic music). In general, for creating our composition, we thought it was essential to think of the sound as a single element that always dialogues with others according to the user's locomotion.

Notably, each sound object created for the composition is a mono signal, which is then encoded via the virtual Ambisonics approach and decoded in binaural within Wwise, as explained in Section 5.2. Moreover, each sound is also adjusted within Wwise via parameter-related attenuation curves, as explained in Section 5.1.

4.1. Foyer

In the foyer, we leveraged the gamification approach (Seaborn and Fels 2015). We implemented a *sound game* where the user is invited to find all the *hidden* sounds through locomotion. We composed various amounts of simple and complex electronic sounds that were displaced in the space. Each user has to move to find and trigger the distinguishable sounds positioned in different places. For the realization of the simple and complex sounds and the organization of the sound material in the foyer, we followed the subsequent composition design choices:

1. To verify and listen to the simultaneous activation of several sounds. During this phase, it was essential to work with sounds in the Digital Audio Workstation (DAW) regarding filtering (equalization) and dynamics to avoid reaching too high a loudness;
2. To avoid using reverb or panning effects within the DAW, as these were implemented, with due precautions, within the Unity-Wwise software, which we explain in detail in Section 5;
3. To create groups of sounds with different simple and complex sounds more or less similar in timbre, allowing us to replace one with the other to create an interactive aleatory musical system. This point is helpful to achieve a sound bank that is always ready for variation and also avoids the perception of a loop by the user.

This third point could probably be equal to the concept of 'Aléa' (Trenkamp 1976), as in the example of the composition *Serenata per un Satellite* by Bruno Maderna. In that composition, each performer has freedom of choice and plays a central role within the composition and performance itself because each performer can select and play different parts of Maderna's score. In our case, the performer is the user who must find and *activate* the sounds by walking or running near or over the different sounds staged in the foyer; consequently, the experience and the produced sounds change every time based on the different movements of the user.

4.2. Auditorium

In the main auditorium, we developed specific internal compositions, one for each loudspeaker, each formed by the union of sound textures created by different electronic sounds dialoguing with each other. We referred to these internal compositions with the term *sound object*. Every sound object can dialogue in musical terms with the other sound objects positioned in the same area. In this way, there are two different musical dialogues to consider in our composition: internally (e.g. the musical dialogue that happens from the different sounds that form a single sound object) and externally (e.g. the musical dialogue that happens between different sound objects). By the term musical dialogue, we refer to the parallel stream of spectromorphologically similar sound objects. The idea of musical dialogue in this work is inspired by the work of Denis Smalley (1997) and the book of Bruce Ellis Benson (2003).

The user can freely move, explore the space, and listen to various sound objects by approaching and moving away from the virtual loudspeakers we placed inside. From a composition perspective, there are two sound layers: one consists of the octophony positioned at the 3-meter height, which has the function of always playing a sound object with the role of remaining as a background sound layer in the composition.

The other layer is related to the sounds produced by the 16 loudspeakers and the octophony positioned in the arcades. Each of these loudspeakers positioned on the floor has the function of reproducing its specific sound object. The challenging aspect of our approach was to make each sound object related to the other sound objects that the loudspeakers emit to create a fluid, spatialized and well-amalgamated acousmatic composition. To overcome such a challenge, we organized and created sound objects by taking into account the following two composition design choices:

1. Each virtual loudspeaker must have its timbre, mix, and unique sound object. Each sound object may last from a minimum of 1′ minute to a maximum of 2′30″ minutes;
2. Every relationship between sound objects should try to make the listener perceive a sort of union and continuity so that the listener understands that these sound objects are all part of the same *unicum* and that they are all tiny fragments of the same piece. This criterion refers more to musical genres where the synchronization between the various sound layers must not be essential or depend on the precision of temporal control like other genres (e.g. rock, folk, hip-hop, techno).
3. Each sound object must have its internal composition form; when more than one is reproduced, they form the global composition form. This form’s dual classification is fundamental because it refers to the user’s freedom, placing the user active with different abilities to choose from within the experience.

In this manner, each user can interact with each sound object in two ways (see Figure 3). First, the user can listen and interact with it individually, positioning near each loudspeaker to listen to each specific sound object. Second, the user is free to move around the space and discover the global composition, listening to the sum of multiple sound objects emitted from multiple loudspeakers. The relationship between multiple sound objects must, in turn, contribute to creating the composition’s global form. However, the global form is entirely shaped by user locomotion, while the user

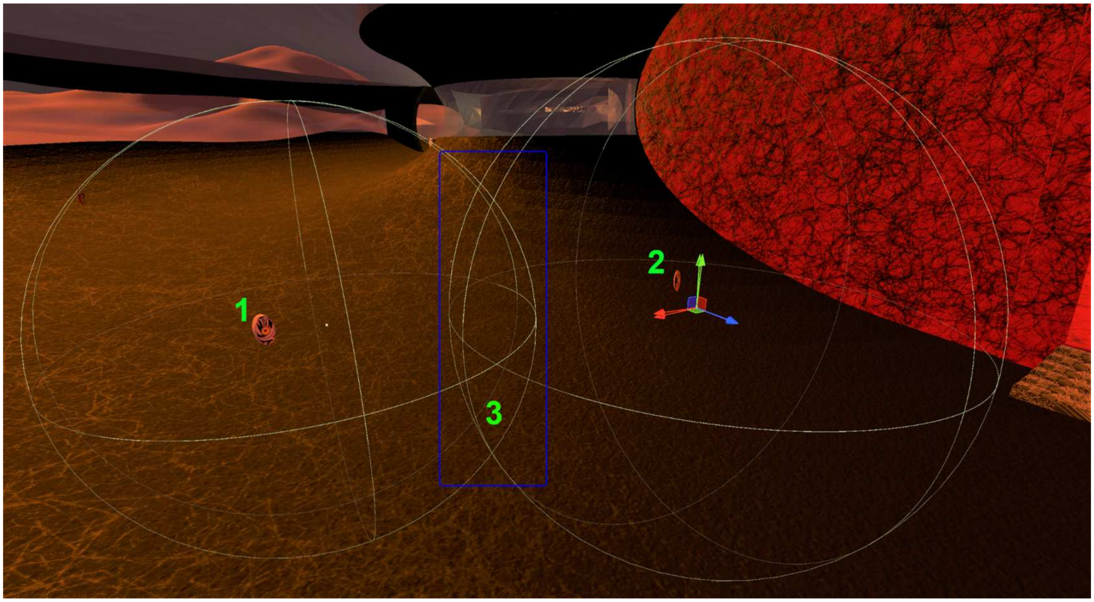

Figure 3. Screenshot explaining user interaction concerning the two different types of compositions. If the user is in front of or near loudspeaker 1 or 2, the user hears that specific loudspeaker's sound object (internal composition). Suppose the user is in the intersection between the two loudspeakers, which in this figure is represented by the number 3, i.e. the area bounded by the blue line. In that case, the user hears the global composition (sum of both sound objects). This rule can be extended to all other loudspeakers in the auditorium.

cannot shape the internal form of each sound object. On the other hand, with these modes, the user shapes the temporal structure of the whole composition. Furthermore, each user hears the global composition differently because the sum of the different sound objects depends on the user's position and the sound parameters that the user attenuates and modifies, which we explain in Section 4.

5. Technical implementation

This section describes all the technical steps we performed to develop our system. Firstly, we developed the 3D model of the nautilus-shaped auditorium with the Autodesk Maya software. Subsequently, we imported it as an FBX file into the Unity software, where we managed all the VE, such as the terrain, the sky, the virtual loudspeakers, and all the textures and lights. An FBX file is a kind of 3D model file made utilizing Autodesk FBX software, such as Autodesk Maya or Blender. Secondly, we developed the 3D models of the different loudspeakers we used in the auditorium within Blender and imported them as FBX files within Unity. After creating the VE within Unity, we linked it with Audiokinetic Wwise, a software mainly used by composers and sound designers to organize and manage interactive composition, sounds, and spatial audio in video games and interactive media.

5.1. The use of the attenuation curves for controlling amplitude, filtering, spread, and focus in Wwise

User locomotion within the VE also triggers a series of attenuations concerning amplitude, filtering, spread, and focus that vary from zone to zone.

The user controls the parameters of these attenuations when moving near or far from a sound source and around it. We concentrated on these parameters because, after several trials, they seemed to be the ones we needed most for our composition needs. In Wwise, the attenuations are based on the following properties:

1. Distance attenuation conditions different audio signal parameters according to the distance between the sound emitter and the user. The distance attenuation is defined using a series of curves;
2. Cone attenuation conditions the amplitude of an audio signal based on the orientation of the sound emitter concerning the user. Cone attenuation is determined using a series of angles defining areas in front of, beside, and in the back of the sound emitter.

Our approach utilizes consolidated game audio methodologies, whose description falls beyond the scope of this paper. Many of the concepts related to the implementation of sound attenuations in the field of nonlinear sound design and music, which are used primarily in the video game domain, have been described accurately in the literature. The reader is referred to the works of Zdanowicz and Bambrick (2019), Schütze and Irwin-Schütze (2018), Robinson (2019), and Phillips (2014) for an in-depth overview of the different methodologies.

Our work utilized distance and cone attenuation curves to adjust parameters such as amplitude, filtering, focus, and spread based on the user's position (see Figure 4). These adjustments were made in real-time using Wwise's Real-Time Parameter Controls (RTPC), allowing real-time dynamic changes as the user moves in the VE. Based on the user's position, messages are sent from Unity to Wwise that control the values of the different cone and distance attenuations. In more detail, amplitude faded out naturally as the user moved away from sound sources, mimicking a natural distance-based decay. We also developed distance attenuation into the filtering parameters, explicitly modifying the cutoff frequency of a low-pass filter, to simulate the attenuation of high frequencies with increasing distance from the sound source. Cone attenuations were applied to mimic the directionality of each of the auditorium loudspeakers, with amplitude decreasing as the user moved away from and to the sides of the loudspeakers. Focus and spread attenuation curves were also implemented to enhance directionality, with spread increasing as the user moved closer to a sound source, and focus condensing the sound sources when close. This approach and the different attenuations we utilized aimed to replicate how humans perceive sounds in reality.

5.2. The use of virtual Ambisonics

We implemented spatial audio within Wwise and linked it with Unity in our VR system. Spatial audio includes all the technologies, such as Ambisonics and binaural, that can produce the feeling of different sound sources positioned in space, just as we perceive the directionality of sounds in real life. In general, Ambisonics is a method that allows the tridimensional recording and playback of audio signals, enabling the production of spatial sound content in any playback system of loudspeakers by applying appropriate decoders (Frank, Zotter, and Sontacchi 2015). Ambisonics consists of several orders and the dimensions for which the sound field is accurately reassembled increase using high-order ambisonics (Daniel and Moreau 2004). However, this type of technology needs extensive refined sound system installations.

For this reason, the most widely adopted system for the reproduction of three-dimensional content is binaural audio using head-related transfer functions (HRTFs), which can be appreciated by any listener simply with a pair of headphones (Nicol et al. 2014). The directionality of a sound source concerning

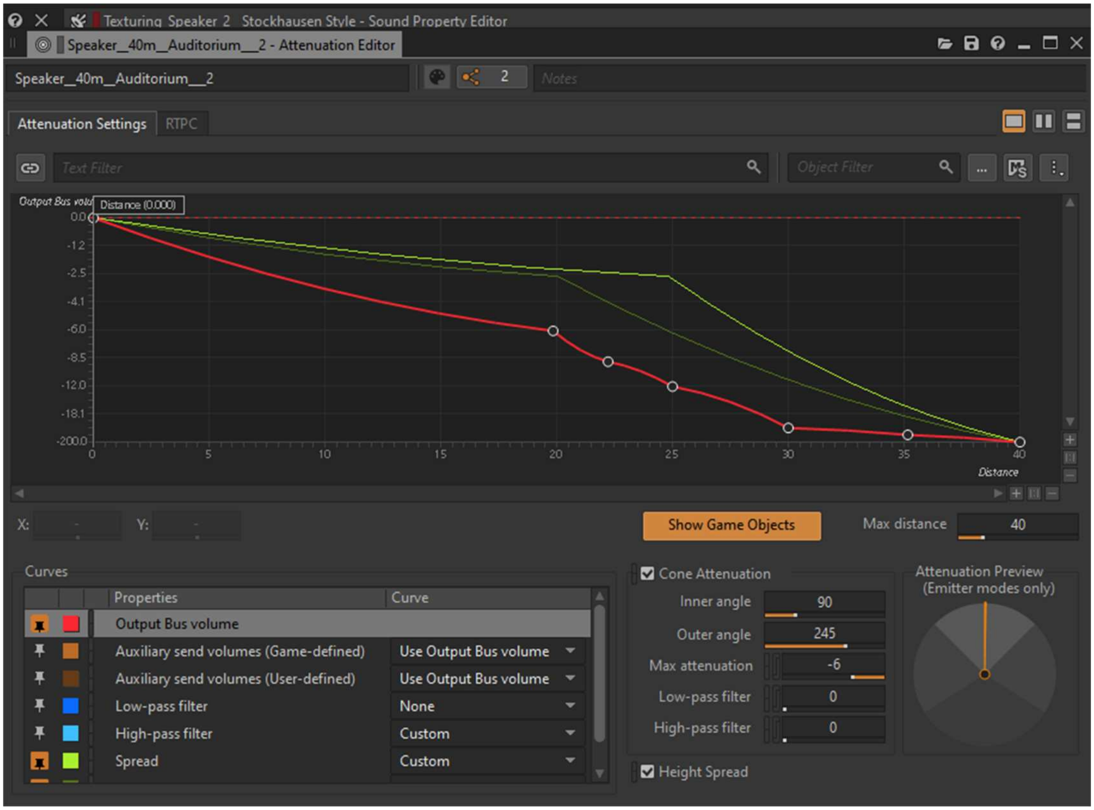

Figure 4. A screenshot of one of the different attenuation curves and angles. Each curve has a different colour because it refers to a different parameter, as in this example where the red curve is related to the amplitude of the audio signal (output bus volume) and the green curve to the spread. In addition, the cone attenuation, which the composer can fully configure, is displayed at the bottom right of this figure.

the listener's eardrum depends on a set of HRTFs, which encodes acoustic information used by humans to localize sounds in space and is essential in the perception of sound localization in headphones (Kulkarni, Isabelle, and Colburn 1999). We used the virtual Ambisonics approach to decode Ambisonics signals in virtual loudspeakers. The convolution of these loudspeaker signals with HRTF creates left and right ear signals, allowing binaural reproduction (Noisternig et al. 2003). We utilized a virtual Ambisonics approach to encode all the monophonic signals we employed to create our piece with third-order Ambisonics. Then, we decoded them for binaural reproduction using the Atmoky Ears binaural plugin within Wwise.

5.2.1. Implementation in Wwise

Our VR system utilized each sound we encoded in Ambisonics by managing the 'Positioning' property in Wwise, enabling 'Listener Relative Routing' and '3D Spatialization', and setting the latter into the 'Positioning plus Orientation' configuration, which is the action of panning each sound based on the relative user's position and the different sound emitters' orientations. The Meta Quest 2's head-tracking provided the relative user's head position in real-time, which was sent from Unity to Wwise to manage spatial sound perceptions. This enabled users to hear sounds in tridimensional surround as they move their heads. Moreover, we designed 3D sound trajectories for certain sound objects in the auditorium's composition using the automation options available in Wwise's 'Position Editor (3D Automation)'. The 'Position Editor' enables us to write and automate sound trajectories within a classic surround panner interface. Finally, each sound was routed to a binaural bus, where the Atmoky Ears binaural plugin was inserted, allowing for the rendering of all sounds in binaural. With this setup, every time the user moves in any direction and moves the head, s/he has the perception of hearing the sounds in a three-dimensional way.

5.3. The use of convolution reverbs

In our system, we utilized 1st-order Ambisonics convolution reverbs, which are reverberations created by algorithms where a dry audio signal is convolved with an Ambisonics room impulse response. We chose to use 1st-order Ambisonics convolution reverbs because it has been shown that there is no perceivable advantage in using those with higher orders (Engel et al. 2019). We used two convolution reverbs: one for the foyer and one for the auditorium. For example, when a user goes from the foyer to the auditorium, the user hears a different reverberation that simulates the passage from one virtual place to another. In detail, we utilized the 'Darkside-50' convolution reverb (Adriensen and Monacchi 2013) for the Foyer. We used the 'Hamilton Mausoleum' impulse responses (Murphy 2006) for the central auditorium part. The choice of reverberation used for our study was based on our own aesthetic choice after trying different types of convolution reverbs.

5.4. Linking musical expressiveness with user locomotion in Wwise and Unity

To link user locomotion with our composition approach, we used AkEvents within Unity, which connects Unity's events with Wwise's sound events. We implemented invisible collision spheres that trigger real-time sound events when the user enters. These spheres, adjustable in size, send control messages for attenuation curves, angles, reverberations, and spatial audio techniques, as described in Section 5.2. Each AkEvent trigger uses the AkTriggerCollisionEnter function, ensuring Unity triggers specific sounds and manages the described sound parameters to Wwise. The naming of AkEvents and their corresponding sound events in Wwise must match to facilitate this communication.

We called these collision spheres *composition spheres* (see Figure 5) and worked with them as follows:

1. In the foyer, we created a small size of all the *composition spheres*. A specific simple or complex sound is triggered every time the user enters these spheres. We created several random containers in Wwise, which allowed us to continuously vary the different sounds triggered by the user so that the user avoids the perception of a loop of the same sound. A random container is a function that allows for continuous variation of the different sounds played. To avoid the loop perception by the user in our sound game, we composed an extensive sound bank with many simple and complex sounds;
2. In the auditorium, we created a larger size of all *composition spheres*, as seen in Figure 5. We placed each composition sphere in each loudspeaker, matching precisely the sphere's center with the sound emission zone of the loudspeaker, which corresponds precisely with the loudspeaker's cone. Each time the user enters the sphere and approaches a loudspeaker, the user hears more of the internal and less of the global composition. On the other hand, if the user moves away from a loudspeaker, s/he remains in an intersection zone between different spheres. Thus, s/he hears the sum of the sound objects produced by individual loudspeakers, as seen in Figure 3.

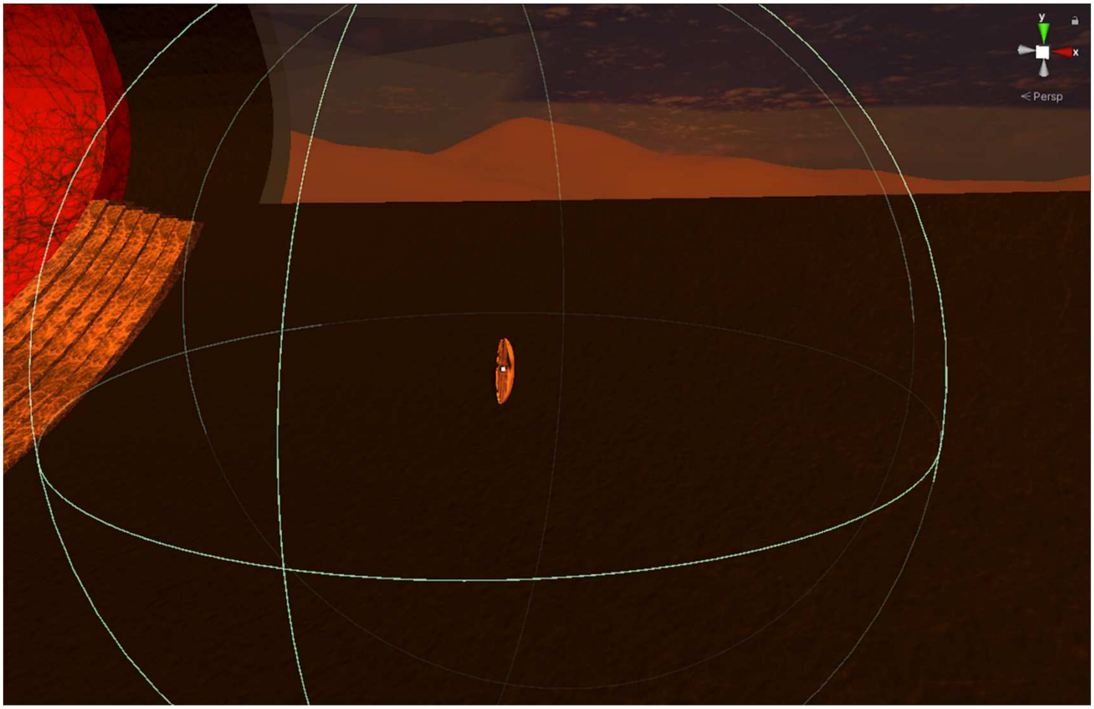

Figure 5. Screenshot of one of the *composition spheres* in one of the virtual loudspeakers located in the main auditorium area.

This approach allowed us to create an interactive composition based on user locomotion. In addition, thanks to the distance and cone attenuations, each user hears the sounds differently because these curve and angle attenuations control different parameters, allowing us to create an ever-changing composition experience in terms of timbre.

5.5. Replicability of our approach in other works

In this section, we describe how a composer or sound designer interested in our approach can replicate it with their sounds and music and how they can make changes to the parameters relating to the attenuation curves to make it customizable based on needs and aesthetic choices, the type of music and sounds used.

The Wwise and Unity projects are made available in an online repository[2], where it is possible to download and install them following a set of step-by-step instructions. Interested practitioners can adopt our approach to create their pieces inside the nautilus-shaped VR auditorium. Below, we try to summarize the fundamental steps for correct replication:

- If the user wants to adopt our approach as presented in this paper but would only like to upload the personal sounds instead of using the ones we composed for the creation of our piece, the user just needs to replace the monophonic files in Wwise and check them in Unity by verify the AkEvents in Unity (regarding the different *composition spheres*) and the sound events in Wwise. Each Wwise sound event containing each sound is linked to Unity via AkEvents, and the names of the Wwise sound events and the Unity AkEvents names are the same.This means that it is enough to find the same name in Unity to understand which sound event and which sound file will be triggered when the user passes through the *composition sphere*. Regarding the foyer, however, if the user wants to use the random container, as we utilized it in our approach, simply replace the sounds with ours within the Wwise project alone, and verify the same concept described above of having the same names in Unity and Wwise.

- Regarding the attenuation curves and the parameters used, the user is required to modify each single AkEvent of the project and open the 'Attenuation Settings' in Wwise, as shown in Figure 4. By opening that window, the user can play and modify all the parameters based on their aesthetic and can do this in all the events created in Wwise that communicate with Unity.

However, suppose a user wants to modify the size of the *composition spheres* or change the settings relating to the XR Interaction Toolkit. In that case, the user must accomplish these actions manually in the Unity editor.

6. Evaluation

In this Section, we describe how we carried out the evaluation phase to compare the three locomotion techniques utilized and collect feedback and insights from composers regarding their interaction and our compositional approach.

6.1. Apparatus and Participants

Fifteen composers participated in the evaluation (11 males and 4 females, aged between 23 and 48, mean age = 32, standard deviation = 6.7). Participants were recruited via a mailing list of acousmatic composers interested in VR. Eligibility criteria consisted of having (i) over five years of experience in acousmatic composition; (ii) previous experience with at least five Musical XR applications using handheld controllers; and (iii) right-handed preference. Recruitment also depended on the availability of the necessary experimental setup. All recruited participants were of Italian nationality. We decided on these eligibility criteria to ensure that all participants were expert acousmatic music composers and had experience with previous Musical XR projects. This was necessary to obtain evaluations aligned with the scope of this work, as we expected that participants' expertise and familiarity with the medium would have provided more relevant and insightful feedback.

Participants took, on average, one hour to complete the experiment. The approved procedure, authorized by the local ethics committee, adhered to the ethical standards outlined in the 1964 Declaration of Helsinki. The experiment was conducted online via a Zoom video call for each participant. To perform it online, we prepared 3 different Desktop-VR applications (one for each technique), sent them, and asked each participant to download them before starting the experiment and verify that the applications ran with 60 frame-per-seconds on their computer's desktop. Being a Desktop-VR application, participants connected to their computer's desktop via the Meta Quest Link cable. We asked each participant to use their studio's headphones and their own Meta Quest 2 HMD and to remain seated during the experimental phase.

6.2. Procedure

The evaluation procedure consisted of the following steps. First, participants were briefed about the experiment and signed a consent form. Afterward, we explained how to use the three techniques with the handheld Meta Quest 2 controllers and showed some photos (screenshots) of the system, particularly the auditorium. The experiment aimed to investigate the composers' preferred locomotion technique for exploring the interactive composition we developed and to understand the rationale behind their preferences. It also aimed to collect feedback and insights from composers regarding their interaction and experience with our system. The experimental phase consisted of trying while sitting, the VR system and the composition for 4 minutes for each technique, for a total of three trials (one with the *continuous* locomotion technique, one with the *discrete* locomotion technique, and one with the *mixed* locomotion technique).

Furthermore, during the pilot study, we decided to start the experiment directly from the beginning of the part of the auditorium where our composition begins, thus avoiding the sound game we developed in the foyer. Each trial used the first 30 s as a familiarization phase to allow the composer to adapt to our tested techniques. After the first 30 s of the familiarization phase, each participant had to explore our musical piece in VR for 3.30 minutes. This procedure was conducted for each of the 3 conditions (i.e. for the 3 different techniques tested). Neither during the familiarization phase nor the experiment participants were informed about the study's purpose. The 3 conditions were presented in randomized order across participants. Before starting each trial, participants were asked to complete the Cybersickness in Virtual Reality (CSQ-VR) questionnaire (Kourtesis et al. 2023) reported in Table 1, which is based on the items to be assessed on a 7-point Likert scale:

After each trial, participants were asked to fill out the CSQ-VR again to analyze whether there were differences regarding cybersickness before and after each trial. Moreover, after each trial, participants were invited to fill in an administered ad-hoc questionnaire formed of the items reported in Table 2 to be assessed on 11-point Likert scales.

After completing all trials and questionnaires, participants underwent a semistructured interview involving the open-ended questions reported in Table 3.

Table 1. Questions of the CSQ-VR questionnaire.

1) Nausea A	Do you experience nausea (e.g. stomach pain, acid reflux, or tension to vomit)?
2) Nausea B	Do you experience dizziness (e.g. light-headedness or spinning feeling)?
3) Vestibular A	Do you experience disorientation (e.g. spatial confusion or vertigo)?
4) Vestibular B	Do you experience postural instability (i.e. imbalance)?
5) Oculomotor A	Do you experience a visually induced fatigue (e.g. feeling of tiredness or sleepiness)?
6) Oculomotor B	Do you experience a visually induced discomfort (e.g. eyestrain, blurred vision, or headache)?

Table 2. Questions of the ad-hoc questionnaire.

1) Agency clarity	How clear was it that it was you who turned on (triggered) the different sounds you heard?	Not at all, very clear
2) Agency control	How in control did you feel in this aspect?	Not at all, very in control
3) Agency parameters	How clear was it that you also controlled some parameters? (e.g. amplitude, filtering) of the different sounds?	Not at all, very clear
4) Utility	How useful is this technique for the exploratory purposes of interactive composition?	Not at all, very useful
5) Localization	How precisely were you able to localize sounds in space?	Not at all, very able
6) Enjoyment	How much did you enjoy this experience?	Not at all, very much

Table 3. Open-ended questions.

1) Which of the three locomotion techniques did you prefer? Why?
2) What did you like about this system? Why?
3) What did you not like about this system? Why?
4) How would you improve the system?
5) Did you realize that the spatialization and the sound trajectories depended on the movement of your head?
6) What are the added values of this system? If there are some, please list them.
7) What are the limits of this system? If there are some, please list them.
8) What was the overall experience of using this system? (e.g. relaxing, boring)? Why?
9) Would you use this system to create interactive composition experiences using locomotion alone? Why?
10) Is locomotion enough as a form of interaction with sound to create interactive compositions in VR environments?

7. Results

7.1. Results of the CSQ-VR Questionnaire

The results of the responses to the CSQ-VR questionnaire questions are reported in Figure 6. An ANOVA was performed on generalized linear mixed-effects models created for each response variable. Each model incorporated the response variable ('Nausea A', 'Nausea B', 'Vestibular A', 'Vestibular B', 'Oculomotor A', and 'Oculomotor B') and the condition (*continuous*, *discrete*, *mixed* techniques) as fixed factors while treating the subject as a random factor. For each model, the assumption on the normality of residuals was verified. Post-hoc tests were performed on the fitted model using pairwise comparisons adjusted with the Tukey correction. No significant main effect emerged from these tests.

7.2. Results of the ad-hoc questionnaire

The results of the responses to the administered ad-hoc questionnaire are reported in Figure 7. An ANOVA was performed on generalized linear mixed-effect models created for each response variable. Each model incorporated the response variable ('Agency clarity', 'Agency control', 'Agency parameters', 'Utility', 'Localization', and 'Enjoyment') and the condition (*continuous*, *discrete*, *mixed* techniques) as fixed factors while treating the subject as a random factor. For each model, the assumption that the residuals follow a normal distribution was thoroughly assessed. Subsequently, post-hoc tests were performed on the fitted model and were executed through pairwise comparisons, with adjustments made using Tukey's correction.

Regarding 'Agency clarity', we found a significant main effect between the different responses among the techniques ($\chi^2 = 6.8904$, $df = 2$, $p = 0.0319$). Post-hoc tests indicated significant differences between *discrete* and *mixed* techniques (respectively, $p < 0.005$, $p =$ 0.0429).

Regarding 'Agency control', we found a significant main effect among the different

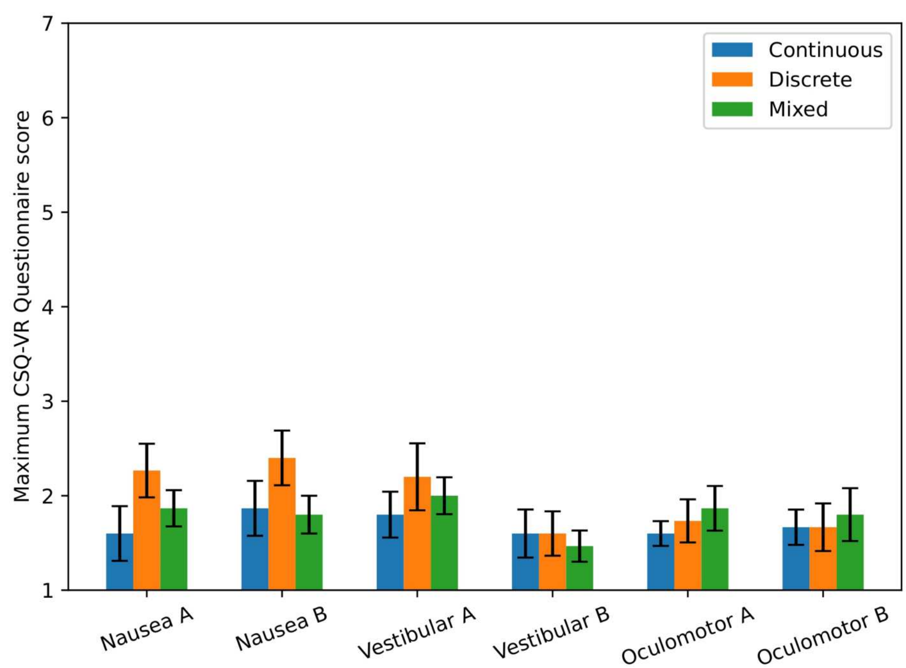

Figure 6. Mean and standard error for each technique concerning the individual questions of the CSQ-VR Questionnaire.

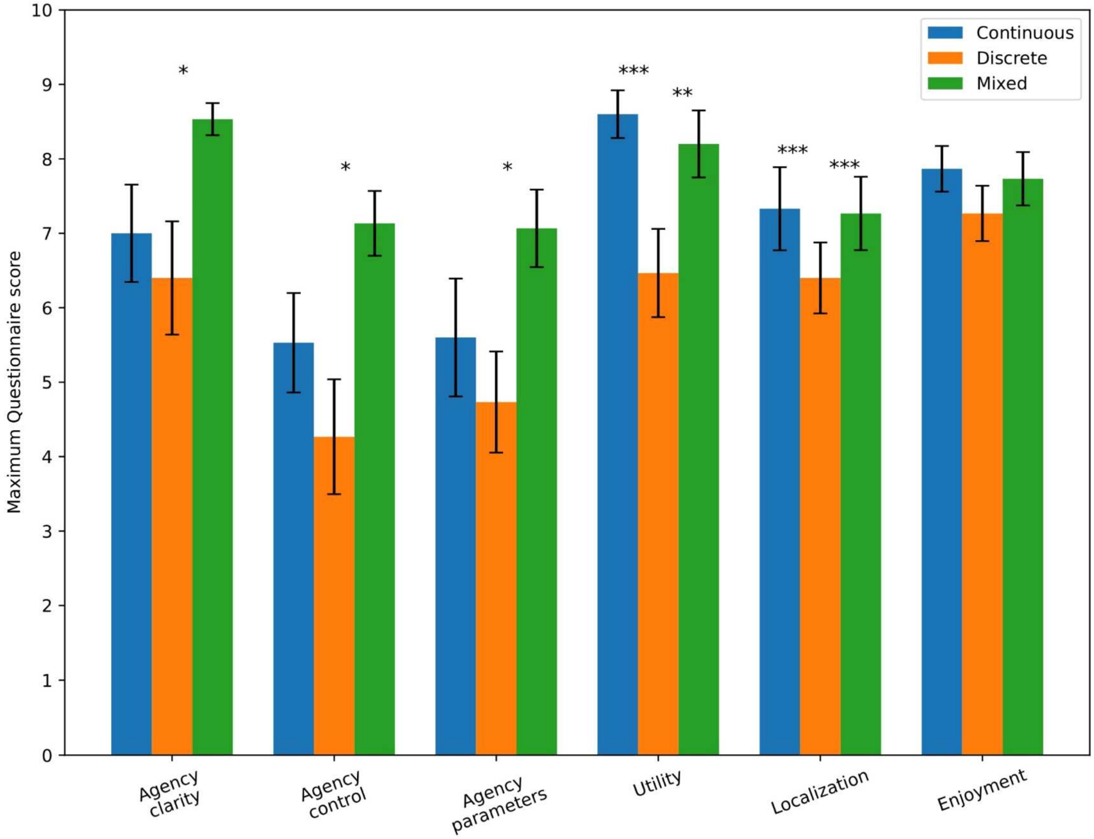

Figure 7. Mean and standard error for each technique concerning the individual questions of the administered ad-hoc questionnaire. Legend: * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

techniques' responses ($\chi^2 = 10.063$, $df = 2$, $p = 0.006529$). Post-hoc tests indicated significant differences between *discrete* and *mixed* techniques (respectively, $p < 0.005$, $p = 0.0101$).

Regarding 'Agency parameters', we found a significant main effect between the different responses among the techniques ($\chi^2 = 6.1401$, $df = 2$, $p = 0.04642$). Post-hoc tests indicated significant differences between *discrete* and *mixed* techniques (respectively, $p < 0.005$, $p = 0.0526$). For the result of that particular question, the main effect is roughly around the $p = 0.005$ limit. Nonetheless, we have considered it and reported it in the figure, as it aligns with the results from the open-ended questions described in Section 7.3.

Regarding 'Utility', we found a significant main effect among the different techniques' responses ($\chi^2 = 19.032$, $df = 2$, $p = 7.366\text{e-}05$). Post-hoc tests indicated significant differences between *continuous* and *discrete* techniques (respectively, $p < 0.001$, $p = 0.0009$) and between *discrete* and *mixed* techniques (respectively, $p < 0.001$, $p = 0.0066$).

Regarding 'Localization', we found a significant main effect between the different responses among the techniques ($\chi^2 = 19.191$, $df = 2$, $p = 6.803\text{e-}05$). Post-hoc tests indicated significant differences between *continuous* and *discrete* techniques (respectively, $p < 0.001$, $p = 0.0014$) and between *discrete* and *mixed* techniques (respectively, $p < 0.001$, $p = 0.003$). No significant main effect was found for the 'Enjoyment' dimension.

7.3. Results of the open-ended questions

We collected the responses provided by composers to the open-ended questions and analyzed them with a reflexive thematic analysis (Braun and Clarke 2006). We conducted such an analysis by generating codes and then organized them into themes that reflected patterns.

7.3.1. Themes regarding the continuous technique

Preference. Six participants reported that they prefer to use the *continuous* technique to explore an interactive composition and to use it for the creation of their composition works (e.g. 'The gradual technique (continuous) allowed me to enjoy the composition and the experience, because with that technique I heard all the variations more gradual, and therefore I liked it better from a musical's perspective'.; 'I preferred the continuous technique because I was able to understand the management of sound parameters (filters, etc.) much better, and therefore, I could concentrate more on the composition and related sound design. I would use it for my work'.).

Open-form composition feature. Four participants emphasized that the composition is open-form only with this technique and found this feature very interesting from an experiential point of view (e.g. 'I noticed that I enjoyed the open-form composition feature more with the continuous technique. I decided whether to hear a part endlessly, and I could also modify it by moving in space. This feature is the one that interested me the most'.; 'With the continuous technique, I could enjoy the characteristic of the composition without a beginning and an end because it was I who gradually decided what to listen to'.).

Spatialization. Eight participants reported that they still preferred this technique for the aim of sound localization because it is the technique that allowed them to localize the sounds more precise in the space, being the most gradual (e.g. 'Even if the sound in the space seemed a bit too diffuse, I found it easier to localize them better with the continuous technique, perhaps due to its characteristic of making one listen to everything more gradually, based on the space's locomotion'.; 'With the continuous technique I have sometimes managed to hear much more precisely the origin of a sound from a loudspeaker. This feature also made me feel more genuinely present in the virtual environment'.).

7.3.2. Themes regarding the discrete technique

Preference. Two participants prefer the *discrete* technique for exploring and composing in a VE (e.g. 'The discrete technique made me understand more quickly that based on where I was, I turned on a different sound, and consequently, if I liked it, I stayed still and listened to it. Otherwise, I moved forward more quickly'.; 'I preferred the discrete technique because in the continuous movement, also because it was slow, I wasn't very aware of triggering the sounds. I didn't feel like I was influencing the sounds and dynamics! What I liked about the discrete technique was the speed; in the experience, I gave a lot of importance to speed!').

Immediacy and jerkiness as compositional elements. Three participants liked that this technique allows one to move much faster in the VE and, therefore, reported that even if it does not allow gradual listening, they would use it for their compositions, using immediacy and jerkiness as compositional elements (e.g. 'On a compositional level, I would choose the discrete technique because the immediacy of the shot is more useful to me if I were to compose something, while the gradual technique is more advantageous for exploring'.; 'I would use the jerkiness given by this technique as a compositional element, in the sense that I would match each jerk with a specific gesture or a specific musical passage'.).

7.3.3. Themes regarding the mixed *technique*

Preference. Seven participants reported that they prefer to use the *mixed* technique to explore and create their interactive compositions (e.g. 'The mixed technique convinced me the most because sometimes I like to experience a gradual musical situation (like the one I experienced with the continuous technique), sometimes, instead, I like to be faster and more snappy, and for this reason, I use the discrete technique'.; 'If I had to choose, I would use the technique where I can move (walk) and where I can also teleport because teleportation is very useful for moving from one scene to another in a very short time, or if I want to feel an effect faster'.).

Sound organization feature. Five participants referred to us that this technique has numerous advantages in exploring a musical piece and in organizing the sounds in a composition (e.g. 'This technique convinced me because I found it new, in the sense that it gave me the possibility to choose, to move quickly or not, to direct the composition, to return to a sound if I liked it'.; 'With this technique, I have much more control over the sound organization because I can stop, listen, and walk around a loudspeaker for as long as I want, but simultaneously, I can teleport myself to another place and completely change the musical situation'.).

'**Play the space' feature.** Five participants reported that the mixed technique is the one that best allows one to 'play the space' because it allows the most excellent flexibility of movements in the VE (e.g. 'What I liked about this technique is the fact of playing space, playing with space and seeing it as an exploratory and compositional element. I could move slowly, play a loudspeaker, then teleport to the window and hear the filtered sound. This feature is unique to this technique'.; 'With this technique, I felt an integral part of the space; I had the feeling and a sensation that I was playing the space myself, determining it'.).

7.3.4. General themes

More musical interactions. Six participants reported that locomotion as a form of interaction to create their musical compositions is insufficient. They considered it interesting to explore compositions in VEs. Still, according to them, since VR is used, new gestural interactions with sound can be introduced (e.g. 'I would add interactions with objects to make the experience more alive. I would like to have tools with which I can actually modify the sounds in real-time, or I can draw them in space as if I were playing a musical instrument within the soundscape'.; 'If I have the possibility of touching sound objects or seeing animations or visual reactions, the experience would undoubtedly be more engaging, and I would also have more opportunities to use these elements in a composition'.).

Lack of animations and visuals. Ten participants reported that the visual part is too static. For this reason, they would like some visual feedback in the VE (e.g. 'I would greatly improve the visual component because everything is too dull and therefore after a while the experience becomes boring'.; 'Without the visual reaction, we cannot speak of an audiovisual composition for this experience. As it is, it might have the risk of being boring. I would work a lot on the visual aspect. For example, it would be enough to show that the loudspeaker cone vibrates when a sound is reproduced'.).

The nautilus-shaped auditorium design adds value. Nine participants reported that what they noticed as an added value of the system was the design of the VE itself and the related listening system made up of this sort of spatially distributed virtual loudspeaker system (e.g. 'The thing that impressed me most about the system was the design of the virtual environment itself. I really enjoyed moving between these acousmonium-style loudspeakers inside this completely virtual auditorium. It impressed and involved me a lot, and I would use it for many other purposes'.;

'Compared to other VR experiences that seemed stylistically much closer to a video game with poorly defined textures, I found great definition in this auditorium. The composition I explored made me feel part of a magical system I had never experienced. I would expand this auditorium and listening system with other activities, such as organizing ambient concerts in VR!'.)

Site-specific composition. Eight participants reported that it is necessary to address this approach as one addresses compositions made especially for a particular place (site-specific) because, in the act of composing, one must think about how it will be experienced and, therefore, the organization of the sound is not linear or canonical enough about the timeline. Timeline, a characteristic element of music, no longer has meaning in this type of experience (e.g. 'If a composer wants to insert a strong gesture, such as a climax, one must keep in mind that the user triggers that gesture and therefore, it is not like in linear compositions, where the composer controls the time. The user has control of the tempo in these types of compositions. The composer must consider the composition when thinking about a site-specific work and how the work will be experienced by the user in that particular space'.; 'This type of composition refers to a particular place and way of experiencing. A composer must, therefore, think about how to organize the space, the place, and the organization of sounds takes on a completely different meaning'.).

8. Discussion and Conclusion

Overall, our results suggest that the *mixed* technique is the one most preferred by the participants because it gives them more control over the musical parameters and, at the same time, allows them to experience musical composition gradually. It is most preferred compared to the *continuous* technique because, through the use of teleporting, a user can also move quickly within the VE, allowing greater agility in choosing where to move. These conclusions are supported by the results of the conducted thematic analysis. We believe that this result is fundamental, especially in those VR environments that could be much larger (in terms of size) than ours that we have proposed. Nevertheless, the *mixed* and *continuous* locomotion techniques are much preferred in comparison with the *discrete* technique, which would seem not to be an ideal technique for musical purposes because it does not allow for the perception of gradual musical listening.

Regarding 'agency clarity', a statistically significant difference emerged between the *discrete* and *mixed* techniques, which means that participants noticed more with the *mixed* technique that they were themselves triggering the different sound objects with locomotion. From our perspective, this technique made the users more cautious and attentive to the composition's events, and this meant that the users realized that they were triggering the different sound objects in the composition.

Regarding 'agency control', the same significant difference emerged between the *discrete* and *mixed* techniques. The participants seemed to feel more in control over the triggering aspect of the sounds with the *mixed* and *continuous* techniques compared to the *discrete* techniques. In line with what emerged from the qualitative results, it seems that these techniques allow users to gradually perceive the variations in the music (such as the fact of triggering sounds). Therefore, these techniques allow users to better notice all the changes in the sound. The *mixed* technique is probably preferred for this aspect because it also allows for greater flexibility and agility in moving within the VE.

Also, regarding the 'agency parameters', the same significant difference emerged between the *discrete* and *mixed* techniques. From our point of view, for this dimension, users realized that they were not only triggering the different sound objects but also controlling the different parameters because the *mixed* technique allowed them to gradually perceive the variation in the sound parameters, compared to the *discrete* technique. Furthermore, it can be seen from Figure 7 that participants noticed this variation more with the *continuous* technique compared to the *discrete* technique. This is because both techniques (*mixed* and *continuous*) allowed them to gradually experience the variation of the parameters of the musical composition. Still, the *mixed* technique was preferred because it allows for considerable variations to be experienced when teleporting is also used.

Regarding 'utility', however, it would seem that participants prefer the *continuous* technique over the others. However, both techniques (*continuous* and *mixed*) show a statistically significant difference compared to the *discrete* technique. This preference does not seem to precisely align with what emerged from the qualitative results, where the *mixed* technique was the most preferred. However, this locomotion technique seems to be the most suitable and stable for gradually experiencing the musical piece we created and experiencing all the nuances and details of the sounds, such as the variation of parameters. Furthermore, it would seem to be the technique composers would use to create their interactive compositional works.

Regarding 'localization', which means the precision in localizing sounds in space, it is evident that *continuous* and *mixed* techniques are preferred and show statistically significant differences compared to the ability to perceive sounds in space with the *discrete* technique. This preference is because *continuous* techniques allow spatialization to be experienced more precisely than *discrete* techniques, where instead, a user must sprint (teleport) from one point of the VE to another to move, creating a sort of distraction in the sound perception. Unfortunately, with the *discrete* technique, these peculiarities of the sound are lost due to the nature of the technique itself. At the same time, these perceptive characteristics are preserved with the gradual movement in the VE.

As regards the 'enjoyment' dimension, no statistically significant differences emerged between the various techniques, and in general, the experience seems to have been enjoyed. However, it would seem that the fact that the visual part remains static throughout the experience could make our system boring if it is experienced more than the time shared in the experiment. This consideration is very relevant and requires future investigations because there is a need for more research in investigating how much the visual part impacts the quality of the experience even in VR systems more linked and developed around sound and composition.

Regarding Cybersickness, the quantitative results indicate that participants suffered more with the *discrete* technique than with the other two. However, no statistically significant results emerged between the techniques. Furthermore, from the quantitative results of the CSQ-VR Questionnaire it emerged that the *continuous* technique led to the smallest amount of Cybersickness, although very close to the scores of the *mixed* technique. Overall, the total Cybersickness score that emerged from the CSQ-VR Questionnaire is meager, indicating that the participants did not suffer from Cybersickness so much as to compromise the experience. More studies should be carried out on the impact of Cybersickness while standing, given that our experiment was carried out while sitting.

Below we propose critical reflections regarding our compositional approach derived from analyzing all the participants' quotes and our critical observations regarding the usage of our system. The reflections concern the proposed approach, its overall structure and open-form, some aesthetic choices made in creating the piece, and the use of locomotion in this approach.

- **Interaction and composition**: Basic interactions in our VR system, like user locomotion, can lead to complex sound organization within the composition and allow participants to experience the composition's open-form characteristic, as reported from some quotes of our participants in the thematic analysis. However, in our opinion, the sole usage of locomotion could limit musical expressiveness and user-controlled musical interaction. Future developments could include functionalities allowing direct control over the composition's form (e.g. introducing silence or musical crescendos via handheld controller buttons).
- **AkEvent and sound events generation**: In our opinion and also according to some participants, the use of *composition spheres* and AkEvent in Unity and Wwise, together with parameterized attenuations and random sound event generators within our approach, have allowed us to create dynamic musical situations, making the experience different each time.
- **Parameters**: During the realization of the piece utilizing our proposed approach, we noticed that critical parameters like amplitude attenuation, filtering, and convolution reverbs are pivotal and facilitate the musical dialogue between sound objects.
- **Binaural sound localization and reverberations**: In our opinion, implementing virtual Ambisonics and binaural decoding, coupled with the head-tracking system, enhances performance and realism in the VR environment, as it is possible to visualize in the ad-hoc questionnaire results for all three locomotion techniques. However, from our point of view, in creating the piece, we mixed particular sound objects with too much convolution reverbs, giving it excessive significance to the point of creating too much sound diffusion, as reported in the thematic analysis. This aspect created challenges in accurately localizing the sound sources in the headphones. Moreover, while crucial, convolution reverbs lack parameter control and result in uniform sounds. Using algorithmic reverbs with parameter control based on VE textures and topology might improve the localization. Furthermore, we believe that, due to the complexity of the sound objects we have used, a composer must try to simplify the sound material and avoid timbral homogeneity to improve localization accuracy. However, this aspect regarding timbral complexity and localization of sounds in space in VR needs more research.

It is worth noticing that this study presents some limitations. Firstly, the majority of participants were Italian, and most of them were male. Including a broader group of composers from various countries and cultures, along with a more balanced gender distribution, would enhance the generalizability of our findings. Secondly, the confounding variables of this study, which are the walking speed and the length of the raycast, were not accounted for because their values were solely based on the outcomes of the preliminary pilot study we conducted. Consequently, the final results might vary if users were given the freedom to control these variables. Notably, the interaction radius of 15 meters might seem substantial. However, this radius is relative to the VR environment's size/scale, and the settings tested in the preliminary pilot study were set based on the composition and design of our VE. This choice about the radius affected the intensity of cybersickness and general confusion of the user. The same concept can be applied to the chosen movement speed and the absence of continuous camera rotation, coupled with the relatively low exposure time. All of these aspects potentially significantly reduced the motion-induced effect of nausea. Consequently, the preferences we identified for mixed locomotion and cybersickness may only be applicable to this specific implementation and study. Further research is needed on these aspects to enhance the generalizability of our results.

However, it would seem to emerge from our study that the technique that mixes *continuous* and *discrete* locomotion is the most appropriate for creating musical environments linked to the composition. Nevertheless, it is also essential to present that this result is significantly related to our study and our design of the VE. A plausible reason is that it allows one to experience a composition gradually and, therefore, to listen to all the particularities in the variations that happen to the sound, but at the same time, it allows for greater flexibility and agility in moving quickly within the VE thanks to teleporting.

Moreover, to answer the first research question, it would seem that, also from what emerged from the qualitative results, the design of our system seemed to be liked by participants as they identified the aesthetic of the nautilus-shaped auditorium and the sort of spatially distributed virtual loudspeaker system as an added value of the experience. Furthermore, several participants defined our approach as very similar to what happens with site-specific works. However, some reported the need to have more interactions to create a greater sense of immersion and involvement with VR technology, but these aspects need further research.

Notes

1. https://youtu.be/Py9x1LRxsgk
2. https://drive.google.com/drive/folders/1BCEp0YtVcVFKIeix5hyJVhtqzZ5ZCcdR?usp=sharing

Acknowledgments

We acknowledge the support of the MUR PNRR PRIN 2022 grant, prot. n. 2022CZWWKP, funded by Next Generation EU. Furthermore, for realizing this work, we would like to thank Prof. Valerio Murat, Prof. Franz Rosati, and Prof. Alessandro Cipriani of the *Centro di Ricerca ed Elaborazione Audiovisiva* (CREA) of the 'Licinio Refice' Conservatory of Frosinone (Italy), who supported us by giving us constant feedback regarding the technical and artistic design of the proposed system. We also thank Gianluca Rotondi for the technical support in creating the 3D model of the nautilus-shaped auditorium and Axel Ferrari for the technical help in the 3D modeling of the auditorium's loudspeakers. We also thank Dr. Alberto Boem for the support provided during the different phases of the research.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by MUR PNRR PRIN 2022 [grant, prot. n. 2022CZWWKP], funded by Next Generation EU.

Notes on contributors

Matteo Tomasetti is a PhD candidate at the Department of Information Engineering and Computer Science of the University of Trento. He received his B.A. in Electronic Music at the Conservatory of Pesaro and completed his M.A. in 3D Audio, Virtual Reality, and Audiovisual Composition at the Conservatory of Frosinone and the IEM (*Institut für Elektronische Musik und Akustik*) in Graz (AT). His research interests are extended reality, music composition, immersive audio, new interfaces for musical expression, and human–computer interaction.

Luca Turchet is an Associate Professor at the Department of Information Engineering and Computer Science of the University of Trento. He received his Ph.D. from Aalborg University Copenhagen. His scientific, artistic, and entrepreneurial research has been supported by institutions and funding agencies such as the EU Commission and ESA. He is co-founder of Elk and is Chair of the IEEE Emerging Technology Initiative on the Internet of Sounds. He serves as an associate editor for IEEE Access and the Journal of the Audio Engineering Society.

ORCID

Matteo Tomasetti http://orcid.org/0000-0001-5951-0302
Luca Turchet http://orcid.org/0000-0003-0711-8098

References

Adriensen, F., and D. Monacchi. 2013. "The Sound of Science Spaces." Accessed 15 February 2023. www.rwdobson.com/sspaces/sciencespaces.html.

Atherton, J., and G. Wang. 2020. "Doing vs. Being: A Philosophy of Design for Artful VR." *Journal of New Music Research* 49 (1): 35–59. https://doi.org/10.1080/09298215.2019.1705862.

Barri, T. 2009a. "Lought." Accessed 02 December 2023. http://tarikbarri.nl/projects/versum.

Barri, T. 2009b. "Versum: Audiovisual Composing in 3D." In *Re: New – Digital Arts Forum*. Copenhagen, Denmark, 18–21 May 2009.

Benson, B. E. 2003. *The Improvisation of Musical Dialogue: A Phenomenology of Music*. Wheaton College, Illinois: Cambridge University Press.

Berthaut, F. 2020. "3D Interaction Techniques for Musical Expression." *Journal of New Music Research* 49 (1): 60–72. https://doi.org/10.1080/09298215.2019.1706584.

Bozgeyikli, E., A. Raij, S. Katkoori, and R. Dubey. 2019. "Locomotion in Virtual Reality for Room Scale Tracked Areas." *International Journal of Human-Computer Studies* 122:38–49. https://doi.org/10.1016/j.ijhcs.2018.08.002.

Braun, V., and V. Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101. https://doi.org/10.1191/1478088706qp063oa.

Cahen, R., X. Rodet, and J. P. Lambert. 2005. "Study of Haptic and Visual Interaction for Sound and Music Control in the Phase Project." In *Proceedings of the 2005 International Conference on New Interfaces for Musical Expression (NIME05)*, Vancouver, BC, Canada.

Çamci, A. 2019. "Some Considerations on Creativity Support for VR Audio." In *IEEE Conference on Virtual Reality and 3D User Interfaces*, 1500–1502. Osaka: IEEE.

Costa, W., L. Ananias, L. Barbosa, B. Barbosa, A. De' Carli, R. R. Barioni, L. Figueiredo, V. Teichrieb, and D. Filgueira. 2019. "Songverse: A Music-Loop Authoring Tool Based on Virtual Reality." In *21st Symposium on Virtual and Augmented Reality (SVR)*, 216–222. Rio de Janeiro: IEEE.

Daniel, J., and S. Moreau. 2004. "Further Study of Sound Field Coding With Higher Order Ambisonics." In *116th Convention of the Audio Engineering Society*.

Engel, I., C. Henry, S. V. A. Garì, P. W. Robinson, D. Poirier-Quinot, and L. Picinali. 2019. "Perceptual Comparison of Ambisonics-Based Reverberation Methods in Binaural Listening." In *EAA Spatial Audio Signal Processing Symposium*, 121–126.

Frank, M., F. Zotter, and A. Sontacchi. 2015. "Producing 3D Audio in Ambisonics." In *Proceedings of AES International Conference: The Future of Audio Entertainment Technology –Cinema, Television and the Internet*, 1–8.

Freitag, S., D. Rausch, and T. Kuhlen. 2014. "Reorientation in Virtual Environments Using Interactive Portals." In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, 119–122. Minnesota: IEEE.

Hamilton, R. 2008. "Maps and Legends: Designing FPS-Based Interfaces for Multi-User Composition, Improvisation and Immersive Performance." In *4th International Symposium on Computer Music Modeling and Retrieval. Sense of Sounds*, Copenhagen, Denmark, 478–486.

Kermit-Canfield, E. 2016. "A Virtual Acousmonium for Transparent Loudspeaker Systems." In *Proceedings of the Sound and Music Computing Conference*, Hamburg, Germany, 233–237.

Kourtesis, P., J. Linnell, R. Amir, F. Argelaguet, and S. E. MacPherson. 2023. "Cybersickness in Virtual Reality Questionnaire (CSQ-VR): A Validation and Comparison Against SSQ and VRSQ." In *Virtual Worlds*, vol. 2, no. 1, pp. 16-35. MDPI.

Kulkarni, A., S. K. Isabelle, and H. S. Colburn. 1999. "Sensitivity of Human Subjects to Head-Related Transfer-Function Phase Spectra." *The Journal of the Acoustical Society of America* 105 (5): 2821–2840. https://doi.org/10.1121/1.426898.

Kwok, K. K. K., A. K. T. Ng, and H. Y. K. Lau. 2018. "Effect of Navigation Speed and VR Devices on Cybersickness." In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, Munich, Germany, 2018, 91–92.

Liu, J., H. Parekh, M. Al-Zayer, and E. Folmer. 2018. "Increasing Walking in VR using Redirected Teleportation." In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, 521–529.

Lohman, J., and L. Turchet. 2022. "Evaluating Cybersickness of Walking on an Omnidirectional Treadmill in Virtual Reality." *IEEE Transactions on Human-Machine Systems* 52 (4): 613–623. https://doi.org/10.1109/THMS.2022.3175407.

Mayor, J., L. Raya, and A. Sanchez. 2019. "A Comparative Study of Virtual Reality Methods of Interaction and Locomotion Based on

Presence, Cybersickness, and Usability." *IEEE Transactions on Emerging Topics in Computing* 9 (3): 1542–1553. https://doi.org/10.1109/TETC.2019.2915287.

Moore, A. G., M. J. Howell, A. W. Stiles, N. S. Herrera, and R. P. McMahan. 2015. "Wedge: A Musical Interface for Building and Playing Composition-Appropriate Immersive Environments." In *IEEE Symposium on 3D User Interfaces (3DUI)*, 205–206. Arles, France: IEEE.

Murphy, D. T. 2006. "Archaeological Acoustic Space Measurement for Convolution Reverberation and Auralization Applications." In *Proceedings of the 9th International Conference on Digital Effects*, 221–226.

Naef, M., and D. Collicot. 2006. "A VR Interface for Collaborative 3D Audio Performance." In *Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06)*, Paris, France.

Nicol, R., L. Gros, C. Colomes, M. Noisternig, O. Warusfel, H. Bahu, B. F. G. Katz, and L. S. Simon. 2014. "A Roadmap for Assessing the Quality of Experience of 3D Audio Binaural Rendering." In *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, 100–106.

Noisternig, M., T. Musil, A. Sontacchi, and R. Holdrich. 2003. "3D Binaural Sound Reproduction using a Virtual Ambisonic Approach." In *IEEE International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, 174–178.

Phillips, W. 2014. *A Composer's Guide to Game Music*. MIT Press.

Riecke, B. E., and D. Zielasko. 2021. "Continuous vs. Discontinuous (Teleport) Locomotion in VR: How Implications can Provide both Benefits and Disadvantages." In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 373–374.

Robinett, W., and R. Holloway. 1992. "Implementation of Flying, Scaling and Grabbing in Virtual Worlds." In *Proceedings of the 1992 Symposium on Interactive 3D Graphics*, 189–192. https://doi.org/10.1145/147156.147201.

Robinson, C. 2019. *Game Audio with FMOD and Unity*. Routledge.

Schütze, S., and A. Irwin-Schütze. 2018. *New Realities in Audio: A Practical Guide for VR, AR, MR and 360 Video*. CRC Press.

Seaborn, K., and D. I. Fels. 2015. "Gamification in Theory and Action: A Survey." *International Journal of Human-Computer Studies* 74:14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006.

Slater, M., M. Usoh, and A. Steed. 1995. "Taking Steps: The Influence of a Walking Technique on Presence in Virtual Reality." *ACM Transactions on Computer-Human Interaction (TOCHI)* 2 (3): 201–219. https://doi.org/10.1145/210079.210084.

Smalley, D. 1997. "Spectromorphology: Explaining Sound-Shapes." *Organised Sound* 2 (2): 107–126. https://doi.org/10.1017/S1355771897009059.

Steinicke, F., Y. Visell, J. Campos, and A. Lécuyer. 2013. *Human Walking in Virtual Environments*, vol. 56, no. 7, 976–985. New York: Springer.

Trenkamp, A. 1976. "The Concept of "Alea" in Boulez's "Constellation-Miroir." *Music & Letters* 57 (1): 1–10. https://doi.org/10.1093/ml/LVII.1.1.

Turchet, L., P. Burelli, and S. Serafin. 2012. "Haptic Feedback for Enhancing Realism of Walking Simulations." *IEEE Transactions on Haptics* 6 (1): 35–45. https://doi.org/10.1109/TOH.2012.51.

Turchet, L., R. Hamilton, and A. Çamci. 2021. "Music in Extended Realities." *IEEE Access* 9:15810–15832. https://doi.org/10.1109/ACCESS.2021.3052931.

Wozniewski, M., Z. Settel, and J. Cooperstock. 2006. "A Spatial Interface for Audio and Music Production." In *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 271-274.

Zayer Majed A. I., MacNeilage Paul, and Folmer Eelke. 2018. "Virtual Locomotion: A Survey." In *IEEE Transactions on Visualization and Computer Graphics*, 2315–2334.

Zdanowicz, G., and S. Bambrick. 2019. *The Game Audio Strategy Guide: A Practical Course*. Focal Press.