MASTER THESIS | TIME-DEPENDENT MEDIA: SOUND/VISION

Flow, intuitive interaction design and music production in an audio-first VR experience

submitted November 28, 2022

Anca-Stefania Tutescu
Matriculation Number [redacted]

First Supervisor: Prof. Dr. Eike Langbehn
Second Supervisor: Prof. Dr. Konstantina Orlandatou

HOCHSCHULE FÜR ANGEWANDTE WISSENSCHAFTEN HAMBURG
Department Medientechnik
Finkenau 35
22081 Hamburg

Zusammenfassung

Diese Abschlußarbeit beschreibt die künstlerische Forschung, die mit dem Entwurf und der Implementierung eines räumlichen Audioerlebnisses für einen einzelnen Benutzer in der virtuellen Realität verbunden ist. Die Anwendung wurde unter der Prämisse entwickelt, dass Audio die zentrale Rolle in einer virtuellen Erfahrung spielen kann und dass Audio-Interaktionen mit der Umgebung, die einer kollaborativen Musikkomposition gleichkommen, intuitiv sein können und dazu beitragen, dass der Nutzer in einen Zustand des Flusses und der Verspieltheit eintaucht. Die Erkenntnisse und Schlussfolgerungen dieser Forschung werden in diesem Beitrag vorgestellt, und es werden weitere Möglichkeiten und die Extrapolation auf Multi-User-Anwendungen untersucht.

Stichwörter: Virtual Reality; Sound Design; Crossmodale Korrespondenz; Verkörperung; Immersion

Abstract

This thesis outlines the artistic research involved in designing and implementing a single-user spatial audio experience in virtual reality. The application was designed with the premise that audio can carry the central role in a virtual experience, and that audio interactions with the environment, which amount to a collaborative music composition, can be intuitive and help immerse the user into a state of flow and playfulness. The learnings and conclusions of this research are presented in this paper, and further possibilities and extrapolation to multi-user applications are explored.

Keywords: Virtual Reality; Sound Design; Crossmodal Correspondence; Embodiment; Immersion

Contents

List of Figures

1 Introduction

"Endogeny" is a VR application developed between June 2020 and August 2021 as a cooperation between the HAW and HfMT Hamburg Universities with €30.000 in funding awarded by HamburgInnovation. The goal of this application was to explore and develop technical and aesthetic methods of music production for and in virtual worlds.

The project was supervised by Prof. Thomas Görne (HAW), Dr. Konstantina Orlandatou (HfMT) and Maik Helfrich (HAW). Counseling and technical support was provided by Doubleshot Audio (Made Indrayana and Benjamin Gallagher). The development team consisted of the following students:

- Anca-Stefania Tutescu (the author - Concept, Production, Programming, Sound / HAW)
- Jan Wegmann (Programming, Sound / HfMT)
- Charly Preisig (Programming / HAW)
- Artjom Fransen (Level Design, Tech Art / HAW)
- Valentin Fischer (3D Modeling, Tech Art, Programming / HAW)
- Maria Weninger (Concept Art, Concept / HAW)

One of the most interesting aspects of VR is the participant's ability to interact with the virtual environment from a first person's point of view. Endogeny was designed to give the participant multiple roles: explorer, audience, performer, and composer all in one. The space we created was an effort in designing and curating an intuitive experience which takes advantage of the specific affordances of the VR medium, where the participant would be compelled to explore, discover, interact, play and listen.

The experience was created for single player, but with the intention to make it easily expandable for multi-user participation.

Chalmers, 2018, as cited in Atherton and Wang, 2020, points to *virtual realism* - the idea that what we experience in a virtual reality is capable of as much meaning as a real life experience. The creators of such experiences are therefore encouraged to come up with poignant ways to use and enrich what is currently technologically achievable. What are the best approaches when it comes to designing such an experience that feels natural and engaging to the participant, and what are the possibilities of current VR technologies for a sound artist?

In an interview with Vinzent Preuß in 2022, who researches the cultural anthropology of audio-centered VR at The University of Freiburg, we summarized the main question that stands at the center of the aesthetic research surrounding my project: What are the properties of VR and the interactions within the medium that are timeless, that really bring value to the human experience once the novelty of the technology wears off?

2 Literature Review

In this chapter we take a look at the scope of an artistic research endeavor in general, scan available literature on the topic of VR, and investigate what constitutes achieving a state of flow in VR, what are the principles to follow when designing a VR experience with a focus on audio, how other artists have gone about implementing them, and how users may respond to them.

2.1 Aesthetic Research

Lüneburg, 2019 defines artistic research as the knowledge gain and communication performed through the arts, through the process of making and presenting artwork. In this process, the artists, researchers and audience all come to new insights and conclusions, which can be used as starting point for further research. On p. 130, the several types of artistic research are listed:

1. ***Arts-based inquiry***: *where the artistic process is used as research by artists, researchers and participants in order to understand the art itself or understand a phenomenon through the artistic process.*
2. ***Arts-informed inquiry*** *that is of two types:*
    a) *where art is used to represent the findings of a study*
    b) *where art is used to represent a response to the findings of an issue or situation studied.*
3. ***Arts-informing inquiry***: *where art is used in order to evoke a response from an audience (in the broadest sense) made to a situation or issue: the response may or may not be captured.*

Endogeny VR is a digital application that emerged from both an arts-based inquiry (aiming to understand what can be achieved with sound and interaction in VR from the artist's perspective) and from an arts-informing inquiry (whereby it was desired to provoke and investigate the response of the users and learn from their feedback).

2.2 The History of Virtual Reality

The concept of an artificial reality that attempts to envelop an individual's senses is not new, but dates back centuries. The first approaches to creating a virtual reality were large panoramic paintings where the viewer could feel they were actually experiencing from within the pictured scene (Bown et al., 2017 as cited by Rauschnabel et al., 2022).

According to Bown et al., 2017, panoramic paintings are an early example of VR technology, built in a time where film and television did not exist and photography was in its early days.

Woeste, 2009 states that panoramas first became known in 1787, when Robert Barker patented his plans for a cylindrical building that was to be erected around a large, panoramic mural painting, with the aim to produce the perfect illusion of a real scene. Barker erected his first permanent circular panorama building (or ”rotunda”) in London in 1792, where one or two new panoramas a year were exhibited for the next half-century. Panoramas as a mass-media phenomenon reached their peak popularity in the middle of the 19th century (Woeste, 2009).

In such a building, the viewer would stand inside a large room atop a platform and all around them gaze upon a vast 360 degree painting. In this space, the sunlight from above was concealed by a roof as to not break the illusion. The spectator had the opportunity to feel more involved in the scene than if they were merely viewing the static imagery on a flat surface (Woeste, 2009).

As this new form of entertainment became more popular, more such circular buildings were built around the world, displaying exotic landscapes, famous battles or important cities. Artists eventually started adding various items in front of their paintings to give them additional depth, which enabled a more realistic experience for the viewer (Woeste, 2009).

Stereoscopic photo viewers employed this principle more directly to create a realistic perception of immersion into a different reality. These were viewing devices that presented slightly different images to each eye to create the illusion of depth. Charles Wheatstone designed them in 1838, laying the groundwork for how VR headsets would function by expanding on the concepts of binocular vision and parallax[1]. The stereoscope was the 19th-century version

[1]Definition of parallax (from https://www.merriam-webster.com/dictionary/parallax)
: the apparent displacement or the difference in apparent direction of an object as seen from two different points not on a straight line with the object

of the first cheap home VR system (Bown et al., 2017).

Morton Heilig took immersion one step further and he is often seen as the father of VR (Carlson, 2007). He aimed to revolutionize how people watched film, so he created created the Telesphere Mask, a VR headset with stereo sound, smells, air flow, and the Sensorama, which featured a vibrating chair, smells and wind simulation (Bown et al., 2017).

Ivan Sutherland, in 1965, created the first ”head-mounted display” (HMD) to incorporate computer technology to mediate a VR system. This was the first time that computers were used to display a real-world environment whose elements were augmented by a computer. The system displayed on cathode ray tube monitors a 3-D wireframe cube that moved and tilted according to the user’s movements (Bown et al., 2017).

Today VR primarily represents a head mounted system that conceals the environment the user is located in, while depicting a digital virtual environment to the them. These HMDs were initially designed for gaming and entertainment, but usage has gradually broadened to include areas like job training, prototyping, marketing, tourism, commercial applications such as retail outlets and supermarkets, the fashion industry, manufacturing, healthcare, and as a research tool (Rauschnabel et al., 2022).

The number of degrees-of-freedom (DoF), or the number of system parameters that may vary independently of one another, is a factor that distinguishes various VR systems from one another (Bown et al., 2017). 3DoF systems only track rotational movement, while 6DoF tracks both rotational and translational movement (Pan and Hamilton, 2018). Other factors, such as per-eye resolution, field of view, refresh rate, display type, whether it features in-built speakers or headphones, and the type of controllers also vary between the VR headsets currently available on the market.

One of the most widely-available headsets was made by HTC in cooperation with Valve (the company that owns game sales Steam), namely the HTC Vive, whose first consumer version was released in April 2016 with dual 6 DoF touch controllers (Corporation, n.d.). In January 2018, HTC unveiled an upgraded Vive model known as HTC Vive Pro, with a higher resolution of 1440×1600 resolution per-eye. Standalone versions which don’t require a computer to operate are Vive Focus released in 2021 and Vive Cosmos released in 2019. Vive Pro Eye has in-built eye tracking, which can make user interaction with the objects they are looking at more natural (Statt, 2019). Some headsets have in-built cameras which can track the user’s fingers and spatial location, such as the Oculus Quest and Quest 2 (although tracking errors due to oclusion can occur). The Valve Index, developed by Valve independently, features handheld controllers that can also track the finger movements of the user.

Rauschnabel et al., 2022 lists a fraction of the big companies operating in the field of virtual and augmented reality: Microsoft (developer of the Hololens ”Mixed Reality” glasses), Oculus

VR bought by Meta Platforms, Inc. (previously Facebook, Inc.), PTC (developer of ”Assisted Reality” software solutions for warehousing businesses).

2.3 Presence, Perceived Realism, Embodiment and Immersion in VR

Presence is most commonly described as a subjective experience of being bodily located in a mediated environment (Hartmann et al., 2013). Weber et al., 2021 specifies that some researchers also stress the additional importance of agency/ability to interact with the virtual environment for describing presence, noting how Sanchez-Vives and Slater, 2005 (on p. 9) claims that the feeling of ”being there” in a virtual environment is tied to the ability to ”do there,” to act in that space.

According to Weber et al., 2021, the allocation of attention is frequently a crucial component of this concept of ”being there” — the experience of being absorbed by a virtual environment with one’s attentional resources directed at it. They adopt the phrase ”illusion of non-mediation” from Lombard and Ditton, 1997, which describes the illusion of losing sense that a mediated environment is being exhibited by a media device.

Weber et al., 2021 argues that the sense of ”being there” is distinct from perceived realism in the research of presence, and should not be confused with each other in order to better assess users’ experience of virtual reality in questionnaires. They describe perceived realism as an individual’s assessment of the virtual environment’s realism in terms of virtual objects, sounds and sceneries, the believability and plausibility of its story and characters, and the naturalness and simplicity of interaction.

> *”The user evaluates the plausibility and naturalness of the depicted world as well as the ease of interaction within the VE by answering questions such as: is there a shadow cast? Are the proportions of objects correct? Does the environment correspond to my own movements? Does my virtual body match the proportions of my real body? [...] are the consequences of actions plausible? Is the story coherent in itself? Does the causal sequence of events make sense? The answers to these questions define the degree of perceived realism.”* (Weber et al., 2021: p. 1)

Embodiment refers to the experience of identifying one body as our own:

> *”SoE (n.b. sense of embodiment) toward a body B is the sense that emerges when B’s properties are processed as if they were the properties of one’s own biological body.”* (Kilteni et al., 2012: p. 375)

Atherton and Wang, 2020 calls embodiment *"a way to experience our own bodies through the lens of a medium"* (p. 36).

Kilteni et al., 2012 also describe the following three principles of the sense of embodiment:

- **Sense of self-location**: identifying one's self as being located inside a virtual body. This experience describes the relationship between one's self and one's body, and it is distinct from the feeling of "being there" inside a world (be it with or without a body) which refers to the relationship between one's self and the environment (see "presence" above). Self-location highly correlates to having a first-person viewer perspective and to experiencing tactile stimuli on one's body that synchronize with visual stimuli at corresponding locations on the virtual body. The article also notes that modulation of the feeling of balance and self-location could be achieved by experimenting with exposure to caloric and galvanic vestibular stimulation[2].

- **Sense of agency**: the sense of having motor control and intention in over how the virtual body's actions correspond to one's own actions. The sense of agency is sensitive to time differences between the execution the participant's movement and received visual input. In order to ensure this sense, visuomotor correlations should therefore occur within crucial temporal bounds (realtime or near-realtime translation to visual response), which is possible in VR via motion capture systems or rigid body tracking translated into animation of the virtual avatar via inverse kinematics.

- **Sense of body ownership**: the virtual body feels and looks like the participant's own body, and it is perceived as the source of the experienced sensations. Sensory information (visual, tactile, and proprioceptive - arising from suitable haptic feedback or by synchronizing participant's passive movements and appropriate avatar animations), as well as cognitive processes that interpret the sensory stimuli (such as identifying enough human likeness to suspend disbelief that the artificial body can be one's own body), contribute to the development of this sense.

  This illusion can be boosted by enhancing the likeness between the biological body and the virtual one, and it may be further augmented by personalized avatars that improve the appearance similarity between the participant and the avatar.

The modalities of achieving embodiment according to Pan and Hamilton, 2018 are **visual-proprioception synchrony** (i.e., the virtual body or body parts are where you expect your own to be), **visual-motor synchrony** (as you move your body, the virtual body moves in

[2]These are both methods for manipulating a person's sense of balance via inner ear vestibular system stimulation. Wilkinson, 2021 describes them: *"Thermal, or caloric, stimulation is traditionally achieved by irrigating the external ear canal with warm or cool water (or less commonly by air)."*; *"The [n.b. galvanic stimulation] technique involves the application of low amplitude (< 2 mA), transcutaneous current to the mastoid processes, the bony protrusions located just behind the ears."*

the same way), or **visual-tactile synchrony** (as you experience touch on part of your body, you see the same virtual body parts being touched at the same time).

VR has been used to study embodiment by visual-tactile synchrony, for example by streaming live video feed to the HMD from a mannequin that is stroked at the same time as the participant's own body, or by motion-capturing the participant and displaying them in the VR world, optionally with aids such as virtual mirrors in which participants can see their avatars in VR. Once embodiment is established, the participant's sense of body can be manipulated by changing parameters such as its spatial location, age or the race of the body. A one-to-one mapping is not necessary to achieve embodiment: Pan and Hamilton, 2018 mentions a study by Lenggenhager et al., 2007 that proved how even a displaced virtual avatar can still induce this sense in VR.

2.4 Self-determination Theory, Player Motivation and Connections to Wellbeing

Self-determination theory (SDT) provides an explanatory framework for why individuals engage in and persist in a given activity (Ijaz et al., 2020).

Within this framework, intrinsic motivation is defined in R. Ryan and Deci, 2000 as the individual's natural propensity to seek out novelty and challenges, to exercise and develop one's abilities, to explore, and to learn. According to the paper, this propensity is crucial to cognitive and social growth and is a major source of joy and energy throughout life. This contrasts with extrinsic motivation, which refers to engaging in an activity to achieve a distinct goal, and which might vary substantially in terms of the degree of relative autonomy felt by the individual.

SDT theory claims that all individuals have three psychological needs (autonomy, competence, and relatedness needs) which must be satisfied for optimal function. These needs do not vary in the degree to which individuals possess them, but in the extent to which the surrounding environment facilitates their satisfaction (Cerasoli et al., 2016). The paper explains the needs as follows:

- **The need for autonomy** signifies the desire humans have to enact change upon their environment and self-drive their behavior.
- **The need for competence** refers to the desire to demonstrate mastery and improve one's skills and abilities. Satisfaction of one's need for competence correlates to performance because exhibiting and improving one's abilities is inherently fulfilling (Deci and Ryan, 2000 and Harter, 1978, as cited by Cerasoli et al., 2016).

- **Relatedness needs** are the desire to have meaningful relationships with others and impact the degree to which individuals actualize innate tendencies for growth and exploration (Deci and Ryan, 2000 as cited by Cerasoli et al., 2016).

R. Ryan and Deci, 2000 notes that pursuing intrinsic aspirations (whose practical implementation and manifestation can vary across cultures) in order to fulfill these basic needs tends to be positively associated with wellbeing, while attainment of extrinsic aspirations is counter to it. According to the article, giving individuals more autonomy via choices, acknowledging their feelings, and chances for self-direction has been shown to increase intrinsic motivation. The innate self-actualizing inclinations of human nature are disrupted by excessive control, inadequate challenges, and a lack of social connection, leading not just to a lack of initiative and responsibility but also to anguish and harmful psychological symptoms.

The player experience of need satisfaction (PENS) model was a model subsequently derived from the SDT framework to investigate video games, developed by R. M. Ryan et al., 2006. According to Ijaz et al., 2020, the research by R. M. Ryan et al., 2006 showed that games that provided more experiences of autonomy, competence, and presence were more engaging, motivating, and likely to be played again.

Player need satisfaction as well as player background, habits and gaming experience are all aspects that need to be taken into account when designing games (such as ones for exercising), according to Ijaz et al., 2020.

Specifically for VR, to explore the role of immersion in the effectiveness and enjoyment of VR exercise games, the study by Ijaz et al., 2020 compared providing their participants a static user interface (UI) condition to an open world (OW) condition, both in a VR biking application. Between the UI and OW settings, the study found substantial differences in autonomy, immersion, competence, and enjoyment. Additionally, the latter was most closely associated with perceptions of autonomy and presence in both circumstances.

Giving the participants a sense of competence was not a good predictor of enjoyment or future play, which was interpreted as possibly due to the design of the application which allows brief exploration. This prompted Ijaz et al., 2020 to observe that users could be categorized as entertainment-focused and exercise-focused. Those entertainment-focused participants exercised less, paid more attention to exploring the open world situation and ignored physiological cues regarding their physical activity. The exercise-focused participants were more attentive to the data showing their physical exertion and were more involved in their physical activity.

The research by Ijaz et al., 2020 also indicated that, in the UI condition, participant physical activity habits inversely impacted both autonomy and enjoyment, likely because users who regularly work out may have higher expectations for the system and may not find some VR exercise platforms with UI conditions any more more stimulating than typical gym equipment that displays a comparable interface with exercise data. The significance of the participants' prior video gaming experience, which negatively impacted their enjoyment and immersion in the OW condition, was confirmed by negative participant reviews of how plain the OW environment was compared to mainstream commercial video games.

2.5 Definition and Properties of Flow

Hassan et al., 2020 refers to Csikszentmihalyi and colleagues (eg. Csikszentmihalyi, 2014, Nakamura and Csikszentmihalyi, 2014) to define flow: an experience whereby we are fully absorbed in an activity which is performed for its own sake (autotelic[3]), to the exclusion of most of other stimuli and of our physical existence and state.

Nakamura and Csikszentmihalyi, 2014 as cited by Abuhamdeh, 2020 specify the following principles of flow:

1. The preconditions of flow:
    - Presenting challenges to the user that neither overwhelm nor underwhelm them in terms of skill
    - Offering the user clear goals that precisely communicate what they need to do. This allows them to pay undivided attention (Csikszentmihalyi, 2014) and improving how well they perform in the tasks given (Egbert, 2004).
    - Giving immediate feedback to the user about their progress. This informs the user to maintain or correct their behavior in line with the goals that they are trying to attain from an activity (Csikszentmihalyi, 2014).
2. The charactheristics of flow:
    - Focus on what one is doing in the present moment
    - The merging of one's actions and awareness
    - Losing one's sense of self-consciousness
    - The user gains a sense of being in control of their actions. This is not the same as allowing the user to be unchallenged and perfectly in control of their experience, which may elicit boredom (Egbert, 2004).
    - Temporal distortion, where one does not feel the passing of time
    - Autotelicity of the experience

[3]According to Merriam Webster, autotelic means "having a purpose in and not apart from itself."

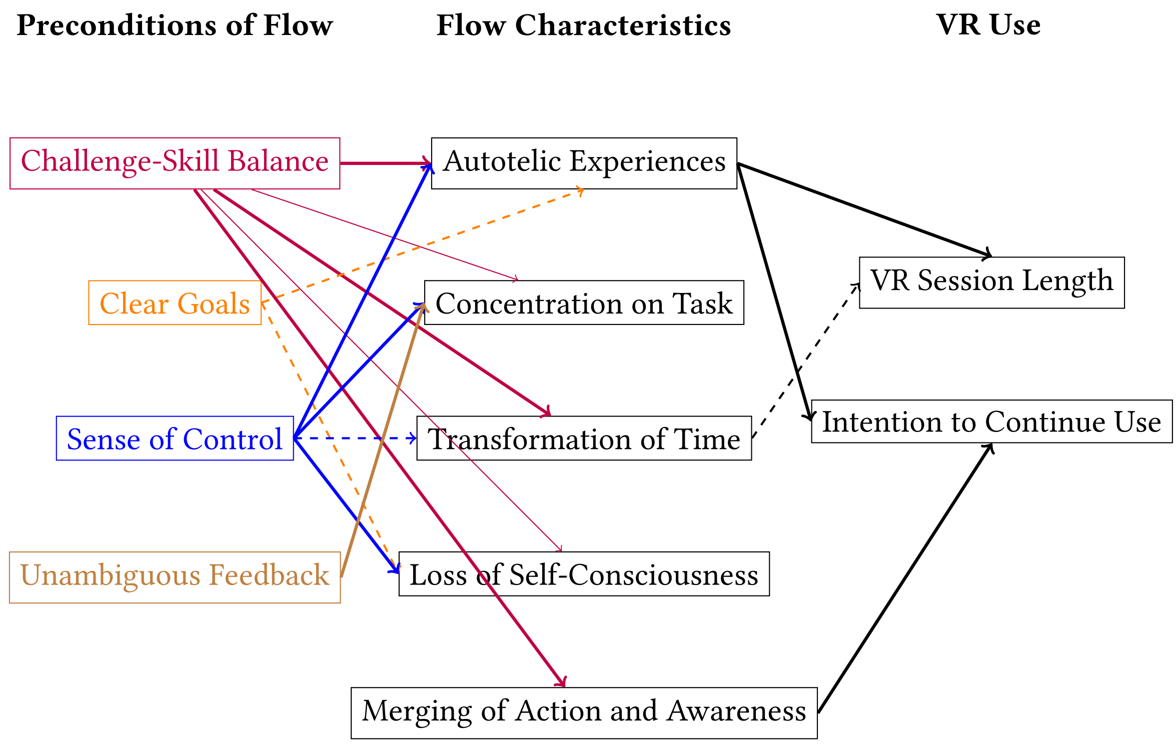


Figure 2.1: Correlations between flow parameters and VR use. Figure adapted from Hassan et al., 2020: p. 1200-1201. The type of connecting line indicates the correlation strength (continuous line: $p<0.05$, dashed: $p<0.1$).

The exact ways and circumstances that flow manifests in are highly subjective, impacted by factors such as personality, the virtual and physical environment in which individuals find themselves, age, mood, or gender (Nakamura and Csikszentmihalyi, 2014 as cited by Hassan et al., 2020).

2.5.1 Achieving a State of Flow in VR

What are then the ways in which we can design VR experiences that can absorb the participant, despite the current limitations of the medium?

Hassan et al., 2020 claims that the biggest potential of VR is that of telepresence (the feeling of being immersed in another reality) and its ability to capture users' attention.

The findings of Hassan et al., 2020 show that achieving a state of flow in VR is positively associated with the intention to continue VR use and with longer VR sessions.

Hassan et al., 2020 surveyed 681 participants with prior VR experience to determine the relationships between flow preconditions and outcomes in VR as well as likelihood to use VR again and length of use. The study concluded the correlations illustrated in Figure 2.1 on page 11.

The authors interpret the findings of their study as follows:

- Challenge-skill balance seems to be a strongly linked precondition to experiences of flow in VR, but they stress that while creating a cognitively challenging environment may induce flow, VR use in itself often presents a challenge with regards to its set-up and possible discomfort of use.
- Giving clear goals to the participant shows positive associations to concentration on task and merging of action and awareness in VR. Giving a clear purpose directs users' attention towards their next action, allowing them to easily become unaware of anything other than the activity itself (Nakamura and Csikszentmihalyi, 2014 and Novak et al., 2000 as cited by Hassan et al., 2020). Hassan et al., 2020 also writes that even exploration in VR can count as a goal (even if it is a more loosely defined one) that allows for experiences of flow to occur.
- Sense of control is positively associated with all flow experiences except merging of action and awareness. VR users express favorable views of their sense of being in control in VR over non-VR mediums. The authors interpret these findings as because VR enables exploration that may not be feasible in real life, therefore it may create a sense of control distinct from what may be experienced in physical situations. Additionally, it is unlikely that feeling a complete loss of control would lead to positive experiences, but rather discomfort and disorientation which would negatively impact flow.
- Unambiguous feedback was correlated in the study to concentration on task. The authors note that such feedback can direct users towards performing further exploration and actions, and that the feedback given doesn't have to necessarily relate to accomplishment of goals: the feedback can simply be an outcome of any user action (visual cues, sound effects, movement responses) that aligns with the user's expectations of the world they are in.
- A key element of flow in virtual reality is autotelic experience, which has a positive correlation with the duration of VR sessions and users' intentions to keep using the technology. The authors note that in general hedonic experiences tend to positively associate with lengthy use of systems per session and return to use of the system.
- No significant associations were observed between concentration on task and intentions to continue VR use or to longer VR sessions. The authors theorize that it's likely that exploratory activities, which don't typically call for intense attention on a task in the traditional sense, are what people value most about VR.

- An unexpected result found by the authors is the lack of correlations between loss of self-consciousness and having longer VR sessions or the intention to continue using VR. They postulate that, since the loss of self-consciousness in VR is accompanied by a dissociation from reality and a loss of awareness of one's immediate surroundings, it can be uncomfortable rather than enjoyable.

2.6 Principles of Good Design for VR

This section explores what it means to design a good VR experience, and what can be defined as good values and guidelines in this regard.

Atherton and Wang, 2020 believes that good VR design expresses human values and assists wellbeing and flourishing, a concept the ancient Greeks called Eudaimonia. An extension of this concept is the Capability Approach created by Amartya Sen, which investigates wellbeing and how to create societies that enable all people to live well by focusing on "capabilities" (opportunities) rather than "functionings" (something a person has achieved) (Robeyns and Byskov, 2021). In this approach, instead of requiring certain actions or states of being that everyone must fulfil in order to flourish, one should instead provide the opportunities that everyone must have access to, should they wish to pursue them (Sen, 1999 as cited by Atherton and Wang, 2020). Good design becomes, then, about supporting human potential.

Atherton and Wang, 2020 goes through the specific ways in which a VR experience can achieve these goals:

1. by providing access and accessibility to experiences that people otherwise wouldn't be able to take part in, in ways that are at least as complex and rich as physical reality.
2. by making sure that the end justifies the means — that there are unique properties of VR from which the application benefits uniquely (otherwise it could simply be made for eg. the desktop platform).
3. by what the paper calls *"total systems"*: creating experience that address all human senses equally, with a balance between audio, visuals and interaction, that employ a mix of realistic and fantastical elements, and that support play for individuals and groups.
4. by striking a balance between doing and being (allowing the user the time and space to reflect as well as to take intentional action), which reinforces embodiment in the medium. By observing how we act, feel and interact in virtual reality, we can learn more about both the virtual world and the real world.

5. by supporting creative expression, emotional development, social connection through community and play, rich living, and holicipation[4].

The paper details these ideas into the following guiding principles for designing a compelling VR experience (from Atherton and Wang, 2020: p. 36):

Lens 1 Don't forget about audio.

- *Principle 1.1: Audio should be dynamically generated.*
- *Principle 1.2: Audio should be immersive.*
- *Principle 1.3: Audio should be interactive.*

Lens 2 Designing to the medium

- *Principle 2.1: Don't Port (Corollary): Make things that would be impossible in the physical world.*

Lens 3 Doing vs. Being

- *Principle 3.1: Design to balance doing (action) and being (reflection).*
- *Principle 3.2: Look up! Use gaze to modulate between doing and being.*

Lens 4 Interaction

- *Principle 4.1: Drive interaction design with aesthetics.*
- *Principle 4.2: Multimodality is a virtue.*
- *Principle 4.3: Make space for being alongside doing in interaction.*

Lens 5 Immersion

- *Principle 5.1: Create worlds that enhance doing and being through animus*[a]*.*
- *Principle 5.2: Balance stylisation and realism.*

Lens 6 Designing for the body

- *Principle 6.1: Design for virtual embodiment.*
- *Principle 6.2: The body is an implicit medium where being supports doing.*
- *Principle 6.3: Movement matters.*

[4]What Killick, 2006 calls a solitary performance done for one's own enjoyment. As the performer and listener are one and the same, the music-maker experiences the *whole* musical event (in contrast to participation / being a part of a group).

Lens 7 Designing for play

- *Principle 7.1: Play is both an activity and a state: a synthesis of doing and being.*

Lens 8 Designing for social

- *Principle 8.1: Replicate baseline social interactions; redesign the rest.*
- *Principle 8.2: Support many kinds of social engagement.*
- *Principle 8.3: Design for social doing and social being.*

[a] *"the feeling that the environment is alive and full of creatures that have their own feelings, desires and goals beyond responding to the user"* (Atherton and Wang, 2020: p. 40)

The application *'12 Sentiments for VR'* - developed by Atherton and Wang, 2020 - applies these principles in several ways, including in how they use the gaze of the user. At key moments in the experience, the player is directed to look up (associated with growth, skygazing, and "being" in the moment), forward (to which hopeful music plays; associated with a positive outlook) or back (associated with "being", contemplation, melancholy).

Another application of the principles above is in the choice of when to give the user the ability to interact with the virtual world, and when to take it away - in order to give the user a break to just be in the moment, experience a narrative and get in touch with the feelings that arise from it. *'12 Sentiments for VR'* also has the user at times embody a plant and catch sunrays with their leaf-hands in order to grow. Each interaction with the sunrays triggers a swelling musical note.

2.7 Audio-first VR

From the outset, Endogeny VR was meant to be an experience centered around sound. Such an endeavor has begged questions such as, how exactly one defines audio-first VR, how do users interact with such applications, what are the available tools and methods to create them, and what possible subsequent developments for musical expression can arise from the practice of audio-focused VR? (Çamcı and Hamilton, 2020)

As discussed in section 2.6, an application for VR designed around audio should make use of the affordances of the medium, specifically of those which are unique to VR in terms of experiencing sound.

So what is audio-centric VR? Çamcı and Hamilton, 2020 gives some examples of what shapes such an experience could take, which will be explored in this section.

For example, *”Points Further North”* is a 360-degrees documentary film about human-altered landscapes, which primarily relies on immersive audio and spatialisation, alongside haptic feedback, to deliver information and imprint the viewer with the characteristics of each different location. This is defined as an acoustemological approach - sound as a means of knowledge acquisition (Trommer, 2020). Another example listed by Çamcı and Hamilton, 2020 is a sonification of the periodic motion of astral bodies using musical systems and relationships.

By combining the digital model of the virtual instrument with a physical object made from the same model, *”Hyperreal Instruments”* by Çamcı and Granzow, 2019 focuses on building ”impossible instruments” for virtual reality, which can still give the musician a sense of scale and control through the haptic feedback. Meanwhile, Hamilton, 2019 with *”Trois Machins de la Grâce Aimante”*, a string quartet composed for *Coretet* (a virtual stringed instrument and networked multiuser performance application for VR), focuses on giving users control over procedurally generated sound.

Çamcı and Hamilton, 2020 advocates for finding workflows that allow programming and creating audio directly in the VR medium: *”Alleviating the need for switching between a desktop device to create content and a VR headset to experience it, VR-based creativity support tools can exploit the immediate action-perception feedback loops that have been available to electronic music composers since the 1950s.”* (Çamcı and Hamilton, 2020: p. 6)

Another point of note is referenced by Çamcı and Hamilton, 2020 from Buckley and Carlson, 2019, who advise using a combination of predetermined and interactive elements in a VR composition and guiding the listener through the immersive audio environment by modulating the density of sound elements in the virtual soundscape.

Interestingly, Çamcı and Hamilton, 2020 lists several VR projects that have shown benefits for music and audio education: whether it is for teaching students in primary or secondary school rhythm or acoustics, or as exposure training to reduce music performance anxiety in front of an audience to and improve musicians’ performance.

A very interesting project for the author of this thesis, which takes the ”music jam” in VR concept forward, has been PatchXR, recently released as PatchWorld on the Quest 2 („Surreal Soundscapes: Explore Musical Worlds + Build Your Own in ’PatchWorld’ on Meta Quest 2“, 2022) with very good ratings on the Oculus app store. This application is heavily inspired by LEGO and Max MSP and allows building modular sound tools entirely within the VR application itself (or mixed reality for the Meta Quest Pro headset). The article lists the following application features[5]:

[5] Multiplayer is intended for the application, but not yet released at the time of this thesis („PatchWorld - Make Music Worlds“, n.d.)

- *"Three unique ready-to-play worlds ('EPs') to explore or remix*
- *Block-based custom world creation*
- *Over 30 instruments, with more on the way*
- *Whimsical sound toys like rubber duckies, glass bottles, skulls, frying pans, and more*
- *Powerful studio-quality tools, including recreations of acid synthesizers, vintage drum machines, loopers, and chord synths*
- *Playable, physics-based sounds based on wind, rain, marbles, gravity, and more*
- *A microphone so you can record samples or sing along to your creations*
- *Built-in bands and characters*
- *Avatar customization*
- *3D spatial audio and stereo mixing"*

Other projects that employ interesting sound interactions in VR and were an inspiration for this thesis project are described by Atherton and Wang, 2020: *"**Canyon Drum** is an exploration of many ways to play a drum that is very large and very far away in VR. [...] **Shred Head** is a set of two objects that can be played either as wind chimes with a virtual hammer or by placing the objects on one's head and swinging them around as if they are wigs. [...] **Twist Flute** is a virtual flute that users play by blowing into their headset; hand distance controls pitch and hand twisting controls timbre."* (Atherton and Wang, 2020: p. 40-41)

2.8 Sound Perception

2.8.1 Sound Perception and Localization

Marks, 1978 indicates that spatial localization of an object through sound alone is prone to errors, and that humans are more reliably able to detect relative sound displacement than the absolute position of individual sounds. Accuracy is improved by rotating our heads, which changes the Interaural Time Delay (ITD) and Interaural Level Difference (ILD) at which a specific sound hits the ears, helping in better determining the position of a sound source. These interaural cues only aid in human detection of the azimuth of the sound source (Hofman et al., 1998).

The pinnae (outer part of the ear) is also involved in sound localization, specifically where ITD and ILD are insufficient (sound elevation and front-back direction). Hofman et al., 1998 explains that the human pinna amplifies or attenuates sound frequencies, having a sound filtering effect ("pinna filters"). An experiment by the same author demonstrated that sound elevation detection was disrupted by modifying their pinnae with molds (although the subjects were gradually able to "learn" to hear better with the ear molds over time, and this neural training did not interfere with their ability to localize sound elevation once the pinnae molds were removed).

2.8.2 Sound Perception and Room Acoustics Terminology

In assessing the perceptual sound quality and fidelity of an application, a terminology needs to be established. The perceptual acoustic characteristics of a space can be attributed to objective physically-measurable properties. Ziemer, 2020 lists many such characteristics that can be taken into account for concert halls (and, by extension, when assessing the sound simulation quality of a VR application). We focus here on the most important ones, based on their appearance in other sources as well.

van Dorp Schuitman et al., 2013 proposes a binaural model of mathematical formulas concerning only the following perceptual auditory parameters:

1. **Reverberance** is the amount of reverberation perceived by listeners. This value relates to the physical reverberation time (the time it takes for the sound pressure to decay by 60 dB after the sound has stopped).
2. **Clarity** is the degree to which discrete sounds in a signal can be perceptually distinguished from one another. If clarity is high, individual notes in a musical piece or individual speech sounds can be heard.
3. **Apparent Source Width (ASW)**, related to spaciousness, is the perceptual apparent broadening of a sound source as a result of early lateral reflection and is most often assessed using the early interaural cross correlation.
4. **Listener Envelopment (LEV)**, related to spaciousness, refers to the environment instead of to the source. A sound field is enveloping when a perception of being surrounded by the sound occurs, because it is coming from all directions.

Additionally, Ziemer, 2020 also defines the following:

1. **Intimacy** describes how close to the listener the sound sources and surfaces seem to be, and how small or big the room feels. This parameter is related to the ITDG (initial time delay gap[6]). Ziemer, 2020 notes that an ITDG of under 21ms is usually measured in the best-rated concert halls.

[6] the delay between the arrival of the direct sound and the first reflection (Ziemer, 2020)

2. **Liveness** is the impression that there is more sound than just the direct sound and its repetitions: a ”live” space features a long reverberation, in contrast to a ”dead” or ”dry” space. The objective measurement of this perceptual attribute roughly corresponds to the reverb decay time in the frequency range of 500 Hz—1 kHz.
3. **Loudness** is the perceived volume of sound. Room reflections increase the loudness compared to the direct sound alone. Sound strength, which defines how the acoustics of the space boost the sound, is the objective measure of loudness.

2.8.3 Multimodality and Sensory Integration

Humans typically corroborate (”integrate”) information from different senses to construct a unified (”multimodal”) representation of reality. Marks, 1978 references an experiment from Auerbach and Sperling (1974) which found that, when asked to distinguish between the spatial location of visual and auditory stimuli, the test subjects seemed not to have different perceptual spaces for each separate sense.

The so-called *”McGurk Effect”* is what many scientists view as a famous proof of this multimodal sensory corroboration, according to Tiippana, 2014. This effect is a multisensory illusion triggered by listening to a sound while watching someone speaking something else, first reported by McGurk and MacDonald, 1976. In this experiment, the acoustic speech signal was accurately identified by study participants when used on its own, but pairing it with incongruent visual speech made them hear a different sound: *”on being shown a film of a young woman’s talking head, in which repeated utterances of the syllable [ba] had been dubbed on to lip movements for [ga], normal adults reported hearing [da]. With the reverse dubbing process, a majority reported hearing [bagba] or [gaba]. When these subjects listened to the soundtrack from the film, without visual input, or when they watched untreated film, they reported the syllables accurately as repetitions of [ba] or [ga]”* (McGurk and MacDonald, 1976: p. 1).

Research summarized by Spence, 2011 has shown that crossmodal correspondences are important to human information processing, as they influence their performance in many kinds of experiments such as direct crossmodal matching, speeded classification tasks, speeded simple detection tasks, spatial localization, and perceptual discrimination tasks. One way in which they might assist information processing could be through crossmodal grouping, since the presence of multiple stimuli will mean that some relative information is likely available.

Between the senses there is however a hierarchy in how humans make use of sensory information to interpret their surroundings. Vision has been found to take the dominant place in object localization, above hearing and touch, and it appears to bias the localization of sounds in space when we see can their visual counterparts, whereas the opposite is not true (Marks, 1978).

2.8.4 Crossmodal Correspondences with Audio

Spence, 2011 notes that research into crossmodal correspondences is still fuzzy in terms of how these phenomena occur and function in humans. They summarize findings on the topic and propose a categorization as follows:

1. **Structural correspondences**: correspondences resulting from the particularities of our neural systems and sensory organs (either innate or resulting from neural development). This can happen, for example, because the specific sensory inputs might be processed by neighboring areas of the brain, or due to what the author terms *"neural economy"* — the brain might use similar mechanisms for different sensory inputs, which might in the process become associated. Found such correspondences:
   - **Loudness—Brightness** (brighter stimuli are associated with louder sounds)
2. **Statistical correspondences**: correspondences indicating an adaptive reaction by our brains to patterns statistically encountered in our environment. These correspondences are learned based on experience, rather than innate. Found such correspondences:
   - **Pitch—Elevation** (higher pitch correlates with higher elevation)
   - **Pitch—Size** (higher pitch correlates with smaller object size)
   - **Loudness—Size** (louder sounds correspond to larger objects)
3. **Semantically mediated correspondences**: associations occuring when linguistic terms are used to describe stimuli applicable to different senses. A notable investigation the author refers to is Stumpf, 1883, who found that most languages use the same linguistic terms usually describing elevation, "low" and "high," to describe stimuli that vary in pitch. These correspondences are learned, an outcome of language development as certain terms come to be associated with more than one perceptual continuum. Found such correspondences:
   - **Pitch—Elevation**[7] (higher pitch relates to higher elevation)
   - **Pitch—Spatial Frequency**[8] (high pitch relates to high spatial frequency)

[7]Though Spence, 2011 points to research that may also allow to classify this correspondence as statistical.

[8]Definition from the APA Dictionary of Psychology: *"the number of repeating elements in a pattern per unit distance. In a simple pattern of alternating black and white vertical bars (an example of a square-wave grating), the spatial frequency is the number of pairs of black and white bars per degree of visual angle, usually expressed as cycles per degree (cpd)."*

In addition to these categories, Spence, 2011 lists the **Brightness—Pitch** correlation (brighter stimuli correspond to higher pitch, found in a study by Marks, 1987), a result which they could not fit in their proposed categorization above. In other words, we can understand this sensory association as such: *"Darkness is the sensation produced by low-frequency components in sounds. The darkness sensation is based on the relative balance of the sound spectrum, independent of the fine structure."* (Holm Pedersen, 2008 p. 16)

Since crossmodal correspondences seem to help humans derive conclusions more easily about their surroundings, such pairings were assumed to be user expectations in a mediated experience like VR and facilitators in adopting a technology which is still foreign and cumbersome to most people. We therefore assumed they should be used as a starting point for designing audio interactions in the Endogeny application.

3 Design and Implementation of Endogeny VR

3.1 Conceptual and Design Considerations

Endogeny VR was conceived as an audio journey for the participant, therefore the it was constructed to offer the user many distinct ways to experience sound, including in terms of spatial layout, element density, and overall narrative concept.

As a digital artwork which offers the user a level of interaction with its environment and soundscapes, Endogeny VR aims to afford access to people of varying backgrounds to a performative virtual experience in a sound garden or "sandbox." The specific properties of the VR medium were the starting point for designing user experiences, and the music created in VR served as basis for further iteration and composition for the developers, as well as the basis for the user themself to improvise on.

Endogeny VR has no difficulty settings. Designed as an open-ended exploration, it is up to the user to decide how far to go in terms of involvement and time spent in the experience.

3.1.1 Design Pillars

It was necessary, as with any artistic endeavor, to define some limitations and guiding principles, in addition to following the lenses defined by Atherton and Wang, 2020 (see p. 13).

Our specific design pillars were:

1. Offer Possibilities, Don't Set Requirements
2. Imply, Don't Tell Directly
3. Give Simple Tools to Make Sound
4. Single Player That Could Extend to Multiplayer
5. Animate and Humanize the Space
6. Make it Strange and Whimsical
7. Focus on Listening and Being in the Space

Offer Possibilities, Don't Set Requirements

In the quest to offer the participant autonomy, as well as musical holicipation and to follow the "Capability Approach" as discussed in section 2.6 on p. 13, the experience was designed without verbal instructions, high scores or demands of any kind, being more of an invitation for the user to engage as they wish.

As a helping hand, some guidance was still implemented: lighting was made brighter at parts in the level where potential points of interest are located, a guiding creature which emits visual ripples and sound and walks ahead of the user leads them out of the maze at the beginning of the level, and interactions were described visually through "cave painting"-like illustrations that glow in the darkness of the cave.

3.1.2 Hypotheses

The following hypotheses were used in the way the experience was designed and subsequently tested:

- H1 Audio interactions help with user immersion.
- H2 SteamAudio is a good tool for physics-based audio implementation in games.
- H3 Separating the virtual spaces into visual and auditive "hotspots" helps the user with orientation and guidance.
- H4 Keeping the controls in VR simple helps with navigating a relatively new consumer technology.
- H5 Physics-based interactions with the environment are an intuitive way for creating sound in a VR application.
- H6 Avoiding translational movement in VR prevents motion sickness.
- H7 Using human elements in a virtual space (eyes, heads, tongues) and animated vegetation and fauna helps with familiarity of the environment, playfulness and engagement via unusual combinations, and aids against the sense of isolation in VR.
- H8 In dim virtual spaces, the participant can rely on spatial audio for orientation.
- H9 Physics-based audio spatialization helps the participant with localization of sounds in space.

3.1.3 Narrative Concept

Although the experience features no traditional narrative, its concept is centered around the idea of getting in touch with the human body, the human psyche and the natural environment, even their strange or unsightly aspects. Its name itself means, in biology, *"growth from within; endogenous formation of cells"* („Endogeny“, n.d.).

The experience initially was going to have an ending: the floating guide that leads the player out of their starting area would reappear (since the user would already know they are meant to follow it) and lead the participant once again, this time to the exit. However, this was never implemented.

3.1.4 Level and Game Design

The experience takes place inside a cave with diversely-shaped labyrinths, small inner rooms, large open areas and buildings.

The space was subdivided visually, acoustically, spatially and thematically (see Figure 3.3).

Figure 3.1: Concept art of the large cave opening by Maria Weninger

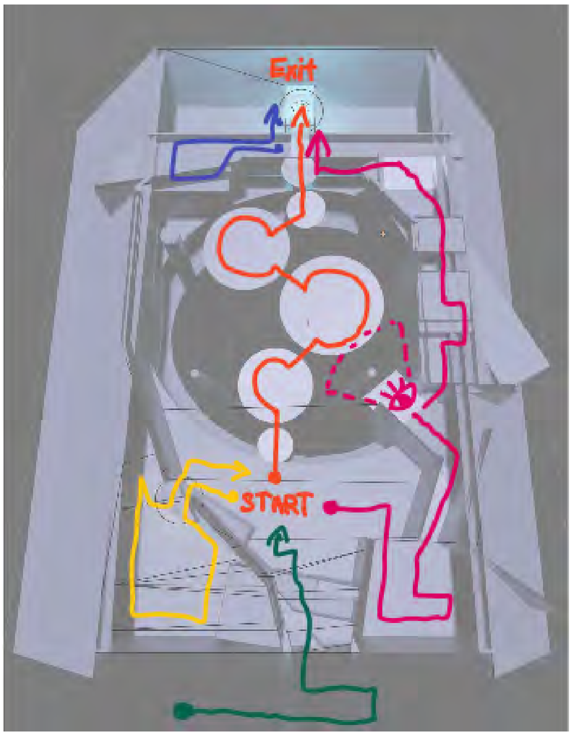


Figure 3.2: Level layout, early greyboxing

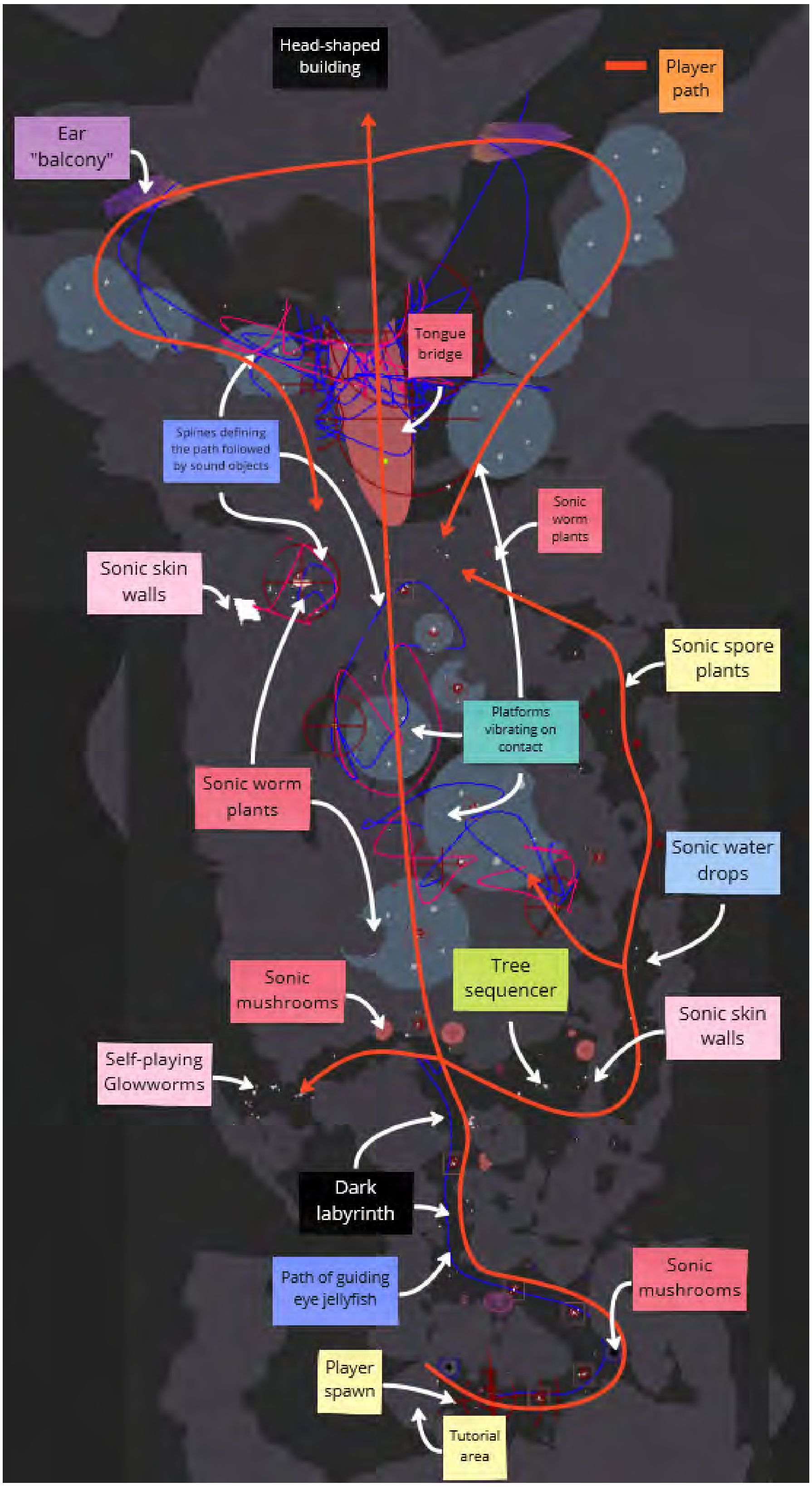


Figure 3.3: Final level layout (unlit screenshot from above, in game engine)

Head-shaped Building

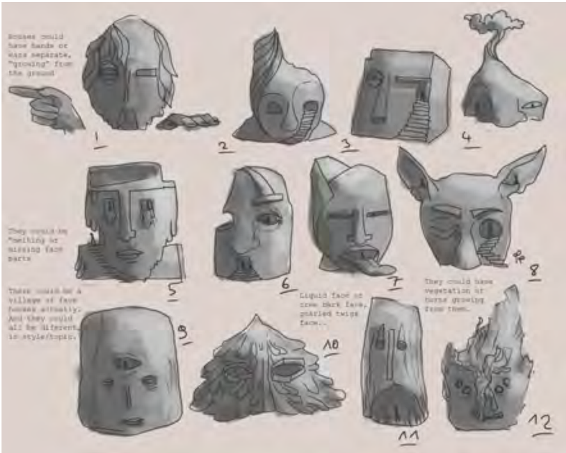


Figure 3.4: Early "Head-shaped Buildings" concepts by Maria Weninger

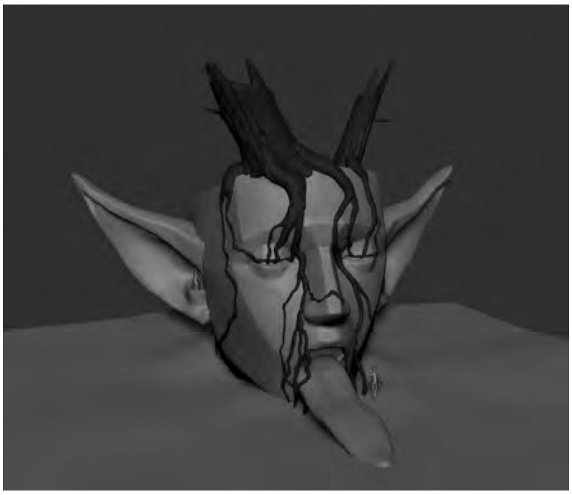


Figure 3.5: Final 3D model of the head-shaped building by Valentin Fischer

Variable Stimulus Density

One manner in which Endogeny presents its sonic environment is by modulating stimulus density.

The water drops are an example of this approach: their spawn rate is modulated by a periodic timeline, cycling between more or less dense over time (within a randomness factor). The experience overall features a range of areas, some quieter or almost entirely silent and others trimming with activity, and the player is able to relocate some of the audio sources.

Although we tried to keep stimulus density in mind, Endogeny, in retrospect, should have introduced the high sonic density of the dim labyrinth at a later point in the level, in order to give the participants some time to adjust to the experience and its specific technical setup.

Player Movement

To avoid motion sickness, the application only makes use of walking and teleporting as means of locomotion (no flying, no translational movement). These movement modalities were employed in the desire for simplicity of implementation and use and due to our lack of other equipment.

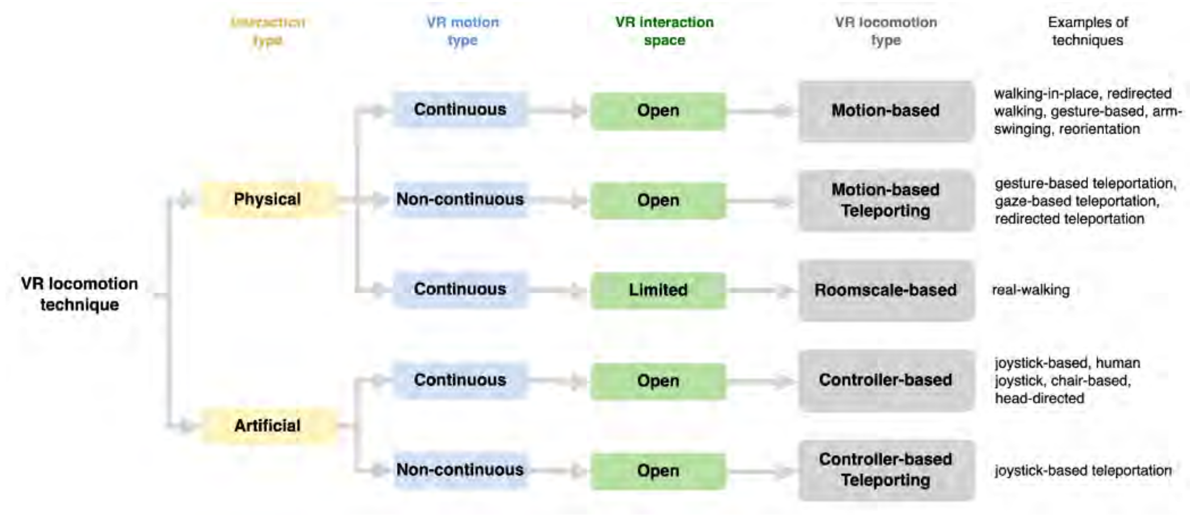


Figure 3.6: Typology of VR locomotion techniques, taken from Boletsis and Chasanidou, 2022

However, there are many other ways to enable user movement in VR. Boletsis and Chasanidou, 2022 analyzed 42 scientific papers from 2021 to investigate the types of locomotion employed in VR and to devise a way of categorizing them, and their results are depicted in Figure 3.6.

In the hopes of preventing teleporting too frequently and to encourage normal movement instead (what Boletsis and Chasanidou, 2022 calls *"real-walking"*), an audiovisual effect was added when the player performs this action: the overall sound is pitched down and gradually returns to normal, a dissonant sound effect is played, and a visual wave ripples around the player.

3.2 Technical Implementation

3.2.1 Technical Requirements

Endogeny VR has the following recommended technical requirements:

Windows 10
32 GB RAM
NVidia GeForce RTX 2080 Ti
Intel Core i7 @3.5GHz

3.2.2 Software Tools and Plugins

The Endogeny VR application was developed in Unreal Engine 4.26 with the in-engine audio system (no middleware) and tested primarily on the HTC Vive Pro and Oculus Rift (due to their wider availability at the university facilities and generally on the consumer market compared to newer headsets). Spatialization to simulate physics-based sound propagation and reverberation was performed using SteamAudio. The AVRFramework was used to implement VR mode and interactions.

The source code is available at https://bitbucket.org/ancatut/vr-music-experience/.

Sound Spatialization with Steam Audio

Steam Audio, EVERTims, VA by the RTWH Aachen University and Wwise are listed as the primary software options for physics-based modelling of audio by Fırat et al., 2022.

Steam Audio was chosen to perform binaural rendering and sound spatialization in engine, as it was at the time easy to integrate with the interface of Unreal Engine 4 and didn't require installing external software.

Steam Audio is a tool that uses Head-Related Transfer Functions (HRTFs) and physics-based sound propagation to render positional audio with accurate reverb, reflection and occlusion effects based on the properties of the sound's virtual environment. The sound propagation and reverb can be calculated dynamically during gameplay, or pre-rendered via a process called baking. Using baked propagation effects reduces the CPU overhead at runtime at the cost of increased memory usage (Valve-Corporation, n.d.).

Steam Audio computes two main variants of attenuation, Geometrical Divergence Attenuation and Atmospheric Absorption Attenuation, to simulate the sound propagation under different physical conditions (Fırat et al., 2022).

3.2.3 Sound Tools in Unreal Engine 4

Sound Cues

The Sound Cue object in UE4 encapsulates sound design operations in a node graph, like frequency modulation, playing a random clip out of a set, fading between two or more input sounds based on the distance from the Sound Cue's origin to the listener and more.

Figure 3.7: Sound cue interface

Mixing with Submixes

Figure 3.8: Submix Interface. In this case, sound can be routed to the submix based on the object's distance from the listener.

„Unreal Engine 4.26 Documentation: Submixes“, n.d. specifies:

"A submix is a DSP (digital signal processing) graph that is always running, even when no audio is being sent.

A fundamental component of the audio renderer in Unreal Engine, submixes have a dual purpose:

- *To mix audio generated from individual sources into a single output buffer, and*
- *To optimize the application of digital signal processing (DSP) effects to multiple sound sources simultaneously."*

Different audio sources can be sent to a specific submix at varying rates to apply sound effects in bulk. This process can be done in the Submix Editor in Unreal or via blueprints.

Quartz System

With the update to Unreal Engine 4.26, we had access to a new audio feature for playing quantized audio events with accurate timing on the audio thread, independently of the game thread which can introduce latency.

The „Unreal Engine 4.26 Documentation: Quartz Overview“, n.d. specifies:

"Quartz is a system that works around the issues of variable latency and game-thread timing incompatibility by providing a way to accurately play any sound sample. Sample accuracy refers to the ability for a sound to render audio at an arbitrary sample (point in time) within an audio buffer rather than at the beginning of the buffer. Instead of rendering a sound at the beginning of an audio buffer, Quartz cues the sound to play on the desired musical value (bars or beats) or time value (seconds), independent of the buffer size, game-thread timing, or other sources of variable latency."

"A clock is the object in charge of scheduling and firing off events on the audio rendering thread. A clock is created with the Quartz Subsystem, and modified via a Blueprint using Clock Handles. Each clock has a Quartz Metronome. The metronome is the audio render thread object that tracks the passage of time, and decides when upcoming commands need to be executed from user-provided information such as BPM (beats per minute) and a time signature.Gameplay logic can subscribe

to events on the metronome to be notified when musical durations occur."

The Quartz system was used to implement synchronization of music layers and various sound events in the experience meant to occur in time with the music tempo, such as the notes played by the tree sequencer.

Music Scale Mapper

Unreal Engine supports reading MIDI files. We have programmed a way to map MIDI values to notes of a predefined scale that can be selected in the backend — any musical scale (including microtonal scales[1]) can be easily selected.

The *BP_ScaleCalculator* blueprint is in charge of calculating the note value that is sent to the synthesizer for playing. Given a specific scale (for example, the twelve-tone equal temperament which is the most popular music scale in use, or the Blackwood 9 scale, a microtonal scale whose intervals are depicted in 3.2.3), this component calculates the frequency and cent[2] value that the specific MIDI note maps to within that scale.

Row	Intervals
1	10/9
2	9/8
3	5/4
4	21/16
5	4/3
6	3/2
7	5/3
8	15/8
9	2/1

Figure 3.9: Blackwood microtonal scale

SteamAudio Acoustic Materials

The different sizes, shapes and materials that the walls and objects in this space are made of were taken into account when computing sound occlusion and reverberation by the game's audio plugin for the space the user is traveling through. The idea was to experiment with the capabilities of a sound spatialization plugin like SteamAudio and investigate how participants experience an acoustic architecture, including transitioning between spaces and experiencing them from different vantage points.

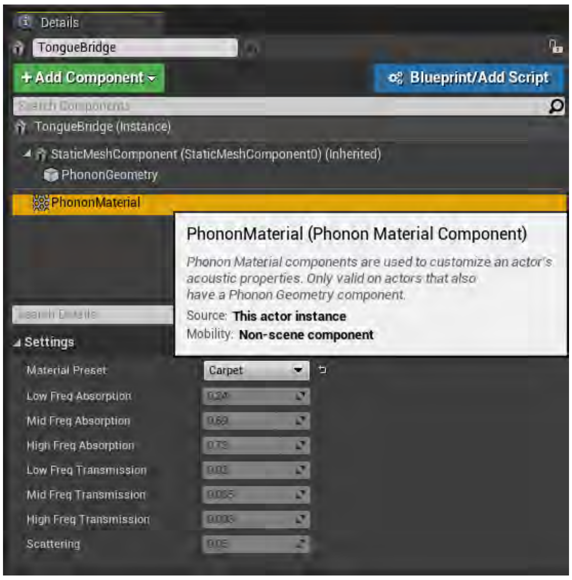


Figure 3.10: SteamAudio acoustic materials added to a level component

To include an object in the acoustic scene geometry, a Phonon Geometry and a Phonon

[1] *"Microtonal music or microtonality is the use in music of microtones—intervals smaller than a semitone, also called 'microintervals'. It may also be extended to include any music using intervals not found in the customary Western tuning of twelve equal intervals per octave."* („Microtonal music", n.d.)

[2] The cent is a logarithmic measurement unit used for the interval between two consecutive notes within a scale. For example, the twelve-tone equal temperament divides the octave into 12 semitones of 100 cents each.

Material component must be added to it before computing the SteamAudio bake, as seen in Figure 3.10. The object must be set as *Movable* in the editor if dynamic spatialization is to be computed for it.

Modular Synthesizers

Modular synthesizers generate dynamic sound via sound oscillators. They are made up of separate modules for different functions, which can be connected and assigned dynamically. Figure 3.12 illustrates all the settings that can be modulated for this device in Unreal.

For playing tones on the modular synth in Unreal Engine, we can call the *Set Osc Cents* method followed by the *Note On* method (Figure 3.11).

The *Note On* takes a MIDI note value as input and actually activates playing the synthesizer, while the *Set Osc Cents* takes the oscillator target (since the modular synth features two oscillators) and the cents value (which will define the relative frequency to play to the MIDI note passed to the oscillator).

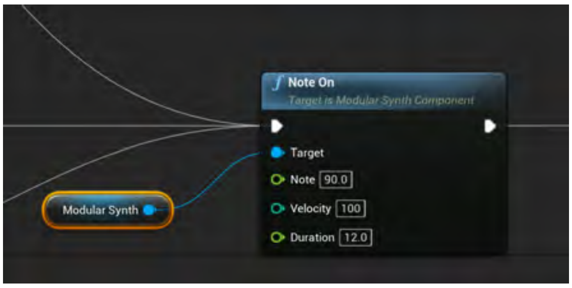


Figure 3.11: Modular synth *Note On* method

Figure 3.12: Modular synth settings preset

3.2.4 Sound Interactions Designed for VR. Implementation Details

This section provides Endogeny VR's list of main sonic interactions and features and a brief implementation rundown of some of the features.

Sequencer Tree

The tree construct is a polyphonic sound sequencer designed for 3-D space and organic interaction. Each of its components and dimensions fulfills a specific role.

The top of its trunk is the **origin point of an orthogonal system**, and in relation to this point several sound parameters are calculated. The tree's **branches** snap to a grid aligned with this orthogonal system.

The tree's **leaves** are grab points that can be picked up and dragged by the user. When the user drops the leaf at a point in space, the leaf snaps to the grid, it spawns an **eye sound node** attached to the branch indicating where a sound will play, and two new leaves grow near it, from which the process can be repeated. In this way, a power of 2 number of sound branches can be grown from each of the original 2 leaves the player finds in the tree.

The tree's system of coordinates is a translated version of the XYZ spatial coordinates of the level. The coordinates of a tree sound in the tree's system of coordinates corresponds to sound pitch, delay, and ring modulation respectively. These values could be mapped to other parameters, as the implementation we created is modular.

Every time the player performs an update action on the tree (spawning a new note or grabbing and relocating an existing note), the list of nodes to be played and their timing is updated.

A **loop** is the length of the longest branch in the tree that was grown from one of the two initial leaves. There are 2 possible loops that can play in parallel in the current application build, but more could be added in the editor (but could have a performance impact).

Several nodes can play sound concurrently when they are located at the same distance from the origin point, as the modular synthesizer we used is polyphonic. However, the predefined number of synth voices is a limit in this regard, and the maximum value that can be defined per synth is 32. Each loop was set to 2 voices each in the current build, to mitigate performance.

Initially, the plan was to add the ability to disable tree nodes by "closing their eyelids" and to destroy branches by cutting them, but sadly there was not enough time to implement these features.

Additionally, the big grabbable eye, sitting on the ground next to the tree, controls the **frequency modulation** of the tree sound. The closer the big eye is carried to the tree, the faster the modulation. This eye is connected to the tree by a spline.

The function of the eye could have been illustrated by its connector spline wobbling at the same rate as the frequency mod, to make it more clear what purpose it served.

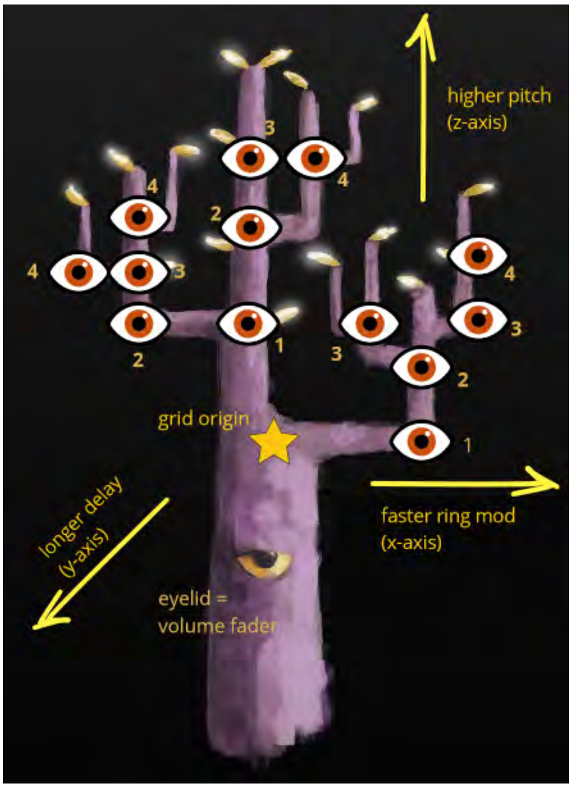


Figure 3.13: Sequencer tree concept with axes and grid explanation. The numbers indicate the sequence in which sounds play at the location of the respective eye node, depending on their distance along the grid from the origin point.

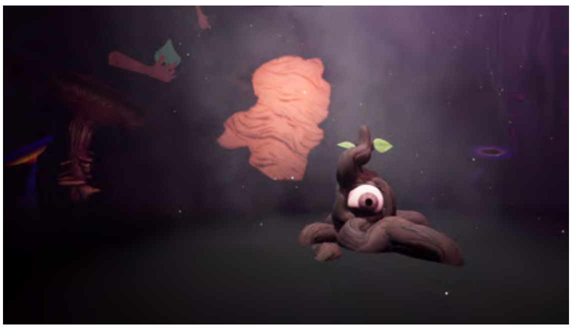


Figure 3.14: Sequencer tree in-level

Figure 3.15: Grown sequencer tree

The Dark Labyrinth and the Floating Eye Guide

Figure 3.16: Concept art for floating eye jellyfish by Maria Weninger

The labyrinth that the user is guided through by the floating eye-jellyfish is darker than most areas in the cave. This space was envisioned to challenge the user to rely on sound more than on visuals to find their way through to the other side.

In retrospect, the way this concept was implemented was not ideal. Since the desire was to see if the user could rely primarily on sound to navigate a dark area in VR, this challenging area should have appeared later in the level, after the user had enough time to get used to the environment, controls, and audio properties. Instead, this section is the *first* that the user enters after they spawn in the tutorial area.

The choice to put this labyrinth section in the beginning of the experience came from the desire to give the user an introductory experience where they would have to follow a guiding element through a sonic landscape in a linear fashion. But because we committed to this approach, it was decided to make the labyrinth much brighter so as to not confuse the player, and to make the guiding actor itself very visible (adding a ripple effect that is visible in the dark when sound plays from it). These choices, in return, made the labyrinth no longer serve its intended purpose of guiding the player through sound alone.

An earlier version of the dark labyrinth that we experimented with featured a shader which transformed the visuals, showing the player only faint outlines of the environment's edges. This visual deprivation would have likely been better suited to the original intention of this challenge.

Flying Sound Emitters

Wobbly Skin Walls

Worm Plants

- **Worm Plants Activated on Touch**. These tall plants can be hit by the player, and depending on the touch velocity, they either play a gentle sound, or a louder *"blub"* sound that also modulates the wobble of the plant's mesh, using the audio envelope values.

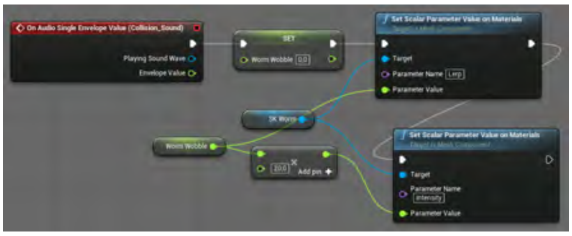


Figure 3.17: Section of the worm blueprint that modulates its rigged mesh based on its audio envelope

- **Self-playing Glowworms**. Some worm plants additionall play modular synthesizer tones by themselves, like a self-playing organ. These plants track the position of the user, so as you move through their area (covered by an invisible volume), the plants change in height and pitch. The corners of the invisible volume correspond to a certain pitch, and the note played by a glowworm is interpolated between these 4 pitches. This indirect interaction correlates height with pitch (higher pitch - taller plant).

- **Random Colorful Worms**. These have the same functionality as any default worm plant, but they have a random pitch and color each time they are hit.

Sonic Water Drops

See blueprint in Figure 3.18 on p. 41.

Step-activated Platforms

Teleportation Sound Ripple

Resizable Sonic Spheres

Orbiting Musical Layers Lead by Player

3.2.5 Music Production

A dynamic soundtrack was composed in an experimental style that uses granular synthesis, randomization, recorded percussion and string instruments, non-verbal vocalizations, and dynamic in-engine sound synthesis.

It was decided to go with non-verbal vocalizations for the composed melodies, to give the space a more ”animated” and human feeling (see *Lens 8* on page 15) while not assigning any explicit meaning via lyrics.

The soundtrack was split up in layers (stems) by instruments and each layer was assigned to different objects moving in space. This resulted in constantly-changing position and loudness, and sometimes in complete occlusion of the layers. The result was a varying composition that the user could hear at any point in time. Because there could not be a fixed mix for such a variable music composition, the approach was to make the music very ambient- and soundscape-like.

3.2.6 Application Features in Detail. Development Timeline

This section provides the full detailed list of the features the team members worked on during development, grouped by month of execution. The team used Miro for project management and worked in 2-week sprints. Some experimental features and ideas had to be dropped from the final build, where specified.

July-Sept. 2020

- Setting up the technical framework and the VR and 3D desktop application modes in Unreal Engine 4.25
- Implementing the ability to grab objects
- Integration of musical scale remapping (including microtonal scales)

- *(not included in final build)* Dynamic ambient music composed for test map
- Objects that respond to the player's touch (cloth and rubber-like materials, tall "worm" plants with manipulated bones)
- *(not included in final build)* Sphere that lights up and emits particles when the player steps inside it
- *(not included in final build)* Ability to detach the Unreal audio listener from the player's position and have it follow another object in the world ("listening from a different perspective")
- Sound objects that move along a random path
- Sound synthesis and modulation in Unreal
- Integration of a MIDI note player into Unreal Engine that can record a MIDI sequence and play it back on a synthesizer
- *(not included in final build)* Grabbable mushroom lamp model, interactive flower concepts
- *(not included in final build)* Environment shader that turns edges and environment into a single color
- Tree sequencer structure:
  - Appearance and functional design for a musical tree with eyes as notes
  - Implementation of an extensible branch that grows with and fits to a grid and connects music event nodes
  - Ability to create sound events and map them to a grid that corresponds to pitch in the musical scale, timing, and sound effect parameters of choice
  - Ability to reposition the sound nodes along the invisible grid
  - Modulation of the tree sound depending on its distance to the player
  - Various sounds emitted from the branch (sample triggers, synthesizers)
  - Volume control for the whole tree structure

September 2020

- Sound spatialization, occlusion and dynamic reverberation with SteamAudio
- Random sounds triggered by a worm plant; the sound depends on where the plant is hit and at what speed
- *(not included in final build due to engine bug that we later read about: the granular synth would not play in the build)* Granular synthesizer
- Standalone sound sample player

October 2020

- *(not included in final build)* Creation and completion of concept and layout for a map with head-like buildings as hotspots that the player can explore and travel between
- *(not included in final build)* Mechanics for moving platforms that can carry the player around the map

- Concept drawings for head-shaped buildings
- Ability to draw a path that a sound object can follow (not in game, only in game editor)
- Sound path can be interpolated between two predefined drawn shapes
- Sound path can be marked as an open or closed curve
- Sound path can be tracked by a single object or a group of objects with offset
- Plant that emits glittery spores when touched by the player
- 3D models for cave stalactites
- Soundtrack for cave environment
- Abstract sounds for water drops
- Recorded lyre and electronic sounds for cave ambience
- Added a persistent level and several sublevels (containing local assets) that can load and unload as the player moves through the space, for performance optimization

November 2020

- Worm plant movement partially controlled by its audio amplitude envelope
- Model and material for cave walls/columns and ground
- Sound guide function (moves in front of the player when the player moves, waits when player stops and backtracks if the player moves back)
- Multiple sound paths that react to the player coming closer by changing movement direction
- Control spore plant's particle emission depending on sound parameters
- Water drop system: system that can cyclically create a "water drop"-like object that makes sounds on collision with the ground or the player's hands
- Labyrinth section at the beginning of the map with the effect of losing visual clarity to encourage players to rely on sound for navigation
- Clarify concept for the head building
- Skin walls that ripple when touched
- *(not included in final build)* Worm plant spawners, from which worm plants grow and eject particles when the player approaches
- Platforms with moving holes that the player can walk onto
- Control size and wobble speed of the platform holes by audio parameters
- Visual effect on the sound guide that sends "sonar" waves through the environment to attract the player's attention
- Concepts for flying creatures, including the idea of an audio-visual drop.
- Music layers manager that can loop music layers and turn music layers on and off based on triggers (such as when the player enters a predefined volume)
- Update to UE4.26

- Improved quantization of audio events by switching to the Quartz system
- Sound and object moodboard and location map for the first building in the world (cave and head-shaped exit)

December 2020

- Detect when player teleports and manipulate the music: teleporting ”scares” the world around you, so all sounds get quieter and lower in pitch and then gradually returns to normal
- Tracking the number of player actions by the music layers manager
- Tracking how many times a certain music layer has been played

January 2021

- Addition of audiovisual ”sonar” ripple effect at the end of teleportation (sound becomes quieter and gradually returns to normal, visual wave effect that spreads across the environment)
- Making flying eye jellyfish model
- Modeling of cave systems - walls, columns, floor
- Modeling and integration of wobbly skin wall sections into cave walls
- Model of organic floating platforms with adjustable hole size
- Sound visualizer implemented for the music manager to show audio amplitude via particle effects
- Music manager now slowly moves to a predefined area when the player enters it, so the player can partially direct the location of the music

February 2021

- Dynamically-synthesized sounds to match wobbly wall movement and amplitude based on player’s touch velocity
- Collision sounds with worm plants
- Sounds for wobbly floating platforms
- Model inside and outside of the head-shaped building
- Music layers spin along splines (orbit shapes) and have a visual counterpart (spheres)
- Rotation movement ”look back to the player” added to sound guide
- ”Adaptive” glowworm plants: height and pitch of worms react to player’s location inside a predefined volume
- *(not included in final build)* Come up with several concepts for the way the experience could be ended

March 2021

- Sounds for organic floating platforms control the movement and size of holes
- Dynamic synthesizer sound for the trampoline effect on organic floating platforms - sound plays when player comes into contact with them by eg. teleporting
- Implementation of god-ray lights
- Interaction sounds with spores plant
- Floating crystals and origami birds with granular synthesis sound around the head-shaped building
- Distributed additional soundtrack layers between several eye jellyfish creatures that fly around the head-shaped building
- Adaptive glowworms play sequentially within the predefined level area

April 2021

- Optimization: reorganize SteamAudio sampling volumes (used for baked reverb calculation) into several large volumes connected by many smaller ones
- Finish atmospheric lighting
- Water drop splash effect
- Glow effect on tree sequencer nodes when they play sound
- Finalized layout of the level

May 2021

- Mushrooms can now be touched to play triggered samples in a loop
- Cleanup of map and project files
- Audio programming troubleshooting with Double Shot Audio: sound occlusion settings, looking into crashes, and replacing granular synthesis in the engine with prerecorded samples

June 2021

- 3D meshes optimization and LOD setup
- Added synthesizer sound effects for the X and Y coordinates of the tree eye nodes
- Added glowing tutorial cave paintings
- Audio programming advice and troubleshooting with Double Shot Audio

July 2021

- Worm plants that change to random color and pitch on impact
- Resizable sphere that changes sound based on its size (bigger = louder & lower pitch)
- Sound mushroom hit speed being used
- Consultation with Double Shot Audio: discussion on whether it is possible to play the sound of the application in real time through HAW’s loudspeaker dome (but no viable solution found)

- Added a glowing indicator on the floor to the areas that the player can enter, which causes the music manager playing soundtrack layers to move there

August 2021

- Audio programming consultation and final troubleshooting with Double Shot Audio

3.2.7 Technical Challenges

Overall, it was very difficult to achieve a well-performing VR application for several reasons.

Despite the optimization we performed, the application could not achieve anywhere near the recommended 90FPS value for VR, even on the high-end computer provided by the university. The high number of dynamically-generated sound sources in the level, either dynamically synthesized or computed with dynamic reverb, impacted performace. Additionally, the object meshes could have been reduced to even fewer polygons, and more loading and unloading of areas should have been performed.

Some users reported that the teleportation had issues: the user would not always teleport to the spot where they pointed, even though the indicator was green. This is a possible collision mesh issue or a bug in the Advanced VR Framework plugin we used.

The application build only succeeded with errors despite a lot of debugging, and the build crashed consistently around 20—30 minutes after start. The source of this crash was never discovered due to running out of time and funds, but it is theorized to be caused by memory leaks and overflow: the high number of audio resources used by the project likely were not cleared from memory correctly and kept accumulating until the program would crash. The crash log is indicative of this:

```
Assertion failed: InCapacity < (uint32)TNumericLimits<int32>::Max()
[File:D:\Build\++UE4\Sync\Engine\Source\Runtime\SignalProcessing\Public\DSP/Dsp.h] [Line: 608]
Max capacity for this buffer is 2,147,483,647 samples. Otherwise our index arithmetic will not
work.
```

It must also be noted how challenging it was to work with sound for a VR application. The ideal workflow would indeed be if one could have a sound design interface directly in VR, or added recording functionality to sound created in the virtual space. Instead, we worked in digital audio workstation software to do composition as well as in the Unreal Editor in desktop mode, and occasionally tested the outcome in VR (which ran very slowly and didn't allow for direct changes).

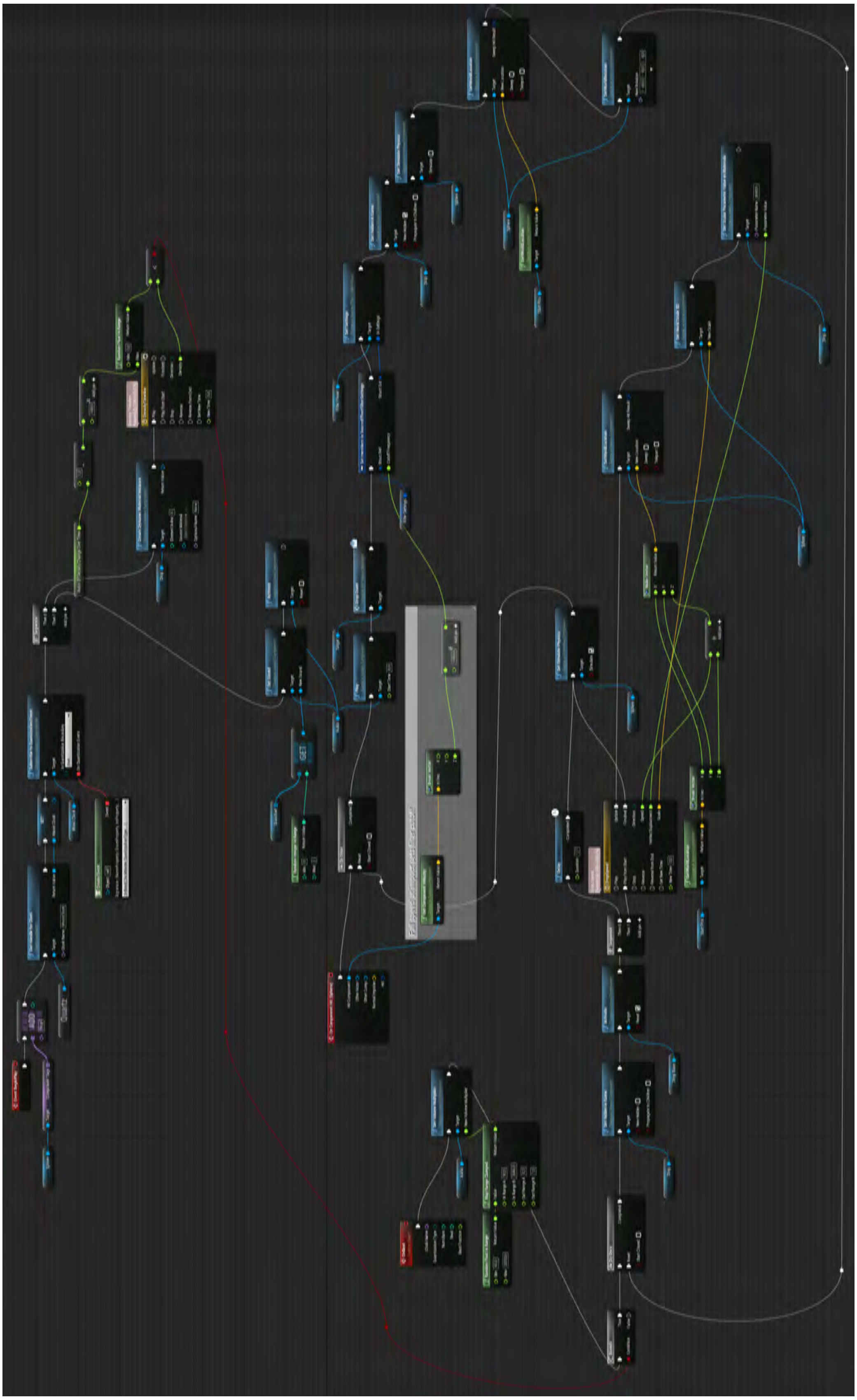


Figure 3.18: Blueprint that controls the behavior of the water drops and the sound they emit on impact

4 Study

A study with 20 participants (9 reported female, 11 male) was conducted in August 2021 to gather feedback on the application via a self-report questionnaire. Participants spent an average of around 20 minutes using the application and displayed various levels of engagement with the virtual medium. The results of this study are discussed below.

4.1 Experimental Setup and Methods

4.1.1 Measures and Equipment

The study was conducted to assess how users found the application in terms of sound quality, interaction, immersion and overall design. It was conducted in English to a majority of non-native speakers of the language (so to a certain extent, a language barrier must be assumed).

The experiment was performed in the sound studio at HAW Hamburg. A set of AKG K702 open-ear headphones was used in place of the default HTC Vive Pro headphones. The choice of open-ear headphones was dictated by the assumption that more sound from the outside world would give the user more confidence in using the medium. However, a comparison study to using closed-ear headphones would have been interesting.

Refer to Appendix on p. 53 for the questionnaire and the answers received from participants.

4.1.2 Procedure

Each participant was introduced to the HTC Vive Pro controllers and shown how to perform the only interactions possible in the experience: grabbing, teleporting and touching. They were asked if they could be filmed while using the application and helped with putting on the HMD. They were then told to explore and discover for themselves what they could do in the virtual world, with minimal intervention from the supervisor, and encouraged to stay in the experience only as long as they wished to.

The questionnaire was given to the participants right after finishing the experience, with the supervisor leaving the room.

Unfortunately, the experiment had to be conducted during the COVID-19 pandemic, and an air filtration system had to be kept on during the entire experiment. It was only realized after the conclusion of the experiment that the ventilator noise might have interfered too much with the application's sound, due to leaking through the open-ear headphones.

4.2 Results and Discussion

In this section we analyze the outcome of the performed survey. It should be noted that, for most of the participants, sadly the experiment could not be concluded at the test subject's will, as originally intended, as the application crashed for the vast majority of the participants somewhere between 20 and 30 minutes in.

General Questions and User Background

The average participant age was of 28,7 years. Only 3 of them reported no prior experience with virtual reality, while 13 reported some experience, and 4 a lot of experience. This outcome makes sense, as the participants were mostly university colleagues who would be likely to come into contact with the technology. 50% of the participants reported a lot of experience with video games and 30% some experience, which reflects an established preference in this test group for this type of media. At 30% of participants being very experienced in working with sound, and 50% having some experience, this test group was rather familiar with audio aesthetics and techniques.

At 50% of the test subjects reporting they have no access to a VR headset, the technology is as expected even in this young group not easy to come by.

Most participants reported that, in their opinion, the most interesting aspect of VR is *"experiencing three-dimensional space like in real life"* and being immersed in another world, whether *"realistic or artificial."* This seems reflective of the desire to engage with a digital medium in a way that is intuitive, physical and imaginative. One participant even described the feeling of flow very well in response to this question: *"completerly loosing the sense of the space and sense of your own body."*

The primary complaints about VR in general were the physical limitations (discomfort or sickness, cabling, the size and weight of the hardware, the movement hindrance, blurriness and unfriendliness to users with vision disabilities). Five users specifically complained about movement in VR - of course, this highly depends on what type of VR locomotion they've experienced in the past, as not all cause the same level of motion sickness. Three users also mentioned that the experience is isolating or hard to share with others — as perhaps virtual reality cuts people off from their physical surroundings while not being able to provide close connection to other humans yet.

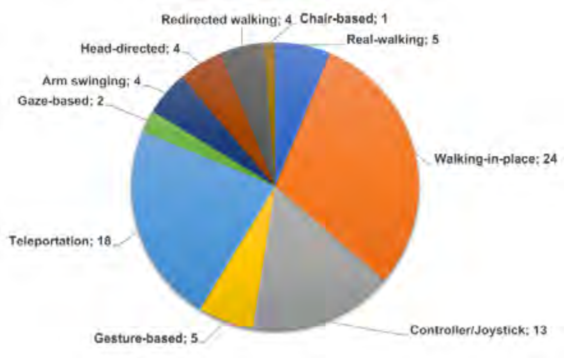


Figure 4.1: Occurrence of VR locomotion techniques encountered in 42 scientific studies from 2021, taken from Boletsis and Chasanidou, 2022.

According to Figure 4.1, walking-in-place was encountered most often in scientific studies in 2021, of which the author notes a large number used treadmills. Teleportation placed second in terms of frequency of occurrence. Of course, in practice, consumer VR applications are likely to have a different distribution of locomotion types used.

Interactivity

85% of users reported that it was clear that in Endogeny they could interact with objects. However, they were biased to do so, as this was part of the short instruction before starting the experiment. A better question to ask in this section would have probably been, *"How easy was it to realize which objects you could interact with in the experience?"* — as this question was trying to assess how good the audiovisual guidance implementation was in the experience. Overall though, the interactions that the users did engage in were reported to be rather intuitive and easy, and the controllers for HTC Vive were found easy to handle.

Sound

Interestingly, 55% of the participants reported they could not reliably tell where the sounds were localized in space in the application. Many factors could have contributed to this outcome: the fact that the binaural rendering was done with a stock HRTF instead of a personalized one, the open-ear headphones used in the experiment may have introduced too much outside noise, or some spatialization settings may have not been set correctly, as it was quite difficult during development to understand the SteamAudio parameters.

The users were asked to also rate the quality of several sound criteria in the main points of interest in the cave space: sound clarity, reverb response, envelope, loudness and intimacy. Overall, the users responded with high ratings regarding all criteria, indicating good correspondence between expectation and outcome.

Sound clarity was reported least positively in the big open area with the wobbling sound platforms, possibly because here many sounds were being played concurrently. It is difficult to assess this feedback though, as the sound clarity could have also been influenced by our choice of sound material to work with — abstract, synthesized, heavily filtered, drone.

Envelopment was rated highest on the tongue bridge and in the head-shaped building (where many soundtrack layers were flying above the user, while the tongue bridge produced sound as the user walked over it) and lowest in the tree sequencer area, which focused on the individual notes played by the tree. Loudness also rated the lowest in this section.

Loudness was reported least fitting in the tree area by those who explored it. Here it is surmised that the sounds were mixed more loudly and were therefore found to stand out too much compared to the overall loudness of the experience.

Intimacy rated lowest in the corridor with the falling water drops, followed by the tree area and the dark first corridor. In these spaces, the users perhaps expected the sound sources to feel closer, as the visual space was tighter, however the sound design chosen for the sounds there was perhaps too muffled, suble, reverberant and lacking clarity. How dark the areas were could have also factored in the illusion that the sound sources didn't appear close enough to the user.

Sound clarity was reported the least fitting in the tree area followed by first corridor. A user stated that *"I expected a stronger surround sound experience and clearer, louder sounds"* while another said *"I didn't get the question about the clarity of the sounds"*, which suggests that perhaps the question warranted more explanation.

Reverb was also reported the least fitting in the tree area and first corridor. It cannot be conclusively said why the reverb here didn't match user expectation. According to one participant, *"i did not really listen the reverberation, but i could imagine that in a sort of 'cave' like this, there should have way more reverberant environment."* It is likely that the typical user expects a cave to sound exaggeratedly reverberant, a quality not matched by the physics-based reverb produced by SteamAudio.

One user reported on the poignant role of the sound in the experience: *"Since the sound broke a bit before the crash, I could experience, how big a role the sound played in creating the immersion in the space - it felt kinda magical throughout, but once the sound was gone it turned somewhat distant and artificial."*

Reception

Most participants reported that exploration was the most significant aspect of Endogeny VR. Artistic expression through sound scored second, followed by interaction and listening to spatial audio. Storytelling, as expected, scored lowest.

The fact that listening to the spatial cues scored rather low compared to exploration can be interpreted through the previous finding that the spatial audio aspect of the experience was not noticeable enough to many users.

Questions on this VR Application

In terms of orientation and navigation of the space, participants were split. Most users found it more or less clear where they could go and what they could do in the space, typically needing some time to learn their their way around the application. Six noted big general difficulties and one player never managed to make it past the starting area.

A good percentage of the participants didn't explore or discover certain areas of the cave (sequencer tree room: 6 didn't discover it; water drops corridor: 8). This can be attributed to the level design possibly not offering enough clues to direct the user to those spaces neither via sound nor visuals, while the path towards the head building was well-lit, large, and displayed motion and direct audio activity.

Four users reported that the teleportation pointer had issues: the player would not always get teleported to the spot where they pointed, even though the indicator was green. This is a possible mesh collision issue or a problem with the VR plugin we used and it frustrated some participants.

Similarly, some users reported controls and collision quirks. One user claimed that *"touching things was not as easy, like as if there was a hotspot inside the things as supposed [sic] to the whole object being responsive to my touch"*. The collision meshes of the 3D objects were generated with a low level of detail for performance optimization purposes, which is a possible cause for this. Additionally, the bendable worm plants had an internal mesh skeleton made up of only several bones, and collision could only occur at the bone joints (as the surface was dynamically changing, there was not a pre-generated collision mesh).

Surprisingly, feedback showed that participants thought the audiovisual effect triggered by teleportation was suitable to the action (ie. associated with spatial distortion) and were not deterred by it from moving around mainly by means of teleporting. It's likely that the audiovisual effect was still too harmonious and not annoying enough to make players want to avoid this action, or that another approach, for example simply reminding users to favor walking, might have been more effective.

In terms of sound interactions, often users reported themselves wishing for more abilities and control, for example by having more range and finesse over the sound (*"The interactions could have more dynamics (they were more like trigger based)"*), by being able to interact with objects which were out of reach (for example flying objects that the user couldn't catch) or move vertically (eg. *"flying", "have a ladder to properly 'grow' the tree"*). A couple users also wished for more traditional ways of music creation. A few others mentioned that, while the sound was perceived as the primary goal of the experience, it was often obscured by the complexity of the visuals.

Many users disclosed confusion and disappointment for the experience lacking a clear endpoint. This does indincate that generally a VR sandbox experience should still offer the option for the user to exit it, especially if it offers guidance in the beginning.

How the application employed game design and progression was in some cases reported on: a couple users would have wanted a goal or a reward from certain interactions, while outside of the survey a participant suggested that the tree sequencer was too complex and that the experience should have had several, more simple trees, where the player could learn what growing them means and which exact parameters can be manipulated.

The users reported very little nausea from VR use and a good general sense of immersion and bodily presence in the virtual space. Overall, the intriguing sound design, the dreamy atmosphere and visuals, and the ability to interact and play were reported to be the best parts of the experience.

5 Conclusion

All in all, developing Endogeny VR was a very interesting creative challenge, whereby interesting solutions and novel ways of expression were found and exercised with state-of-the-art virtual reality and game engine tools. The general feedback of the testers was positive and their engagement was good, making this artistic research experiment an overall success.

As with any artwork that one produces, there are many lessons to learn from the experience, and things that in hindsight could have been achieved or done differently.

The project was ambitious for the limited resources available and for the size of the team. In hindsight, the scope could have been reduced to focus more deeply on specific aspects. Some features could have been cut down, while focusing on the essential ones and polishing them instead. Formulating the research question more clearly before development would have also helped in this regard.

Designing an experience audio-centrically while aiming to make all other aspects of gameplay well-rounded can lead to a result that feels lacking in one of its facets, in this case of progression and story. There was not enough time to invest into game design, goals, rewards, variations on an established theme, or multi-step challenges (where your actions at a step translate to effects multiple steps down) which would engage the user in hour-to-hour gameplay.

In the future, Endogeny VR could be improved and expanded to include more features that enhance the experience. Personal user HRTFs could be integrated to make for a more accurate binaural experience. It can also easily be conceived how an application like Endogeny VR could integrate multi-user participation, as the experience is already designed as a collaboration between the pre-composed musical background and what the participant produces.

Adding more participants would increase the level of audio diversity in this collaborative soundscape, whether the users help each other create, or on the contrary, destroy or tamper with each other's creations. More interaction and manipulation mechanics could be implemented to allow users to do this. Moreover, virtual sound ”toys” with more degrees of control could also be implemented to satisfy users with more advanced musical training. Adding modular sonic elements that can be combined into more complex composition tools could exponentially increase what can be done and experienced in the virtual world.

Bibliography

Abuhamdeh, S. (2020). Investigating the “flow” experience: Key conceptual and operational issues. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00158

Atherton, J., & Wang, G. (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

Boletsis, C., & Chasanidou, D. (2022). A typology of virtual reality locomotion techniques. *Multimodal Technologies and Interaction*, *6*(9). https://doi.org/10.3390/mti6090072

Bown, J., White, E., & Boopalan, A. (2017). Chapter 12 - looking for the ultimate display: A brief history of virtual reality. In J. Gackenbach & J. Bown (Eds.), *Boundaries of self and reality online* (pp. 239–259). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-804157-4.00012-8

Buckley, Z., & Carlson, K. (2019). Towards a framework for composition design for music-led virtual reality experiences. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 1497–1499. https://doi.org/10.1109/VR.2019.8797768

Çamcı, A., & Granzow, J. (2019). Hyperreal Instruments: Bridging VR and Digital Fabrication to Facilitate New Forms of Musical Expression. *Leonardo Music Journal*, *29*, 14–18. https://doi.org/10.1162/lmj_a_01056

Çamcı, A., & Hamilton, R. (2020). Audio-first vr: New perspectives on musical experiences in virtual environments. *Journal of New Music Research*, *49*(1), 1–7. https://doi.org/10.1080/09298215.2019.1707234

Cerasoli, C., Nicklin, J., & Nassrelgrgawi, A. (2016). Performance, incentives, and needs for autonomy, competence, and relatedness: A meta-analysis. *Motiv Emot*, (40), 781–813. https://doi.org/https://doi.org/10.1007/s11031-016-9578-2

Chalmers, D. (2018). The virtual and the real. *Disputatio*, *9*, 309–352. https://doi.org/10.1515/disp-2017-0009

Corporation, H. (n.d.). Vive now shipping immediately from htc, retail partners expand demo locations [Accessed: 2022-07-31]. https://www.htc.com/us/newsroom/2016-06-07/

Csikszentmihalyi, M. (2014). Toward a psychology of optimal experience. *Flow and the foundations of positive psychology*, 209–226. https://doi.org/10.1007/978-94-017-9088-8_14

Deci, E. L., & Ryan, R. M. (2000). The ”what” and ”why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Egbert, J. (2004). A study of flow theory in the foreign language classroom. *Canadian Modern Language Review 60*, 549–586. https://doi.org/10.3138/cmlr.60.5.549

Endogeny [Accessed: 2022-11-23]. (n.d.). https://www.collinsdictionary.com/dictionary/english/endogeny

Fırat, H. B., Maffei, L., & Masullo, M. (2022). 3d sound spatialization with game engines: The virtual acoustics performance of a game engine and a middleware for interactive audio design. *Virtual Reality*, *26*(2), 539–558. https://doi.org/10.1007/s10055-021-00589-0

Hamilton, R. (2019). Trois machins de la grâce aimante: A virtual reality string quartet.

Harter, S. (1978). Effectance motivation reconsidered. toward a developmental model. *Human Development*, *21*, 34–64. https://doi.org/10.1159/000271574

Hartmann, T., Wirth, W., Vorderer, P., Klimmt, C., Schramm, H., & Böcking, S. (2013). Spatial presence theory: State of the art and challenges ahead. Oxford: Routledge. https://doi.org/10.1007/978-3-319-10190-3_7

Hassan, L., Jylhä, H., Sjöblom, M., & Hamari, J. (2020). Flow in vr: A study on the relationships between preconditions, experience and continued use. https://doi.org/10.24251/HICSS.2020.149

Hofman, P. M., Van Riswick, J. G., & Van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nature Neuroscience*, *1*(5), 417–421. https://doi.org/10.1038/1633

Holm Pedersen, T. (2008). *The semantic space of sounds.*

Ijaz, K., Ahmadpour, N., Wang, Y., & Calvo, R. A. (2020). Player experience of needs satisfaction (pens) in an immersive virtual reality exercise platform describes motivation and enjoyment. *International Journal of Human-Computer Interaction*, *36*(13), 1195–1204. https://doi.org/10.1080/10447318.2020.1726107

Killick, A. (2006). Holicipation: Prolegomenon to an ethnography of solitary music-making. *Ethnomusicology Forum*, *15*(2), 273–299. https://doi.org/10.1080/17411910600915414

Kilteni, K., Groten, R., & Slater, M. (2012). The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments*, *21*(4), 373–387. https://doi.org/10.1162/PRES_a_00124

Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: Manipulating bodily self-consciousness. *Science*, *317*(5841), 1096–1099.

Lombard, M., & Ditton, T. (1997). At the Heart of It All: The Concept of Presence [JCMC321]. *Journal of Computer-Mediated Communication*, *3*(2). https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Lüneburg, B. (2019). *Transcoding - from 'highbrow art' to participatory culture: Social media - art - research.* transcript Verlag. https://doi.org/doi:10.1515/9783839441084

Marks, L. (1978). 2 - the doctrine of equivalent information. In *The unity of the senses* (pp. 11–48). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-472960-5.50007-X

Marks, L. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of experimental psychology. Human perception and performance*, *13*, 384–94. https://doi.org/10.1037/0096-1523.13.3.384
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0
Microtonal music [Accessed: 2022-11-23]. (n.d.). https://en.wikipedia.org/wiki/Microtonal_music
Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. *Flow and the foundations of positive psychology*, 239–263. https://doi.org/10.1007/978-94-017-9088-8_16
Novak, T., Hoffman, D., & Yung, Y.-F. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science*, *19*, 22–42. https://doi.org/10.1287/mksc.19.1.22.15184
Pan, X., & Hamilton, A. F. d. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. https://doi.org/https://doi.org/10.1111/bjop.12290
Patchworld - make music worlds [Accessed: 2022-11-19]. (n.d.). https://www.oculus.com/experiences/quest/3715150718552632/
Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H., & Alt, F. (2022). What is xr? towards a framework for augmented and virtual reality. *Computers in Human Behavior*, *133*, 107289. https://doi.org/https://doi.org/10.1016/j.chb.2022.107289
Robeyns, I., & Byskov, M. F. (2021). The Capability Approach. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University.
Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American psychologist*, *55*, 68–78. https://doi.org/10.1037/0003-066X.55.1.68
Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, *30*(4), 344–360. https://doi.org/10.1007/s11031-006-9051-8
Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, *6*, 332–339.
Sen, A. (1999). *Commodities and Capabilities*. Oxford University Press. https://ideas.repec.org/b/oxp/obooks/9780195650389.html
Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7
Statt, N. (2019). Htc announces new vive pro eye virtual reality headset with native eye tracking. https://www.theverge.com/2019/1/7/18172700/htc-vive-pro-eye-tracking-virtual-reality-headset-features-pricing-release-date-ces-2019
Stumpf, C. (1883). *Tonpsychologie i*. Leipzig: VERLAG VON S. HIRZEL. https://pure.mpg.de/rest/items/item_2389927_3/component/file_2456998/content

Surreal soundscapes: Explore musical worlds + build your own in 'patchworld' on meta quest 2 [Accessed: 2022-11-19]. (2022). https://www.oculus.com/blog/surreal-soundscapes-explore-musical-worlds-and-build-your-own-in-patchworld-on-meta-quest-2/

Tiippana, K. (2014). What is the mcgurk effect? *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00725

Trommer, M. (2020). Points further north: An acoustemological cartography of non-place. *Journal of New Music Research*, *49*(1), 73–87. https://doi.org/10.1080/09298215.2019.1704020

Unreal engine 4.26 documentation: Quartz overview [Accessed: 2022-11-19]. (n.d.). https://docs.unrealengine.com/4.26/en-US/WorkingWithAudio/Quartz/

Unreal engine 4.26 documentation: Submixes [Accessed: 2022-11-22]. (n.d.). https://docs.unrealengine.com/4.26/en-US/WorkingWithAudio/Submixes/

Valve-Corporation. (n.d.). *Steam audio api 2.0-beta.19* [Accessed: 2022-09-24]. https://web.archive.org/web/20201205010343/https://valvesoftware.github.io/steam-audio/doc/capi/index.html

van Dorp Schuitman, J., de Vries, D., & Lindau, A. (2013). Deriving content-specific measures of room acoustic perception using a binaural, nonlinear auditory model. *The Journal of the Acoustical Society of America*, *133*(3), 1572–1585. https://doi.org/10.1121/1.4789357

Weber, S., Weibel, D., & Mast, F. W. (2021). How to get there when you are there already? defining presence in virtual reality and the importance of perceived realism. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.628298

Wilkinson, D. (2021). Caloric and galvanic vestibular stimulation for the treatment of parkinson's disease: Rationale and prospects. *Expert Review of Medical Devices*, *18*. https://doi.org/10.1080/17434440.2021.1935874

Woeste, H. (2009). A history of panoramic image creation [Accessed: 2022-07-31]. https://web.archive.org/web/20180114065712/http://www.graphics.com/article-old/history-panoramic-image-creation

Ziemer, T. (2020). *Psychoacoustic music sound field synthesis: Creating spaciousness for composition, performance, acoustics and perception*. Springer Nature Switzerland. https://doi.org/https://doi.org/10.1007/978-3-030-23033-3

Appendix

.1 Questionnaire Contents

The questionnaire featured the following questions:

1. **General Questions**
   1.1. Age: ____________
   1.2. Gender: ____________
   1.3. Agree to be filmed (Yes / No)
   1.4. How experienced are you with VR prior to this project?

No experience	Some experience	Very experienced
○	○	○

   1.5. How experienced are you with working with sound?

No experience	Some experience	Very experienced
○	○	○

   1.6. How experienced are you with video games?

No experience	Some experience	Very experienced
○	○	○

   1.7. Do you have access to a VR headset at home/uni/work? (Yes / No)
   1.8. What would you say are the most interesting benefits of VR? ____________
   1.9. What would you say are the most frustrating constraints of VR? ____________

2. **Interactivity**
   2.1. Was it clear in this application that you can interact with objects? (Yes / No)
   2.2. How intuitive was it to interact with objects? (Unintuitive 1 — 5 Intuitive)
   2.3. How easy was it to interact with objects? (Difficult 1 — 5 Easy)
   2.4. How easy was it to handle the VR controllers? (Difficult 1 — 5 Easy)

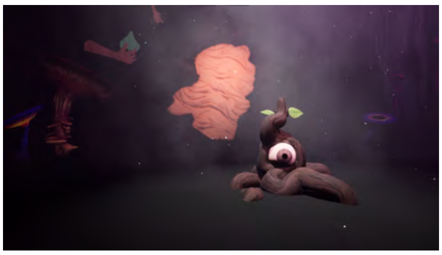
3. **Sound**
   3.1. Did you notice different spaces in the environment with different acoustic properties? (Yes / No)
   3.2. Could you reliably tell where the sounds were localized in space? (Yes / No)
   3.3. Did your perceived sound sources correspond to their visual object counterparts? (Yes / No)

3.4. Rate the quality of the sound criteria in Table 3.4. from 1 — 5 in the following environment sections:

	Very poor	2	3	4	Very good	Didn't explore it
Clarity of sound detail	○	○	○	○	○	○
Reverberant response: how suitable to the space was the perceived sound reverberation?	○	○	○	○	○	○
Envelope: how well did you feel enveloped by the sound?	○	○	○	○	○	○
Loudness: was it adequate in terms of distance to the object and in the space?	○	○	○	○	○	○
Intimacy: the feeling of being physically close to the sound sources	○	○	○	○	○	○

3.4.1. In the first corridor

3.4.2. In the tree area



3.4.3. In the big open area with the floating platforms

3.4.4. In the pink glowing worms area

3.4.5. In the corridor with the falling water drops

3.4.6. On the tongue bridge

3.4.7. Inside the head-shaped building

4. Reception

4.1. What was the experience about? Rate on a scale of 1 — 5 the significance of the following elements in the experience:

	Very insignificant	2	3	4	Very significant
Exploration	○	○	○	○	○
Listening, spatial audio experience (more technical approach)	○	○	○	○	○
Interacting with objects/environment	○	○	○	○	○
Music (artistic sound expression)	○	○	○	○	○
Storytelling	○	○	○	○	○

4.2. How would you describe this project?

○ Game

○ Interactive installation

○ Immersive experience

○ Interactive music video

○ Other:

5. **Questions on this VR Application**

5.1. Did you feel during the experience that you knew what to do next? If no, can you elaborate why?

5.2. What do you wish you could do in the application that you were not able to do?

5.3. Did you feel that overall the application made good use of the potential of the VR medium?

5.4. What were the best parts about this application?

5.5. What were the worst parts about this application?

6. **Immersion**

6.1. Did you experience sickness or nausea from being in VR?

6.2. Did you have a sense of bodily presence in the virtual space?

6.3. Did you feel immersed in the virtual environment?

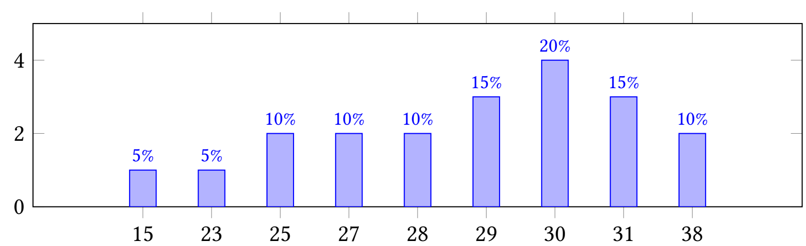
7. **Additional Feedback**

7.1. If you have additional feedback please comment here: ___________

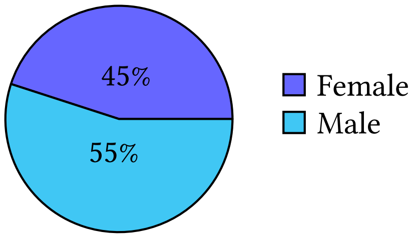
.2 Participant Answers

General Questions

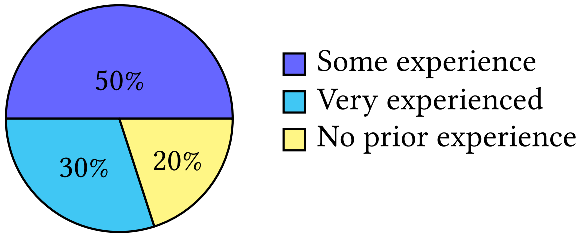
Age *(20 responses)*



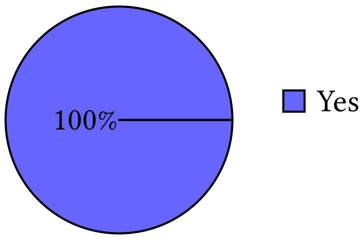
Gender *(20 responses)*



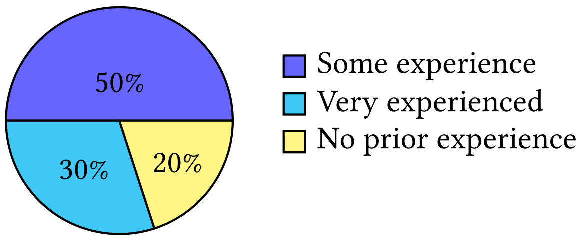
How experienced are you with working with sound? *(20 responses)*



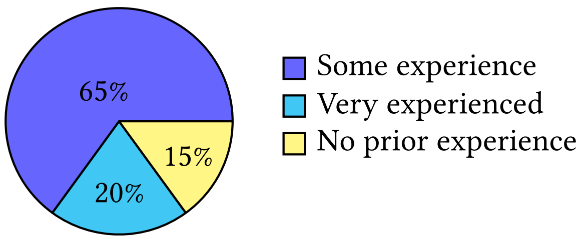
Agreed to being filmed *(20 responses)*



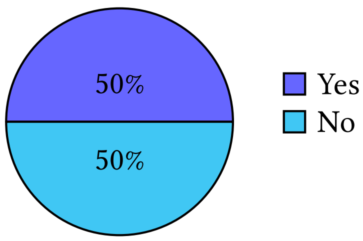
How experienced are you with video games? *(20 responses)*



How experienced are you with VR prior to this project? *(20 responses)*



Do you have access to a VR headset at home/uni/work? *(20 responses)*



What would you say are the most interesting benefits of VR? *(20 responses)*

"Potential of training and learning applications"

"Simulation"

"try somethings out, experience things, gamification elements, show things that are not real"

"the immersion of completely diving into another world/scenario"

"experiencing three-dimensional space like in real life"

"Creeating virtual environmets and explore them organically"

"Obviously the immersion. To be moving in RL while being in VR does a lot to the experience"

"Intuitive interaction, audiovisual immersion"

"possibilities to look around and move 'like normal"'

"Creating virtual worlds"

"immersion"

"It's very immersive"

"Totally imersive experience"

"Creating new spaces that could not exist in reality"

"The possibility to explore 'new' worlds"

"being completely emerged into the game;"

"to be interactive with sound"

"To be able create any kind of environment, realistic or artifical, with or without following the laws of physics (eg gravitiy)"

"immersive experience, you can get lost and forget the real surrounding for a while"

"completerly loosing the sense of the space and sense of your own body"

What would you say are the most frustrating constraints of VR? ***(20 responses)***

"Price of the headsets and the hardware to run it, cable on the back, putting headset on and off while while developing for it"

"get addicted maybe"

"if the teleportation doesn´t work right or if you getting sick because of to fast changes after the teleporting"

"1. the tech: low field of view, still not perfect resolution, heavy headset
2. no good locomotion system yet"

"low resolution, bulky headset"

"Not beeing able to walk, danger of hitting something, very isolating experience"

"tempo (usage of the app is usually pretty slow; compared to for example modern games). then it's the cables and the constraints you get from being in a physical room."

"locomotion is difficult to achieve, hard to share the experience with someone else"

"headset hurts my head; stuff often gets unfocused when it's not right in the middle of the screen; traveling via joystick input makes me feel sick immediately (teleport is fine, tho); cables"

"Uncomfortable headsets, motion sickness"

"heavy gear, spacial constraints, blurry vision"

"not real enough, at some point you realise it's just a projection as in you can't feel stuff physically"

"the hardware is too big"

"learning how to properly interact with it, and how little access there is to it"

"I got a little bit tired (eyes tired)"

"Being physically limited due to the cables/ the room + loosing your orientation can be frustrating"

"not knowing what to do"

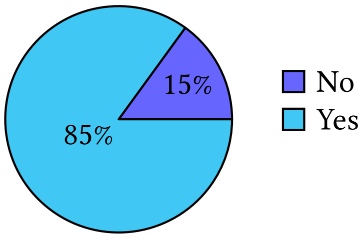
"That the timing often lags, so reaction to an action are slow and late"

"you can't use your 'normal' hands, interaction with others is missing"

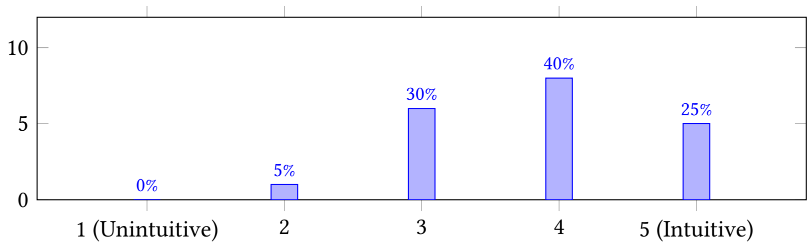
"still the weight of the classes.. and people with vision disabilities cannot really find the best settings, but nothing really to do about that.."

Interactivity

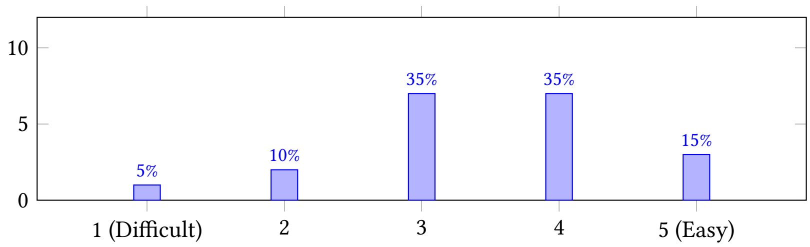
Was it clear in this application that you can interact with objects? *(20 responses)*



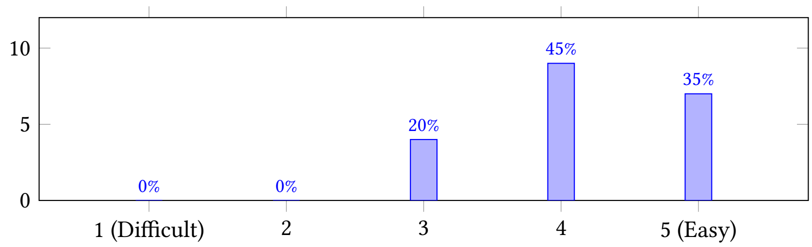
How intuitive was it to interact with objects? *(20 responses)*



How easy was it to interact with objects? *(20 responses)*

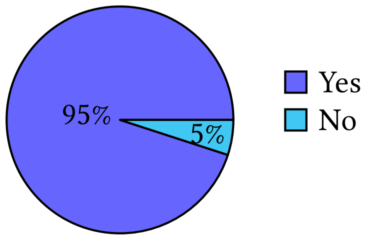


How easy was it to handle the VR controllers? ***(20 responses)***

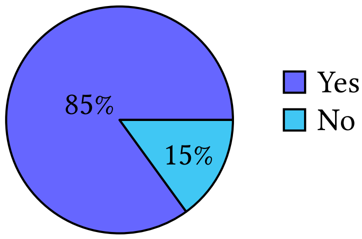


Sound

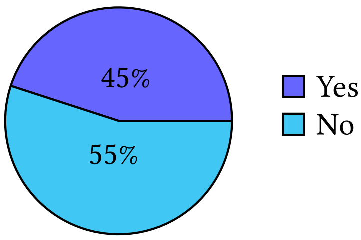
Did you notice different spaces in the environment with different acoustic properties? ***(20 responses)***



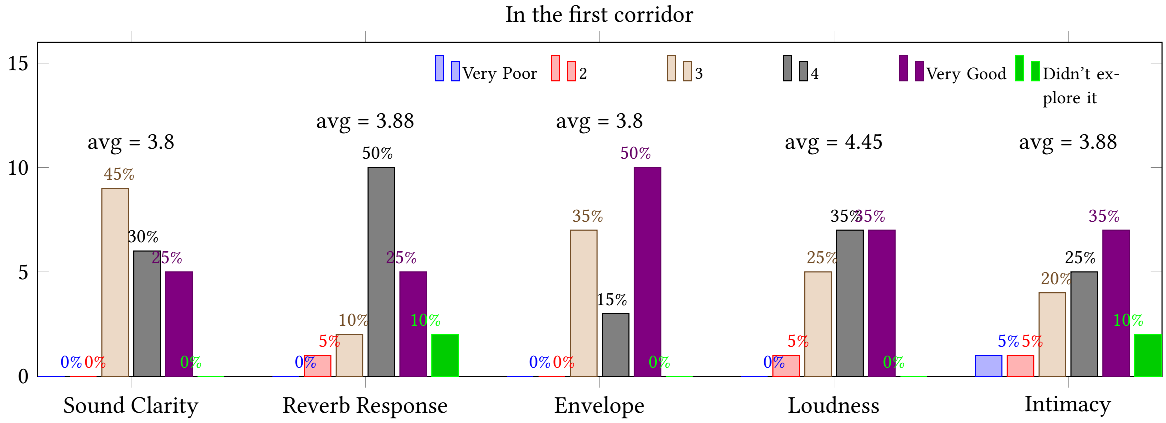
Did your perceived sound sources correspond to their visual object counterparts? ***(20 responses)***

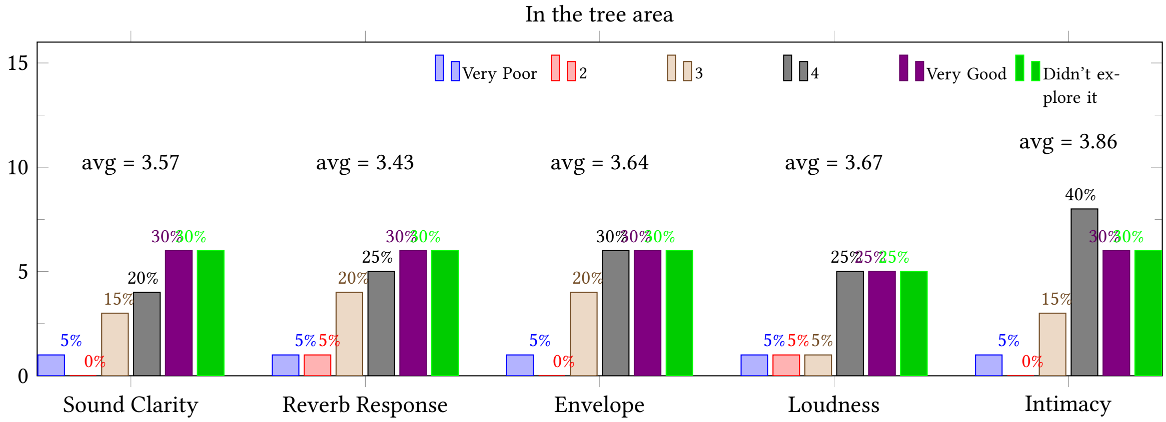


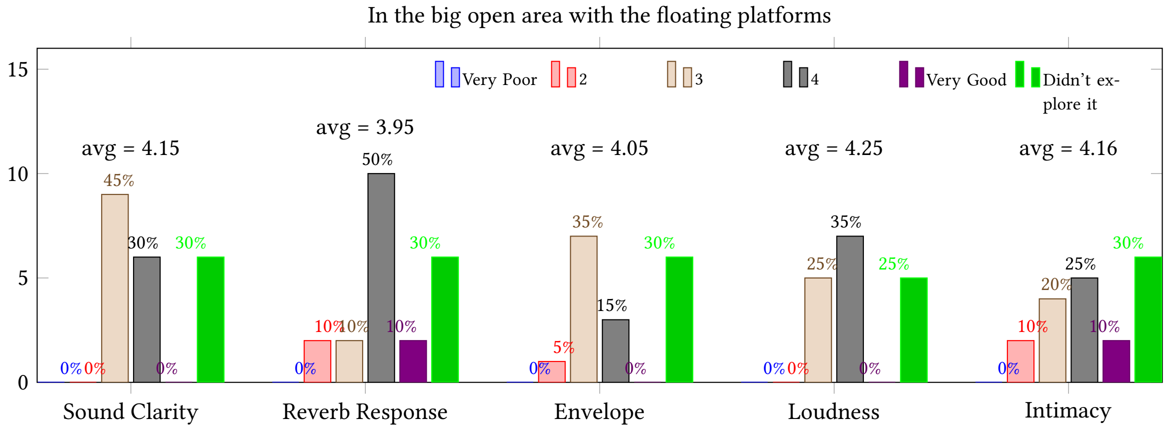
Could you reliably tell where the sounds were localized in space? ***(20 responses)***



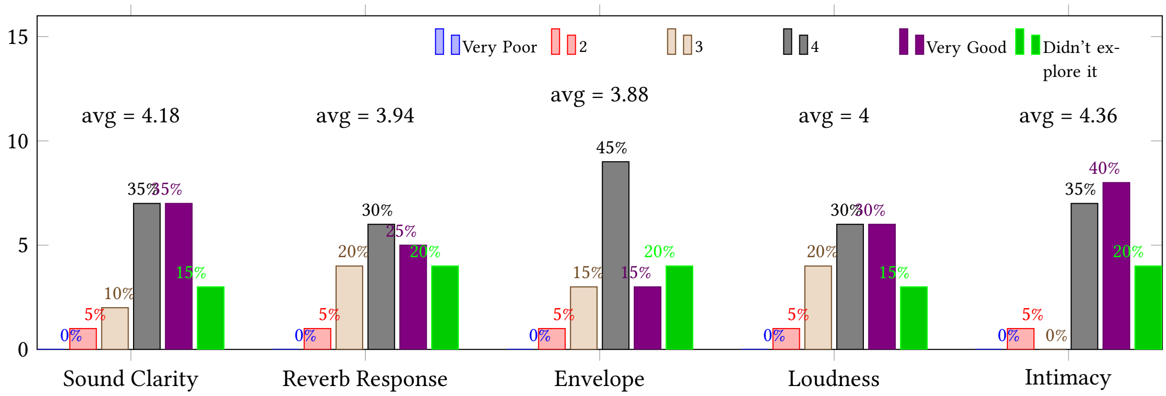
Rate the quality of the following sound criteria from 1—5 in the following environment sections:



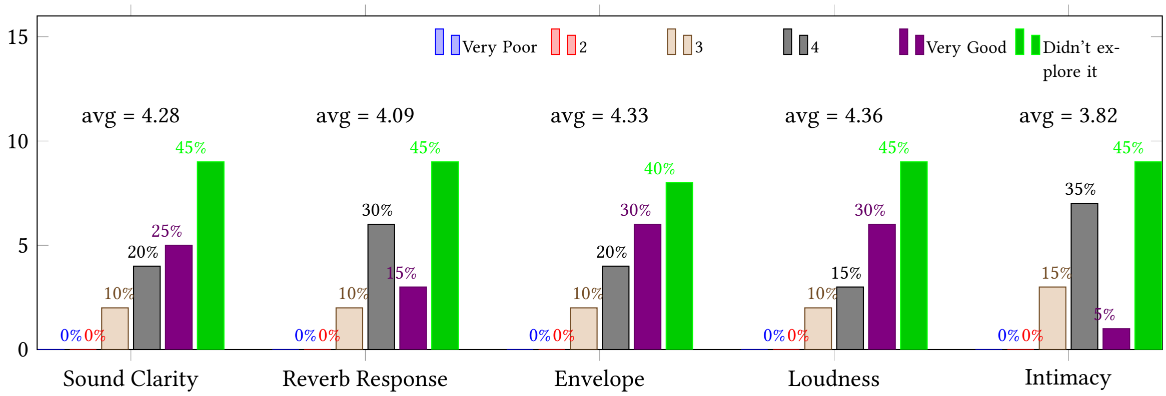




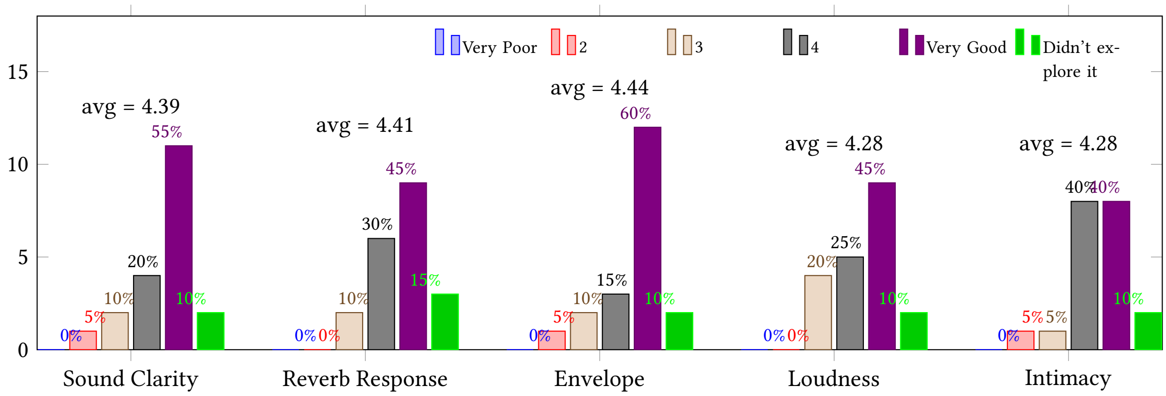
In the pink glowing worms area



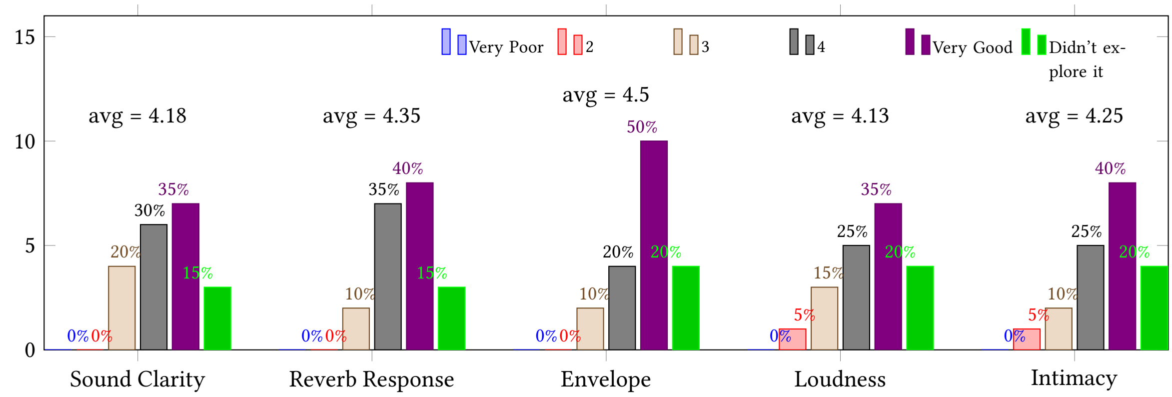
In the corridor with the falling water drops



On the tongue bridge

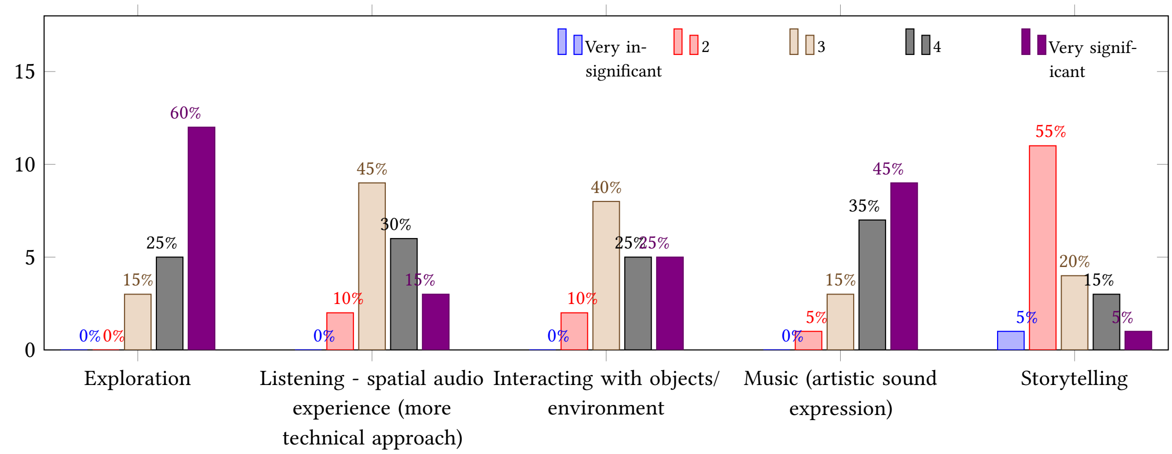


Inside the head-shaped building

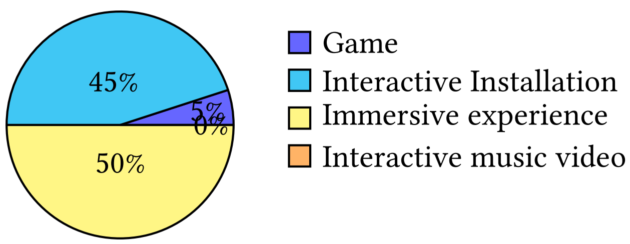


Reception

What was the experience about? Rate on scale 1—5 the significance of the following elements in the experience:



How would you describe this project? *(single choice, 20 responses)*



Questions on this VR Application

Did you feel during the experience that you knew what to do next? If no, can you elaborate why?

"No, I didn't feel like I was lead to any specific area or interaction by visual or acoustic indicators"

"Firstly, i got no idea, but then i follow the 'ghost"'

"sometimes I was a little lost where to go and which parts I saw, because the different locations looked very similar"

"It felt rather linear, with the floating eye guiding me through the corridors. There were some side paths to explore, which could have been lighted better to make clear you can go there. The ending in the skull was a bit unspectacular"

"Yes - the floating eye character and the semi-linear level architecture told me where I could go"

"I did'nt know I had reached the end of the path and there was nowhere else to explore"

"It was clear to follow the eye to the 'playground'. From there it was somehow clear that this is for exploration but latest on the 'tongue' I had difficulties finding out where I 'have to' go next. Everything felt like leading towards a (bigger) 'end' but then you leave the 'head shaped building' again and you are lost."

"At some point I felt I had explored every area and didn't know how to proceed"

"I wsa guided by the eyeball to cricles on the ground and there were hints on the walls, but I kept feeling that I was missing something or that I could be performing 'better'?"

"Its very dark, and the controlls woudn't always do as I intendet."

"i didn't know exactly. from the instructions at the beginning, i assumed you should follow the eyeball, but i lost track of it after some exploring, and that you could grab hold of it, but i didn't manage to do that"

"It felt like a free world so as to be explored. it felt like an adventure so it was immersive and inviting in that regard"

"yes, after a while"

"At the start i was very confused about how to interact with the objects and where to go, but once i got the hang of it i pretty much never run out of things to explore or move"

"Yes, although the controllers didn't respond the same way always"

"It took me a while to figure out how to coreccly follow the flowting eye"

"plenty of things you could see but still not being able to decide what to do"

"It wasn't super clear what to do, especially in the beginning (and what not to do), which objects are objects to interact with, where to go. Also the controlling lagged quite a bit, so teleporting and sound creation always took a while, which was a little confusing"

"No, I didn't find my way through the installation. I think it was too much for me discover in the space"

"i did feel like i knew what to do next"

What do you wish you could do in the application that you were not able to do?

"reset to some kind of checkpoint or the nearest 'walkable' area; have more posibilities in interacting with objects or more objects to interact with"

"Flying"

"interact with to things and my two hands at the same time, slap on the mushrooms like a drum or something"

"The scalable eyes in the beginning suggested a more direct control of the sound sources, I wish that would've been possible. To shape the sound in the corresponding areas"

"I wish the 'wall of tubes' behind the tongue had a reward (sound effect, VFX, etc.) if you touch them all. Also I would have liked to teleport a little bit farther."

"Interact with all the flying objects somehow, or with objects just out of reach Tear out the spongy pillars, reset the tree, have a ladder to properly 'grow' the tree, safe/harvest sounds."

"Throw the eyeballs; activate things that keep on producing sound, so as to create lasting harmonies in space"

"I wanted more control over the sounds that were produced and to create more 'musical' music, but the sound felt random to me so I couldn't combine them in a sensible manner. Also it would have been nice to use two hands to create diffent sounds (like pulling one of the strings longer to change the sound you create with the other hand or something like that)"

"Play an instrument or something!"

"grab hold of the eyballs and origami birds"

"Interact with the moving and flying objects"

"touch the mushrooms"

"interact with the paper cranes"

"The interactions could have more dynamics (they were more like trigger based)"

"-"

"interact with more objects"

"Sound creation could have been a more (or I didn't find it - I know it's not easy to implement either)"

"ask for direction"

"there could have been more objects to touch and throw around, also objects that would touch each other. would have also been nice to jump directly up."

What were the best parts about this application? *(20 responses)*

"Nice sound design and visuals"

"X"

"i really liked the worm in the head and the tree with the eyes"

"The visual effects and the mysterious mood, sounds were sometimes good but could use improvement (maybe just the volume, or the spatial sound wasn't too impressive)"

"overall dreamy mood, eye tree"

"The tongue bringe and the climb out of the ear"

"I loved the vocal samples in the head, the feeling of finding a hidden treasure in the dark when I found the mushrooms with glitter, and the nice bass on the one of the hanging plants."

"very nice, intriguing, atmosphere"

"I like exploring and the looks were engaging, my favourite part was the colour changing worm things at the end, especially the leftmost one which was singing."

"Sound Design, Visuals"

"for me, the visuals and the atmosphere in the cave bit"

"The sound, Moving eye things, big open world experience, visuals"

"touching objects and getting sound response"

"The aesthetic was quite inmmersive and consistent, this applies to all, graphics, music, sound design etc."

"The tonge. And the raindrops room"

"The colorchanging worms were very fun, the tree and the tongue bridge"

"being able to interact"

"The great visuals, I really loved just watching the different rooms and get emersed in the color and also sound space"

"the interaction with objects, surprised what sound I'll hear"

"the best was the playfulness and a feeling of exploration. i also found interesting feeling of danger when i tried to touch something outside the blue grid"

What were the worst parts about this application? *(20 responses)*

"No guidance"

"X"

"for me the mushroom had mostly the same function, it would be great if i could interact with them in different was or on the different levels of the mushroom"

"no meaningful interactions and no proper ending"

"the tongue sounds"

"Having to position yourself in a way to be able to reach the objects that you want to interact with"

"Was frustrated with not finding an ending to the experience. I expected there to be one because of having a clear beginning and guide at the start. A tree reset would have been nice. I also couldn't tell where the sounds were coming from in the head or from the birds"

"it crashed :P"

"I missed some kind of goal"

"at some point I felt like I went into a deadend of blackness and it took forever to get out, that controls (which I am used from robo recall) didn't always feel like they put me where I wanted to be and also touching things was not as easy, like as if there was a hotspot inside the things as supposed to the whole object being responsive to my touch"

"Usability"

"the controls were a little wonky sometimes"

"the interaction with the objects could be smoother"

"teleporting failed sometimes"

"getting to learn how to use it"

"I kind of got lost at some point and I'm not so sure how did I came back. And it crashed in the end :c"

"Sometimes I had troubles teleporting; I also sometimes felt like I didn't really know where to go next"

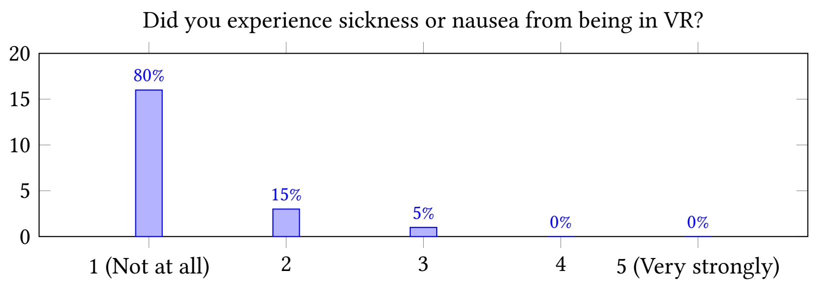
"you have to find out that you should follow the eye, takes pretty much most of the time"

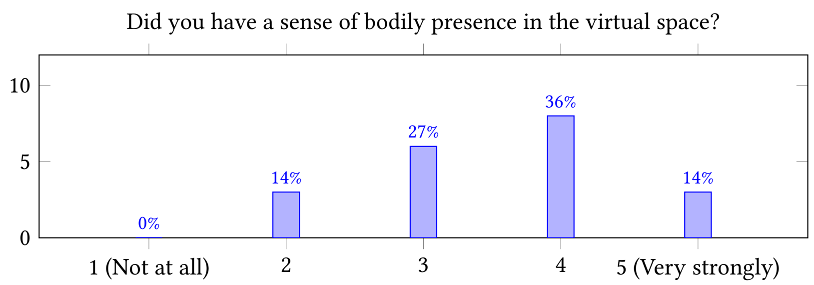
"That the controlling lagged and I didn't know if I successfully teleported or not. Maybe a little tutorial would be nice so that participants would be less 'lost'?"

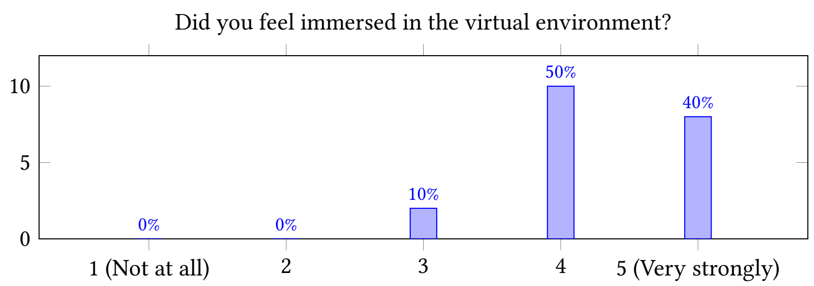
"the space seemed very small for me, I often saw the grid"

"I found sound not interacting enough - the different spaces were not really clearly distinguishable in terms of sound."

Immersion







If you have additional feedback please comment here:
(15 responses)

"Nice work!"

"I really liked the expericence, there are great features and a lot to explore. I think I wish a little bit more different elements or functions within the elements. Great job!"

"overall enjoyable and 'artsy' in a positive way, i feel like the main goal - the sound - could still use improvement, but maybe that's just me. I expected a stronger surround sound experience and clearer, louder sounds. Also the ability to control the sound more directly could be fun"

"wish list:
- 'end': goal/exit point at the end of the linear path
- resolution for the floating eye
- reward for climbing inside the head
- different tube types (red and blue) make different sounds"

"I didn't get the question about the clarity of the sounds"

"A story about my Uncle. Would be a good reference, since the world transmit the same feeling."

"I like the approach of building surreal worlds. Since the sound broke a bit before the crash, I could experience, how big a role the sound played in creating the immersion in the space - it felt kinda magical throughout, but once the sound was gone it turned somewhat distant and artificial."

"it was certainly strange. please continue your good work."

"fun stuff .. gg xD"

"did the first corridor wall drawing mean I should follow the eye (or that I should try reshaping the toucheable eyes)? thanks!"

"I think it was a suuuuper cool experience. I think it would be nice to be able to walk instead of teleporting."

"great experience"

"Thanks for the experience!"

"What a pity that I saw just a few parts of it! The objects I interacted with were very impressive (design and sounds)"

"the sounds could change more its character in terms of the actions and different spaces. it would be nice if the overall sound would change in different rooms - this would also be more playful and suprising for the listener. also the jumping from one room to another could produce some kind of sound to give a sense of actual 'jump'. In general i think that sound enviroment could be more dominant as the visual cues take so much space that it takes time to even listen what is going on with sounds, on the other hand i think that more dominant sound environment could even give the listener stronger feeling of being immeresd in the space. also i did not really listen the reverberation, but i could imagine that in a sort of 'cave' like this, there should have way more reverberant environment."

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Masterarbeit mit dem Titel

Flow, intuitive interaction design and music production in an audio-first VR experience

selbstständig und nur mit den angegebenen Hilfsmitteln verfasst habe. Alle Passagen, die ich wörtlich aus der Literatur oder aus anderen Quellen wie z. B. Internetseiten übernommen habe, habe ich deutlich als Zitat mit Angabe der Quelle kenntlich gemacht.

Declaration of Independent Work

I hereby declare that I have written this Master's thesis titled

Flow, intuitive interaction design and music production in an audio-first VR experience

independently and only with the indicated aids. All passages that I have taken verbatim from the literature or from other sources, such as internet pages, I have clearly marked as a quotation with indication of the source.

Hamburg, November 28, 2022