Chair for Computer Science 10 (Media Computing and Human-Computer Interaction)

RWTH AACHEN UNIVERSITY

Exploring Audio-Derived Immersive Environments in Virtual Reality for Electronic Music Producers

Bachelor's Thesis at the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Aaron Sergeev

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Ulrik Schroeder

Registration date: 13.06.2025
Submission date: 13.10.2025

Zentrales Prüfungsamt/Central Examination Office

Eidesstattliche Versicherung
Declaration of Academic Integrity

Sergeev, Aaron Jonas

Name, Vorname/Last Name, First Name

434442

Matrikelnummer (freiwillige Angabe)
Student ID Number (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/ Masterarbeit* mit dem Titel
I hereby declare under penalty of perjury that I have completed the present ~~paper~~/bachelor's thesis/~~master's thesis~~* entitled

Exploring audio-derived immersive environments in extended reality for electronic music creators

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt; dies umfasst insbesondere auch Software und Dienste zur Sprach-, Text- und Medienproduktion. Ich erkläre, dass für den Fall, dass die Arbeit in unterschiedlichen Formen eingereicht wird (z.B. elektronisch, gedruckt, geplottet, auf einem Datenträger) alle eingereichten Versionen vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.
independently and without unauthorized assistance from third parties (in particular academic ghostwriting. I have not used any other sources or aids than those indicated; this includes in particular software and services for language, text, and media production. In the event that the work is submitted in different formats (e.g. electronically, printed, plotted, on a data carrier), I declare that all the submitted versions are fully identical. I have not previously submitted this work, either in the same or a similar form to an examination body.

Aachen, 09.09.2025

Ort, Datum/City, Date

A. Sergeev

Unterschrift/Signature

*Nichtzutreffendes bitte streichen/Please delete as appropriate

Belehrung:
Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt
Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.
§ 156 StGB (German Criminal Code): False Unsworn Declarations
Whosoever before a public authority competent to administer unsworn declarations (including Declarations of Academic Integrity) falsely submits such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment for a term not exceeding three years or to a fine.
§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt
(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.
(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.
§ 161 StGB (German Criminal Code): False Unsworn Declarations Due to Negligence
(1) If an individual commits one of the offenses listed in §§ 154 to 156 due to negligence, they are liable to imprisonment for a term not exceeding one year or to a fine.
(2) The offender shall be exempt from liability if they correct their false testimony in time. The provisions of § 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:
I have read and understood the above official notification:

Aachen, 09.09.2025

Ort, Datum/City, Date

A. Sergeev

Unterschrift/Signature

Contents

List of Figures and Tables

Abstract

The landscape of modern music production is characterized by an increasing number of home producers. This development has introduced challenges to focus and inspiration through the absence of dedicated studio environments providing a work space only related to musical creation.

VR technology provides new possibilities in this regard by giving producers the opportunity to work in any virtual environment they desire according to their preference. In this thesis we investigate some of the possibilities of this technology by exploring music-reactive immersive environments as a creativity support tool, to see how they influence the creative process of electronic music producers, if they can provide any benefit to producers and lastly how different levels of user-control over environments affect the creative experience.

To this end we developed ImmersiveEnGen, a novel system that generates immersive VR environments based on real-time audio analysis of live music production and conducted a mixed methods within-subjects user study with five experienced electronic music producers. We observed participants during their workflows and collected data through questionnaires and interviews.

We found immersive environments to be able to limit distractions, generate atmospheric support, but only rarely cause direct inspiration. The affordance of VR technology to enable workspace mobility was characterized as a valuable future use-case scenario. Control was shown to be an important mediating factor for improving the experience, however definitions for optimal control varied substantially between participants and have to be explored more deeply in future research.

This work contributes the development of a novel prototype capable of generating diverse music-reactive immersive environments, early insights into the effects and potential benefits of music-reactive immersive environments and grounded future system design recommendations in the artistic domain of electronic music production.

Überblick

Moderne Musikproduktion ist zunehmend von Produzenten Zuhause geprägt. Diese Entwicklung bringt Herausforderungen für Fokus und Inspiration mit sich, da Studioumgebungen fehlen, die ausschließlich dem kreativen Prozess dienen.

VR-Technologie eröffnet neue Möglichkeiten, indem sie Produzenten erlaubt, in jeder beliebigen virtuellen Umgebung zu arbeiten – angepasst an ihre Vorlieben. In dieser Arbeit wird das Potenzial von VR zu diesem Zweck untersucht, indem immersive Umgebungen als System zur Unterstützung von Kreativität erforscht werden. Ziel ist es herauszufinden, wie sie den kreativen Prozess elektronischer Musikproduzenten beeinflussen, ob sie einen Nutzen bieten und wie sich unterschiedliche Grade an Nutzerkontrolle auf das kreative Erlebnis auswirken.

Dazu wurde ImmersiveEnGen, ein neuartiges System zur Erzeugung immersiver VR-Umgebungen auf Basis von Echtzeitanalyse der laufenden Musikproduktion, entwickelt. Eine Mixed-Methods-Studie mit fünf erfahrenen Produzenten wurde durchgeführt. Diese wurden in ihren Arbeitsprozesse beobachtet, es wurden Daten über Fragebögen und Interviews erhoben.

Die Ergebnisse zeigen, dass immersive Umgebungen Ablenkungen reduzieren und atmosphärische Unterstützung bieten können, jedoch nur selten direkt Inspiration auslösen. Die Möglichkeit, Arbeitsräume flexibel zu gestalten, wurde als vielversprechendes Zukunftsszenario beschrieben. Kontrolle erwies sich als wichtiger Faktor zur Verbesserung der Erfahrung, wobei sich die Vorstellungen optimaler Kontrolle deutlich zwischen den Teilnehmenden unterschieden, etwas was zukünftige Forschung untersuchen sollte.

Diese Arbeit leistet einen Beitrag durch die Entwicklung eines neuen Prototyps zur Erzeugung vielfältiger musikreaktiver VR-Umgebungen, durch erste Erkenntnisse zu deren Wirkung und Nutzen sowie durch fundierte Designempfehlungen für zukünftige Systeme im künstlerischen Kontext der elektronischen Musikproduktion.

Acknowledgments

I would like to thank Prof. Dr. Jan Borchers and Prof. Dr. Ulrik Schroeder for examining my thesis.

I would like to very much thank my family for their continued support and help in all matters that concern or go beyond science. It is very much appreciated.

I would like to thank my friends, the amazing individuals and students at chair i10 and all other people, who took an interest in this thesis, each of them giving me valuable insight into the ways of importance and improvements that could be made to this work.

I would also like to thank the dedicated creative individuals that took some of their important time to participate in our interviews and study. Their insights and ideas give this exploration credibility that would have never been possible without their valuable help.

Most of all I want to express my sincere gratitude to Paul Preuschoff, who played a integral part in me being able to write about this topic, which is very much important to me in personal regards and further provided me with insights and knowledge that I will always be able to carry with me. Far beyond being a dedicated and insightful advisor that guided me through the process of this thesis, he opened my eyes to the sometimes cumbersome, but nevertheless beautiful process of proper scientific investigation.

Conventions

Throughout this thesis we use the following conventions:

- The thesis is written in American English.
- The first person is written in plural form.
- Unidentified third persons are described in female form.

Short excursuses are set off in colored boxes.

> **EXCURSUS:**
> Excursuses are set off in orange boxes.

Where appropriate, paragraphs are summarized by one or two sentences that are positioned at the margin of the page.

This is a summary of a paragraph.

Chapter 1

Introduction

In this thesis introduction chapter we will cover the motivation behind our research, by introducing environmental challenges music producers face today. Afterwards, we hypothesize how VR technology and especially immersive environments might be helpful to overcome these challenges. Furthermore, we will establish our research questions, highlight our prototype and methodology, give a brief overlook on the results and then lastly finish the chapter with a short overview of the structure of this thesis.

Introduction covers motivation, research questions, prototype, method and overview

The modern world of music production is characterized by an increasing number of music producers that create music within their personal living spaces [Thompson, 2023][Chimbudzi et al., 2022]. Technological advances such as the widespread adoption of Digital Audio Workstations (DAWs) [Lerch, 2018] and improved accessibility to professional music equipment through price reductions have shifted the landscape to a point where creating music outside a professional studio environment may very well be the norm rather than the exception. This shift represents a fundamental transformation in how music is created and the production process conceptualized.

Today many music producers create music at home, technological advancements and accessibility are the reasons

While this development has likely enabled broader participation in music creation [Lerch, 2018] and diversified the range of available sounds and genres, it has also introduced a number of challenges for music producers today

Broader participation today leads to diverse creation, has also introduced challenges

[Meier, 2015]. One aspect that may often be overlooked in this context is the challenging reality of doing creative musical work inside your own home.

Studios provide dedicated work environment

Back when professional studio creation was the norm, artists were provided with a dedicated work environment in which the primary or singular focus was musical creation. Since providing this type of environment is the only purpose of most studio spaces they may often also be built to facilitate the creative process and aim to create an atmosphere that is conducive to creativity.

Young producers lack financial ressources to access studios, home environments are distracting

On the other hand, home-based producers, particularly younger or emerging artists who may often lack the financial resources to access studios or create similarly optimized spaces, have to work in bedrooms, basements or shared living areas that offer only little environmental separation to everyday tasks. This limits the ability of these environments to serve as distraction-free creative spaces. This environmental constraint carries weight, as physical environments are shown exert measurable influence on creative output [J. H. Lee, 2023].

Environment monotony presents challenge, lack of access to natural environments another

The inherent familiarity and unchanging nature of home and professional studio spaces could present another challenge in this regard. Producers often spend hours looking at the same walls and surroundings, which may provide only limited novelty to spark inspiration through new impressions in the immediate context of a production session. Furthermore research found that exposure to natural environments can enhance creative thinking abilities and imagination [C.-W. Yeh and Chang, 2022]] — a form of stimulation that is hard to immediately access during music production at home. Going beyond these everyday limitations the possibility of creating an environment around the musical process could enhance the experience by providing an additional visual dimension to it and thus making it more cohesive overall.

VR technology capable of transporting users into any environment, could be helpful in music production context by mirroring music

Recent advances in virtual reality (VR) technology have the potential to fit exactly this task. Modern VR headsets offer high-resolution immersive experiences capable of transporting users to virtually any environment without requiring physical space alterations.

Unlike traditional displays, VR provides spatial presence and environmental immersion [Cummings and Bailenson, 2016] and offer capabilities that are impossible to realize in physical spaces. They can dynamically respond and be adjusted to creative content or adapt in real-time to changing musical characteristics.

VR offers possibilities of dynamic music-reactive environments, can offer unlimited environment variation

This environmental responsiveness could be helpful in the context of musical production by creating an environment that mirrors the current feeling of the music, thus possibly enriching the creative experience. VR environments could offer a large variety of different locations that are not bound by normal confines of reality in their appearance or style. This aspect may be especially interesting in the context of inspiration as this unlimited access to different environments could help to provide a wide range of stimuli for inspiration.

This technological potential raises some intriguing questions: can immersive environments meaningfully support and enhance the creative process of music producers working from home? What would they need to look like and how can a producer make sure that they fit his creative vision?

Our investigation in this matter is guided by three core research questions:

Research questions

RQ1: What effects do immersive environments have on the creative process of electronic music producers?

RQ2: In what ways does the level of user control over immersive environments affect the creative experience?

RQ3: Can immersive environments be helpful to the workflow of electronic music producers and if so, how?

In order to explore these questions, we developed the Immersive Environment Generator or short "ImmersiveEnGen", a novel system that generates music-reactive immer-

ImmersiveEnGen: novel system able to generate music-reactive immersive environments, 5 different modes for environments sive environments which can be displayed through a VR headset during music production. The system leverages machine learning algorithms and AI image generation to create virtual environments that respond to any musical content. ImmersiveEnGen features five operational modes offering different levels of user control: A Baseline mode shows a virtual environment of a recording studio, random mode generates dynamically changing random environments, music-reactive mode analyzes the live musical content during a music production session and generates fitting environment, modify gives users the ability to change the dynamic music-reactive environments according to their preferences and Manual hands over full control to the user.

Prototype modes enable systematic invesitagtion of RQ2 (effect of user control)

This prototype design enables systematic investigation of our second research questions (RQ2): "How does the level of user control over immersive environments affect the creative experience" and separates VR production effects from environmental ones.

User study with 5 experienced producers, Conditions: 1 No-VR, 5 VR, done at participants homes

This investigation took place in a subsequent user study with five experienced electronic music producers as our participants. We asked them to create a musical piece throughout six consecutive 15 minute production sessions, first without VR and then while using the different modes of our system. We conducted the study inside their own homes in their usual production environments as to make our findings reflect actual usage scenarios and maintain practical relevance to real-world music production.

Mixed-methods study combining quantitative measures and qualitative interviews

Our mixed-methods approach combined quantitative measurements through validated questionnaires with rich qualitative data gathered through in-depth semi-structured interviews conducted at the end of each study. We put our focus on a deep qualitative investigation for exploring this novel space and to generate actionable insights for future system development and usage scenarios.

immersive environments have potential to meaningfully support creative music production, user-agency played a key role in all data

In our study we found that immersive environments have the potential meaningfully support creative music production and that user agency played a key role in the way the environments were perceived. Analysis of questionnaire data showed consistently high flow states across all conditions and a high correlation between perceived control and overall system rating.

This thesis makes the following contributions:

We design and implement ImmersiveEnGen, a novel system that combines music information retrieval algorithms, natural language processing, and AI image generation to create dynamic music-reactive virtual environments. The system employs the Apple Vision Pro headset to deliver an immersive experience and features multiple operational modes offering varying levels of user control. These modes range from fully manual environment specification by users to fully automatic generation based only on musical analysis.

Contribution: novely system design ImmersiveEnGen

We provide one of the first empirical investigations of how music-reactive immersive environments impact creative music production workflows. Our mixed-methods approach yields insights into both the potential benefits and challenges of such systems.

Contribution: early ingestigation of effects of music-reactive immersive environments

We contribute evidence-based actionable design insights for AI-assisted creative support system for music producers in the dimensions of system behavior and user control.

Contribution: evidence-based design insights for AI-assisted CSS

This thesis is structured as follows: after this introduction we begin by reviewing existing literature and providing an overview over the current scientific landscape in real-time music visualization systems, VR and immersive environments and their creative potential and also the current work on creativity support system in Chapter 2. Afterwards in Chapter 3 we offer some background information on electronic music production, introducing some key concepts and terminology that may be relevant in the context of this thesis.

Structure: related work, ImmersiveEnGen, method, results, discussion, conclusion

We then move on to our prototype in Chapter 4 and discuss the ImmersiveEnGen system in detail. We begin by giving some insight into our preliminary interviews we conducted with electronic music producers that informed our design decisions and then provide a comprehensive system overview and detailed examination of the architecture itself.

The system we present in Chapter 4 was later used for our user study - in Chapters 5 and 6 we describe our methodology and findings. We detail the study design, participant characteristics, experimental conditions and data collection procedures before presenting both quantitative results from questionnaires and qualitative insights from our interview analysis.

Chapter 7 offers interpretation and discussion of our findings from Chapter 6, in relation to our research questions. Furthermore, it examines implications for the design of creative immersive systems, connects our finding to current literature and acknowledges the limitations of our approach. Finally, Chapter 8 summarizes our key contributions and outlines promising directions for future research in immersive creative environments and music-responsive systems.

Chapter 2

Related Work

In this chapter we will review and highlight related work and supporting literature in the fields of real-time music visualization systems, immersive VR environments with a particular focus on creative contexts and lastly some of the recent work and design insights in creative support system.

Related work: real-time music visualization systems, immersive VR environments, creative support systems

2.1 Real-Time Music-Reactive Visualization Systems

A growing body of work examines different ways of how music can be visualized. This visualization can occur a from a purely technical standpoint [Lima et al., 2021], however more closely related to our thesis, involves systems that are able to respond dynamically to musical signals in real-time. This is done to provide visual feedback to users in the contexts of improving listening experiences [Li, 2022] or performance [Shao-Ying Lin, 2024].

Growing work on real-time music visualization system, different motivations like improved listening or performance

Real-time Music visualization systems vary in sophistication of their musical analysis. Recent work has begun to go beyond the simple analysis of instantaneous audio signal components and started to integrate AI into their approaches [Huang et al., 2025],[Meng Yang, 2024] thus en-

Varying sophistication of musical analysis, recent advancement: AI integration abling a different semantic type of analysis that traditional visualization systems were unable to perform before.

AI-mediated, dual-layer visualization architecture, visualizes instantaneous changes and deeper semantic meaning

Huang et al. [2025]] proposed a novel architecture using this technology in their work, capable of generating abstract visual representations of music, by combining mappings of instantaneous change in the music, through the visual based programming language Touch Designer, with a deeper semantic analysis done through AI. This analysis is enabled by using the python music information retrieval (MIR) library Essentia and the resulting analysis is interpreted by the AI model GPT4 in order to generate prompts for a lightweight Image Generation Model. This approach of using AI as an interpreter was first proposed by Chakrabarty et al. [2023] and shown to be effective in visualizing more complex and abstract subject matters. Huang et al. evaluated their approach with deaf individuals in an attempt to make music more accessible to them, it remains unexplored if this system or a variation of it can be effective in visualizing music in a creative context.

Music-to-image system for creative inspiration proved helpful in some cases

Meng Yang [2024]'s real time music-to-image system exemplifies another approach that uses AI for musical analysis, this time by capturing MIDI input of piano players and converting it into standard ABC note notation for GPT-4 to analyze. The language model interprets musical structure and perceived emotion, generating text prompts for an image generator to create corresponding imagery. Particularly relevant to our context is their user study, where they tested the system with 5 piano players. They conducted a qualitative analysis and found that users reported the system to be helpful to their improvisation tasks and experience. However their system only looked at direct musical instrument input. Research as to whether this approach can also be applied in a computer-based production process remains unclear.

Real-time music visualization systems are feasible and promising, research gap: VR connection

Both of these systems establish that real time music visualization system leveraging AI are feasible mechanism for visualizing music and that they are worth investigating further. However none of these system use the novel technology of VR to see, whether this would enhance benefits or effectiveness, or offered users with any control mechanisms

- something that was acknowledged by Yang et al. to be important to incorporate into future research - a clear research gap our thesis aims to fill.

2.2 VR and Immersive Environments - Potential Spaces of Creativity

A different research area we want to explore in this chapter examines the possible effects of Immersive Environments and VR on creativity. Studies have shown that the physical environment plays a role in mediating the creative process and J. H. Lee [2023] systematic review on the effects of creativity in physical environment acknowledged a research gap for the creative potential of VR and immersive environments.

environments affect creativity, review acknowledges VR gap

In this context Hagedorn et al. [2023] investigated whether VR environments can enhance creativity by measuring both brain activity and subjective experience during creative problem-solving. Participants generated ideas for environmental challenges in either VR or on 2D screens. VR participants reported higher enjoyment, stronger immersion, and temporal dissociation, which is a typical characteristic marker of flow states. EEG data revealed increased frontal beta and gamma band activity during VR tasks, patterns associated with heightened attention and information integration. Hagedorn et al. argued these neural signatures indicate that VR facilitates cognitive states conducive to creative thinking. However in this work participants only spent minutes generating ideas for hypothetical problems. Whether the same benefits emerge during sustained, technically-focused creative work like music production remains unclear and needs to be investigated further.

VR environments may improve creativity, but evidence is inconclusive

Related to the musical context Abdou et al. [2024] investigated presence as a mediating mechanism between immersion in VR environments and musical emotion during passive listening. Participants experienced music in both VR headsets and on laptop screens while presence and emotional responses were measured. Results showed that pres-

Immersive environments help to enhance musical listening experiences ence fully mediates the relationship between immersion and musical emotion - technical immersion affects emotional experience primarily by creating presence, which then shapes emotional engagement. This finding suggests that for music-related VR applications, the subjective sense of "being there" matters more than hardware capabilities alone. However, Abdou et al. examined brief music listening experiences rather than active music-making. Whether presence influences creative work in the same way, and how immersive environments support the dual demands of creative exploration and technical execution during extended production sessions, remains unexplored.

Immersive environments face challenges in wellness applications

As a base for AI related immersive environment generation Ney et al. [2025] recently explored AI generated immersive virtual environments for wellness applications. They conducted a study where three participants were able create personalized relaxation environments using natural language prompts with SkyboxAI and Meta Quest 3. The study demonstrated generative promise for personalized immersive environments, although with some important limitations in effectiveness of the system to adhere to user preferences. This leaves a gap for research to explore if immersive environments face similar changes in contexts other than relating to wellness and what controls seem to be effective in order to create satisfactory user experience

Research gap: studies that examine effects of immersive environments on creative process

The related work in this section shows that while immersive environments and VR may be promising in the field of music and creativity, any form of concrete studies that examine individuals who are doing creative musical work are lacking and should be explored.

2.3 News in Creative Support Systems

creativity support systems can be effective

Our last area for examination in related work focuses on related work looks at creativity support systems (CSS), which are specialized tools designed to enhance creative work through technology, that have long been shown to foster creativity [Massetti, 1996]. Traditional CSS approaches typically focus on providing task-specific stimuli to trigger new perspectives or offering structured processes to guide creative thinking [Wang and Nickerson, 2017].

music-reactive immersive environments as a form of context-aware CSS

Recent research has expanded beyond these paradigms to consider how environmental and contextual factors can shape creative performance. Sielis et al. [2009] proposed that adding context awareness to CSS can enhance the creativity process by providing real-time support during problem preparation and ideation phases. Their framework suggests that CSS should adapt to user context, recommending relevant resources and methods based on the current creative task and user state. This context-aware perspective highlights that creativity emerges from the interplay between user, environment, system, and task. A system offering creative support through music-reactive virtual environments serves a direct extension based on this approach.

VR environments can be designed to support musical creativity

Building on this emerging perspective, Schlagowski et al. [2022] investigated how virtual environments can be designed to support musical creativity in VR. Through a human-centered design process, they developed and evaluated a virtual reality drum-sequencing system that allowed musicians to compose within either fixed or customizable virtual environments. Their findings emphasize that users have highly individual needs regarding their creative surroundings, with a strong preference for environments that can be customized in lighting, ambience, and visual detail. This work reframes the virtual environments itself, to not only function as a background, but as an active design dimension of a creativity support system.

Research gap: Music-reactive immersive environments as a new form of CSS

Despite these advances, research on CSS specifically related to music production contexts remains limited. While Schlagowski et al.'s work highlights how customizable VR environments can influence musical creativity, it focuses primarily on structured composition within VR interfaces rather than dynamic, music-reactive environments. The potential of immersive, responsive spaces that evolve in real time with musical performance or production, has yet to be systematically explored.

2.4 Research Gaps & Summary

The literature review in this chapter reveals three interconnected gaps that limit our understanding of how immersive environment technology might support creative music production.

Real-time music-reactive visualization systems serve as base, lack creativity studies and VR exploration

1. While real-time music-reactive visualization systems have demonstrated the technical feasibility and creative potential of connecting audio analysis with generative AI, none of these systems address whether these systems can support the extended, iterative workflow characteristic of computer based music production and how this potential support might be extended to the VR dimension.

VR environments could be suitable for creative process, lack of evidence should be adressed

2. Research on VR, immersive environments and creativity has established that VR might potentially be a suitable platform for creativity and music related activities. However research is lacking work on any actual experienced creatives using this technology in a concrete musical production workflow, which should be addressed.

Immersive Environments as new form of CSS has not been explored, studying control with automated system also big gap

3. Research in CSS relating immersive environment and its benefits remains understudied, with only a few single papers we could find on this topic. While important first advances of understanding the value of user control in a musical CSS in VR have been made by Schlagowski et al. further research into the relationship of dynamically changing and automated environments is warranted.

ImmersiveEnGen addresses these gaps by investigating whether music-reactive immersive environments can function as a novel form of CSS in VR and what design decisions relating to user control are essential in this novel design space.

Overall the literature provides technical foundations for music-reactive generation that we can use, evidence that immersion and VR environments can enhance creative experience and first indications that control preferences matter in important ways in this domain.

Chapter 3

Background

Some readers may not be familiar as to what exactly a music producer does, how they commonly work, or what kind of tools they use. In this chapter, we will give a brief overview of the musical creation process and introduce some of the key terminology used throughout this thesis.

Chapter provides background into music production and creative process

Note: Music is subjective - The information in this chapter is the authors personal interpretation of a good explanation of the concept of musical production and should be interpreted as such.

3.1 The Role of Music Producers

Most music available today — whether on the radio or streaming platforms — likely had some sort of music producer involved in the creation process. Their job is to oversee the process of musical creation by creating, arranging, and shaping musical recordings and coordinating between different individuals like recording artists and sound engineers who are involved in the production process. In the space of electronic music production, it is relatively common for a producer to also work alone on entire projects. They are thus responsible for most, or all steps that are necessary for creating a finished musical piece.

Producers are essential in the creation process of music

3.2 The creative music production process

Music production is an individual and iterative process

The process of creating music is a highly individual and iterative process and approaches differ widely from individual to individual. Producers will often continuously experiment, listen back, and refine their work in cycles that may span hours, days or even weeks and longer.

Convergent vs. divergent phases make up essential parts of the process

For an easier understanding the process can be divided conceptually into two different phases that stand opposite to each other. Converging phases during production are characterized by a producer trying to get to a single point. An example for this could be trying to make the sound of a kick-drum to sound just like the producers wants it. He would try adjust the length of the sound, put effects on it or try other methods to converge onto a final end result, that would mark the end of this phase of the process. A divergent phase may be characterized more by an ambiguity as to what may come next in the process. These phases often consist of trying new ideas, changing things around, until it becomes clear what has to be done next specifically (converging again), ending this phase of the process. The patterns, frequencies and times for when each phase occurs is by no means obvious. Generally the process might start with a divergent phase, if the producer had no previous idea in mind and just wants to try out what is gonna happen after he starts working. If there was however previous inspiration involved or the producer has a rough idea in his head for what he wants to do that day, the process may start out with more of a convergent phase pattern.

What is flow

Often during long production sessions, when individuals become fully immersed in what they are doing at that moment, producers may occasionally enter what is often described as a 'flow state' — a state of deep focus and engagement with the creative task, where all other things seem to slip away. This is often characterized as a pleasant feeling afterwards and may be analogous to a form of effective meditation, where all thoughts unrelated to the activity itself seem to slip away.

Specific worfklow routines

More specifically the general production workflow may often involve the producer first picking out a sound like a percussion or instrument. If the sounds does not seem to be quite right - this always involves personal taste - it can be changed through various methods like effects and timing manipulation.

The sound is then "arranged," which is the common reference for the practice of placing a sound in a timely pattern that creates rhythm and energy. By creatively using different sounds which differ in their frequency range (bass sounds are typically found on the lower end of the frequency spectrum and high-pitched sounds like whistling are on the higher end) a producer is able create a musical landscape that feels coherent and carries some sort of emotion for him.

3.3 Tools in Music Production

The main tool used in the creative process of music production today is called a Digital Audio Workstation, or DAW, which is the name for any sort of software application that serves as the centerpiece for digital music creation. Inside a DAW, it's possible to record, edit, arrange and also mix audio among other things. DAWs present music in a visual timeline where producers can see their arrangements laid out horizontally across time. This visual workspace typically includes multiple windows like an arrangement view, mixing console, instrument controls and effects which producers frequently navigate during their workflow.

DAW is the main tool and serves as digital centerpiece in production

Common examples of DAWs used by beginners and professionals alike include Ableton Live, Logic Pro, and FL Studio. A musical piece can be created entirely inside the software alone, e.g., with the help of digital synthesizers that are able to generate any sound that can be perceived by humans.

Producers often work with MIDI (Musical Instrument Digital Interface), a protocol that allows them to control virtual instruments and edit musical notes independently from the audio they produce. Through a MIDI keyboard for exam-

MIDI is a protocol for virtual instruments ple, it's possible to record a sequence of piano notes and later change the sound to be a guitar that plays the exact same notes and melody.

Many producers also use samples, which are names for short musical recordings of, e.g., a single drum hit or bass note that can be laid out in repetitive or non-obvious patterns to create different grooves and melodies.

To listen back to their music, producers often use either monitors (professional loud speakers) or studio headphones, characterized by their precise output of sound.

Chapter 4

Immersive EnGen

This chapter provides an overview of our prototype the Immersive Environment Generator or short ImmersiveEnGen, a novel system which is capable of generating music-reactive immersive environments through real-time audio analysis. We first present some insights from our preliminary expert interviews that informed some of our design decisions, followed by a detailed examination of the system architecture and each of its key components.

Immersive EnGen: novel system for music-reactive immersive environments, chapter structure

4.1 Preliminary Interviews

Before developing the ImmersiveEnGen system, we conducted preliminary semi-structured interviews with four experienced electronic music producers to better understand how their creative processes looked like and what kind of preferences and expectations for an immersive music visualization system the had.

we conducted interviews with experienced music producers

We found large variability in the answers, ranging from abstract worlds to realistic exotic natural environments. Multiple participants mentioned forms of visualizations comparable to a waveform that would be shown somewhere in their field of view and react to their music.

desire for user control

A consistent theme that emerged across interviews was the desire for some type of user control over the environments although there was much large variability in these answers due to the fact that everyone imagined the system to function very differently.

background support

Participants also emphasized that environments should remain in the background rather than demanding too much active attention, with visualizations positioned peripherally where they could be sensed without requiring focus.

We took theses answers into due consideration during the development of our prototype. The Design principle of background support and the emergence of user control as a central theme in our exploration were shaped in important ways by these initial interviews.

4.2 System Architecture Overview

system primary components: audio analysis engine, environment generation pipeline, immersive VR app

ImmersiveEnGen is a modular system designed to be able to generate music-reactive immersive environments in real-time. The system is made up by four primary components: an audio analysis engine that extracts musical information, an environment generation pipeline that translates these characteristics into visual descriptions and images and an immersive VR application, natively running for the Apple Vision Pro headset, that displays the immersive environments to the user. Additionally the system makes use of a control unit that facilitates communication and provides endpoints for web interfaces for study management and user interaction.

High level overview shows overall flow from user to environment

Figure 4.1 shows a high level overview of the system's generation pipeline. Live audio input from a producer's DAW enters the **audio analysis engine**, which then extracts musical features like tempo and key together with more high level descriptors such as mood, tags and genre. These features are then passed on to the **environment generation pipeline**. In a first step, the large language model (LLM) Chat GPT-4 interprets the musical input and based on this, generates corresponding descriptive prompts for visual environments that fit the music. These prompts are then used in a second step to generate high-resolution images through the AI image generator DAll-E 3 and then transmitted in a last step to the Apple Vision Pro to be displayed as an immersive virtual environment, where users are able to perceive them.

VR environment has four components: virtual display, curved backdrop, dome environment, real world display

The **VR environment** itself is based on a fully immersive space - meaning that it allows no pass through of real world surroundings. It consists of four primary visual components (a model for reference can be found in Figure 4.2). The first is the virtual display floating in front of the user, which displays their normal computer screen and can be used normally with a mouse and key board. This virtual display feature is enabled by the Apple Vision Pro technology itself and can be adjusted in size, as well as be moved in space anywhere the user desires. This process happens through gesture control, where users pinch their fingers and then execute the desired movement-gesture. A curved backdrop is positioned behind the virtual display, which wraps around the user's peripheral field of view while they work inside their virtual computer screen in front of them. This curved backdrop together with the surrounding half-dome, which extends the visual space above and behind the user, are updated whenever the system receives a new environment update by the generation pipeline. Both display the same generated environment as a texture on their surface. The fourth component inside the immersive environment space features a window to the real world on the level of the users desk, in order for users to still able to perceive some of their real-world surroundings. This feature helps to ground the experience and also gives users the necessary opportunity to look at their keyboard or adjust the volume of their speakers during the use of the system.

Technical implementation: python backend, essentia audio analysis, Open AI generation pipeline, VisionOS VR application

From a technical standpoint this multi-platform architecture combines a Python backend for audio analysis (Essentia) and AI environment generation (DALL-E 3) with a native VisionOS application for immersive rendering, connected through a Flask-based server that manages communication and provides web interfaces for both administrative control and user interaction.

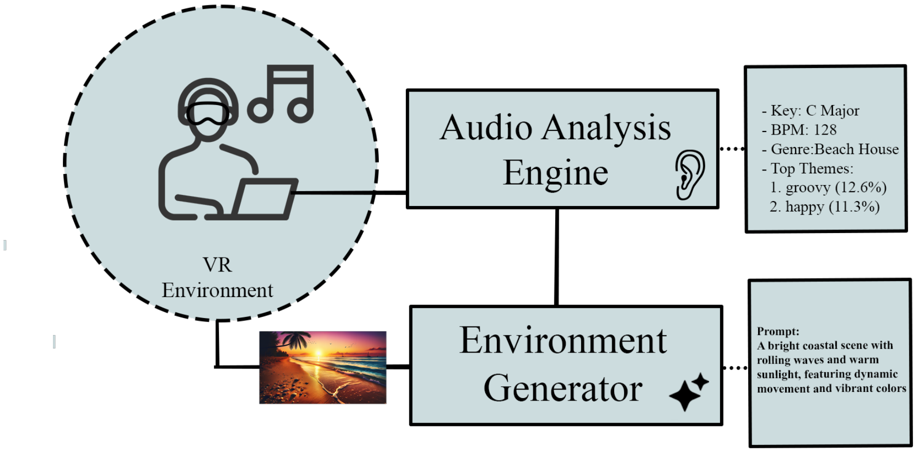

Figure 4.1: High level system overview of ImmersiveEnGen. The system is able to listen and analyze audio in real-time, generate fitting environments based on the results and then display them to the user in an immersive VR space.

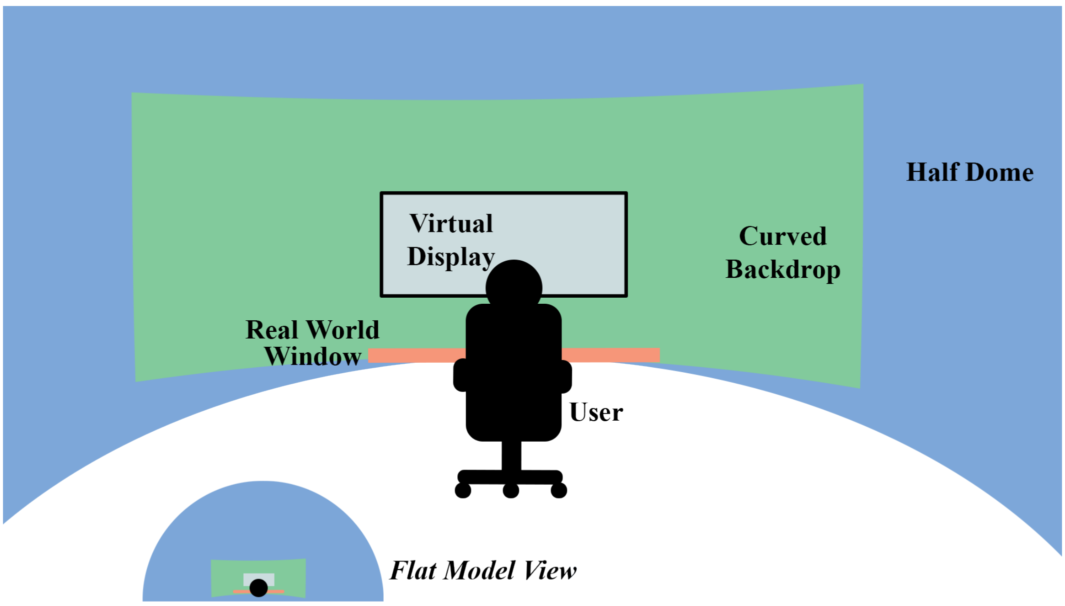

Figure 4.2: Model of the immersive VR environment. In front of him, the user can see a virtual display of his computer. Behind this lies a curved backdrop. The environment is enclosed by a half-dome environment. Both the backdrop and half-dome environment display the generated images of the environment generator as a texture

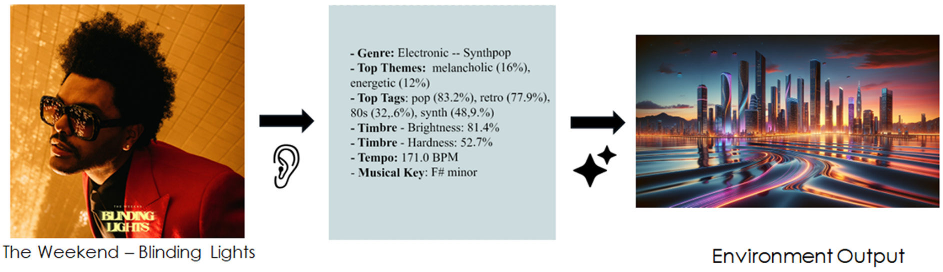

Figure 4.3: Example Generation of ImmersiveEngen on the popular Pop Song "Blinding Lights"

4.3 Audio Analysis Engine

The audio analysis engine serves as the musical "ear" of ImmersiveEnGen, responsible for extracting representative musical characteristics from a live audio signal in real-time. Traditional music visualizers often respond to amplitude or frequency - the instantaneous changes in a stream of music. Our system however follows a different design principle by attempting to understand the semantic meaning of music by trying to analyze both low level descriptors such as the tempo and key and more high level and abstract descriptors such as mood, tags, genre and energy estimations. This semantic understanding enables the generation of coherent visual environments that reflect the character of the music rather than simply reacting to its volume.

Audio analysis by analyzing semantic meaning instead of instantaneous signal changes

Note: Our system is not able to understand any semantics of lyrics incorporated into the music, since only the entire audio stream is analyzed by the system. A separation of individual elements such as voices was not taken into consideration, due to the scope of this thesis being an exploration focused on electronic music.

Audio analysis extracts different musical characteristics with ML models

Our approach leverages the Essentia framework - an open-source python library for audio analysis - using some of their pre-trained tensorflow machine learning models. These models were trained on large datasets of labeled music and can classify various musical attributes with reasonable accuracy. The analysis pipeline examines multiple parameters of musical content simultaneously: structural features like tempo and key, different "characters" of sound (timbre) and more abstract characteristics such as mood & themes, genres and tags. Figure 4.4 can be referenced for an example of an analysis output for an Electronic House track.

We chose these particular analysis parameters for the model because they provided a wide range of data that characterized the music. They were also observed to be accurate in their representativeness, relative to some of the other models we tested during our initial development phases.

Analysis based on 15 second sliding window,

The analysis of a live audio signal operates on a sliding window approach to maintain real-time performance while capturing sufficient musical context. The engine analyzes 15-second windows of audio as a default parameter, updating its analysis every 5 seconds as new audio becomes available. This hop interval creates overlapping analysis windows that help smooth transitions and avoid abrupt changes in the extracted features. During testing that tried to mirror an actual production context, we found some cases, where we had to digress from these standard parameters. We go into more detail on this in our study management section (4.6).

Silence detection to accommodate music production context

An important feature of the audio analysis engine is silence detection. When producers pause audio playback to think or make adjustments, the system should not continue generating environments based on empty analysis data. For this reason we implemented energy thresholding to detect when audio signals falls below a certain level, indicating silence or near-silence. During these periods, the system

Musical Analysis Report:

- **Primary Genre (Discogs)**: Electronic - House
- **Top Genres (Jamendo)**: house (87.3%), electronic (76.2%), dance (64.8%)
- **Top Moods & Themes**: energetic (82.1%), party (74.5%), uplifting (68.9%), groovy (63.2%), summer (58.7%)
- **Top Tags**: dance (89.2%), electronic (84.5%), rhythmic (78.9%), energetic (73.4%), vocal (65.8%)
- **Timbre - Brightness**: 72.3%
- **Timbre - Hardness**: 38.6%
- **Tempo**: 122.5 BPM
- **Musical Key**: A minor

Figure 4.4: Example for a musical analysis report. The system extracts tempo in beats per minute, identifies the musical key, classifies the primary genre, and generates confidence scores for various high level descriptors such as mood and themes, timbre and musical tags. Additionally, each of these attributes is accompanied by a confidence score, allowing the system to weight more certain classifications more heavily in subsequent processing.

pauses environment generation and maintains the current environment, resuming only when audio playback continues.

The final output of the analysis engine is a structured data object containing all extracted musical features, which serves as input for the environment generator. This rich musical profile provides the raw material for translating abstract sonic characteristics into concrete visual environments.

4.4 Envrionment Generator

Fundamental translation challenge from musical input to representative visual environments, AI as solution

The environment generator of our prototype faces a fundamental translation challenge: How can abstract musical characteristics be converted into concrete visual descriptions of environments that somehow represent the music they were made from? Musical features like "groovy," "120 BPM," or "minor key" do not directly specify visual attributes such as color palettes, spatial layouts, or atmospheric qualities of an environment. To bridge this semantic gap, we make use of modern AI technology and employ the large language model (LLM) Chat GPT 4o as a "creative interpreter" with the task to translate the musical analysis into rich visual environment descriptions.

Two stage generation pipeline: first ChatGPT "interpreter", then Dall-E 3 generation

The pipeline operates in two stages. First, the structured musical analysis output is sent to GPT-4, along with a number of specific system instructions. These instructions first specify the role of the LLM as a creative assistant and that its task is to translate the musical data it receives into a rich descriptive prompt for an AI image generator to later be displayed as a virtual environment to music producers. This instruction is followed by multiple rules that further guide the model to generate concise visual scene descriptions that capture the mood and character of the music without directly referencing any musical concepts in the final prompt.

Prompt engineering important to keep output consistent and limit big instruments in environments

In this phase of the process Prompt engineering proved important for ensuring that the system would consistently generate appropriate environments. Early system versions produced descriptions that mentioned big instruments in the final output, even with the initial rules stating that the final prompt should make no reference to any musical construct. In these cases we decided to add additional rules that further constraint the models output prompts, even if this lead to losing some variance in the final environment outputs. Other rule-based constraints we included

Nena - 99 Luftballons

simple pipeline generation | additional rule based generation

Figure 4.5: Prompt engineering proved to be an important step in ensuring the environment genrator produced results that were appropriate for the context. Initial environment prompts included mentions of large instruments inside the environments - so we added additional rule based constraints that limited the AI model in this specific way. The images show examples illustrating the difference in output when running the pipeline on the popular pop song "99 Luftballons" by Nena

for similar reasons included the depiction of humans, as they often looked distorted in the final generation and also cities and crystals as their mention dominated all generation prompts, which led the pipeline to lack any form of environmental diversity. Figure 4.5 illustrates the difference we saw between naive and refined prompt generation with additional rule constraint with the popular song example "99 Luftballons". Again it should be noted that the model is not able to understand any semantics of speech ;)

In the second stage of our generation pipeline these textual descriptions serve as the input for DALL-E 3, OpenAI's flagship image generation model. The system requests images in the maximal 1792×1024 resolution and each generated image typically requires 20-30 seconds to produce, which informed our decision to implement an additional 15 second generation cool down, which was found to be a reasonable middle ground between too quickly changing and long standing single environments during testing.

Dall-E for generation with maximum quality, 35-45 seconds music to environment response time

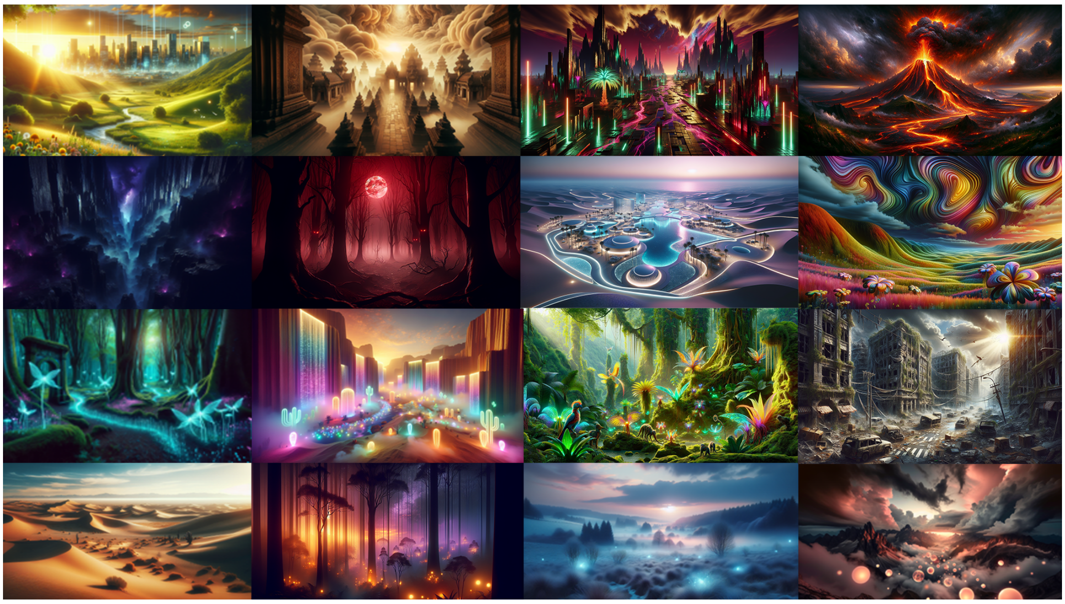

Figure 4.6: Shown are some examples for environments our system generated over the the course of this thesis, displaying a variety of vastly different locations. Standard DALL-E 3 image generation showed a clear stylistic direction that seems to be characteristic of the model.

Figure 4.6 shows a range of examples of environments that were generated during the span of this thesis. Some readers may notice some apparent aesthetic and stylistic directions of environment generations. This was not something we specifically intended, but rather came with using DALL-E 3 as our base image generation model. We considered trying out different models to see if this would diversify the range of aesthetic styles, but decided against it in the end due to time constraints and marginal impact this would have on our research questions.

4.5 VR Environment

VR environment transforms images into spatially surrounding experience

The VR environment our prototype transforms the generated environment images into a spatial experience that surrounds producers during their creative workflow. The application was developed in SwiftUI using RealityKit, Apple's framework for building 3D and augmented reality experiences.

An important design consideration from our interviews in this part of the process was to ensure the environment remained supportive rather than distracting. The visual content needed to provide sufficient interest to create a sense of place and mood without becoming so visually complex that it competed for attention with the music production task.

design consideration from interviews: background support

After some experimentation with different setups we settled on a basic visual architecture consisting of two geometric components that complement each other in order to create a cohesive immersive space. A curved backdrop serves as the primary visual element, positioned approximately eight meters in front of the user and wrapping horizontally across their field of view with a radius of curvature of six meters. This curved geometry creates a sense of spatial depth while the user remains focused on their virtual display DAW interface floating in front of them. The curvature was calibrated through testing to be large enough to feel immersive but not too big in order to avoid any visual distortion of the displayed environment imagery.

Final visual architecture: display, curved backdrop behind, dome around, real world window for orientation

Surrounding this curved backdrop is a hemispherical dome environment with a 30-meter radius, extending above and behind the user to provide complete environmental coverage. Both surfaces display the same generated environment image, mapped as a texture to their geometries to maintain a visual coherence between the two elements. Removing either component resulted in the space either feeling too empty or not providing enough visual interest to our researchers who tested them.

backdrop, dome have images mapped as texture for environment

The application maintains a virtual display for the producers computer screen, which can be freely positioned and scaled according to user preference. Producers can adjust the display size and distance using standard Vision Pro gestures, allowing them to optimize their workspace for comfort during production sessions.

virtual display enabled by Vision Pro, can be positioned and moved freely with gesture control

Figure 4.7: Visual examples of the VR space. Top left and right display the edges between the curved backdrop and the dome environment behind it. Bottom left shows the virtual screen display and standard user POV. Bottom rights shows a view of the real world pass through window. **Note:** Bottom right screenshot was taken at a different time than the other ones.

Examples for how the VR environment space looked like when it had an environment inside it can be found in Figure 4.7

4.6 Operational Modes

Systematic investigation of control preferences (RQ2) enabled by five distinct modes

In order to systematically investigate our second research questions of how different levels of user control affect the creative experience, ImmersiveEnGen was designed to operate in five distinct modes. Each mode represents a different balance between system automation and user agency, ranging from completely static environments to fully automatic music-reactive generation. Here in this chapter we only focus on the technical implementation of these modes. We will discuss the exact reasons for our choices in greater detail in our methodology chapter (5.3).

VR Baseline uses an image of a recording studio and maps it to the environment space throughout the entire production session. This mode completely bypasses both the audio analysis engine and environment generator.

VR baseline: recording studio

Random generates new environments at regular intervals using the same timing constraints as other conditions (respecting image generation times and silence detection), but without any connection to the music being produced. Instead of analyzing audio, the system constructs visual prompts by randomly selecting words from predefined theme components: settings, atmospheres, color palettes, elements and mood descriptors—and combining them into one coherent scene description.

Random: random environments

Music-Reactive analyzes the producer's music in real-time and generates environments based solely on the extracted musical characteristics, without any user input regarding the visual content. This mode activates the complete pipeline described in previous sections: audio analysis extracts musical features, the language model interprets these features into visual descriptions, and DALL-E generates corresponding images. Producers have no control over what environments are generated - the system makes all creative decisions about visual content based on its own interpretation of the music.

Music-Reactive: automatic pipeline, no user control

Modify combines automatic music-reactive generation with user modification input. Like the Music-Reactive mode, the system analyzes audio and generates base environment descriptions from musical characteristics. However, producers are able to specify up to five modifications that constrain or guide the generation process. These modifications might include color preferences ("use dark colors with blue accents"), spatial settings ("set in a natural outdoor environment"), or atmospheric qualities ("create a mysterious, foggy atmosphere"). The system incorporates these user-specified modifications into every environment generation by intelligently combining them with the base music-reactive environment description. Environments continue to update based on musical changes, but always respect the user's guiding constraints. Additionally, this mode includes a regeneration feature that allows

Modify: automatic pipeline, user able to constrain output with modification inputs

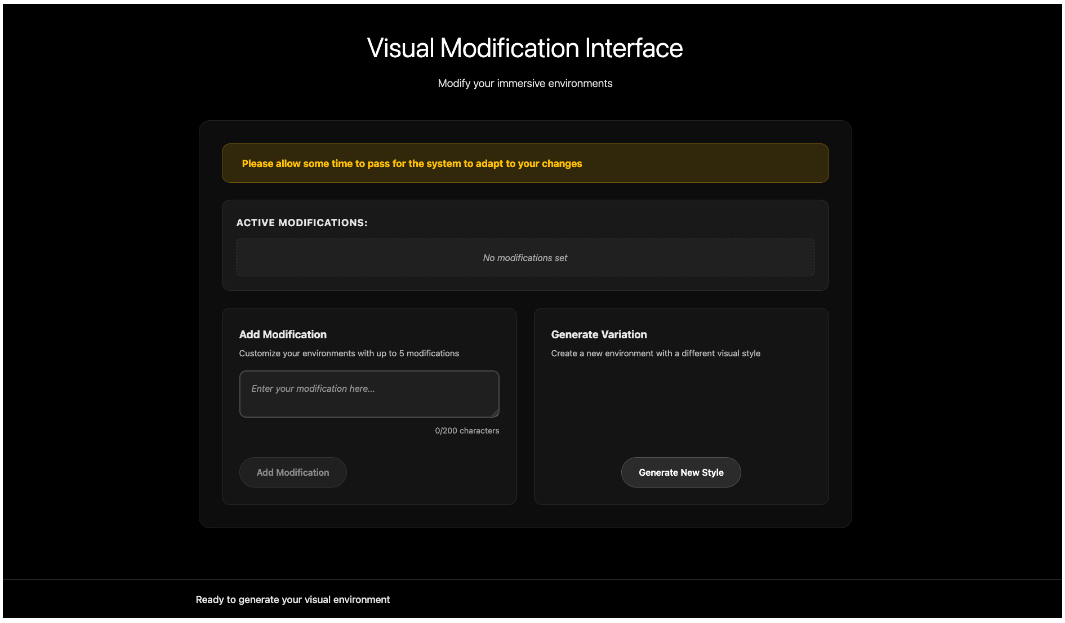

Figure 4.8: Web interface used by participants in the Modify condition featuring an active modification display, sections for adding modifications and generating variations plus a status bar for dynamic feedback responses.

producers to request a stylistically different variation of the current environment without this changing their current modifications. Figure 4.7 shows the modification web interface producers had access to, displaying how producers could add, edit, or remove their guiding constraints during the session..

Manual: full user control, no system intervention

Manual gives producers access to a different web interface where they are able to input their own environment description. The exact prompt they type is used for generation. Audio analysis is disabled in this mode and environments do not change until a user submits a new prompt via the web interface.

4.7 Study Management and Data Collection Features

user control enabled by web-interface, research interface for study control,

To facilitate flexible study administration and real-time

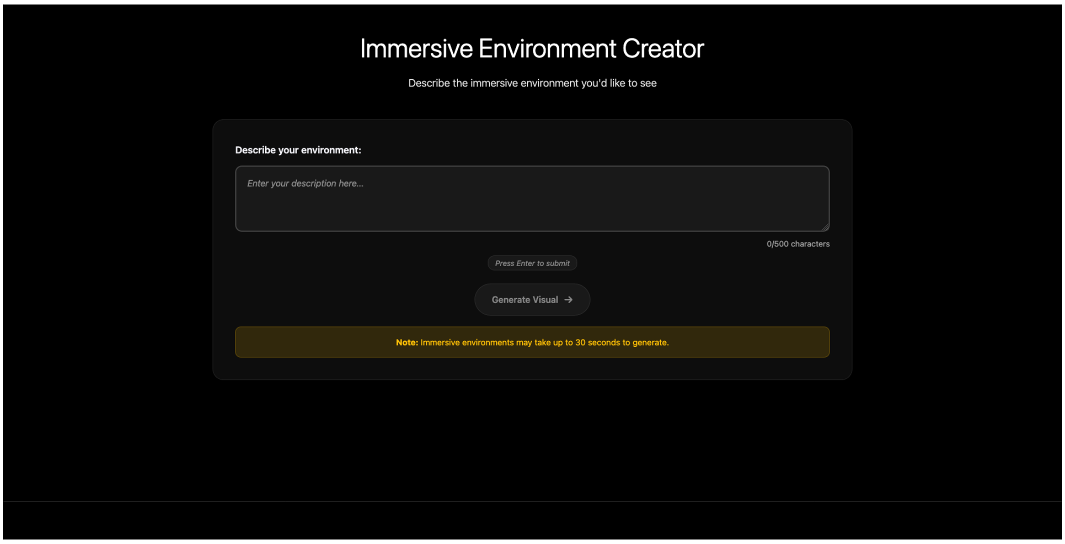

Figure 4.9: Prompt Interface used by Participants in the Manual Condition. A loading status bar would indicate remaining response time for environment generation.

system monitoring, ImmersiveEnGen includes a web-based control interface accessible from any device on the local network. This interface served as the primary tool for managing study sessions and monitoring system status during production sessions.

The control panel provides master controls for starting and stopping the system and switching between the operational modes. A force new image button clears the current immersive environment from the users VR space and triggers immediate generation of a new one. This control was implemented as backup for a scenario where a generated environment might cause visible discomfort, enabling a quick replacement and safety precautions without having to interrupt the session. Fortunately, it did not have to be used in a single case.

control panel for switching modes, force new image button as safety backup

Furthermore, the interface also displays real-time system status, including the current operational mode and a live preview of the current environment image, allowing researchers to verify what environment participants are experiencing at that moment without direct access to the Vision Pro headset. The most essential part of the interface is shown in Figure 4.10 as reference

Control interface shows real-time system status, operational modes, live environment preview

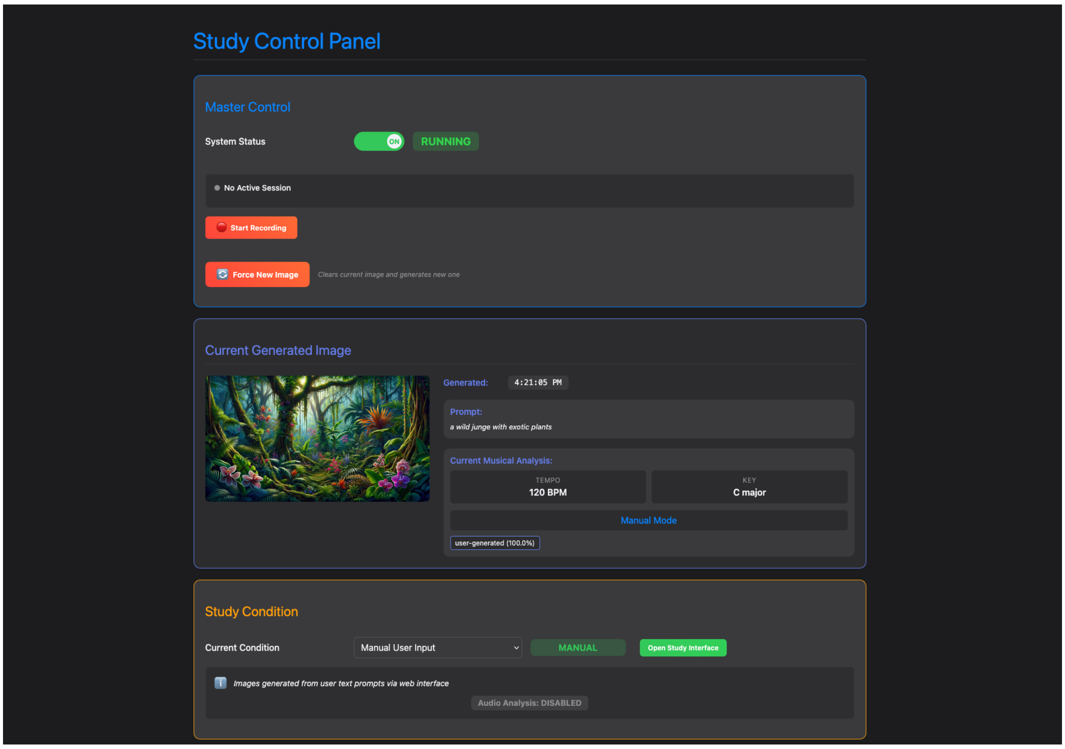

Figure 4.10: Study control interface that enabled remote study management without needing direct access to the apple vision pro. It showed the current study condition, generated environment and music analysis and provided access to audio analysis parameters.

Control panel manages analysis timing parameters, important for workflow variability of producers

Additionally, as mentioned previously in this chapter, the control interface manages analysis timing parameters. We implemented this control because music production workflows can vary widely and producers may listen back to their music at different lengths during sessions according to their habits. The interface allowed dynamic adjustment of these parameters, ensuring the system remained responsive and could capture enough analysis data consistently without requiring session interruptions. This was necessary in a few cases during our later study, when some producers workflow would lead them to only enable playback for a few seconds and the system needed to be adjusted accordingly. In all other cases the system used the standard values of a 15 second sliding window together with a 5 second hop mentioned in the audio analysis engine section.

This centralized control approach provided the flexibility to adapt to individual producer workflows while overall maintaining consistent system behavior across participants.

Lastly, we implemented some interaction logging that captures detailed records of user behavior throughout each session. The system records every environment generation event, including the musical analysis that triggered it, the generated prompt text, the resulting image path, and precise timestamps. For modes involving user input (Manual and Modify) the system logs all text and button inputs entered by participants along with when they submitted it.

Implementation of interaction logging for study capturing environment and user data with timestamps

Chapter 5

Method

This chapter gives an overview of our methodology and the procedure of our user study. We provide details and motivation behind our study design, characterize the demographics of our participants, describe the individual steps we took while conducting each study, what instrument we used for measurement and lastly how we analyzed the resulting data.

Chapter provides overview of method, procedure, study design motivation, instruments and analysis

5.1 Study Design Overview

We conducted a within-subjects user study with five electronic music producers to investigate what effect immersive environments have on creative music production, how different levels of user control influence this experience and if and how immersive environments can be helpful to electronic music producers workflow. The study employed a mixed-methods approach, combining quantitative measurements (a mixture of validated and study specific questionnaires) with qualitative insights from think-aloud protocols and semi-structured interviews.

The experimental design examined six conditions that systematically varied the presence of VR, the source of visual content and the level of user control over environment gen-

Experimental design varied VR stimulus, contol levels and source of visual content eration. For the duration of the whole study participants were asked to try to create a somewhat "complete" musical piece starting from a blank project.

first no VR, then counterbalanced VR conditions, 15 minute sessions

All participants began with a No-VR session. This was followed by a counterbalanced run through of the five VR conditions VR Baseline, Random, Music-Reactive, Manual, and Modify, which explored different balances between system automation and user agency. Each production session lasted 15 minutes. To control for order effects as participants became familiar with VR production, we used a Latin square design, ensuring that each condition appeared once in each position across participants. To prevent bias in participants' perceptions of the music-reactive conditions, we used pseudonyms during the study: the Random condition was called "Dominik" and the Music-Reactive condition "Anton," obscuring the process behind generation in these conditions.

Conducted study at participant homes to make results more generalizable to actual music-production

We specifically conducted the study at participants' homes in their usual production environments. This somewhat unconventional approach allowed us to better understand how exactly immersive environments might support real-world music production, with participants working in familiar spaces using their own equipment, DAW preferences, and acoustic setups. Simultaneously, this limited any confounding effects that would result from using unfamiliar tools during the study or behavioral changes that occur in an unknown lab environment.

5.2 Participants

Participants: 5 electronic music producers, 1-5 years experience, regular music production habits, no VR experience

The participants of our study were five male electronic music producers, aged 21-28 (mean 23). All had active music production practices with 1-5 years of experience (mean 3 years) and self-reported skill levels ranging from beginner (1) to advanced (4). This experience range ensured participants had sufficient technical competency to focus on creative work during sessions while being able to characterize and identify differences between the VR conditions and their normal workflow.

Most participants identified their musical style as a variation of a Techno subgenre like Trance or Bounce. Three used Ableton Live Studio and two used FL Studio for their preferred and standard DAW software. All participants produced music regularly, ranging from a few times per month to several times per week, primarily as a hobby or for personal projects, although two also characterized themselves as semi-professional producers - one of them in a commercial context for other artists. All participants stated that they normally produce music at home in their bedroom or in a home studio space.

Critically, none of the participants had regular VR experience. Four had never used VR before and one had used it less than once per month. This lack of familiarity meant the study captured genuine first-time experiences with VR-based music production. Relevant in the VR context: two of the participants normally wore glasses. During the study one of them used contact lenses and the other wore the headset over his glasses.

5.3 Experimental Conditions

For detailed technical descriptions of how each condition operates in the ImmersiveEnGen system, see Section 4.6; this section focuses on the study-specific implementation and rationale.

No-VR was the initial condition in which participants produced music in their normal physical environment without the VR headset, using their own computer setup and DAW interface. This condition served as our primary baseline, establishing each participant's typical creative experience and workflow habits. Completing this condition first, before any VR exposure, ensured participants had a clear reference point for evaluating subsequent VR conditions.

No-VR: normal workflow, initial reference point

VR Baseline presented a single static environment (a neutral recording studio) for the entire 15-minute session. This condition isolated the effect of wearing the VR headset and

VR Baseline: controls for VR effects

having any visual environment present from the effects of dynamic or music-reactive environments.

Random: controls for Music-Reactive audio analysis

Random generated new environments at regular intervals matching the timing of music-reactive modes without any connection to the music being produced. Presented under the pseudonym "Dominik" this condition controlled for novelty and visual variety and was designed to help us understand whether simply having changing environments, regardless of musical relevance, affected the creative experience.

Music-reactive: automated pipeline activation, tests effectiveness of prototype

Music-Reactive activated the complete pipeline described in Chapter 4, analyzing participants' music in real-time and generating environments based on extracted musical characteristics. Participants had no control over what environments were generated; all creative decisions about visual content were made by the system. Presented under the pseudonym "Anton," this condition tested whether our automatic music-reactive generation could meaningfully support creative work.

Manual: represents full user control

Manual gave participants the opportunity to type in a description of their desired environment through a web interface. The environment was determined entirely by participants intent. New environment requests could be submitted at any point during the session. This condition represented maximum user agency, allowing producers to deliberately craft environments that aligned with their creative vision.

Modify: represents hybrid user control based on user-contraints

Modify combined music-reactive generation with user guidance and served as a hybrid model between the two extremes of full user control and no user control. The system continuously analyzed audio and generated environments based on musical characteristics, but participants could specify up to five text modifications through a web interface that would constrain every subsequent generation.

5.4 Procedure

Each study session took place at the participant's home in their usual music production environment and lasted approximately three to four hours from arrival to departure, including setup, the six production sessions, questionnaire completion, breaks, and the final interview.

Introduction and Setup. Upon arrival, we provided participants with an overview of the study structure, explaining that they would complete six 15-minute music production sessions, starting with their normal workflow and then continuing with VR-based production. Participants signed informed consent forms after having the opportunity to ask questions about the study procedures, data collection, and their rights as participants. We emphasized that there were no correct answers or preferred outcomes and that we wanted their honest reactions and creative work.

First briefing, then informed consent

We then set up the technical infrastructure for data collection. An Elgato Cam Link 4K capture device connected the participant's computer to our recording laptop, routing both the DAW screen output and system audio into OBS (Open Broadcaster Software) for synchronized capture. This setup allowed us to record exactly what participants saw on their DAW interface and heard in their music throughout each session. Notably, the screen capture caused noticeable delay in participants' mouse behavior (implications of this are mentioned in the Discussion) We also prepared the Apple Vision Pro headset for subsequent VR sessions.

technical setup, Elgato Cam Link adapter to our study device with system

Participants received a task for the study: we instructed them to create a complete piece of music during the six sessions, something with a beginning, middle and end, even if it should be short or simple. They started with a blank project in their preferred DAW and then could work in any genre or style. We encouraged participants to think aloud during production, verbalizing their creative decisions, reactions, and thoughts about their workflow.

Task instruction: "complete" musical piece from scratch during the whole study

No-VR Session. Participants always began with the 15-minute music production session in their normal workflow without the VR headset. After 15 minutes, participants completed a post-session questionnaire measuring flow state, enjoyment, mental workload, creativity support, and focus.

VR introduction, video examples, calibration

VR Introduction and Acclimatization. Before beginning the VR production sessions, we introduced participants to the Apple Vision Pro headset through a familiarization process lasting approximately 5-10 minutes. We first showed participants videos demonstrating the system setup, required gestures, and control mechanisms. We then fitted the headset to each participant, adjusting straps for comfort and weight distribution, and performed eye tracking calibration.

Note: Participants mentioned that the headset would be uncomfortable after extended use in some cases. We switched the head strap to a different version which provided better weight distribution after the third study. This reportedly alleviated much of the general discomfort in relation to the Headset experience

short technical acclimatization face

Participants then entered the immersive environment for the first time, normally spending 3-5 minutes in the application to acclimate. They looked around the virtual space, activated the virtual display showing their DAW, and confirmed that everything worked correctly. This acclimatization phase was purely technical familiarization with the VR hardware and interface; participants did not produce any music. We did not proceed to the first VR session until participants confirmed they felt comfortable with the controls and interface.

actual production sessions, pseudonyms for automatic conditions to prevent bias

VR Production Sessions. Participants then completed five 15-minute VR production sessions, one for each VR condition, in their assigned counterbalanced order.

Before each session, we briefly explained the specific condition using the pseudonyms "Dominik" (Random) and "Anton" (Music-Reactive) for those conditions to prevent bias in the experience. Participants then put on the headset, positioned their virtual DAW window comfortably and continued working on their project where they left off. The 15-minute timer started once they confirmed readiness. We asked participants at the start whether they wanted the timer visible during their session; responses were mixed, with some choosing to see it and others preferring to hide it, though those who kept it visible rarely checked it except near the end of sessions.

For the Manual condition, participants accessed a web interface through their browser on their own device to write environment descriptions before putting on the headset. For the Modify condition, participants used a different web interface to add, edit, or remove modification constraints at any point during the session.

web interfaces for user control conditions

Throughout all sessions, we remained in the room and encouraged think-aloud verbalization. We occasionally prompted participants to continue verbalizing when they went silent, though we avoided disrupting their focus too much. After each VR session concluded, participants removed the headset and immediately completed a post-session questionnaire measuring flow, presence, enjoyment, system response quality, mental workload, creativity support, perceived control, and focus. These questionnaires consistently took 2-3 minutes to complete.

Researcher remained in the room, encouraged think aloud without forcing it

Breaks and Comfort Management. Between sessions, participants had natural breaks while completing questionnaires. After the third VR session (the fourth overall session including No-VR), we explicitly offered participants a longer break of 10-20 minutes. Three participants accepted this extended break, while two felt comfortable continuing without one. On two occasions across all participants, brief breaks were needed mid-session. In these cases we paused the timer and continued afterward.

Optional break after thrid VR session

Final Questionnaire and Ranking. After completing all six production sessions, participants filled out a final questionaire which asked them to rank all six conditions from best (1) to worst (6) based on personal preference. We clarified which condition occurred in which session to ensure accurate ranking. They quesionaire asked whether they

Final questionnaire with ranking of all sessions could imagine using any of the tested systems in their own creative workflow and, if so, which specific conditions and their general attitude towards using AI for environment generation.

Semi-structured interview 20-30 minutes, flexible question protocol

Semi-Structured Interview. The study concluded with an in-depth semi-structured interview lasting approximately 20-30 minutes. We asked participants to reflect on their overall experience, compare and contrast the different conditions, explain their ranking decisions, and discuss whether and how immersive environments affected them and their creative process. The interview followed a flexible protocol that allowed us to pursue interesting threads and probe deeper into individual experiences. At the very end of the interview, we revealed the generation process behind ambiguous systems, discussed reactions and answered any remaining questions. Interviews were audio recorded and later transcribed for qualitative analysis.

Data collection: screen recordings DAW, AVP, logging,

Throughout the study, we collected multiple data streams: screen recordings of the DAW interface, Apple Vision Pro recordings showing what participants saw in VR, head tracking analysis data and system logs capturing all environment generations and user interactions with timestamps, think-aloud audio, questionnaire responses and interview recordings. This multi-modal data collection enabled triangulation across quantitative measures, behavioral observations and self-reported experiences.

5.5 Measurement Instruments

Quantitative assessment through validated questionnaires mixed with study related questions

We assessed participant experience using multiple validated questionnaires, strategically shortened to reduce burden on participants, together with questions relating specifically to our study and research questions. Rather than administering full-length scales that would cause fatigue, we selected the most relevant items from established instruments to capture key constructs efficiently. This section describes what we measured and why its relevant in the context of immersive environments.

Core Experience Measures. We measured flow state using five items adapted from the Flow State Scale, assessing whether participants achieved that optimal state of absorption where creative work feels effortless. Flow is particularly relevant to music production, where deep engagement correlates with productive, satisfying sessions. We measured this across all six conditions to understand whether immersive environments supported or disrupted creative absorption. For the five VR conditions, we assessed spatial presence using three items from the Igroup Presence Questionnaire, capturing whether participants felt integrated into the virtual environment. We also included two straightforward enjoyment items asking whether participants had fun and would recommend the experience, providing a simple measure of immediate satisfaction independent of creative performance.

Measuring flow, enjoyment

System Quality and Control. Three items assessed participants' perceptions of the generated environments: whether visualizations matched the music, were visually appealing, and whether the system responded appropriately. These items directly addressed whether participants perceived meaningful connections between their musical output and the visual environments, a core question for evaluating music-reactive systems. For VR conditions, three items measured perceived control over the environment: whether participants had sufficient control, felt like active shapers of the environment, and found it helpful. These items captured the dimension of user agency that varied systematically across our conditions, from no control in Music-Reactive to complete control in Manual.

Measuring system rating

Workload and Focus. We used three items from the NASA Task Load Index to assess cognitive workload: mental demand, self-evaluated performance, and frustration. This helped us understand whether VR environments added cognitive burden or, conversely, reduced mental demand by eliminating external distractions. Two additional items assessed attention management directly: whether the environment distracted from music production and whether participants could concentrate well on the task.

Measuring NASA TLX workload

Creativity Support. Four items assessed whether partic-

Measuring creativity support through environment ipants felt the environment supported their creative process: feeling immersed in their music, whether the environment stimulated creativity, inspired new musical ideas, and supported their creative workflow. These items directly addressed our research questions about how immersive environments might enhance creative work, distinguishing between creative absorption, active inspiration, and practical workflow utility.

Think aloud encouragement

Throughout each production session, we encouraged participants to think aloud, verbalizing their thoughts, reactions, and decisions as they worked. This concurrent verbalization captured in-the-moment experiences that might not be recalled accurately in post-session questionnaires. We audio recorded all think-aloud data for later analysis.

The study concluded with a semi-structured interview lasting 20-30 minutes. We asked participants to reflect on their overall experience, compare conditions, explain their ranking decisions, and discuss whether and how immersive environments affected their creative process. The flexible interview protocol allowed us to pursue participant-specific themes and probe deeper into unexpected findings. Interviews were audio recorded and transcribed for thematic analysis.

After all production sessions, participants completed a ranking task ordering all six conditions from best to worst based on personal preference. We also asked whether they would consider using any of the systems in their real workflow and assessed their general attitude toward AI-generated immersive environments for creative work.

This multi-instrument approach captured both broad patterns through standardized measures and nuanced individual experiences through qualitative methods. The questionnaires provided quantitative data for comparing conditions on specific constructs, while think-aloud and interview data revealed the lived experience behind those numbers.

5.6 Data Analysis

This section describes the analytical procedures for each data type and how we integrated findings across methods.

Data aggregation and mean analysis

Quantitative Analysis. Questionnaire data from all participants and conditions were compiled into a structured dataset with each row representing one participant's response to one condition and columns representing individual questionnaire items and computed subscale scores. We calculated mean scores for each construct (flow, presence, enjoyment, workload, creativity support, control, focus) by averaging the relevant items within each subscale. For reverse-scored items like "The environment distracted me," we inverted the scale before computing subscale means.

descriptive statistics

Given our small sample size (n=5), we prioritized descriptive statistics and visualization over inferential testing. We computed means and standard deviations for each construct across conditions, examining patterns of differences rather than testing for statistical significance. With five participants, traditional null hypothesis significance testing would have very low statistical power and risks both Type II errors (missing real effects) and Type I errors (finding spurious significant results due to small sample variability).

correlation assessment

We examined correlations between perceived control and overall satisfaction measures (enjoyment, recommendation, helpfulness) to explore whether user agency related to positive evaluations across conditions. We also looked at within-participant patterns, examining how each individual's ratings changed across conditions to identify consistent preferences and individual differences that might be obscured in group-level averages.

Ranking data analysis

Ranking data from the final questionnaire provided ordinal preference information. We examined the distribution of ranks across conditions—which conditions received the most first-place votes, which were most often ranked last—and looked for patterns in individual ranking profiles. Since rankings are inherently comparative, they complement the absolute rating scales from post-session questionnaires by revealing relative preferences when participants must make explicit trade-offs.

Reflexive thematic analysis approach for qualitative data

Qualitative Analysis. We followed Braun and Clarke's approach to reflexive thematic analysis for analyzing interview transcripts Braun and Clarke [2006]. This approach emphasizes researcher interpretation and active engagement with the data rather than purely mechanical coding procedures. We began with familiarization, reading through all interview transcripts and listening to think-aloud recordings while taking initial notes on striking patterns, unexpected comments, and recurring themes. This immersion phase helped us develop intuition about the dataset as a whole before systematic coding.

coding phases, code generation

The coding phase involved close reading of transcripts and annotating segments with descriptive codes capturing their content—for example, "finds manual control empowering," "frustrated by slow generation," "environment fades to background during focus," "novelty effect from condition order." We used an inductive coding approach, allowing themes to emerge from the data rather than imposing predetermined categories. Initial codes were refined iteratively as we moved through the dataset, splitting codes that captured multiple distinct ideas and merging codes that proved redundant.

emerging themes

From these granular codes, we identified broader themes that captured patterns across multiple participants. For example, codes related to "control enables creativity," "wants to shape environment," and "frustration with passive systems" coalesced into a theme about user agency and creative control. Codes about "environment supports mood," "visuals underline emotional state," and "atmospheric rather than inspirational" formed a theme about the nature of creative support from immersive environments. We developed these themes through iterative refinement, checking that they accurately represented the coded data and that the boundaries between themes were clear and meaningful.

attention to contrary patterns

Throughout analysis, we paid attention to contradictions and exceptions—instances where participants expressed views contrary to dominant patterns or where individual experiences diverged significantly from group trends. These outliers often revealed important nuances or boundary conditions for our findings. For instance, while most participants valued control, one participant actively preferred systems that made decisions for him, revealing that optimal control levels might vary with individual working styles.

Integration of Quantitative and Qualitative Findings.

Using mixed-methods for contextualizing different approaches

We used qualitative data to explain and contextualize quantitative patterns and conversely, quantitative patterns sometimes prompted targeted qualitative investigation. This iterative movement between quantitative patterns and qualitative explanations enriched our understanding beyond what either method alone could provide. The combination of standardized questionnaires and open-ended qualitative data collection allowed us to both measure specific constructs consistently across participants and conditions while also capturing the rich, contextualized experiences that reveal how and why immersive environments affected creative work. Quantitative measures identified what changed across conditions; qualitative analysis revealed what those changes meant to participants and how they manifested in actual creative practice.

Chapter 6

Results

In the following chapter we will report the quantitative results we collected in questionnaires, followed by the qualitative themes that emerged from our analysis of the semi-structured interviews.

6.1 Quantitative Results

Quantitative data from post-session questionnaires provided initial insights into how different conditions affected participants' experiences across multiple dimensions. Given our small sample size (n=5), we prioritized descriptive statistics and pattern identification over inferential testing.

6.1.1 Ratings Across Conditions

Table 6.1 presents mean ratings and standard deviations for key constructs across all six conditions.

Flow State. Flow ratings remained consistently high across all conditions (M = 4.12-4.36), with minimal variation between No-VR and VR conditions. Music-Reactive and Baseline achieved the highest flow ratings (both M = 4.35+),

Construct	No-VR	Baseline	Random	Music-Reactive	Modify	Manual
Flow	4.12 (0.55)	4.36 (0.45)	3.95 (0.84)	4.35 (0.55)	4.20 (0.55)	4.24 (0.45)
Enjoyment	4.8 (0.45)	5.0 (0.0)	4.8 (0.45)	4.8 (0.45)	4.8 (0.45)	4.8 (0.45)
Presence	—	3.53 (1.10)	3.73 (1.79)	3.47 (0.84)	3.60 (1.10)	3.73 (0.84)
System Rating	—	—	3.2 (1.0)	3.3 (0.45)	3.6 (0.45)	4.3 (0.82)
Control	—	2.7 (1.48)	2.7 (1.30)	2.9 (1.22)	3.4 (1.14)	3.9 (0.84)
Mental Demand	4.4 (0.89)	4.2 (0.71)	4.2 (0.55)	3.67 (0.71)	3.67 (0.55)	3.33 (0.84)
Creativity	2.45 (0.89)	3.1 (1.14)	3.05 (1.58)	3.75 (0.84)	4.07 (0.45)	3.53 (0.45)
Focus	4.4 (0.55)	4.7 (0.45)	3.8 (0.55)	4.3 (0.55)	4.3 (0.55)	4.2 (0.84)

Table 6.1: Mean ratings (standard deviations) for key constructs across conditions on 5-point Likert scales

while Random showed the lowest among VR conditions (M = 3.95).

Enjoyment. Participants reported uniformly high enjoyment across all conditions (M = 4.8-5.0), with Baseline reaching the ceiling of the 5-point scale.

Perceived Control. A clear gradient emerged in perceived control ratings, ranging from Baseline and Random (both M = 2.7) through Music-Reactive (M = 2.9) and Modify (M = 3.4) to Manual (M = 3.9).

System Response Quality. Among conditions where participants evaluated system-generated environments, Manual received the highest ratings (M = 4.3), followed by Modify (M = 3.6), Music-Reactive (M = 3.3), and Random (M = 3.2). A strong positive correlation emerged between perceived control and system rating (r = 0.973, Figure 6.2).

Creativity Support. Creativity support ratings varied notably across conditions, with Modify receiving the highest score (M = 4.07), followed by Music-Reactive (M = 3.75) and Manual (M = 3.53). No-VR scored lowest (M = 2.45), while Baseline and Random fell in the middle range (M = 3.1 and 3.05 respectively).

Mental Demand and Focus. Mental demand showed a decreasing trend from No-VR (M = 4.4) to Manual (M = 3.33), with intermediate values for other VR conditions. Focus ratings remained high across most conditions (M = 4.2-4.7),

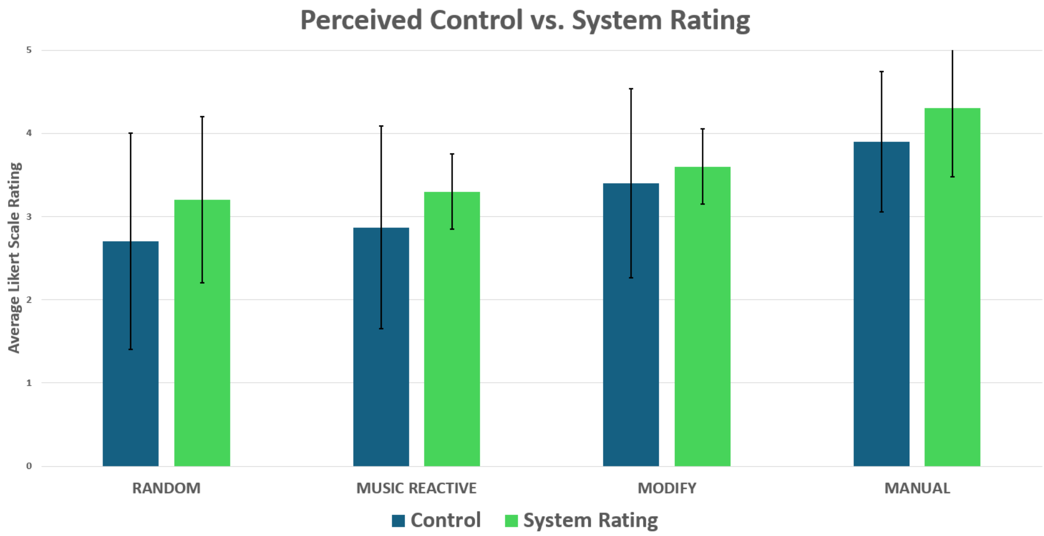

Figure 6.2: Strong positive correlation (r = 0.973) between perceived control and system response quality ratings across VR conditions. Whiskers denote standard deviation.

with Random showing notably lower focus (M = 3.8) compared to other VR conditions.

Presence. Presence ratings across VR conditions showed minimal variation (M = 3.47-3.73).

Implementation error in latin square design, but minimal calculated order effects

6.1.2 Order Effects

Analysis of session order revealed minimal impact of position on most ratings. Due to an implementation error, the Latin square design did not achieve perfect counterbalancing across all condition-position combinations. To assess potential order effects, we calculated mean ratings for each measure by session position (averaging together the data from all participants and conditions). Flow ratings showed a slight upward trend from the first VR session (Position 2, M = 4.28) through mid-study (Position 5, M = 4.40) while enjoyment remained perfectly stable across all positions (M = 4.8).

6.1.3 Preference Rankings

Participants' final rankings of conditions (Table 6.3) revealed their preference patterns when forced to make comparative judgments. To preserve participant anonymity, the table presents rankings in randomized order rather than by participant number.

Participant	No-VR	Baseline	Random	Music-Reactive	Modify	Manual
A	1	6	5	4	2	3
B	3	2	5	4	6	1
C	4	1	5	2	6	3
D	5	4	3	1	2	6
E	6	4	5	3	1	2
Mean Rank	3.8	3.4	4.6	2.8	3.4	3.0

Table 6.3: Condition preference rankings (1=most preferred, 6=least preferred)

Music-reactive mean first in rankings, manual and modify followed, random worst

Music-Reactive emerged as the most consistently preferred condition (mean rank 2.8), receiving one first-place rankings. Manual followed closely (mean rank 3.0) with also one first-place ranking, while Random was least preferred (mean rank 4.6) and No-VR second-least (mean rank 3.8). Baseline received mixed rankings (mean rank 3.4) together with Modify (mean rank 3.4) which polarized heavily with three rating in the top two and two rating as worst condition. The ranking data revealed substantial variability. All conditions except for random were chosen as a favorite once.

6.1.4 Summary of Quantitative Results

The quantitative data revealed several key patterns. Participants maintained high flow and focus while enjoying the experience across all conditions. A strong correlation emerged between perceived control and system response quality (r = 0.973). Creativity support ratings varied notably, with Modify receiving the highest scores and No-VR the lowest. Mental demand decreased in conditions offering greater user control. Individual ranking preferences varied substantially, with no single condition universally preferred across all participants.

6.2 Qualitative Results

Our qualitative analysis of participant interviews illuminates their subjective experience during our study. Through reflexive thematic analysis following Braun and Clarke's approach (Braun and Clarke [2006]), we identified several themes as to what effects the immersive environments had on participants creativity and workflow, what level of control participants preferred and in what ways the environments were helpful or distracting.

themes with focus on research questions

This section presents our findings organized by major themes, emphasizing the insights most relevant to our research questions. For better readability we provide our results with direct context interpretation of the data. We try to clearly differentiate between speculations and actual observations. Furthermore, quotes by participants are denoted with the letter P together with the number of participant (P1, P2, ..., P5). The numbers have been randomized but are consistent across quotes. All quotes were translated from German into English while trying to keep the content as close to the original as possible.

Analysis and results notations

6.2.1 Occlusion - Limiting Distractions and Enhancing Focus

A recurring theme mentioned across multiple interviews was that one of the main benefits of the experience likely consisted not of the immersive environments content itself but their occlusion effect of the physical world. Although not directly mentioning this as the cause of the effect they identified the resulting absence of visual distractions to have a positive effect on their flow and focus.

theme of occlusion effect limiting outside world distractions

> **P1:** "...and yes, the one positive thing for sure: nothing distracts you. There is, I don't know, a mess in the room... gone. Someone walking past your window... gone. To just be in this focus, that definitely helped a lot"

> **P2:** "What I though was cool is that everything is enclosed. You really only have this screen and the environment."

P4 explicitly mentioned this effect in context to his flow state and went on make a comparison to having a dedicated work out environment.

> **P4:** "The point that stood out to me most positively is that I got into flow more easily, I would say. For one, because you're really cut off from any distractions. You don't just look at other things in your room that you need to do, it's really, yeah, as if you go to a studio space, specifically to do that, kind of similar to when you want to work out, you also go somewhere else, because it's easier there, because you give yourself a commitment to actually work out then."

Going even further this separation may also extend beyond visual distraction to behavioral patterns:

> **P4:** "...especially no distraction, because I wouldn't think of the idea, of like checking my phone or something, when I was in that space and I think that really helps to get into the flow in thase case, full concentration and so on. Tha's why this definitely surprised me positively. I wouldn't have thought that it would make such a big difference"

P1 already envisioned a concrete use case scenario where this occlusion effect could be valuable.

> **P1:** "I think, if I were in a distracting environment, if I imagined myself working on a train with headphones - this thing, awesome. There are, i don't know, kids running around, doing something that distracts me. Put the Apple Vision Pro on, only see that [the environments], great. So the main reason for me to use it, would be to block out environments."

Naturally the questions arises as to what extent the actual visual environment content itself played any role in all this. Although we are unable to definitively answer this question here and it likely warrants future work, an insightful speculation of P1 gives us a first-hand experienced indication, that the environment content might still play an important role in grounding the experience.

> **P1:** "Of course there was also something around you and I think if it had been just black screen it wouldn't have worked as well, because there was something there that didn't really distract me, it was actually more perfect to enable this blocking out. I think with a full black I would have thought, oh damn where am I here..."

Overall, this finding suggests a valuable insight into our third research question as to how immersive environments might be helpful in concrete ways to a producers workflow. The effect of real world occlusion may limit outside visual distractions and simultaneously provide a dedicated portable work environment where music production is primary or even singular.

6.2.2 Environments Fading from Awareness

Possibility of a unique effect of immersive environment production

At this point it might be useful to quickly characterize a possible unique effect of the immersive environments, which we cautiously term here as the "immersed focus effect". It was observed by multiple of our participants in a manifestation of fading attention to the environment background and seemed to appear for most after initial novelty factors had decayed:

> **P1:** "At the very start I did kind of tend to be like: new image, okay let me look around, what does it do, how does it look up there? Later it didn't matter any more, then I sometimes almost didn't even notice that pictures changes. It was really more of an initial thing"

> **P5:** "In the beginning with each change of backgrounds, I was impressed and though it to be really beautiful, but eventually I was really just focused on Ableton and blocked out the rest a little bit"

P1 was specific in elaborating that this had not been a conscious decision:

> **P1:** "Nah, that really happened more by itself. At some point for me the visuals weren't.. I didn't actively perceive them anymore"

P4, who was observed as being very focused on his DAW software immediately and throughout the whole study, noted a similar occurrence when asked how he had perceived the environments.

> **P4:** "Actually, I perceived them more peripherally. I sometimes looked at them a bit when I didn't know what I should do next or something. I noticed that it changed, that something was there, but it was really more actual environment. My focus was definitely still on the software"

It remains unclear whether these quotes describe an actual unified effect or similar descriptions of multiple distinct phenomena.

6.2.3 Control is Key - But How?

control emerged as central theme relating to user experience

We also found some answers to our second research question relating to the effects of different levels of user control. Generally we observed that in the conditions featuring user control (Modify and Manual), users would follow almost uniformly the same approach: An initial system configuration by prompting an environment or establishing modifiers was followed by focused work and rarely ever going back to change anything about the initial configurations.

Afterwards during interviews participants were very consistent in establishing that control over the environment had helped to improve the experience for them. However what level of control was appropriate varied between participants and we found multiple categories of preferences.

More than one participant reported to like or prefer modify the most out of the VR conditions. P4 explained his preference this way:

> **P4:** "I liked modify, because there I was able specify that I would rather have dark, black and white environments, with like a few color accents. I just found that somehow visually more appealing. I thought that integrated the work - the software - better into the environment"

Some examples of environment generations during P4's modify condition can be referenced in Figure 6.4.

preference for automated system featuring user control have aesthetic and variation motives

The preference for modify was likely not only based on successful system approximation of users aesthetic appeal but also on the aspect of its automatic environment changes. This was a fundamental difference compared to the other user control condition - Manual - which had to be prompted from scratch in order to generate new environments during the workflow.

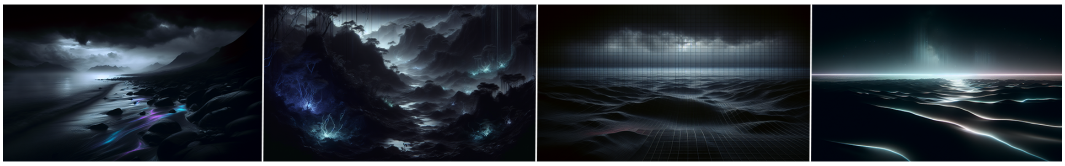

Figure 6.4: Environment examples displayed during the modify condition of P4. His active modifications during the session were: "landscape", "rather dark", "black and white with few color accents".

> **P4:** "... and I chose Modify above Manual because I found it pleasant that it kept changing. Yes, it simply provided good variety."

> **P1:** "That's why I also found the one, where you can completely describe the environment yourself a bit worse, because I did it once and wanted to be in roughly one environment and then kind of expected that it would switch again and that I would get several of this type of environment"

For P1 this ranking preference seemed to also be related to workflow utility:

> There [Manual] I wouldn't want to say during production, man now I'm thinking up a new environment, because then I'm actually completely out of it"

However modify was not everyone's darling and in direct contrast, P3 appeared to rather prefer a single environment that remained present throughout his sessions.

> **P3:** "I think one location would have been better"

This was also backed by his final session rating, where Manual and VR Baseline were among the favorites. He specified his preference for manual:

> **P3:** "I found Manual to be the best of course, because I could control myself what it would be. In retrospective I should have played with it a bit more, but that is the one I would most likely integrate into my workflow, because I am the one choosing what I want to see"

A general observation was that P3 seemed relatively critical of the system and stated on multiple occasions that he had expected something different. When we asked how the environments could have been improved for him, he gave us very specific feedback:

> **P3:** "For example that the colors change, but the the picture stays the same, or that I am in an actual recording studio and behind me, it also looks kind of different. Something like that for example. I kind of expected this and then it was more like static images that changed completely from time to time. I would have changed colors for example or the flow of the structures, or something like that. That wasn't the case."

This exact preference specification for an option to allow P3 to stay in a single environment, which is able to change its style dynamically over time, is something we did not incorporate into our system and study design but would've likely substantially improved the immersive environment experience for P3.

In between these two view P2 might have occupied a middle ground. Initially he reported to prefer Manual control:

> **P2:** "Control-wise I though the one, where you put everything in at the beginning and then just had one picture the whole time, to be the most pleasant"

Modify seemed to have been somewhat disruptive to his workflow and focus:

> **P2:** "That was during Modify, where you always put in different things. During that I was like: Do I want to right now? This doesn't fit quite right - Let me go back again and change that and adjust it a bit, but I don't know yet what exactly is gonna appear, and then it comes eventually and then it's like yeah, okay. I found that a bit disturbing"

However when asked to elaborate on control preferences with static versus changing images he concluded on this:

P2: "...what I would find best: You input something once and then the things change... so not that just one stays"

These seemingly contradictory statements may be interpreted as inability of modify to adapt to P2's preference changes accordingly, but that generally he liked the idea of a dynamically changing system, which responds more accurately and immediately to his input. This interpretation is further backed and expanded in suggestion for advanced control mechanisms here:

> **P2:** "So that something changes is actually better, because then it's a bit like, it's something new and always new impressions and so on. What would be cool is, if you could hold it at a good image. Since it's kind of random. [...] basically like changing to manual.

When asked how an optimal system workflow would look like for him he concluded with this:

> **P2:** "Yes, so I would input once at the beginning and then see what appears, maybe with like a window and then with one click or something, that I want to change it, then it changes and when I'm satisfied, then I can leave it for now and when I don't feel like it anymore then click. So that I have a bit more control over the backgrounds"

These very diverse and divergent preferences suggests a need for advanced control mechanisms, that could go in a similar direction of P2's imagined system behavior, while somehow incorporating preferences of other users like P3, P4 or P1. In any case it seems reasonable to say that the simple options our system provided to users may likely not be sufficient, if the goal is to build a system that accommodates for a wide range of creative individuals and their unique preferences and workflows.

diverse control preferences suggest need for advanced control mechanisms

However we did not only test systems that gave users the ability to customize their environment but also two automatic variations - random and music-reactive. These systems gave us some insights into the effectiveness of systems that react intelligently to musical input featuring no user control.

6.2.4 Imperceptible Music-Reactivity

Despite the Music-Reactive condition being designed to generate environments based on ongoing audio analysis of participants' music, most participants told us they could not perceive a meaningful difference between Music-Reactive and Random conditions. Both were characterized as feeling relatively arbitrary, with P1 expressing that neither condition created environments that reflected his creative output:

Music-Reactive and Random could not be differentiated

> **P1:** "With the automatic things... I think, it still had a bit... so I think the visuals are actually appealing and don't bother me at all, only what's there is a bit random. I don't have the feeling that it reflects my music."

We observed participants to seem surprised at the fact that some of the automatic systems were supposed to have been analyzing their music, when we revealed this fact to them at the end of our interview. Interesting to know for context however, is the observation, that none of our participants ever chose random above the music-reactive condition in

Surprise of participants, when music-reactivity was revealed their final rankings. Generally random had by far the worst mean preference out of all conditions (ref. Fig. 6.3) and also scored lowest for multiple questionnaire items, although that difference is much less drastic (ref. Fig 6.1).

This inability to distinguish the environments from consciously from each other, but rating them somehow differently was best characterized by P4:

> **P4:** Yes, so with the first one I actually kind of thought that it was supposed to be really random, like not-worldly, but almost like alien worlds or something like that. The thing with the music analysis, I wouldn't have noticed now, but as I mentioned, I also didn't pay close attention to the pictures, because it was really more in the background, but I have seemed to like it better than the random ones. That means somehow it hit the nerve better there.

There are multiple possibilities for this apparent lack of perceptibility to the music-reactive dimension - something we will speculate on in more detail in our discussion in the next chapter.

6.2.5 Atmospheric Support Over Idea Generation

Limited direct inspirational support

We have already discussed some of the ways that immersive environments seemed to affect the flow and focus dimension of the creative processes of our participants. These are however not the only aspect of creativity, which is of course much more complex than that. An initial question and motivation was see whether immersive environments might be able to provide direct inspirational support to the musical creation process. Occurrences of this happening, was something most participants were not able to identify:

> **P1:**"For idea generation... actually not at all. I also tried once like: 'oh, what can I pick up from

> the environment here?' Somehow that didn't lead me to want to do anything"

P3 reported that the environments had not consciously influenced him when asked if the environments had had any influence on him and his creative process or if he had perceived his music any differently through them:

> **P3:** "Same answer, if so, then more subconsciously, rather than me perceiving anything actively. Of course I looked around from time to time and that was of course interesting, when they changed all the time like in the last session for example. But like actively, I would say no."

The inspiration question seemed to become more complex with the answers of P4:

> **P4:** I asked myself the same thing, because I wrote myself prompts that were always kind of nature like with a starry sky. And I think Trance music is already kind of, I don't know it kind of fits into the theme with galaxies and stars and colorful and so on. That's why I would say it definitely feels kind of like a product you are in, that the music you are making fits to the environment."

> **P4:** Now, whether I directly drew inspiration from it, I can't really say. It could be, but I think I couldn't distinguish now, whether that came from me or from the environment. But it definitely felt different than just sitting in front the PC"

What P4 seemed to get at was the fact the immersive environment appeared to function in a different role than being visual objects to directly pull inspiration from. They may more accurately be described as creating the right atmosphere which supports already existing creative directives. P2's assessment might help corroborate this hypothesis:

> **P2:** "Yeah, so they were helpfully underlining, not in the sense, that I somehow got ideas through the things, but it kind of .. yes, underlined my feeling a bit"

There however seems to be more research needed in this regard. P4 gives us first insight as to how this might look like:

> **P4:** "To really make statements about whether it inspired me, I would have to try more whether drastically different environments somehow work better or worse for different music, for example. I don't really know that"

Inspiration rare but happened, in some occasions

Although occurrences of inspiration were observed by participants very rarely, our work would be incomplete mentioning the cases where creative influence happened. P5, who was observed as very enthusiastic about the future possibilities of this technology, reported instances of inspiration relating to something different than his own music:

> **P5: "And if I now for example wanted to produce something playful, like today, the environment also adjusted accordingly. And I found that to be really cool. It just complemented the mood and also gave me ideas, maybe for a potential cover or something like that"**

P1 was the only one that reported an instance of the immersive environment directly influencing his musical direction. He described moments when his music seemed to algin with the current visual environment:

> **P1:** "In some moments something fitted to the visuals and I thought, well yeah, might as well continue there. So it was not like, that I saw the visuals and thought, wow I am doing that, but more coincidentally. And when it fit together like that, it felt right to go into that direction"

This experience raised concerns for him about creative autonomy, while using automated system that are intransparent in their mechanisms of generation. When asked if this was something that could be used in a standard workflow using the immersive environments he answered:

> **P1:** "I can imagine that it might happen. But I personally would have the problem, that if I want to have a finished product that somehow reflects what I like and what I want to do, then I also have to choose my inspiration. And in that case I REALLY to decide what's on there. And I can imagine that it happens and that you let yourself be pushed in that direction, but it's kind of unfortunate if you can't perfectly filter what's on there and what gives you this inspiration"

Given inspiration's nature of being often highly spontaneous and coming through sources that are previously unexpected, it remains unclear whether a system of full control, which seems to be hinted at by P1 here, would actually serve the task of creative support best in this regard. However this statement might clearly highlight the fact that any future forms of immersive environment systems, if they are supposed to be used in a real production context will, will likely need to be highly transparent in their generation processes. It might also further substantiate the evidence for control mechanisms that give artists the opportunity to shape their environment according to personal taste and artistic vision.

Future AI mediated environment systems will need to take questions of creative authorship into account

6.2.6 Practical Affordances and Future Possibilities

workflow enhancements, VR affordance of flexible display and possibility to workspace mobility

Beyond the environmental content itself, participants identified several additional practical benefits of the VR platform that appeared to enhance the workflow experience for them:

> **P1:** "Yes, what's really cool in general with this headset, is that you can lean back, look up a bit and have the screen where you want it"

> **P4:** "I also though it was pleasant, that you could change your head position through the VR Headset. Sometimes it's a little bit annoying when you are sitting at your PC, looking in front of you all the time, that at some point you will feel neck pain [...] At some point point I placed the screen a little further up. Then I could really just look straight ahead, that was really an advantage. Just because it doesn't cause fatigue this way"

This practical accordance was also mentioned in the context of portability by P5:

> P5: So if you're on the go now and you have a small screen. Then it's kind of exhausting, or can be exhausting and annoying. You have to zoom in a lot, zoom out. And with this giant screen that you had in front your eyes, this was really cool.

He seemed intrigued by the possibilities this technology could offer producers in relation to new work space opportunities:

> **P5:** Now, if you have a small Bluetooth keyboard and Bluetooth mouse, then I can very

> well imagine myself sitting outside and producing beats in fresh air. I'd find that really cool. I really like the idea in itself. Especially because I have such a small screen.

> **P5:** "[...]If it should be possible at some point, that you are on the train for example and just have a screen in good resolution. And then you can just be in the train or be in the park and see the people - but at the same time make your beat"

These statements likely go beyond use-cases of our immersive-environment system. They characterize the current possibilities of VR systems to enhance music production workflows in concrete ways today.

Immersive environmets useful for piano improvisations

One more interesting possible use-case scenario was mentioned by P2 who was the only one who used a real instrument during the production session. While he mentioned multiple noticeable constraints like time delay and motion blur resulting from the headset experience, he seemed nevertheless most enthusiastic of the possibilities of immersive environments to be used in this context:

> **P2:** I would definitely sit down now for an hour and play piano and look at different environments, it's definitely cool for that... I also think you can just loose yourself there, because you normally always just see a wall, but with the Apple Vision you can just play everything and then something comes during that... Seeing something cool there, that's really amazing."

This suggests broader possible use cases of the immersive environment that extend beyond the music production process. Something we

As a closing statement by P4 in connection to immersive environments as a concrete production use-case:

> **P4:** I think the best thing would of course be to have a different work space for this, where you only do that, instead of your own desk. Something like a studio. I can still imagine, that even there it would get boring after some time, if you always have the same thing. That's why I think this approach is very interesting, that you can always change environments.

participants recognized value in VR platform, discovered possible new use-case scenarios

Taken all together these statements suggest that participants recognized value in the VR platform itself for music production, but also discovered possible use-case scenarios of the environments that go beyond our initial scope of music-production workflows.

6.3 Summary of Qualitative Results

Our qualitative findings reveal that immersive environments affect the creative process of music producers in multiple ways.

immersive environments enhance focus and flow through distraction limitation

Multiple participants mentioned enhanced focus and flow, which likely resulted from the occlusion of distracting real world surroundings. In this focus and flow context multiple participants seemed to experience a phenomena we coined the "immersed focus effect", manifested through environments fading away from conscious attention. The existence of this effect can not be validated here and should be investigated further for more conclusive evidence.

Control plays and important mediating role in user experience, individual differences for optimal level

Relating to our second research question, control over the immersive environments appeared to play an important role to our participants that shaped the way they experience them. However preferences for the exact level of control and system automation behaviors varied substantially and individually between participants.

automatic music-reactive system failed to evoke meaningful connection to music

Automatically generated music-reactive environments failed to evoke meaningful connections to the music of participants in most cases, but was nevertheless rated higher by all of them than purely random images, suggesting an unconscious influence that could be explored further.

Environments provided only very limited direct inspiration, instead likely functioning as atmospheric support that complements existing creative directions.

limited direct inspiration, rather atmospheric support

Participants indicated multiple concrete use-case scenarios where immersive environments but also the VR platform itself, might help to improve their workflow. This includes work space portability, environmental variety and ergonomic flexibility. Beyond music production, instrumental improvisation was mentioned as a related domain, where this technology might be useful.

concrete use-cases of portability, environment variety and ergonomic flexibility

We discuss and contextualize these findings in our next chapter.

Chapter 7

Discussion

In this chapter we will further discuss our qualitative findings, put them in relation to our quantitative insights and furthermore connect our results to the broader context of music production and related research.

7.1 Creativity Support and Uses Cases

7.1.1 Real World Affordances

Our findings seem to paint a somewhat clear picture. Immersive environments - and VR technology itself for that matter - offer multiple concrete ways in which they could help real-world production workflows.

immersive environments offer multiple ways to enhance production worfklows

The most practical examples for this includes the flexible virtual display function offered by the Apple Vision Pro that afforded producers to change their display setting mid way through their workflow and continue to work in ways that better fit their current desired ergonomics and posture. The possibilities emerging from this conceptually relatively simple feature were cause to much excitement in participants 5, who imagined very practical use cases of outdoor music creation with the augmented display and portable

practical example: virtual display helps ergonomics peripheral input devices like a bluetooth mouse and keyboard.

work space mobility affords production in new real-world environment

This could also enable production in natural environments, which are shown to positively impact creativity [C.-W. Yeh and Chang, 2022]. This insight should also be taken very seriously in the sense that sometimes less is more. Many producers might not, or only sometimes, need fancy virtual environments to transport them into spaces outside their normally experienced reality. What many of them may actually need is a way to get up from their desk and move their music production to another real-world place that provides them with the right atmosphere they desire in that moment.

portable work space accomodates travel-heavy lives of DJs and producers

The possibility of creating a portable work space was also mentioned by P1 in the context of distraction limitation, which was likely caused by the occlusion effect of the immersive environment. This demonstrates a second clear use-case possibility of the technology: Accommodating the often very busy and travel-heavy every day realities of producers and DJs. Trains, airports, lounges or other distracting environments that make it hard or sometimes virtually impossible to concentrate and get any meaningful work done, could be transformed into places of deep focus and flow with the right headphones and a VR headset displaying the immersive environments.

flow benefits might be helpful in home-based production, possibility of sustained use needs to be clarified

The suggested benefits of enhanced flow and focus might of naturally also be very helpful in the context of home production. To this end it would be very important to clarify how the system holds up under long production sessions, something we will discuss in more detail in our future works section. Also quantitative findings did not reveal a substantial increase on flow and focus between No-VR and VR conditions. Although users seemed very convincing that these enhanced effects are real, this needs to be further clarified in future work and studies.

7.1.2 Atmosphere and Inspiration

Beyond the immersive environments creative influence on the two dimension of focus and flow, our findings also suggest influence in other components of creativity. Questionnaire data revealed substantial differences between No-VR and user-controlled VR conditions for example.

More concretely and hard to measure through quantitative means, was the question on the influence on inspiration. Here, direct inspirational influence seemed to be the exception and our results indicate that immersive environment might function more in the role of creating the "right" atmosphere that helps to underline or strengthen the existing creative direction. P4 was most explicit in his mention that it was hard to characterize whether the environments had any direct effect on inspiration, but that he was however able to say that the experience felt perceptibly different to the one he normally had sitting in front his PC. We can only speculate here, but this might result from the fact that in a space where the individual, their music and also visual environment all fit together coherently, it gets harder to differentiate where the one thing ends and the other begins. This would make the task of pointing to the source of inspiration much harder.

influence on overall inspiration is complicated to assess, qualitative differences in perception of music were pronounced

For any of this to be the case however, and also to get more tangible again, our results indicate it to be important that producers perceive the environment to be meaningfully connected to their environment.

7.1.3 AI-Automated Systems

Creating this meaningful connection was something our music-reactive condition and also random images were not able to do. While the indication of a possibility of unconscious preference to the music-reactive condition is interesting, it did definitely not succeed in it's original purpose, which was to actively and meaningfully support producers in their workflow. If anything it proved as a small distraction at times together with the random condition. Most

automated system failed to make meaningful connections to participant's music producers either mentioned the two conditions to not represent anything to them or in some cases that they had been a little much in their display of colors.

possible reason for failure: delay - possible solution: dual layer architecture from related work

There might be a large number of reasons why this conditions failed to elicit a more positive response. The most immediate and obvious explanations that come to mind would be the large delay (30-40s) between analysis and display result, which limits any close temporal perception between music and environment. This problem in the architecture in our system could be improved by adding a more immediate dimension to the system that reacts to instantaneous changes in the music. An example for such a dual-layer architecture can be found in the work of Huang et al. [2025] which he have already discussed in our related work.

possible reason for failure: limitations on data input

Also the musical analysis can be taken as the culprit. Although the analysis would generally point in the right direction of genre, moods etc. the real-time constraint and resulting 15 second analysis window of the system, heavily limited the amount of data it was able to receive. This was further complicated by the fact that during production workflows playback times vary drastically and sometimes only last a few seconds, which further limits the accuracy of the machine learning models, used to infer the characteristics of the music.

possible reasons for failure: AI mediated musical analysis

Another questions arises in this context how well Machine learning models or AI in general for that matter, is currently or will ever be able to understand and characterize music of experienced music producers during a production workflow, so that it would come close enough for them to be able to perceive it as a direct result from their work. Any purely automated system will always face the challenge that its own interpretation of music, however sophisticated its understanding might be, is likely gonna be vastly different than the one from the human that is creating it in the same moment.

concerns about creative authorship in an an automated AI CSS

Another dimension was added to this by P1 when he voiced concerns about creative autonomy in an automatic generation system. The possibility of an arising feeling of non-authorship over creative production work, in a context where an external intransparent support system shaped the creative decision making, without any user-agency in them, should be taken seriously for any future production design applications.

Contrastingly however, this form of automated system design might be very useful in a different scenario where creative authorship plays only a more secondary role. The instance of P2's imagined use-case for piano improvisation, connects very tangibly to Meng Yang [2024] findings about music-to-image system for creative inspiration. During their study they found their system to be particularly valuable for improvisation. Furthermore, our findings also align in the statement of participants that the music-reactive system did not drastically influence the creative direction in composition and that a more immediate temporal feedback response would've likely improved the system in our case.

Automated AI system could be useful for improvisational tasks

A lot of the points we discussed here, together with the questions regarding meaningful music-environment connections, lead us the discussion point and research question about control levels.

7.2 System Design Factors

7.2.1 Control is complicated

As our findings suggest, the level of control mattered very much to our participants for a number of different reasons. Through them participants were able to bring the environments closer to their preferred aesthetic and own vision of what a helpful production environment looked like. It also became clear, that these two things vary substantially between participants and will likely also change during the course of a production session.

Control matters for aesthetic preference

Given the complex dynamics of music production workflows where different phases of the process require dras-

Control requirements change dynamically during production workflows, adds necessity for fine and accurate user controls tically different tasks to be fulfilled, a single control paradigm or system dynamic is likely to always come short. For example during a divergent phase, where ideas might go in many different directions, it would probably be more useful to have a range of environments shown to you, mirroring the process of collecting different input and trying out new ideas. However if the process is situated in a convergent phase, the same environmental variety might be perceived much more distracting, as the producer is trying to converge on a single point or idea. This could be the exact shape of a sound or anything else that requires elongated focus on a single task. Pairing this with concrete mentions of multiple participants that a space to test their music - like generating an accurate dynamic club atmosphere - leaves us with so many different variables to consider, that a system without fine and accurate user control would likely never make it in a real production context. This insight is important for future research to incorporate in study designs and developers to implement it in their system.

7.2.2 Suggestions for System-Control Paradigms

Suggested user control paradigms in dynamic immersive environment systems

The real tasks begins when asking how exactly useful controls should look like. Given the context and scope of this thesis we are only able to speculate about this here, but from our findings we might be able to identify a few things.

Automated system without control

Fully automated system without many options of user control in theory provide the opportunity for users to really experiment and improvise with the mechanism of variable environments, without having to think about the influence they want to take on it. However in the case of our study, which focused on the structural task of production and composition, this type of environment in most cases seemed to distract rather than to support the workflow.

Constrained automated systems: give users ability to shape environments in important ways, don't require detailed control mechanisms

Constrained automated system that feature options for modifying and shaping environments to preference, seem to have a clear use in providing variety and novelty, while still fitting into the expected visual style a user desires.

They give users the ability to occupy a middle ground, where they can shape and influence the environment in important ways, without spending too much time on getting every last detail right. This allows for more time actual task-focused work.

Manual systems with granular and fine control over generation provide users with an option to specify exactly what they want and how it should look like. These environments would, in the best case, also be adjustable afterwards, so that if a generation did not quite fit into the expected vision of the producer/artist/.../ they can go back and tweak what didn't work as expected. This would likely help an artist that is seriously considering using such a system in their ability to feel comfortable and in control inside the environment, so that all focus can go into the creative task at hand. Additionally, through our results we identified, that this category of environment system does not necessarily have to be purely static. P3, who showed clear preference for this type of environment style, communicated that dynamic color shifts or subtle animations would have been a desired feature to him.

Manual Systems: provide fine user control, important for users that need a very specific environment

Overall it seems very clear that these controls and any future immersive environment system designs should operate under the **principle of providing atmospheric support** and **operating in the background** rather than trying to place themselves in the foreground. This seems necessary so that they are not perceived as distracting to the creative processes by the individuals who are using them.

Future design principle of atmospheric support and opperating in the background

7.3 7.3 Limitations

Several limitations of our study have to be put into consideration when interpreting the generalizability and validity of our findings and suggest directions for future research.

Novelty Effects. The responses to immersive environments were almost certainly heavily influenced by novelty, since most participants had no prior experience of producing with VR. This novely factor reportedly took an influ-

Novelty effects because of No prior VR experience in participants ence on some of the participants workflow during the first conditions, where they had to orient themselves how to use the technology. Conditions that took place during this time would probably have been assessed differently by the participants if they took place at a lager stage. Any form of counterbalance is likely unable to account for the large individual differences of novelty decay between the different participants, especially in a study with only 5 individuals. Furthermore participants responses to the technology reflects the experience of a first time user, confronted with exciting new possibilities. How the experience of immersive environments changes with repeated use or experienced users, has to be investigated in future studies.

Experimenter presence had influence on behaviour

Experimenter Presence. Two participants explicitly noted that the researcher's presence influenced their behavior and experience. Knowing someone was watching over their shoulder could very well have influenced their their willingness to experiment or other aspects of the creative process. For this reason the findings should not be interpreted as evidence for a real production scenario. Being in their home however, likely increased the transferability to real world-scenarios compared to a lab study. Future studies should consider leaving the room or conducting a diary study if their goal is to overcome this limitation.

System limitation limited effective interaction with the prototype

System Limitations. Several technical constraints limited the accurate assessment of our system.

Reported to be the biggest influence was the noticeable mouse delay that resulted from the HDMI connection between their personal PC or laptop to our study device with the system on it. Furthermore, the absence of smooth transitions between environments created discrete jumps that participants found to be more or less disruptive. Our somewhat hacky real world window pass-through implementation caused participants to see their hands twice. It was mentioned that controls to enable/disable this feature or to just see the keyboard and mouse would be improvements that could be made in this regard. These technical constraints likely negatively affected the experience of our participants in ways that a more polished system would avoid.

Errors in methodology. Due to an implementation error on our side, the Latin square counterbalancing did not achieve perfect counterbalancing across all condition-position combinations. This potentially introduced order effects we could not fully account for. Additionally due to a print error that went unnoticed, one participant received only incomplete post session questionnaires during the study. We asked him to fill out the missing questions (categories 6,7,8) from memory the next day. Before filling out each remaining questionnaire had the opportunity to look at the VR footage as long as needed to remember how he felt during this time. We then gave him the questionnaire to fill out and repeated this process for all remaining conditions. This likely affected the accuracy of those ratings.

Error in full counterbalance did not control for all sequence effects, print error - one participant filled out some questions in retrospect

Session Duration and Physical Comfort. Study sessions lasted only 15 minutes per condition, while real music production sessions often extend for hours. Participants questioned whether VR headset comfort would remain acceptable during longer sessions. Our findings about flow and focus enhancement may not generalize to extended production sessions where physical discomfort becomes more salient. Future research should examine experiences during multi-hour VR sessions that better approximate real production workflows.

Session duration not representative of real workflow length

Time Constraints as Non-Limitation. Interestingly, multiple participants mentioned the overall time constraint as helpful rather than limiting. The limitation that was put on them to finish their work in the course of the study was reported to prevent overthinking in some cases and also that it presented a nice change to the normal production workflow.

Time constraint was mentioned as a nice change of working style

These limitations suggest that while our findings might provide valuable insights into the way immersive environments can support music production, they represent only an initial exploration in this field.

More robust technical implementations, longer-term studies, and larger sample sizes would strengthen confidence in these patterns and reveal whether the benefits we observed persist under more naturalistic creative conditions.

future work would help overcome limitations

We present a more detailed look into our suggestions for future academic work in the next and last chapter.

7.4 Synthesis

Collectively our results help to shed light on our initial research questions that we posited in our introduction.

RQ1: What effects do immersive environments have on the creative process of electronic music producers?

Participants showed high flow throughout all VR conditions and mentioned the immersive environments limitation of outside distractions through VR occlusion to have a positive effect on their flow and focus.

After initial novelty had decayed, multiple participants reported a phenomena that they would seize to notice the environments, they would fade into the background or disappear.

Immersive environments would cause distraction from the workflow, when the system did not properly align with the expectations of participants.

Immersive environments were reported to be able to add to the atmosphere of the creation process by underlining existing creative directions.

We found only **limited** effects on direct inspiration, these were a few rare exceptions.

RQ2: In what ways the level of user control over immersive environments affects the creative experience?

User control correlated very highly with system satisfaction.

The level of environmental control was mentioned as a key factor in how it shaped the experience of our participants inside the virtual environments

Individual differences regarding preferences varied between individuals and suggests a need for further categorization to better be able to classify patterns of preferences. A first attempt of this can be referenced above.

RQ3: Can immersive environments can be helpful to the workflow of electronic producers and if so, how?

While the nature of this research is preliminary and our study featured many limitations, we want to list the following possible benefits:

VR and immersive environments afford mobility and dedicated portable work spaces, able to limit distractions and enhance focus. This present solutions to some of the everyday challenges music producers and DJ's face in their lives that include a lot of traveling and home-constrained music production.

The positive creative benefits mentioned in **RQ1**

Chapter 8

Summary and Future Work

8.1 Summary & Contributions

This thesis investigated the effects of music-reactive immersive virtual environments on the creative process of electronic music producers, if they have the ability to support their workflow and if yes, how and lastly how different levels of control factor affect their creative experience.

Thesis investigated three research questions

For this reason we built ImmersiveEnGen, a novel system able to generate music-reactive immersive environments, which are displayed through a VR headset during music production, featuring five operational modes that systematically vary user control from fully automatic to completely manual generation.

Built ImmerisveEnGen - novel system generating music-reactive environments

We then conducted a mixed-methods user study with five experienced music producers in their own homes, combining validated questionnaires with in-depth interviews to understand how exactly the immersive environments affect their creative work and gather valuable insights into how the different conditions should alter the experience.

Conducted a mixed-methods user study

Our findings reveal several promising directions for the fu-

Findings reveal promising future directions and multiple benefits ture of immersive creative tools. Overall we found immersive environments to offer several benefits that may concretely help producers in their workflow.

The immersive environments ability to occlude real world surroundings limited distractions and resulted in higher flow and focus experience for multiple participants.

Having a virtual display that can be adjusted to preference was mentioned as an enhancement to existing workflows through improved flexibility and ergonomics and participants suggested future use-cases that enable greater flexibility in workspace location.

Immersive environments did not seem to have a direct inspirational effect for idea generation. Rather they were described to be helpful in creating a fitting atmosphere during music production, receding into peripheral vision when participants were deeply focused on their work.

environments need to reflect music in a meaningful way to be effective - music-reactive mode was not able to achieve this

To be able create this atmosphere and not serve as distractions, the immersive environments needed to reflect the artists music in a somewhat meaningful way. Our music-reactive system mode - although preferred to completely random images in final rankings - was not able to achieve this task. Participants were not able to consciously make a connection between their musical input and the environment, likely as the results of a relatively high music to environment response delay of 30-40 seconds, lacking accuracy of analysis results and inherent problems of AI-mediated environment generation.

modes with higher control were strongly preferred, both in quantitative and qualitative data

In contrast, modes that offered higher user control were strongly preferred. This result was reflected both in our quantitative data in a direct correlation between perceived control and system ratings and also characterized in more nuanced ways by participants in our qualitative results. Preferences for the amount of control however, varied between participants and we found no universal solution that would accommodate for their diverse preference needs. This suggests a need for diverse control mechanisms in future system and warrants further research in this area.

In this thesis exploration we made the following contributions:

contributions

1. We designed and implemented a novel system ImmersiveEnGen with multiple operational modes, capable of generating music-reactive environments by integrating ML audio analysis, large language models, and AI image generation into the Apple Vision Pro's VR platform.

2. We conducted on of the earliest empirical investigations on the effects of music-reactive immersive environments on music production in a mixed methods user study, revealing that immersive environments have the potential to meaningfully support creative workflows in multiple ways.

3. We generated evidence-based actionable design insights for AI-assisted creative tools for music producers in the dimensions of system behavior and user control.

8.2 Future Work

The findings from our study revealed both technical limitations to address and broader research questions to explore. We organize future directions into three complementary paths: refining the current system based on identified shortcomings, exploring unexamined design variations, and finally validating findings through expanded explorations.

three paths for future directions

These directions for future work can be taken as additional insights relative to addressing the limitations of our approach.

8.2.1 System Refinement and Technical Improvements

The most logical next step involves addressing the various technical limitations that were present in our system in

Adressing technical limiations

order to improve user experience and be able to more accurately assess the viability of this approach.

This includes fixing the temporal pairing issue that prevented participants from perceiving music-environment connections. Future systems should consider implementing dual-layer responsiveness: maintaining the slow semantic generation for overall aesthetic direction while adding instantaneous parametric mappings between audio features and visual parameters. We have also seen evidence in our related work that this is feasible. These instantaneous mapping could include but are not limited to real-time brightness modulation, color shifts synchronized to musical energy or subtle animations responding to beat detection - all of which could provide forms of immediate perceptual feedback while AI-generated environments evolve more slowly in the background. This approach would preserve the benefits of sophisticated environment generation while creating the responsiveness necessary for users to experience the system as truly reactive to their creative output.

Our transition mechanism also requires rethinking. Rather than discrete jumps between completely different scenes, future implementations should explore ways of smooth morphing between environments and/or, as P3 envisioned, single coherent spaces that evolve and transform over time. This could involve 3D environments where structures subtly animate, colors shift gradually and spatial elements reconfigure while maintaining continuity. Additionally, implementing P2's proposed "hold and advance" mechanism would give users temporal control over when transitions occur, allowing them to freeze compelling environments while letting others pass.

Lastly, implementing proper desk space pass-through without the use of a somewhat hacky real world window solution, would help to remove small friction points, that may nonetheless be important in future investigation of extended usage scenarios.

These system improvements and additional controls for them may address some the most consistent participant frustrations while also accommodating some of the divergent preferences we observed.

Technical investigation also matters for extended use. Finding ways to test the system without mouse lag resulting from the study setup and somehow trying to improve headset comfort for multi-hour sessions would remove friction points which were noted very often during the study, as a make or break for this technology.

The Vision Pro's weight and pressure created discomfort in our first participants. Even though the dual-strap solution we employed during the last two studies reportedly improved this matter for participants it needs to be clarified how this technology might hold up during longer sustained sessions mirroring real production workflows.

8.2.2 Exploring Unexamined Design Variations

Addressing questions that arised during study

Several design questions emerged from our study that warrant systematic investigation. First, we need to isolate the occlusion effect from environmental content. A condition presenting complete darkness - no environment, just black space with the virtual display and keyboard - would clarify whether the benefits we observed in this regard, came purely from blocking outside distractions or whether environmental presence really matters. P1's reflection that complete blackness would feel disorienting suggests, that the anchoring function of the immersive environments might be important, but empirical comparison would resolve this question definitively.

The evolving single-environment concept also deserves dedicated exploration. P3's vision of a coherent 360-degree space that transforms over time represents a fundamentally different approach than our scene-switching implementation. Developing and testing such a system would reveal whether spatial continuity with temporal transformation addresses the preferences our current system could not accommodate. This could extend to immersive soundscapes that fade in during silence, creating richer atmospheric support when participants pause to think rather than maintaining visual-only presence. Long-term studies examining how experiences evolve beyond initial novelty are critical. Our 15-minute sessions captured first impressions but could not address whether benefits persist or diminish with extended use. Multi-week studies where participants use the system during actual production work would reveal whether benefits remain valuable once novelty fades, whether atmospheric support continues to feel meaningful, and whether new usage patterns emerge as producers integrate the technology into their workflows.

propositions for future studies

From our point of view there is a multitude of directions that future research may face towards. Here we offer suggestions for two distinct approaches.

One involves testing a slightly refined version of our current system in another similar user study approach with slightly longer sessions, but this time only featuring the most preferred conditions modify and manual with additions of a darkness condition and a condition exploring the evolving single environment.

Another approach would be to take the evidence that the evolving environment is likely a desired feature ,in an actually viable system directly into account and start to develop a more sophisticated prototype that is able to incorporate most of the preferences voiced by our participants. This prototype would offer a number of dedicated controls for free use and could then be tested in extended sessions, closer to real production length. This could happen in an expert diary study resulting in real data of professional system usage. Alternatively, if a more controlled approach is needed, a lab or home study that features one long sustained production session where users are able to freely explore the system and its controls, could provide researchers access to a lot more data of how the system and its various controls are being used in this context.

expanding research to new domains and collaborative settings

8.2.3 Expansion

Expanding research on immersive environments beyond electronic music production would test whether our findings generalize to other creative domains. Writers, visual artists, designers and other creative professionals working at computers may face similar challenges in their home or work environments. Adapting the system for content-reactive environments in these domains could reveal whether atmospheric support and occlusion benefits extend across creative practices or remain specific to music production.

Finally, our study examined only individual production sessions, but much creative work occurs collaboratively. Future systems could explore shared immersive environments where multiple producers work together, seeing synchronized or personalized environments, coordinating creative direction through shared visual spaces, or experiencing each other's musical contributions through environmental responses. This social dimension represents entirely unexplored territory for immersive creative support.

8.3 Closing Remarks

This thesis began with a question about whether immersive environments and VR had the potential to meaningfully support home-based music producers.

What we discovered was both simpler and more complex than we first anticipated. Simpler, because some the most powerful possible benefits came not from sophisticated AI-generated imagery but from the fundamental VR affordances of blocking out the physical world or enabling production in more places inside of it, through the augmented reality display.

More complex, because understanding how to design immersive environments that support rather than disrupt creative work requires navigating a wide range of individual differences in control preferences, temporal perception challenges, and the subtle edge between active atmospheric support and developing fancy distractions nobody needs.

ImmersiveEnGen demonstrated that music-reactive immersive environments are technically feasible and can very likely meaningfully support creative work, but our findings also suggest the path forward requires the acquisition of a deep understanding for whom, under what conditions, and through what mechanisms. The producers in our study did not seem to need the environments to inspire them - they needed tools that got out of their way while maintaining a supportive presence, that offered control matching their individual working styles and creative visions and atmosphere that responded to their music in personal ways that mirrored its timing and emotion.

The practical implications of ImmersiveEnGen extend beyond music production. As VR technology matures and AI-assisted tools become increasingly integrated into creative workflows, understanding how to balance intelligent automation with user agency becomes critical. Our findings suggest that effective creative support systems must prioritize user control not as a feature to implement but as a fundamental design principle. They must provide responsiveness at timescales matching human perception and must recognize that individual creators have genuinely different preferences for how their tools should behave.

The future of immersive creative environments lies not in replacing the tools and workflows producers have spent years mastering, but in augmenting them with a tool that supports flow, eliminates distractions, and adapts to individual needs. This thesis provides initial evidence that such systems may be possible and identifies some of the design principles necessary to make them effective. What remains now is to rigorously investigate this matter further, so that we may soon discover, what creative possibilities emerge when producers can truly work anywhere, surrounded by environments that evolve together with their music.

Appendix A

Informed Consent Form

EINVERSTÄNDNISERKLÄRUNG ZUR STUDIENTEILNAHME

STUDIENINFORMATIONEN

Studienleitung: Aaron Sergeev
Institution: RWTH Aachen
Kontakt: aaron.sergeev@rwth-aachen.de

Betreuung:
Paul Preuschoff
paul.preuschoff@rwth-aachen.de

ABLAUF DER STUDIE

Dauer: ca. 2,5 Stunden

Sie werden gebeten:

- 5 verschiedene visuelle Umgebungsbedingungen während der Musikproduktion für jeweils 15 Minuten zu erleben
- Fragebögen zur Bewertung Ihrer Erfahrung auszufüllen
- An einem anschließenden Interview über Ihre Erfahrungen teilzunehmen
- Ihre Gedanken und Denkprozesse anhand der Think Aloud Methode mitzuteilen (Einführung folgt)

Mögliche Risiken umfassen:

- **VR-bedingte Beschwerden:** Übelkeit, Schwindel, Augenmüdigkeit oder Kopfschmerzen durch die Nutzung der Apple Vision Pro

Sicherheitsmaßnahmen:

- Sie können jederzeit Pausen einlegen oder die Studie abbrechen
- Technische Unterstützung ist während der gesamten Sitzung verfügbar

DATENERFASSUNG UND -VERWENDUNG

Erfasste Daten:

- Verhaltensdaten während der Systemnutzung
- Audiodaten Ihrer Musik/Musikproduktion
- Antworten aus Fragebögen und Interviews

- Tonaufnahme aus Studie und Interviews
- System Interaktionsdaten
- Technische Leistungsmetriken
- Bildschirmaufnahme der Apple Vision Pro
- Bildschirmaufnahme ihres Laptops

VERTRAULICHKEIT UND DATENSCHUTZ

- Alle gesammelten Daten werden anonymisiert behandelt und mit Codes versehen
- Ihre Identität wird in keinen Veröffentlichungen preisgegeben
- Daten werden sicher auf verschlüsselten Systemen gespeichert
- Zugang haben nur autorisierte Forschungspersonen
- Sie haben das Recht auf Auskunft, Berichtigung und Löschung Ihrer Daten

FREIWILLIGKEIT UND RÜCKTRITT

- Ihre Teilnahme ist vollkommen freiwillig
- Sie können jederzeit ohne Angabe von Gründen zurücktreten
- Ein Rücktritt hat keine negativen Konsequenzen für Sie

EINVERSTÄNDNISERKLÄRUNG

☐ Ich habe die Studieninformationen vollständig gelesen und verstanden

☐ Ich verstehe, dass meine Teilnahme freiwillig ist und ich jederzeit zurücktreten kann

☐ Ich bin damit einverstanden, dass meine Daten wie beschrieben erfasst werden.

☐ Ich bin damit einverstanden, dass die gesammelten Audiodaten in einem wissenschaftlichen Kontext veröffentlicht werden dürfen.

Falls irgendwelche Daten nicht an jemand anderen als die genannte Studienleitung weitergegeben werden sollen, kann ich dies hier genauer erläutern:

Name des Teilnehmers/der Teilnehmerin: ________________________________

Unterschrift: ________________________________ **Datum:** _______________

Appendix B

No VR Questionnaire

Nachbefragung zur Musikproduktion ohne VR

Teilnehmer-Code:
Bedingung: No VR
Durchlauf NR:

Bitte bewerten Sie die folgenden Aussagen basierend auf Ihrer Erfahrung aus der vorangegangenen No VR Session.

	Stimme gar nicht zu (1)	**Stimme eher nicht zu (2)**	**Neutral (3)**	**Stimme eher zu (4)**	**Stimme vollkommen zu (5)**
1.1 Ich fühle mich optimal beansprucht.	☐	☐	☐	☐	☐
1.2 Ich habe keine Mühe, mich zu konzentrieren.	☐	☐	☐	☐	☐
1.3 Ich bin ganz vertieft in das, was ich gerade mache.	☐	☐	☐	☐	☐
1.4 Die richtigen Gedanken kommen wie von selbst.	☐	☐	☐	☐	☐
1.5 Ich weiß bei jedem Schritt, was ich zu tun habe.	☐	☐	☐	☐	☐

	Stimme gar nicht zu (1)	Stimme eher nicht zu (2)	Neutral (3)	Stimme eher zu (4)	Stimme vollkommen zu (5)
3.1 Ich hatte Spaß während der Erfahrung.	☐	☐	☐	☐	☐
3.2 Ich würde die Erfahrung einem Freund weiterempfehlen.	☐	☐	☐	☐	☐

	Gar nicht (0)	Sehr niedrig (1)	Niedrig (2)	Eher niedrig (3)	Mittel (4)	Eher hoch (5)	Hoch (6)	Sehr hoch (7)
5.1 Wie mental anspruchsvoll war die Aufgabe?	☐	☐	☐	☐	☐	☐	☐	☐
5.2 Wie erfolgreich waren Sie bei der Erfüllung der Aufgabe?	☐	☐	☐	☐	☐	☐	☐	☐
5.3 Wie unsicher, entmutigt, irritiert, gestresst oder verärgert waren Sie?	☐	☐	☐	☐	☐	☐	☐	☐

	Stimme gar nicht zu (1)	**Stimme eher nicht zu (2)**	**Neutral (3)**	**Stimme eher zu (4)**	**Stimme vollkommen zu (5)**
6.1 Ich habe mich in meine Musik hineinversetzt gefühlt.	☐	☐	☐	☐	☐
6.2 Die Umgebung hat meine Kreativität angeregt.	☐	☐	☐	☐	☐
6.3 Ich wurde durch die Umgebung zu neuen musikalischen Ideen inspiriert.	☐	☐	☐	☐	☐
6.4 Mein kreativer Workflow wurde durch die Umgebung unterstützt.	☐	☐	☐	☐	☐

	Stimme gar nicht zu (1)	**Stimme eher nicht zu (2)**	**Neutral (3)**	**Stimme eher zu (4)**	**Stimme vollkommen zu (5)**
8.1 Die Umgebung hat mich von der Musikproduktion abgelenkt.	☐	☐	☐	☐	☐
8.2 Ich konnte mich gut auf die mir gestellte Aufgabe konzentrieren.	☐	☐	☐	☐	☐

Vielen Dank für das Ausfüllen dieses Fragebogens!

Appendix C

VR Questionnaire

Nachbefragung zur Musikproduktion mit VR

Teilnehmer-Code:
Bedingung: _______________
Durchlauf NR: _________________

Bitte bewerten Sie die folgenden Aussagen basierend auf Ihrer Erfahrung mit der gerade durchlaufenen VR-Session.

	Stimme gar nicht zu (1)	Stimme eher nicht zu (2)	Neutral (3)	Stimme eher zu (4)	Stimme vollkommen zu (5)
1.1 Ich fühle mich optimal beansprucht.	☐	☐	☐	☐	☐
1.2 Ich habe keine Mühe, mich zu konzentrieren.	☐	☐	☐	☐	☐
1.3 Ich bin ganz vertieft in das, was ich gerade mache.	☐	☐	☐	☐	☐
1.4 Die richtigen Gedanken kommen wie von selbst.	☐	☐	☐	☐	☐
1.5 Ich weiß bei jedem Schritt, was ich zu tun habe.	☐	☐	☐	☐	☐

	Stimme gar nicht zu (1)	**Stimme eher nicht zu (2)**	**Neutral (3)**	**Stimme eher zu (4)**	**Stimme vollkommen zu (5)**
2.1 Ich war stark in die virtuelle Umgebung eingebunden.	☐	☐	☐	☐	☐
2.2 Meine Erfahrung mit der Umgebung wirkte intuitiv.	☐	☐	☐	☐	☐
2.3 Ich hatte das Gefühl, ein Teil der visuellen Umgebung zu sein.	☐	☐	☐	☐	☐

	Stimme gar nicht zu (1)	**Stimme eher nicht zu (2)**	**Neutral (3)**	**Stimme eher zu (4)**	**Stimme vollkommen zu (5)**
3.1 Ich hatte Spaß während der Erfahrung.	☐	☐	☐	☐	☐
3.2 Ich würde die Erfahrung einem Freund weiterempfehlen.	☐	☐	☐	☐	☐

	Stimme gar nicht zu (1)	Stimme eher nicht zu (2)	Neutral (3)	Stimme eher zu (4)	Stimme vollkommen zu (5)
4.1 Die Visualisierungen haben zur Musik gepasst.	☐	☐	☐	☐	☐
4.2 Die Visualisierungen waren optisch ansprechend.	☐	☐	☐	☐	☐
4.3 Das System hat angemessen auf meine Musik reagiert.	☐	☐	☐	☐	☐

	Gar nicht (0)	Sehr niedrig (1)	Niedrig (2)	Eher niedrig (3)	Mittel (4)	Eher hoch (5)	Hoch (6)	Sehr hoch (7)
5.1 Wie mental anspruchsvoll war die Aufgabe?	☐	☐	☐	☐	☐	☐	☐	☐
5.2 Wie erfolgreich waren Sie bei der Erfüllung der Aufgabe?	☐	☐	☐	☐	☐	☐	☐	☐
5.3 Wie unsicher, entmutigt, irritiert, gestresst oder verärgert waren Sie?	☐	☐	☐	☐	☐	☐	☐	☐

	Stimme gar nicht zu (1)	Stimme eher nicht zu (2)	Neutral (3)	Stimme eher zu (4)	Stimme vollkommen zu (5)
6.1 Ich habe mich in meine Musik hineinversetzt gefühlt.	☐	☐	☐	☐	☐
6.2 Die Umgebung hat meine Kreativität angeregt.	☐	☐	☐	☐	☐
6.3 Ich wurde durch die Umgebung zu neuen musikalischen Ideen inspiriert.	☐	☐	☐	☐	☐
6.4 Mein kreativer Workflow wurde durch die Umgebung unterstützt.	☐	☐	☐	☐	☐

	Stimme gar nicht zu (1)	Stimme eher nicht zu (2)	Neutral (3)	Stimme eher zu (4)	Stimme vollkommen zu (5)
7.1 Ich hatte ausreichend Kontrolle über die Umgebung.	☐	☐	☐	☐	☐
7.2 Ich fühlte mich als aktiver Gestalter der Umgebung.	☐	☐	☐	☐	☐
7.3 Die Umgebung war hilfreich	☐	☐	☐	☐	☐

	Stimme gar nicht zu (1)	Stimme eher nicht zu (2)	Neutral (3)	Stimme eher zu (4)	Stimme vollkommen zu (5)
8.1 Die Umgebung hat mich von der Musikproduktion abgelenkt.	☐	☐	☐	☐	☐
8.2 Ich konnte mich gut auf die mir gestellte Aufgabe konzentrieren.	☐	☐	☐	☐	☐

Vielen Dank für das Ausfüllen dieses Fragebogens!

Appendix D

Post VR Questionnaire

Endbefragung

Teilnehmer-Code: _______________

SYSTEM-RANKING

Bitte erstelle ein persönliches Ranking der Sessions (No VR, Baseline, Anton, Dominik, Manuell, Modify) von 1 (beste) bis 6 (schlechteste):

Rang	Session-Nummer	Name
1. (beste)	____________	____________
2.	____________	____________
3.	____________	____________
4.	____________	____________
5.	____________	____________
6.. (schlechteste)	____________	____________

Hinweis: "Session-Nummer" bezieht sich darauf, wann Du das jeweilige System getestet hast (1. = zuerst getestet, 6. = zuletzt getestet) Du kannst jederzeit nachfragen, wann welches System an der Reihe war.

Kannst Du dir vorstellen, eines oder mehrere der getesteten Systeme in deinem eigenen kreativen Workflow zu verwenden?

☐ Ja, definitiv
☐ Ja, wahrscheinlich
☐ Unentschieden
☐ Nein, eher nicht
☐ Nein, definitiv nicht

3.3 Falls Du nein angekreuzt hast, erläutere bitte kurz deine Entscheidung:

Falls Du ja angekreuzt hast, welche Systeme würdest Du verwenden wollen? (Mehrfachauswahl möglich)

☐ System Nummer 1
☐ System Nummer 2
☐ System Nummer 3
☐ System Nummer 4
☐ System Nummer 5

Wie ist deine generelle Einstellung zur Verwendung von Künstlicher Intelligenz (KI) zur Generation von immersiven Umgebungen?

☐ Sehr positiv
☐ Positiv
☐ Neutral
☐ Negativ
☐ Sehr negativ

Vielen Dank für deine Teilnahme und das Ausfüllen dieses Fragebogens!

Appendix E

Interview Questions

INTERVIEW-LEITFADEN POST-STUDY

Einleitung: "Vielen Dank für Ihre Teilnahme an der Studie! Ich möchte nun gerne ein kurzes Interview mit Ihnen führen, um Ihre Erfahrungen besser zu verstehen. Das Interview wird etwa 10-15 Minuten dauern”

---

1. ALLGEMEINE ERFAHRUNG

1.1 Gesamteindruck

- Wie würden Sie Ihre Gesamterfahrung mit den VR-System beschreiben?
- Gab es Momente, die Sie besonders überrascht haben?
- Was für einen Einfluss hatte das Tragen der VR Brille auf deine Musikproduktion?

---

2. VERGLEICH DER BEDINGUNGEN

2.1 Präferenzen

- Welche der fünf Bedingungen hat Ihnen am besten gefallen (Ranking besprechen)
- Welche Bedingung fanden Sie am wenigsten hilfreich? Aus welchen Gründen?

2.2 Unterschiede in der Nutzung

- Hast du bei verschiedenen Bedingungen unterschiedlich gearbeitet?
- In welcher Bedingung fühltest du dich am kreativsten?
- Wo hattest du das Gefühl, am produktivsten zu sein?

---

3. MUSIK-VISUAL VERBINDUNG

3.1 Synchronisation

- Wie gut haben die Visuals insgesamt Ihrer Musik gepasst?
- Gab es Momente, wo die Verbindung zwischen Musik und Visual besonders stark war? Gab es Momente wo die Visuals überhaupt nicht gepasst haben?
- Hast du das als störend empfunden?

3.2 Einfluss auf den kreativen Prozess

- Hat die immersive Umgebung deine Musikproduktion beeinflusst? Wie?
- Haben Sie aufgrund der immersiven Umgebung musikalische Entscheidungen anders getroffen?
- Konnten die immersiven Umgebungen dir neue Ideen geben?
- Hast du durch die Umgebungen Aspekte deiner eigenen Musik anders wahrgenommen?

---

4. SYSTEMKONTROLLE UND BENUTZERFREUNDLICHKEIT

4.1 Kontrolle über das System

- Wie war es für Sie, in manchen Bedingungen mehr bzw. weniger Kontrolle zu haben?
- Wann fühlten Sie sich am meisten/wenigsten im Kontrolle?
- Was würden Sie sich an zusätzlichen Kontrollmöglichkeiten wünschen?

4.3 Technische Probleme

- Sind Ihnen technische Probleme aufgefallen die nicht vorher besprochen wurden sind?
- Wie haben sich diese auf Ihre Arbeitsweise ausgewirkt?
- Gab es Momente, wo Sie das Gefühl hatten, das System reagiert nicht richtig?

---

6. FLOW UND KREATIVITÄT

6.1 Flow-Erlebnis

- In welcher Bedingung sind Sie am ehesten 'in den Flow' gekommen?
- Was hat Sie aus dem Flow herausgebracht?
- Haben die Umgebungen geholfen oder gehindert, sich zu konzentrieren?

6.2 Kreative Prozesse

- Hat die VR-Umgebung Ihre normale Arbeitsweise verändert? Inwiefern?
- Sind Sie experimentierfreudiger geworden durch die Umgebung?
- Würden Sie sagen, dass die Umgebung Ihre Kreativität gefördert haben?
- Hat Sie die virtuelle Umgebung abgelenkt oder unterstützt?

Zukünftige Nutzung

- Könnten Sie sich vorstellen, so ein System in Ihrem normalen Workflow zu verwenden?
- Unter welchen Umständen würden Sie es nutzen/nicht nutzen?

Verbesserungsvorschläge

- Was würden Sie am System ändern oder verbessern?
- Welche Features würden Sie sich zusätzlich wünschen?
- Was war überflüssig oder störend?

---

8. SPEZIFISCHE NACHFRAGEN

[Basierend auf Beobachtungen während der Studie oder auffälligen Fragebogen-Antworten]

8.1 Beobachtungen ansprechen

- Mir ist aufgefallen, dass Sie in [Bedingung X] [Verhalten Y] gezeigt haben. Können Sie das beschreiben?
- Bei [spezifischer Situation] wirkten Sie [Beobachtung]. Was ging Ihnen da durch den Kopf?

ABSCHLUSS

Zusätzliche Gedanken

- Gibt es noch etwas Wichtiges, was wir nicht besprochen haben?
- Haben Sie noch Fragen zum System oder zur Studie?

Feedback zur Studie

- Haben Sie Verbesserungsvorschläge für zukünftige Studien?

Interview-Ende: *:*
Gesamtdauer: ______ Minuten

Bibliography

[1] Nissoite Abdou, Milana Dzhamalova, Samuel Désirée, Théo Le Goff, Bruno Mesz, and Jean-Christophe Sakdavong. The Mediating Effect of Presence on Musical Emotions in Virtual Environments. In *Extended Reality: International Conference, XR Salento 2024, Lecce, Italy, September 4–7, 2024, Proceedings, Part IV*, page 3–20, Berlin, Heidelberg, 2024. Springer-Verlag. doi.org/10.1007/978-3-031-71713-0_1.

[2] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77 – 101, 2006. doi.org/10.1191/1478088706qp063oa. Cited by: 125114.

[3] S.-H. Hung C.-W. Yeh and C.-Y. Chang. The influence of natural environments on creativity. *Frontiers in Psychiatry*, Vol. 13, july 2022. doi.org/10.3389/fpsyt.2022.895213.

[4] Tuhin Chakrabarty, Arkadiy Saakyan, Olivia Winn, Artemis Panagopoulou, Yue Yang, Marianna Apidianaki, and Smaranda Muresan. I Spy a Metaphor: Large Language Models and Diffusion Models Co-Create Visual Metaphors. In *Annual Meeting of the Association for Computational Linguistics*, 2023.

[5] Weston Chimbudzi, Richard Muranda, and Wonder Maguraushe. The Evolution of the Roles of Producers in the Zimbabwe Recording Industry. *Pop Music, Culture, and Identity*, Part F1533:395 – 411, 2022. doi.org/10.1007/978-3-030-98705-3_23.

[6] James J. Cummings and Jeremy N. Bailenson. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*, 19(2):272–309, 2016. doi.org/10.1080/15213269.2015.1015740.

[7] Laura Josephine Hagedorn, Alwin De Rooij, and Maryam Alimardani. Virtual Reality and Creativity: How do Immersive Environments Stimulate the Brain during Creative Idea Generation? In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23, New York, NY, USA, 2023. Association for Computing Machinery. doi.org/10.1145/3544549.3585848.

[8] Jenny Huang, Christoph Johannes Weber, and Sylvia Rothe. An AI-driven Music Visualization System for Generating Meaningful Audio-Responsive Visuals in Real-Time. In *Proceedings of the 2025 ACM International Conference on Interactive Media Experiences*, IMX '25, page 258–274, New York, NY, USA, 2025. Association for Computing Machinery. `doi.org/10.1145/3706370.3727869`.

[9] S. Lee J. H. Lee. Relationships between physical environments and creativity: A scoping review. *Thinking Skills and Creativity*, Vol. 48, june 2023. `doi.org/10.1016/j.tsc.2023.101276`.

[10] Alexander Lerch. The Relation Between Music Technology and Music Industry. *Springer Handbooks*, page 899 – 909, 2018. `doi.org/10.1007/978-3-662-55004-5_44`. Cited by: 5.

[11] Li Li. Research and Implementation of Real Time Music Visualization. volume 12474, 2022. `doi.org/10.1117/12.2654065`. Cited by: 0.

[12] Hugo B. Lima, Carlos G. R. Dos Santos, and Bianchi S. Meiguins. A Survey of Music Visualization Techniques. *ACM Comput. Surv.*, 54(7), July 2021. `doi.org/10.1145/3461835`.

[13] Brenda Massetti. An empirical examination of the value of creativity support systems on idea generation. *MIS quarterly*, pages 83–97, 1996.

[14] Leslie M. Meier. *Popular music making and promotional work inside the ew' music industry*. Routledge, 2015. `doi.org/10.4324/9781315725437-45`. Cited by: 10.

[15] Jon McCormack Meng Yang, Maria Teresa Llano. Exploring Real-Time Music-to-Image Systems for Creative Inspiration in Music Creation. *International Conference on Computational Creativity*, July 2024.

[16] Rayna Ney, Aydin Javadov, Melanie Baumgartner, Raphael Weibel, Florian von Wangenheim, and Joseph Ollier. Generative AI for Wellness Applications via User-Generated Immersive Virtual Environments. In *GenAICHI: CHI 2025 Workshop on Generative AI and HCI*, Yokohama, Japan, 2025. ACM. Workshop paper.

[17] Ruben Schlagowski, Fabian Wildgrube, Silvan Mertes, Ceenu George, and Elisabeth André. Flow with the Beat! Human-Centered Design of Virtual Environments for Musical Creativity Support in VR. In *Proceedings of the 14th Conference on Creativity and Cognition*, CC '22, page 428–442, New York, NY, USA, 2022. Association for Computing Machinery. `doi.org/10.1145/3527927.3532799`.

[18] Chien-Wen Cheng Shao-Ying Lin. Enhancing Musical Immersion: A Multimodal Framework Integrating GSR Signals with Visual Art Generation. In *2024 IEEE International Symposium on Consumer Technology (ISCT)*, pages 102–106. IEEE, August 2024. `doi.org/10.1109/ISCT62336.2024.10791200`.

[19] George A. Sielis, Aimilia Tzanavari, and George A. Papadopoulos. Enhancing the Creativity Process by Adding Context Awareness in Creativity Support Tools. In Constantine Stephanidis, editor, *Universal Access in Human-Computer Interaction. Part III: Applications and Services*, volume 5616 of *Lecture Notes in Computer Science*, pages 424–433, Berlin, Heidelberg, 2009. Springer. `doi.org/10.1007/978-3-642-02714-7_45`.

[20] Paul Thompson. Any sound you can imagine? The Bedroom Producer, creativity and popular music education. *Journal of Popular Music Education*, 7(2): 205 – 218, 2023. `doi.org/10.1386/jpme_00114_1`. Cited by: 2.

[21] Kai Wang and Jeffrey V. Nickerson. A literature review on individual creativity support systems. *Computers in Human Behavior*, 74:139–151, 2017. `doi.org/10.1016/j.chb.2017.04.035`.

Index

Typeset October 13, 2025