HEC MONTREAL

Sounds in immersive experiences

Par

Farah Abou Jaoude

Annemarie Lesage
HEC Montreal
Codirectrice de recherche

Marie- Claude Trudel
HEC Montreal
Codirectrice de recherche

Science de la gestion
(spécialisation Expérience utilisateur)

Memoire presente en vue d'obtention
du grade de maîtrise es science en gestion
(M. Sc.)

Août 2025

Abstract

This study investigates the emotional effects of different audio rendering technologies on user experience, comparing spatial audio with high-fidelity stereo. The research addresses a gap in empirical evidence regarding the claimed benefits of spatial audio for immersion and emotional engagement.

A within-subjects experimental design was used, in which 38 participants listened to a series of musical and environmental audio clips in a specialized listening room. Data was collected through mixed methods. Participants used questionnaires to rate their subjective levels of immersion, presence, cognitive load, and temporal dissociation after each clip. They also completed a ranking task and took part in semi-structured interviews to describe their perceptions and emotional responses.

The results showed a consistent pattern across the collected data. Participants gave higher ratings for immersion, realism, and physical presence to spatial audio compared to stereo. They also reported a greater sense of losing track of time and feeling disconnected from their physical surroundings with spatial audio. The qualitative data indicated that spatial audio was associated with a listening experience that felt more participatory and enveloping, leading to more detailed mental imagery and descriptions of emotional and physiological reactions.

These findings suggest that the dimensionality of sound is a notable factor in mediated experiences. The study concludes that spatial audio can be used to generate experiences of immersion and emotional connection, which has applications for content creators, technologists, and UX designers in areas such as entertainment and virtual reality.

Keywords: spatial audio, stereo, immersion, presence, user experience (UX), emotional engagement, psychoacoustics, sound design, immersive media

Résumé

Cette étude examine les effets émotionnels des différentes technologies de rendu audio sur l'expérience utilisateur, en comparant l'audio spatial au son stéréo haute-fidélité. La recherche aborde un manque de preuves empiriques concernant les bénéfices revendiqués de l'audio spatial pour l'immersion et l'engagement émotionnel.

Un devis expérimental intra-sujet a été utilisé, dans lequel 38 participants ont écouté une série de clips audio musicaux et environnementaux dans une salle d'écoute spécialisée. Les données ont été collectées par des méthodes mixtes. Les participants ont utilisé des questionnaires pour évaluer leurs niveaux subjectifs d'immersion, de présence, de charge cognitive et de dissociation temporelle après chaque clip. Ils ont également complété une tâche de classement et participé à des entretiens semi-dirigés pour décrire leurs perceptions et leurs réponses émotionnelles.

Les résultats ont révélé une tendance cohérente à travers les données collectées. Les participants ont attribué des notes plus élevées pour l'immersion, le réalisme et la présence physique à l'audio spatial par rapport à la stéréo. Ils ont également rapporté une plus grande sensation de perte de la notion du temps et de déconnexion de leur environnement physique avec l'audio spatial. Les données qualitatives ont indiqué que l'audio spatial était associé à une expérience d'écoute perçue comme plus participative et enveloppante, menant à une imagerie mentale plus détaillée et à des descriptions de réactions émotionnelles et physiologiques.

Ces résultats suggèrent que la dimensionnalité du son est un facteur notable dans les expériences médiatisées. L'étude conclut que l'audio spatial peut être utilisé pour générer des expériences d'immersion et de connexion émotionnelle, ce qui a des applications pour les créateurs de contenu, les technologues et les concepteurs UX dans des domaines tels que le divertissement et la réalité virtuelle .

Mots-clés : audio spatial, stéréo, immersion, présence, expérience utilisateur (UX), engagement émotionnel, psychoacoustique, conception sonore, médias immersifs

Table of Contents

List of figures and tables:

Preface:

Have you ever closed your eyes during a movie and realized the sound alone was telling the story? Or felt a shiver from a song that seemed to wrap itself around you? That feeling is what this thesis is all about. For as long as I can remember, sound has felt like a kind of magic—the invisible architecture of our reality. It's a force that can build entire worlds in our minds, and I wanted to understand exactly how it works.

When I first experienced true spatial audio, it was a revelation. Sound was no longer coming from speakers; it was simply *there*, all around me, with a life of its own. I knew instantly that this technology was more than just a gimmick; it was a doorway to a different kind of human experience. I became fascinated with the "why." Why does one soundscape feel more real than another? How can audio make us lose all track of time, or feel an intimate connection to a voice recorded years ago? This study is my exploration of that magic, an attempt to map the human side of this incredible technology.

Ultimately, my hope is that this work inspires you to think about sound differently. For the creators and designers out there, I hope it offers a new color for your palette. For everyone else, I hope it serves as a simple invitation: to close your eyes, open your ears, and listen a little more closely to the rich, sonic worlds we create and live within every day.

Acknowledgement:

This research, while bearing my name, is truly the product of a community of support, guidance, and encouragement. I am filled with immense gratitude for everyone who played a role in bringing this work to life.

First and foremost, I extend my deepest thanks to my codirectors, Marie-Claude Trudel and especially Annemarie Lesage. Annemarie, thank you for believing in me; your dedication and unwavering support were the bedrock of this project. Your collective mentorship, sharp insights, and endless patience pushed me to refine my thinking and gave me a constant source of strength.

This study would not have been possible without the team at Centre PHI; special thanks to Rennelle Desjardins and Phil Rochefort for making this project come to life and for generously opening the doors to their Habitat Sonore, which allowed the magic of spatial audio to be truly felt.

Also my appreciation goes to HEC Montréal and the Tech3lab for providing an environment of intellectual curiosity and innovation.

I am also deeply grateful to my dear friend and colleague, Maivel Abdelnoor, whose help with data collection was invaluable.

To the 38 participants who generously offered their time and attention, thank you. Your thoughtful feedback and willingness to share your experiences were the heart of this research. You gave this study its voice.

Finally, to my family, thank you for your unwavering support and for being my cheerleaders through it all. And to my friends, thank you for the love and the occasional, much-needed

distractions that kept me grounded and motivated. This journey was made so much brighter because of all of you.

Chapter 1: Introduction

1.1 Context

In the ever-evolving landscape of digital media, the pursuit of realism has consistently driven technological innovation. While visual fidelity has long been at the forefront of this evolution, the role of audio in creating compelling and believable experiences is now a central focus of research and development (Saroka, Volha. 2024). The human auditory system provides a rich, 360-degree sense of our surroundings, offering critical cues for spatial awareness, emotional context, and environmental understanding that modern audio technologies aim to replicate (Zhu et al., 2025). As such, the quality and dimensionality of sound are essential to crafting deeply engaging and present mediated experiences (Saroka, Volha. 2024).

The last decade has seen a fundamental transition in audio technology, characterized by the shift from traditional, channel-based stereophonic sound to advanced spatial audio. Stereo, for all its utility, presents sound as a relatively flat plane, a "window" through which the user listens, a limitation to channel-based standards (Francombe et al., 2017, Kasmenda et al. 2024). In contrast, modern spatial audio aims to create immersive auditory experiences by modeling both the physical acoustics and psychoacoustic perception of sound in three-dimensional space. (Koyama, De Sena, Samarasinghe, Thomas, & Antonacci, 2025). Technologies such as Dolby Atmos and DTS:X, which are grounded in object-based audio principles, aim to replicate real-world hearing. When delivered via headphones, the use of advanced binaural rendering—often through individualized HRTFs—is essential to achieving an accurate and convincing immersive experience (Rafaely et al., 2022)

Audio immersion is an important component for creating believable virtual experiences, as it provides the spatial and emotional context that visuals alone cannot. Scholars argue that sound is fundamental to generating a sense of presence—the feeling of "being there" in a mediated environment—by providing 360-degree environmental cues that ground the listener in a coherent and responsive world (Hameed, Möller, & Perkis, 2024). Unlike visuals, which are often confined to a limited field of view, spatialized audio can direct a user's attention, signal off-screen events, and enhance the perception of space and volume (Gao, Wang, Feng, & Lv, 2022). Furthermore, audio assumes a powerful role in modulating the user's emotional state, with the ability to build tension, create intimacy, or elicit surprise, thereby deepening the engagement and believability of the immersive experience (Saroka, Volha. 2024).

The widespread integration of spatial audio technologies in entertainment, virtual reality, and communication platforms has shown a gap in the existing literature. While the technical capabilities of spatial audio are well-documented, research on its psychological effects remains underdeveloped. Preliminary studies indicate that immersive web-based spatial audio applications can reduce stress and negative mood states in both clinical and non-clinical populations (Greenberg et al., 2021). However, as noted by Stefanowska and Zieliński (2024), while there is growing evidence of a positive influence of spatial audio characteristics on listeners' affective responses, more data is required to build reliable, universal, and useful models explaining this relationship. Does a more realistic soundfield lead to a greater sense of immersion? Does the ability to spatialize sound sources affect the user's emotional connection to the content?

1.2 Statement of the Problem

The central issue addressed by this research is the absence of direct, empirical evidence comparing the psychological effects of contemporary spatial audio to traditional high-fidelity stereo on key user experience metrics. While the advantages of spatial audio are frequently emphasized, claims of enhanced immersion and emotional engagement often rely on technical specifications or anecdotal reports rather than controlled empirical validation (Stefanowska & Zieliński, 2024). The academic literature, though developing, has yet to provide sufficient controlled investigations that isolate audio format as the primary variable (Agrawal et al., 2022). Consequently, content creators and developers often operate on industry assumptions rather than data-driven insights. Given this gap, the present research is positioned as an exploratory study designed to provide foundational, empirical data. It aims to bridge this gap by isolating the audio format as a variable to measure its direct impact on user immersion and emotional response.

1.3 Purpose of the Study and Research Questions

The purpose of this exploratory study is to investigate how differences in audio rendering quality, specifically spatial audio versus stereo sound, influence a listener's perceived levels of immersion and emotional engagement. This will be achieved through a within-subject experiment designed to provide a direct comparison of the two formats.

The study will be guided by the following research questions:

1. How does sound quality, defined by the rendering format (spatial vs. stereo), affect the user's **subjective experience of immersion i.e. the sense of presence**?
    a. And therefore, what is the impact of this difference in sound quality on the user's **emotional experience** and **engagement** with audio content?

1.4 Significance of the Study

This research holds significant value across multiple domains. For content creators and sound designers, it offers empirical evidence to inform decisions regarding the implementation of spatial audio to achieve specific creative objectives. For technologists and software developers, it provides user-centric data to refine rendering algorithms and enhance the overall quality of immersive experiences. From a scientific perspective, this study contributes to our fundamental understanding of human perception, investigating how individuals process complex auditory cues and how these processes influence the feeling of immersion in various experiences.

The appeal of this discovery lies in its potential to deliver realistic experiences by creating three-dimensional soundscapes that envelop the listener. While the technical standards of these technologies are well-documented, their psychological impact on users remains an evolving area of interest (Stefanowska & Zieliński, 2024). A deeper understanding of how sound's fidelity and dimensionality affect human perception, immersion, and emotion is crucial, not only for media creators and technologists but also for the fundamental study of the human-computer interface.

1.5 Structure of the Thesis

This thesis is organized into five chapters. This Chapter 1 has provided an introduction to the research problem, the purpose of the study, and the guiding research questions. Chapter 2 will present a comprehensive review of the relevant literature on immersion, presence, psychoacoustics, and emotional responses to audio. Chapter 3 will detail the methodology used in the experiment, including the research design, participants, materials, and procedures. Finally, Chapter 4 will present the results of the statistical analyses. Chapter 5 will discuss the findings in the context of the existing literature and chapter 6 concludes with possible implications of our findings, acknowledges the study's limitations, and offers recommendations for future research.

1.6 Author's Contributions:

Research process	Student Contribution
Literature review	**100% Literature review:** Read and analysed more that 50 articles to identify concepts and gaps to explore
Research question	**70% Research question formulation:** Formulate a research question to answer the identified gap & problem in literature
Experimental design	**70% Ethics Approval:** Prepare the required documentation for the REB. Help was provided by directors for the final review of the application **80% Clip selection:** Select the clips used for the tests with the help of the sound engineer at centre PHI **80% Questionnaire Development:** Develop the data collection methods and questions on qualtrics and the final interview structure with the help of my directors
Recruitment	**80% Participant Recruitment:** Provide the participant criterias for the recruitment through Panelfox (with the help of Tech3Lab). I was in charge of contacting the participants and the distribution of compensations.
Data Collection	**75% Data Collection:** Moderate the experiment and all the data collection methods. Help was provided for the final interview
Data Analysis	**75% Data Analysis:** Performed statistical analysis for the quantitative data, used AI tools as part of my content analysis
Thesis Writing	**95% Thesis Writing:** Write the thesis details. Review feedback and guidance was provided by research directors through the whole process

Table 1: Author's Contribution

Chapter 2: Literature Review

2.1 Introduction

This chapter synthesizes a broad range of literature from psychoacoustics, media psychology, human-computer interaction, and neuroscience to construct a theoretical framework for the current study. It begins by defining the foundational concepts of immersion and presence, establishing them as the core psychological states that immersive media seeks to create. It then examines the technological and psychoacoustic principles that differentiate spatial audio from stereo, detailing how the human processes auditory cues to construct a sense of space. Subsequently, the review explores the links between audio quality, cognitive load, emotional engagement, and narrative transportation.

2.2 Theoretical Foundations of Immersion and Presence

2.2.1 Defining Immersion in Digital Experiences

Immersion is a central concept in the study of digital media, describing a user's state of deep mental and sensory engagement with a mediated environment. It is crucial to distinguish between technological immersion, which refers to the objective capabilities of a system to deliver a vivid sensory illusion, and psychological immersion, the user's subjective experience of being absorbed into that illusion (Agrawal et al., 2020). The current study focuses primarily on psychological immersion, which is understood not as a binary state but as a graduated process. Foundational research conceptualized immersion as a progression through stages such as engagement, engrossment, and total immersion, a model that continues to inform contemporary views on player experience (Brown & Cairns, 2004). This framework emphasizes a journey from initial interest to a state in which the user's awareness of the real world is significantly diminished. For the purposes of this study, the term *immersion* will refer specifically to psychological immersion unless otherwise noted.

This deep level of engagement is often associated with the psychological concept of "flow," a state of optimal experience in which individuals are so fully absorbed in an activity that they lose self-consciousness and find the activity intrinsically rewarding (Csikszentmihalyi, 1990). The relationship between flow and immersion has been validated in modern contexts such as virtual reality gaming, where a balance between challenge and skill fosters a strong sense of presence (Hassan, Lobna & Jylhä, Henrietta & Sjöblom, Max & Hamari, Juho. 2020). In digital media, the "challenge" can be interpreted as suspending disbelief, while the user's "skill" is supported by a convincing and minimally intrusive system. Key markers of successful immersion include temporal dissociation, such as losing track of time, and a reduction of self-awareness, as the user's cognitive and perceptual faculties are fully dedicated to the mediated environment (Agrawal et al., 2020).

2.2.2 Presence: The Feeling of "Being There"

Psychological immersion is the process through which a user becomes cognitively and sensorially engaged with a digital environment, which, when successful, can lead to presence—the distinct subjective feeling of “being there” within the virtual world (Brown & Cairns, 2004; Witmer & Singer, 1998). Presence was formally defined in foundational work as a “perceptual illusion of non-mediation” (Lombard & Ditton, 1997), a concept that has been refined and integrated into contemporary models of user experience (Latoschik & Wienrich, 2022). This illusion is achieved when the user no longer consciously notices the technology and instead accepts the mediated environment as a real place they are inhabiting.

This concept was further refined by distinguishing between different forms of presence, most notably **spatial presence** (the sensation of being physically located in the virtual space), **social presence** (the feeling of being with others), and **self-presence** (the alignment of the user's identity with their virtual self) (Lee, 2004). The importance and interplay of these dimensions have since been validated by large-scale systematic reviews of the field (Oh et al., 2018). While this study focuses on spatial presence, it acknowledges that a strong sense of it is often a prerequisite for the other forms.

The relationship between immersion and presence is synergistic yet distinct. A seminal argument posits that **immersion** describes the objective technical capabilities of a system (e.g., high-resolution visuals), whereas **presence** is the subjective, psychological response to that environment (Slater, 2009). This distinction remains a cornerstone of current frameworks, which affirm that highly immersive systems are more likely to induce a strong sense of presence (Agrawal et al., 2020). This positions presence as a crucial mediator between technology and user experience and as the psychological outcome that technologies like spatial audio aim to generate.

2.2.3 Multi-Sensory Integration in Immersive Experiences

Human perception is an inherently multi-sensory process. The brain continuously integrates information from different sensory channels—vision, hearing, and touch—to form a single, coherent model of the world, a process known as **cross-modal processing** (Spence & Charles, 2011). This integration is now increasingly understood through the lens of causal inference, where the brain actively determines the most likely cause of its sensory inputs (Jones et al., 2024). In many situations, vision is the dominant sense; a classic example is the "ventriloquism effect," where the perceived location of a sound is captured by a synchronized but spatially offset visual cue (Alais & Burr, 2004). However, recent studies show that as audio fidelity improves, this visual dominance can be challenged, with highly plausible spatial audio cues more strongly resisting visual capture (Schmidt et al., 202).

The auditory system possesses unique advantages that make it critical for immersion. Sound provides 360-degree, omnidirectional information, alerting us to events happening outside our field of view (Blauert,Jeans, 2001). In immersive environments, the congruence and fidelity of audio-visual integration are paramount. When audio and visual cues are perfectly synchronized and spatially aligned, they reinforce one another, strengthening the illusion of **presence**. Conversely, a mismatch—such as the sound of a footstep being out of sync with the visual, or coming from the wrong direction—can immediately shatter the illusion (Slater, 2009). While this principle is well-established, modern systems are now leveraging machine learning to dynamically manage this synchronization in real-time to maximize user immersion (Saha et al., 2024). In cases of sensory degradation, such as a low-resolution visual display, high-quality audio can perform a compensatory function, "filling in" the missing detail and sustaining the user's engagement—a principle now validated in modern applications like streaming 360° video (Shafi et al., 2020).

2.3 The Technological and Psychoacoustic Basis of Spatial Sound

2.3.1 From Stereo "Windows" to Spatial "Worlds"

The evolution of audio reproduction can be understood as a journey toward greater ecological validity—a closer approximation of how we perceive sound in the natural world. In modern applications like virtual reality, this goal is often framed within the concepts of plausibility and immersion (Serafin et al., 2018). Traditional stereophonic sound, while an improvement over mono, primarily presents a soundstage as a flat "window," and even channel-based surround sound remains limited by a fixed speaker layout (Francombe et al., 2023).

The advent of object-based spatial audio, exemplified by systems such as Dolby Atmos and MPEG-H, marks a fundamental shift in audio production. Instead of mixing sound to fixed channels, sounds are treated as discrete "objects" accompanied by metadata specifying their precise three-dimensional location and other attributes. These audio objects are then rendered in real time to match the playback environment and hardware, whether it be a multi-speaker setup or headphones. For headphone listening, this rendering typically uses binaural audio applying Head-Related Transfer Functions (HRTFs) to simulate how sound waves interact with the listener’s head and ears, creating an immersive 3D spatial effect. A key challenge in achieving accurate spatial perception is the personalization of HRTFs to account for individual anatomical differences, which significantly enhances localization accuracy (Oehler et al., 2023). Recent research leverages artificial intelligence and machine learning to generate personalized HRTFs from user-specific data, improving accessibility and precision in spatial audio experiences (Fantini, 2025). This transition from traditional stereo to a personalized, object-based spatial sound environment represents the core innovative variable under investigation in this study (Bleidt, 2014; Object-Based Audio Mixing, 2021).

2.3.2 The Psychoacoustics of Spatial Perception

Our ability to navigate a complex sonic environment relies on a sophisticated set of innate psychoacoustic mechanisms. Bregman's (1990) theory of Auditory Scene Analysis (ASA) provides the framework, describing how the brain parses a continuous stream of sound into distinct "objects" or "streams." Modern cognitive neuroscience has further revealed that this is an active process involving distinct cortical pathways for identifying what a sound is and where it is located (Ahveninen, Huang, Nummenmaa, et al., 2013). This allows us to separate a speaker's voice from background music or follow a single car as it drives past.

The brain achieves this feat by analyzing several key cues. Sound source localization on the horizontal plane is primarily determined by Interaural Time Differences (ITDs) and Interaural Level Differences (ILDs) (Blauert, 1997). Recent research emphasizes how the brain robustly integrates these cues even in complex, reverberant environments where reflections could cause ambiguity (Ivanov et al., 2022). Localization on the vertical plane and front-back discrimination rely on the complex spectral filtering performed by the pinnae (Blauert, 1997), while distance perception is informed by cues such as loudness, the ratio of direct-to-reverberant sound, and high-frequency attenuation. Our ability to selectively focus on one sound source amidst many—the "cocktail party effect" (Cherry, 1953)—is a testament to the power of these mechanisms. It is now understood not just as passive filtering, but as an active process of selective attention, where neural activity dynamically tracks and enhances a chosen auditory object while suppressing others (Mesgarani, N., & Chang, E. F. 2012). The fundamental premise of spatial audio is to recreate these precise cues with a much higher degree of fidelity than is possible with traditional stereo. (Blauert, J. 1997)

2.3.3 Technical Implementation and Its Challenges

While the theory of spatial audio is solid, its practical implementation faces significant challenges. The quality of headphone-based spatial audio is highly dependent on the accuracy of the Head-Related Transfer Function (HRTF) used. Since every individual's anatomy is unique, a generic HRTF may not provide accurate localization for all listeners, a fact confirmed by extensive modern research (Oehler et al., 2023). While creating personalized HRTFs has traditionally required complex measurement processes, this barrier is now being addressed by AI-driven methods that can generate them from user data like photographs or 3D scans (Lee, G. W., & Kim, H. K. 2018). Beyond static accuracy, dynamic cues are also critical; without head-tracking, binaural audio remains "head-locked," a phenomenon known to severely limit realism and the externalization of sound sources in immersive systems (Brimijoin, & al. 2013).

Speaker-based systems, such as those using Ambisonics or Wave Field Synthesis, avoid this particular issue but are highly sensitive to room acoustics and require precise calibration, a challenge being tackled with advanced signal processing (Corteel, 2012). Consequently, the current market is a complex ecosystem of competing formats and rendering solutions, where the quality of the end-user experience can be highly variable depending on the interplay of hardware, software, and the content itself (Sun, X., 2021).

2.4 The Impact of Sound Quality on Immersion and Emotion

2.4.1 Defining and Measuring Sound Quality

Sound quality is a multi-faceted construct that extends beyond simple technical specifications. While objective metrics like frequency response and signal-to-noise ratio are important, the ultimate arbiter of quality is the human listener. Subjective quality assessment methods, such as those standardized by the International Telecommunication Union (ITU, 2015), are designed to capture this perceptual experience. Key perceptual attributes include fidelity, clarity, and freedom from artifacts, which are central to the modern evaluation of audio systems (Sun, X. 2021).

In the context of this research, the most critical attribute is spatial quality, which itself comprises several sub-dimensions: localization accuracy, envelopment (the sense of being surrounded), and source externalization (the perception of sound sources as being outside the head) (M.L. Franco Rojas, D.M. Rivera Pinzón 2022). This study proceeds from the premise that a higher degree of spatial quality will correlate with higher ratings of presence and immersion. (Johnston, D., Egermann, H., & Kearney, G. 2022)..

2.4.2 Audio Quality, Cognitive Load, and Emotional Engagement

According to Wickens' (2002) Multiple Resource Theory, humans possess limited pools of attentional resources. When an audio signal is of poor quality or spatially incompatible, the brain must expend more cognitive effort to decode it. This increases the "listening effort," a concept that extends Lavie's (2005) perceptual load theory, describing the deliberate allocation of finite mental resources to overcome challenges in perception (Neeman, et al. 2022). This effort diverts resources away from higher-level cognitive tasks, such as understanding narrative, appreciating aesthetic detail, or experiencing an emotional response. Prolonged exposure to such demanding conditions can lead to listening-related fatigue, a state of mental exhaustion that impairs cognitive performance (Kansegaran et al. 2023). Conversely, high-quality, intelligible, and spatially coherent audio is processed with less effort, freeing up cognitive resources and allowing for deeper engagement.

This principle has a direct connection with emotional experience. The brain's emotional centers are deeply intertwined with perceptual processing, forming networks that mediate the affective quality of sound (Koelsch, 2014). An audio experience that is perceptually rich and easy to process allows for a more direct and powerful emotional response. In virtual environments, presence is a core mechanism for evoking emotion; the stronger the feeling of "being there," the more intense the resulting experience (Susindar et al., 2023). Research has shown that spatial audio amplifies emotional responses precisely by enhancing this sense of presence. For example, studies have shown that spatial audio in virtual reality significantly enhances emotional engagement and immersion. Warp et al. (2022) found that head-tracked spatial audio increased physiological indicators of emotional response and improved users' ability to localize sound sources, thereby intensifying the sense of presence. Similarly, Yang et al. (2025) demonstrated that the spatial attributes of sound, such as direction and distance, can modulate emotional reactions—sounds originating from closer or more threatening positions elicited stronger arousal and negative affect. These findings suggest that spatialized audio cues, including musical scores and environmental sounds, can evoke profound feelings of fear, surprise, awe, or intimacy that are difficult to achieve with traditional stereo mixes.

2.5 Immersive Storytelling and Narrative Engagement

2.5.1 Audio's Role in Narrative Transportation

The power of a compelling story lies in its ability to transport an individual into its world. Transportation theory, as developed by Green and Brock (2000), describes this as a state of deep narrative immersion where real-world awareness recedes and the individual becomes fully engaged. This concept is a cornerstone of how we understand immersion in modern media, from interactive games to virtual reality (Wang, J. 2024). High-quality, dimensional audio is a powerful catalyst for this process. By creating a believable and evocative soundscape—a "storyworld"—audio has been empirically shown to enhance listener immersion and reduce the perceived distance between the user and the narrative (Isak de Villiers Bosman, Oğuz 'Oz' Buruk, Kristine Jørgensen & Juho Hamari 2023).

The distinction between diegetic sound (sounds within the story world, like dialogue or footsteps) and non-diegetic sound (external sounds, like a musical score) is crucial here (Chion, 1994). Spatial audio dramatically enhances the impact of diegetic sound by giving it a physical location and presence, making the storyworld feel more tangible. This heightened realism facilitates narrative transportation more effectively than stereo. For example, a study directly comparing the two formats found that a spatial audio narrative produced significantly higher scores on the Narrative Transportation Scale than its stereo counterpart (Vicente, P.-N., & Pérez-Seijo, S. 2022). This effect is particularly potent in genres like horror, where localized diegetic cues are used to maximize psychological impact and presence (Langiulli et al., 2023). Ultimately, spatial audio transforms the sound from a "window" onto the story into an environment the user can truly inhabit.

2.5.2 Interactive and Dynamic Audio Systems

In interactive media like video games and virtual reality, audio must be dynamic and responsive to the user. Adaptive audio systems are designed to change in real-time based on user actions, location, and narrative state. For instance, a game's musical score might seamlessly transition from calm exploration to intense combat music as an enemy approaches, a technique that has been extensively developed in modern games (Broderick, James & Redfern, Sam. 2018). This audio interactivity is crucial for reinforcing the user's sense of agency—the feeling that their actions have a meaningful impact on the environment (Slater, M., & Sanchez-Vives, M. V. (2016). Furthermore, procedural audio—the real-time synthesis of sound based on algorithmic and physics-based models—provides a method for creating infinitely variable and highly responsive soundscapes without relying on pre-recorded files, enhancing the dynamism and interactivity of virtual environments (AES Journal Forum, 2023). While this study employs non-interactive stimuli, the principles of responsive and believable sound design underscore the critical role of audio fidelity in maintaining a coherent and engaging user experience.

2.6 Synthesis, Research Gaps, and Justification for the Current Study

The literature reviewed establishes a theoretical pathway. The technical format of an audio system (spatial vs. stereo) directly influences its perceived quality, most notably its **spatial quality** attributes like localization accuracy and envelopment (Conetta, 2014). This enhancement in perceptual quality is a primary driver for inducing a stronger sense of presence—the subjective feeling of "being there" in a virtual environment. A review by Triberti et al. (2025) emphasizes that presence is not merely a byproduct of technological features but is influenced by content, narrative, and user characteristics. This heightened state of presence, in turn, acts as a crucial mediator, facilitating deeper immersion and enabling more intense and nuanced emotional engagement. Research by Liao (2025) supports this, indicating that cognitive absorption in VR enhances emotional engagement, leading to greater behavioral empathy.

Despite the strength of this theoretical chain, a critical gap persists in the empirical literature. While many studies have explored aspects of audio quality or presence in isolation, there is a notable lack of research that directly compares current, object-based spatial audio with high-fidelity traditional stereo in a controlled experimental setting to measure these specific psychological outcomes. Previous work has demonstrated that spatial audio can enhance presence and immersion in virtual environments (Poeschl, Wall, & Doering, 2014), and that immersive audio storytelling can predictably affect users' sense of presence (Verhulst et al., 2024). However, these studies do not directly compare object-based spatial audio with stereo in a manner that isolates its impact on perception and emotional engagement. Consequently, the field requires empirical data to move beyond theoretical postulation and validate the purported benefits of these new audio technologies.

Furthermore, measuring subjective states like immersion is notoriously challenging, and research in this area is often confounded by the significant individual differences in auditory perception, preference, and prior experience (Schwind et al., 2019). A **within-subjects experimental design**, as proposed for the current study, is the most effective methodology to address this challenge (Field, 2018). By having each participant experience both audio conditions, they serve as their own control. This design minimizes the statistical noise created by individual variance and allows for a more precise, powerful, and direct assessment of the audio format's true effect on the user experience.

This study is designed to fill this specific gap in the literature by providing empirical data on the psychological impact of spatial audio.

2.7 Conclusion and Research Questions

As stated previously, this study seeks to address the gap by answering the following primary research questions:

- How does sound quality, defined by the rendering format (spatial vs. stereo), affect the user's subjective experience of immersion, specifically the sense of presence?
    - And therefore, what is the impact of this difference in sound quality on the user's emotional experience and engagement with audio content?

Given the study's exploratory goals, this research is designed to investigate a potential chain of effects. The research will be guided by the assumption that **sound quality directly impacts the level of immersion in an experience**, and will explore whether this enhanced immersion from **spatial audio evokes stronger emotional engagement compared to stereo sound**.

Chapter 3: Methodology

1.Research Design:

This study uses a within-subject experimental design with a purpose of investigating how differences in sound quality, spatial audio versus stereo sound, influence participants' perceived levels of immersion, presence, and emotional engagement during audio experiences. Participants were not informed beforehand about the specific audio formats they would experience to minimize potential bias. They were told that the study's purpose was to understand how people perceive sound in immersive environments. That they will listen to a series of audio clips in the listening room with dimmed lights. Participants were instructed to use the first clip as a baseline for comparison with subsequent clips and were advised that the differences between each experience would be "subtle." They were encouraged to relax, focus on their subjective feelings, and assured that there were no right or wrong answers in the questionnaire.

Each participant was exposed to the same set of six audio clips. The first audio clip served as a baseline/introduction and was always presented in stereo format to all participants. The first clip allowed participants to get acquainted with the procedure and the kind of audio track that would follow. The survey data collected for this 'practice' clip was not included in the quantitative data set, thus taking out data affected by an initial surprise or "wow" effect associated with the novelty of the experience. The subsequent five clips were delivered in a randomized order and alternating audio format (stereo or spatial) to reduce potential bias.

The within-subject format was chosen to enable direct comparisons across audio conditions for each individual. This design ensured that all participants served as their own control, making the influence of the audio format on immersion more directly observable.

This research project has been evaluated and approved in accordance with ethical conduct for research involving human subjects by the Research Ethics Board (REB) of HEC Montréal. 2025-6304

2.Participants:

A total of 38 participants (21 women and 17 men), aged between 18 and 40 years with a median age of 27, were recruited for this study. Inclusion criteria were:

- Must be 18 years or older.

- Must have normal or corrected-to-normal hearing.

- Must feel comfortable listening to immersive and potentially unfamiliar audio environments.

- Must be fluent in French or English to complete the questionnaire reliably.

Participants were recruited from the university network (Tech3lab's Panelfox). Each individual received an information sheet and signed an informed consent form before the experiment began.

Sessions were conducted either individually or in small groups (2 to 4 participants), depending on scheduling availability. Regardless of group size, each participant was given space and autonomy in the room and completed their own questionnaire independently. The shared environment did not permit conversation or interaction during the experiment, maintaining the integrity of individual perception.

Participants were compensated 25$ via Interac transfer upon completion of the study

3.Settings:

The experiment was conducted at Centre PHI's Habitat Sonore, an immersive listening room designed to deliver high-quality spatial audio experiences. The room features 16 hidden and strategically placed speakers, allowing for three-dimensional acoustic diffusion. Its small capacity (maximum 10 persons) created a more personal space while its low-light environment enhanced the perception of sound by minimizing visual distractions, promoting focused listening.

The audio stimuli consisted of five sound clips, each lasting approximately 4 minutes and 30 seconds to 5 minutes. The content spanned a variety of music genres and nature soundscapes, selected to elicit emotional and sensory responses.

The audio tracks were selected to represent diverse content categories, ensuring a comprehensive assessment of how spatial audio affects different types of musical and auditory experiences. The selection strategy aimed to capture varying levels of familiarity, cultural backgrounds, musical complexity, and emotional intensity to provide a robust testing environment for spatial audio effectiveness.

- **Merrouh - by Emel:** Arabic song dwelling on traditional Middle Eastern musical elements (came across as cultural/exotic content) - Selected for its rich instrumental layers, and complex spatial arrangements that would benefit from three-dimensional audio positioning. This track tests how spatial audio enhances culturally unfamiliar music with intricate melodic structures.
- **Get Lucky - by Daft Punk**: Commercial pop song (familiar content) - Chosen as a widely recognizable track with clear stereo production elements, allowing us to see how participants experienced this familiar track as a spatial audio experience (unbeknownst to

them). Its electronic production style and distinct left-right channel separation make it ideal for demonstrating spatial audio enhancements.

- **Rescuer - by Emel:** Alternative/Indie song (unfamiliar musical content) - Selected to eliminate preconceived notions about how the song "should" sound, allowing for pure assessment of spatial audio impact without the bias of prior listening experiences. Its contemporary composition provides a suitable production for spatial processing.
- **Broken - by Ela Minus:** Primarily electronic with elements of indie electronic, techno, and alternative/indie rock song (unfamiliar musical content) - Chosen for its experimental electronic elements and dynamic range, which can showcase spatial audio's ability to create depth and movement in abstract musical compositions. The track's minimalist yet textured approach tests spatial audio's effectiveness with avant-garde content.
- **Lolupuppy Kutiquu - by Sofi Birch:** Nature melody with environmental sounds (ambient/natural content) - Selected specifically for its organic soundscape and environmental audio elements, which naturally lend themselves to three-dimensional positioning. This track tests how spatial audio enhances the realism and immersion of nature-based content, where directional audio cues are inherently important for creating authentic environmental experiences.

(See appendix B for Links)

Track	Genre/ Style	Familiarity	Key Musical Characteristics
Merrouh-Emel	Arabic / Traditional Middle Eastern	Culturally unfamiliar	Rich instrumental layers; complex melodic and spatial arrangements
Get Lucky-Daft Punk	Commercial Pop / Electronic	Familiar	Clear stereo production; distinct left-right channel separation
Rescuer-Emel	Alternative / Indie	Unfamiliar	Contemporary composition; modern production style
Broken – Ela Minus	Electronic / Indie Electronic / Techno	Unfamiliar	Experimental and minimalist; textured, dynamic range
Lolupuppy Kutiquu-Sofi Birch	Ambient / Nature Soundscape	Unfamiliar (natural content)	Organic environmental sounds; nature-based melodies

Table 2 : Clips description summary

Each of the five experimental clips was presented in either spatial or stereo format, randomized across participants. The spatial audio versions were professionally mixed to exploit the unique layout of Habitat Sonore and ensure an immersive distribution of sound throughout the room.

4.Experimental Procedure:

Upon arrival, participants were welcomed and seated in the center area of the room. The researcher gave a brief introduction explaining the flow of the session, without disclosing the focus on sound quality in order to avoid response bias.

The session unfolded as follows:

1. Participants first listened to a fixed stereo baseline clip, identical for all.
2. They then experienced five randomized clips in varying order and format (stereo or spatial), with the sequence predetermined by a randomization matrix.

3. After each clip, participants were given two minutes to complete a short questionnaire on their smartphone using Qualtrics.
4. At the end of the session, participants performed a ranking task, ordering the five experimental clips from most immersive (1) to least immersive (5).
5. A semi-structured interview followed to explore participants' reasoning, emotional responses, and perceptions of immersion.

The entire experimental session lasted approximately 60 minutes, including introduction, audio playback, questionnaire responses, ranking, and interview.

Step	Description	Duration (minutes)	Format
1. Welcome & Briefing	Participants received general instructions (without revealing the study’s focus on audio format).	~5	Individual or small group
2. Listening: Clip 0 (Baseline)	All participants listened to the same stereo baseline clip.	4.5–5	Audio
3. Post-Clip 0 Questionnaire	Participants completed the questionnaire	~2	Mobile survey
4. Listening: Clips 1–5 (Randomized)	Participants listened to five audio clips in randomized order and format (stereo/spatial).	~25	Audio
5. Post-Clip Surveys	After each of the 5 clips, participants answered the same questionnaire (same as for Clip 0).	~10	Mobile survey
6. Ranking Task	Participants ranked the 5 experimental clips from most to least immersive (1–5).	~1-2	Paper for ranking
7. Semi-Structured Interview	Individual interview recorded and transcribed.	~10	Audio recording

Table 3: experimental session

5.Measures:

a. Self-reported Immersion and realism

After each clip, participants completed a short survey designed to assess:

- **Temporal Dissociation**: Did you lose track of time during this audio experience?
- **Cognitive Load**: Did you feel like you had to focus to follow the audio? Did it feel natural?
- **Spatial Realism**: Did the audio feel like it was really coming from the space around you, or just from a speaker?
- **Presence**: Did the audio make you feel physically present in the audio environment?
- **Disconnection from Reality**: Did you feel disconnected from your immediate surroundings while listening?

All items were rated on a 7-point Likert scale (1 = not at all; 7 = completely). A final open-ended question asked participants to describe the clip in one word or a short phrase.

A 7-point Likert scale was used to capture subtle differences in participants' perceptions while remaining easy to interpret. This format provides greater sensitivity and reliability than shorter scales, making it suitable for measuring subjective experiences such as immersion (Lozano et al., 2008).

b. Ranking and debriefing interview:

At the end of the session, participants were asked to rank the five experimental audio clips (excluding the baseline) based on perceived immersion. This exercise served to capture overall impressions and support triangulation with quantitative responses.

A semi-structured interview was then conducted, focusing on:

- Rationale for clip rankings;
- Emotional impact and standout moments;
- Comparison to typical audio listening (e.g., on headphones or mobile);
- Perceived differences between stereo and spatial sound;
- Personal definitions and markers of "immersion";
- A metaphorical exercise: If the most immersive clip were an animal, which one would it be, and why? This question was used to explore through their creativity, what qualified as immersive, using the animal as a metaphor to be unpacked.

All interviews were audio-recorded and transcribed verbatim for qualitative analysis.

6. **Data Analysis:**

A mixed-methods analysis was conducted to evaluate participants' responses. Quantitative data from the Likert-scale questionnaires were analyzed by comparing the average ratings for each question between the stereo and spatial audio conditions, allowing for the identification of general trends in immersion and perceived sound quality. Descriptive statistics were calculated, and comparisons between conditions were used to assess the impact of audio spatialization.

The qualitative data from the semi-structured interview was subjected to a **content analysis**, done in two separate stages of the analytical process:

1. **Researcher-Guided Initial Inductive Coding:**

In the first stage, the principal researcher went over the entire interview transcripts, which included 38 participants' comments on all sound samples that were played in both standard stereo and spatial surround (including the practice clip, which some participants referred to in their comments). In order to produce an initial set of preliminary themes that emerged directly from the raw data, the initial coding process involved line-by-line analysis and sorting. This initial step preserved the interpretive insights essential to qualitative inquiry by ensuring that the initial conceptualization of themes was properly anchored in the participants' complex narratives. (Braun & Clarke, 2006)

2. **AI-Assisted Thematic Refinement and Pattern Exploration:**

After the researcher's initial coding, ChatGPT-4 - a generative artificial intelligence chatbot developed by OpenAI, designed to understand and produce human-like text in a conversational dialogue based on user prompts (OpenAI, 2022) ,- was used as a computational partner to analyze and provide the themes that were found in the verbatim tags. The main reasons this creative hybrid approach was selected were its capacity to speed up the analytical process and investigate a fresh, effective technique for analyzing qualitative data, which is consistent with new developments in human-AI cooperation in research. (Yan et al., 2023; Barros et al., 2024)

The full, unedited interview transcript (including verbatim responses in both French and English) was supplied as input, on ChatGPT-4 (through its official interface). Following ethical research procedures, all transcripts were carefully anonymized and de-identified before being put into the AI system to protect participant privacy and data security. (Lee, 2025; Schroeder et al., 2024).

The following direct prompt was used to start the AI's analysis:

> *"Do a content analysis for the attached file, this data analysis is part of a thesis project about sounds in immersive experiences. The verbatim attached is the feedback of 38 participants that took part in an immersive hearing experience of 5 pieces of sound. Participants were presented some pieces in spatial surround and some in standard stereo*

unbeknown to them. Assign the tags needed to capture how they define immersion. Analyse the whole document, the French and the English."

Since Chat GPT was not provided with explicit coding instructions or prior themes, it conducted an initial data-driven search for emergent patterns on its own. In addition to broader patterns regarding emotional engagement and sound perception throughout the entire dataset, including cross-linguistic analysis of French and English responses, the AI's job was to find and recommend a comprehensive set of "tags" (i.e., codes or themes) associated with participants' definitions of immersion. (see Appendices C & D)

The primary researcher began an iterative refinement process using the AI's generated "tags" and thematic suggestions.

Every recommendation produced by ChatGPT-4 was carefully examined by the researcher. Following a critical evaluation against the researcher's own initial findings, these AI-identified themes and tags were then combined with the initial findings of the researcher. This procedure made it possible to combine related ideas, find subtleties that had not been noticed before, and eliminate AI recommendations that were unrelated or lacked sufficient evidence. While AI models can recognize patterns, they can also generate superficial summaries or overlook more complex, culturally nuanced meanings, so human oversight is crucial (Agarwal et al., 2025).

By comparing the AI's pattern detection across the entire dataset with the researcher's initial inductive coding, the process facilitated a consistency check, particularly valuable for a single-researcher study where consistency can be challenging over large datasets (Zhang et al., 2025).

Although the AI's output is data-driven, the researcher's critical oversight was crucial in ensuring that the final thematic structure appropriately reflected the complex nuanced qualitative essence of the data and in reducing any **potential biases present in AI models** (Zhang et al., 2025). ChatGPT-4's effectiveness in systematic pattern recognition and its capacity to process large, multilingual datasets quickly are combined with the researcher's deeper contextual understanding and interpretive acuity in this **hybrid methodology** (Li et al., 2024).

This hybrid approach strengthens the study's methodological quality in three specific ways. First, it enhances rigor by creating a synergy between the AI's comprehensive, systematic pattern detection and the researcher's deep, interpretive analysis, thereby combining computational breadth with human depth (Malik, 2025). Second, it improves transparency by establishing a clear and auditable trail—documenting the prompts, the AI-generated codes, and the researcher's subsequent refinements (Lumivero, 2025). Third, it increases reliability (specifically, intra-coder consistency) by providing a stable, data-driven baseline that mitigates the risk of researcher drift or fatigue, a particular challenge in single-researcher studies (Delve, 2025).

This methodology thus represents a practical and ethically-grounded application of emerging computational tools in qualitative research. It aligns with the evolution of computer-assisted qualitative data analysis (CAQDAS) by leveraging AI to augment analytical efficiency and scale. This "human-in-the-loop" model strategically maintains the **primacy of human intellect** for ensuring contextual validity, uncovering nuanced meaning, and upholding ethical research standards—a consensus point in the emerging literature on this topic (Hayes, 2025).

Chapter 4: Results

The following results detail the findings of a mixed-methods investigation into the immersive capabilities of surround sound versus stereo audio. The analysis is presented in three parts. First, we present the statistical results from a Qualtrics questionnaire, which measured key aspects of immersion such as time perception, concentration, and realism. Second, we give a descriptive analysis of participants' explicit rankings of the audio presentations to show preferential patterns. Finally, a qualitative content analysis of participant interviews explores the subjective themes found, adding explanatory depth to the quantitative results.

Quantitative data from the Qualtrics questionnaire were analyzed to compare participants' reported levels of immersion and emotional engagement across the different sound conditions. Descriptive statistics (means and standard deviations) and inferential analyses (paired-sample *t*-tests) were conducted to examine differences between stereo and spatial audio formats. The analysis was carried out with the assistance of ChatGPT-4, which was used to perform statistical calculations and verify results. The researcher reviewed and validated all outputs to ensure accuracy and interpretive consistency.

Qualtrics questionnaire:

After each audio clip, participants completed the survey. Each clip was either in surround or stereo format, and responses were later matched by the researcher to the clip and format type.

Temporal Dissociation :

- Stereo Mean: 4.52
- Surround Mean: 5.31
- T-Value: -3.47
- P-Value: 0.0006
- Significance: Significant

Interpretation: Participants felt they lost track of time more with surround sound compared to stereo sound. The difference is statistically significant, indicating that surround sound creates a more immersive experience (See table 4).

Cognitive Load:

- Stereo Mean: 5.01
- Surround Mean: 5.77
- T-Value: -3.74
- P-Value: 0.0002
- Significance: Significant

Interpretation: Participants reported that following the audio in surround sound required *more* concentration than in stereo. This significant difference suggests that stereo sound is more engaging and easier to follow.

Spatial Realism:

- Stereo Mean: 4.82
- Surround Mean: 5.59
- T-Value: -3.51
- P-Value: 0.0005
- Significance: Significant

Interpretation: Surround sound was perceived as more realistic than stereo sound. The significant difference indicates that surround sound better simulates the experience of sound coming from the environment.

Presence:

- Stereo Mean: 4.80
- Surround Mean: 5.43
- T-Value: -2.89
- P-Value: 0.0042
- Significance: Significant

Interpretation: Participants felt a greater sense of physical presence in the sound space with surround sound compared to stereo sound. This significant difference highlights the enhanced immersive quality of surround sound.

Disconnection:

- Stereo Mean: 4.63
- Surround Mean: 5.17
- T-Value: -2.40
- P-Value: 0.0173
- Significance: Significant

Interpretation: Participants experienced a greater sense of disconnection from their immediate environment with surround sound. The significant difference suggests that surround sound more effectively transports listeners into the audio experience.

	Stereo Mean	Surround Mean	T- Value	P- Value	Significance
Loss of time	4.52	5.31	-3.47	0.0006	Significant
Concentration Needed	5.01	5.77	-3.74	0.0002	Significant
Realistic Sound	4.82	5.59	-3.51	0.0005	Significant
Physical Presence	4.80	5.43	-2.89	0.0042	Significant
Disconnection	4.63	5.17	-2.4	0.0173	Significant

Table 4: Summary of Qualtrics survey results

Summary

All questions show significant differences between stereo and surround sound clips, with surround sound generally providing a more vivid experience on all 5 parameters defining the immersive experience. The T-values indicate the magnitude of difference between the two groups, and the P-values confirm that these differences are statistically significant.

Descriptive analysis of the rankings

The analysis included the data from the 38 participants who completed immersion ranking tasks across five audio presentations. A total of 190 rankings were collected, with audio formats distributed as follows: 78 clips in surround sound presentations (41.1%) and 112 in stereo presentations (58.9%). The unequal distribution of audio formats resulted from logistical constraints during data collection. The original experimental design called for many more listening sessions with balanced audio format distribution. However, due to practical limitations, the study had to be reduced to 18 sessions (with 1 to 3 participants per session), which created the observed imbalance favoring stereo presentations. The audio format distribution varied across tracks, with Merrouh receiving the highest proportion of surround sound presentations (26 out of 38, 68.4%), while Get Lucky, Rescuer, Broken, and Lolupuppy each received approximately one-third surround sound presentations (ranging from 31.6% to 36.8%).

Participants' first post-listening task was to rank audio presentations from 1 (most immersive) to 5 (least immersive). In spite of the fact that surround audio tracks were underrepresented in the data set, the ranking distribution shows a pattern whereby surround sound appears to deliver greater immersion. For the most immersive experiences (Position 1), surround sound presentations were overwhelmingly preferred, accounting for 27 out of 38 rankings (71.1%) compared to just 11 stereo presentations (28.9%). For Position 2, surround sound and stereo presentations were equally ranked, each receiving 19 mentions (50.0%) (See Figure 1).

The pattern shifts progressively toward stereo dominance in lower immersion positions. Position 3 showed 15 surround sound rankings (39.5%) versus 23 stereo rankings (60.5%). Position 4 shows further stereo concentration with 10 surround sound rankings (26.3%) compared to 28 stereo rankings (73.7%). Most notably, Position 5 (least immersive) was dominated by stereo presentations, with only 3 surround sound rankings (7.9%) compared to 35 stereo rankings (92.1%).

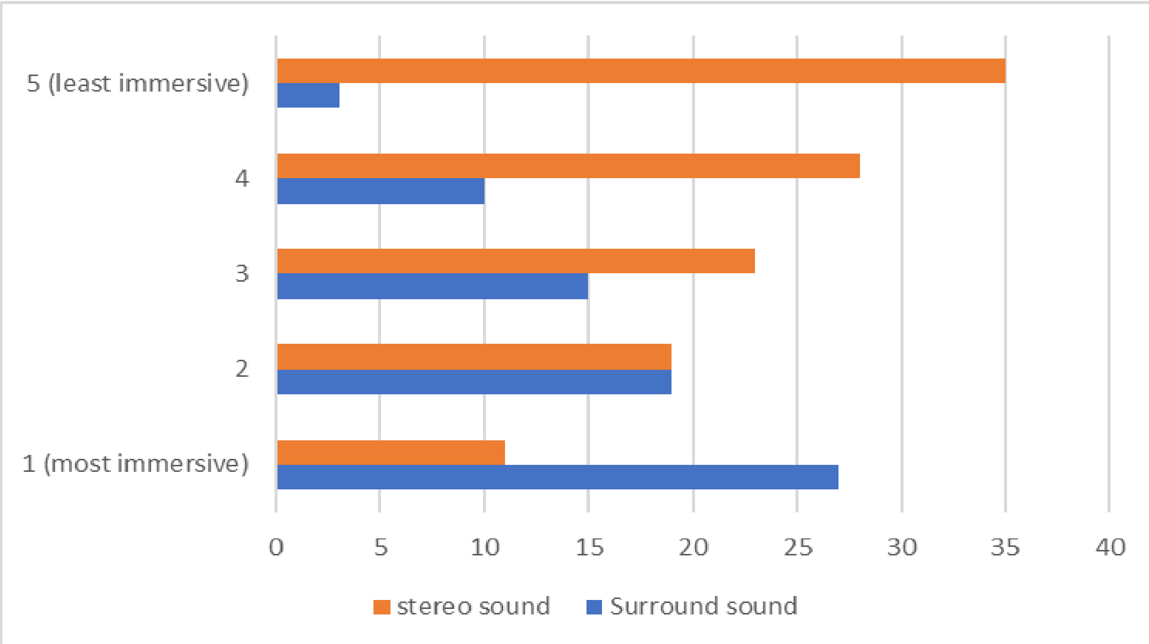

Figure 1. Diagram summary of the ranking results by clip format

the average immersion ranking suggest that participants perceived a stronger immersive experience with surround sound.Surround sound presentations achieved a notably higher mean ranking of 2.00 (SD = 1.23), while stereo presentations received a substantially lower mean ranking of 3.53 (SD = 1.18), representing a difference of 1.53 ranking positions in favor of surround sound. The standard deviations indicate that surround sound rankings showed slightly more variability around the mean, suggesting participants may have been more discriminating in their evaluations of surround sound quality.

All five audio tracks showed consistently better performance with surround sound, suggesting a stronger immersion across diverse content types (see Table 5):

Merrouh : Surround sound presentations (n = 26) achieved a mean ranking of 2.00, while stereo presentations (n = 12) averaged 3.42, representing a difference of 1.42 ranking positions in favor of the surround sound version.

Get Lucky: Surround sound presentations (n = 13) averaged 2.08, compared to stereo presentations (n = 25) which averaged 3.52, showing a difference of 1.44 ranking positions again in favor of the surround sound version.

Rescuer: Surround sound presentations (n = 14) achieved a mean of 1.93, while stereo presentations (n = 24) averaged 3.58, representing the largest difference of 1.65 ranking positions in favor of the surround sound version.

Broken: Surround sound presentations (n = 12) averaged 2.00, compared to stereo presentations (n = 26) which averaged 3.50, showing a difference of 1.50 ranking positions in favor of the surround sound version.

Lolupuppy Kutiquu: Surround sound presentations (n = 13) achieved a mean of 2.08, while stereo presentations (n = 25) averaged 3.56, representing a difference of 1.48 ranking positions in favor of the surround sound version.

Song	Condition	Mean	Std Dev (SD)	n	P-value	Significant
Merrouh	Surround	2	1.1	26	0.00011	Yes
	Stereo	3.42	0.67	12		
Get Lucky	Surround	2.08	1.04	13	0.00018	Yes
	Stereo	3.52	0.65	25		
Rescuer	Surround	1.93	1.07	14	0.000019	Yes
	Stereo	3.58	0.65	24		
Broken	Surround	2	1.04	12	0.00008	Yes
	Stereo	3.5	0.65	26		
Lolupuppy Kutiquu	Surround	2.08	1.04	13	0.00011	Yes
	Stereo	3.56	0.65	25		

Table 5: Average of the ranking results by song

Despite representing only 41.1% of total presentations, surround sound formats show a stronger immersive performance in higher ranking positions. Surround sound collected 71.1% of the most immersive rankings (Position 1), representing 1.73 times overrepresentation relative to its proportion in the dataset. More importantly, surround sound received only 7.9% of the least immersive rankings (Position 5), representing 5.2 times underrepresentation in the poorest performance category. This distribution inequality actually strengthens the findings regarding surround sound higher immersion rate , as surround sound's immersion advantage proved enough to overcome the numerical disadvantage in the dataset.

The descriptive analysis suggests that surround sound delivers higher immersion performance across multiple dimensions of the data. Surround sound presentations consistently achieved more favorable (lower numerical) immersion rankings across all content types, with mean advantages ranging from 1.42 to 1.65 ranking positions over stereo presentations. The distribution of rankings showed a systematic gradient favoring surround sound, with representation decreasing from 71.1% in the most immersive position to only 7.9% in the least immersive position. These patterns demonstrated remarkable consistency across diverse content categories, including culturally specific music, familiar commercial content, unfamiliar musical material, and nature-based audio, indicating that surround sound's immersion advantage is robust and generalizable across different audio contexts.

Content analysis

After the analysis the tags were grouped under four central themes: spatial-temporal perception, narrative imagination, sound quality, and emotional-physical reactions, which provide the framework for the findings presented below.

1. Sense of Space and Time: From External to Enveloping

The most commonly reported difference among participants was a shift in spatial perception. Under stereo conditions, 18 participants described sound as flat, localized, and externally anchored. As P12 explained, "*Felt like the sound wasn't fully around me*," while P34 added, " *Pour moi la voix ne bougeait pas… comme s'il y avait un haut-parleur à ma gauche (For me, the voice wasn't moving… it was like there was a loudspeaker on my left)*." Such spatial stasis reinforced listeners' awareness of the physical environment and maintained a psychological distance from the auditory content, consistent with research on external spatial referencing. (Sheffield, Romigh, Zurek, Bernstein, & Brungart, 2019)

In contrast, spatial audio induced a sense of 360-degree envelopment. P11 described being "*wrapped in this sound*," while P33 confirmed "*on est très enveloppé (We're really enveloped.)*." Mobility in sound was reported frequently: "*Ça tournait autour (it was turning around)*" (P35), "*les voix qui vraiment voyageaient d'un côté à l'autre(the voices that were really travelling from one side to the other)"* (P30), and "*like pinpointing where the sound is*" (P06). These immersive auditory cues generated a heightened sense of spatial presence, aligning with Lombard and Ditton (1997) theory of presence.

Notably, several participants experienced temporal dissociation during spatial audio. P06 stated, "*I felt like I was losing track of time*," and P26 noted*, "I didn't even feel the minutes passing."* These moments resemble the "*flow state"* described by Csikszentmihalyi (1990), characterized by altered time perception, effortless attention, and immersive focus (Shernoff, D. J., Csikszentmihalyi, M., Shneider, B., & Shernoff, E. S. 2003)

2. Storytelling and Imagination: From Passive Listening to Active Narrative Construction

Spatial audio also enhanced imaginative engagement, pushing listeners beyond descriptive reactions into storytelling and metaphorical thinking. While stereo yielded more technical commentary, spatial formats enabled creative mental imagery. For instance, P31 noted, "*Ça me donnait vraiment l'espace d'imaginer, comme si j'étais dans un rêve*,(it was really giving me the space to imagine, as if I was in a dream)" and P01 reflected, "*Every instrument, like, tells a story kinda.*" These responses reflect transportation into narrative worlds (Green & Brock, 2000).

Participants frequently reported embodied cognitive effects. P15 described "*une sensation de liberté… comme un oiseau (a feeling of freedom, like a bird),*" associating sonic immersion with flight. Others experienced submersion: "*Like I am drowning into the environment*" (P04), or "*comme quand on est immergé dans l'eau… on est immergé dans la musique (similar as when we're immersed in the water, we're immersed in the music*" (P25). These reactions correspond to Dourish's (2001) embodiment theory, where sensory cues engage bodily schemas even in stillness.

3. Sound Quality and Complexity: From Monotone to Multi-Layered

Participants also made clear distinctions in sound quality, with spatial audio perceived as richer, more textured, and more authentic. Stereo was often described as flat or monotonous: "*Bored because it was every monotone*" (P17) and "*There's no vibration, no diversity"* (P14).

Spatial audio, by contrast, elicited responses that emphasized perceptual richness and auditory scene separation. P02 reported "*so many layers of the song*," while P10 noticed *"many sounds, like an ecosystem of sounds*." These reactions align with Bregman's (1990) Auditory Scene Analysis, which explains how listeners organize overlapping sounds into meaningful streams.

Participants emphasized their ability to disentangle and localize sounds: "*You could really separate everything"* (P19), and "*I could even hear them breathe*" (P29). This enhanced auditory granularity reflects McDermott's (2012) research on sound source discrimination in complex acoustic scenes.

An interesting phenomenon reported was subjective loudness amplification: "*70 decibels sounded like 90 at least"* (P08). This aligns with psychoacoustic work showing that spatial spread and early reflections can increase perceived intensity and apparent source width (Lokki, Pätynen, Kuusinen, & Tervo, 2016)

4. Emotional and Physical Reactions: From Neutral to Deep Engagement

Spatial audio produced consistently more intense emotional and physiological responses than stereo. Four participants explicitly mentioned experiencing goosebumps evoked by emotionally powerful stimuli. "*What actually gave me goosebumps*" (P38); "*Ça fait des frissons quand tu sens que le son se déplace autour de toi (it give goosebumps when you feel the sound moving around you ")* (P16). These reactions mirror neuroscientific findings linking frisson to dopaminergic activity in reward centers (Ferreri et al., 2019).

The emotional spectrum was diverse. P20 felt "*relaxée en entendant ça*,(relaxed while listening to this)" while P28 described being *"touched, kinda my soul*." Others experienced emotional discomfort: "*On me criait dessus, ça m'a un peu oppress*é*(I was being shouted at, it felt a bit oppressive)*" (P10), and "*It was kinda terrifying for me*" (P18). These extreme responses suggest high affective engagement whether pleasant or unsettling.

Importantly, many participants described the experience as intimate or personal. P22 said it was "*like someone was humming next to my ear*," and P24 added, "*C'était comme si j'avais des femmes autour de moi qui chantaient pour moi*.(It was as if I had women around me singing for me.)" These effects relate to Lombard & Ditton's (1997) theory of proximal auditory presence, where spatial sound simulates social nearness.

Several participants (e.g., P27) even expressed purchase intent: "*I would pay money to have that in my home…,*" demonstrating experiential value in line with Hassenzahl's (2010) model of meaningful UX. The desire for ownership indicates the experience transcended mere **pragmatic utility**, providing strong **hedonic qualities** that fulfilled psychological needs for stimulation and enjoyment, which Hassenzahl identifies as the core of a positive user experience

The within-subjects design enabled direct comparison of individual responses across audio conditions, revealing consistent patterns in how spatial audio enhanced participant experiences compared to stereo conditions. Across all four thematic areas, spatial audio appears to deliver better capacity for creating immersive, emotionally engaging experiences.

Individual differences in response style and preference were maintained across conditions, suggesting that while spatial audio enhances the overall experience quality, it does so while preserving personal variation in how individuals process and interpret auditory information.

Chapter 5: Discussion

This study investigated how spatial audio, compared to traditional stereo, influences immersion, presence, and emotional engagement. Findings from the mixed-methods analysis provide evidence that the dimensionality of sound is a noteworthy determinant of user experience.

This discussion interprets these results by examining the experiential differences between formats, situating them within the broader literature on presence and psychoacoustics.

Comparison with Literature

These findings strongly align with and provide empirical support for key theories in media psychology and psychoacoustics. The consistent preference for spatial audio in creating a sense of ‘being there’ directly supports the concept of presence, defined by Lombard and Ditton (1997) as the ‘perceptual illusion of non-mediation. Furthermore, recent research indicates that object-based audio rendering methods can significantly enhance perceived audio quality and the feeling of immersion in virtual reality environments, thereby reducing the listener's awareness of the technology itself. (Potter et al. 2022).

Furthermore, the way participants described creating mental imagery and narratives ("opened a space for storytelling") under spatial audio conditions is consistent with **Transportation Theory**, which posits that engagement with a narrative world can absorb an individual's mental resources (Green & Brock, 2000). Empirical studies demonstrate that 3D audio can significantly enhance listeners' experience and engagement with narrative content, such as audio dramas, when compared to traditional stereo (Potter et al. 2022).

The study's results do not diverge from the established literature; rather, they address persistent questions in the field. While the technical benefits of spatial audio are well-documented, recent comprehensive reviews have called for more controlled studies on its specific emotional and psychological impacts, moving beyond anecdotal reports to validate its effect on user experience and overall immersion (Stefanowska et al., 2024).

The Power of Presence: Spatializing Diegetic Sound

One of the most pertinent findings from this study is the capacity of spatial audio to heighten listeners' sense of presence and realism. Quantitative results highlights this effect: participants rated spatial audio clips as significantly more realistic (mean score 5.59 vs. 4.82 for stereo) and reported a stronger feeling of physical presence within the auditory environment (5.43 vs. 4.80). These numerical patterns were corroborated by the qualitative interviews, where participants described a shift in how they perceived the origin and behavior of sounds in space.

This finding underscores the importance of **diegetic sound**—those sounds belonging to the world of the narrative, such as dialogue, environmental noises, or footsteps (Chion, 1994). Spatial audio amplifies the immersive potential of diegetic sound by situating it within a three-dimensional framework. In doing so, it acts as a catalyst for **narrative transportation**, the process through which individuals become absorbed into a story world (Green & Brock, 2000). By anchoring sound sources to specific locations and allowing them to move dynamically—as in voices that "vraiment voyageaient d'un côté à l'autre" (really traveled from one side to the other, P30)—participants no longer experienced the audio as a detached recording. Instead, sound was perceived as an embodied element of a lived environment. Prior research confirms that such spatialization significantly elevates listener experience and engagement compared to stereo formats, validating the effect observed here (Verhulst et al., 2024).

Participant descriptions illustrate this transformation. Under stereo conditions, sound was often experienced as tethered to an external point: "Pour moi la voix ne bougeait pas... comme s'il y avait un haut-parleur à ma gauche" (For me the voice didn't move… as if there was a speaker to my left, P34). This aligns with the limitations of stereo, which often fails to create a fully externalized and enveloping soundfield, instead presenting audio through a metaphorical "window" (Rumsey, 2002). By contrast, spatial audio was consistently described as enveloping and embodied: "I felt like I was losing track of time" (P06), and "like someone was humming next to my ear" (P22). Such accounts illustrate how the dimensionality of sound dissolved the perceptual boundary between listener and content, shifting the experience from observation to participation.

This transition from a "window" to a "world" encapsulates the immersive power of spatial audio. By rendering diegetic sounds with precision and congruence, spatial audio creates a **plausible** environment in which sound is not merely heard but inhabited (Turner, Murphy, Pike, & Baume, 2022). Cognitive theories of **predictive processing** suggest that this realism emerges because the brain's expectations about how sound should behave in physical space are consistently confirmed, reducing prediction error and reinforcing the believability of the environment (Denham, 2020). At the same time, the proximity effects participants described—such as voices perceived "next to the ear"—suggest a deep engagement of **embodied simulation** mechanisms, wherein auditory cues trigger visceral, sensorimotor responses associated with physical presence and intimacy (Langiulli, Calbi, Sbravatti, Umiltà, & Gallese, 2023).

In this way, spatial audio does not simply add realism; it restructures the listener's perceptual stance toward the narrative. Rather than being positioned outside the story world, as one might be when listening through stereo "windows," participants felt themselves enveloped within it. This heightened presence is important to the findings of this study and provides evidence that the dimensionality of sound is a critical driver of immersion.

The Conceptual Limit of a Static World: Head-Lock and Dynamic Presence

Head-Lock Phenomenon in Spatial Audio

A significant challenge in spatial audio systems, particularly with headphones, is "head-lock." This occurs when the entire soundscape rotates with the listener's head because dynamic head-tracking is absent. Instead of remaining fixed in the environment, a sound intended to be in front of the listener will follow their gaze as they turn. This issue compromises realism and hinders externalization—the perception of sounds originating outside the head—which is vital for a truly immersive experience (Brimijoin, & al. 2013).

Speaker-Based Systems and Their Limitations

The present study circumvented head-lock by utilizing a multi-speaker array (Centre PHI's *Habitat Sonore*), as these systems project sound into the physical environment. However, speaker-based systems introduce their own set of constraints. They are highly sensitive to room acoustics and demand meticulous calibration to ensure spatial fidelity. The listener's position relative to the speakers also heavily influences localization accuracy (Pulkki et al., 2001). Although participants in this study did not perceive head-lock as a limitation—due to the intentionally stationary listening design—the lack of dynamic responsiveness highlights an important conceptual point. Slater (2009) argues that the plausibility of an immersive environment is maximized when the virtual world adheres to the same physical laws as the user's world.

The Variable End-User Experience in Immersive Audio

This observation underscores a broader issue in contemporary immersive audio: the end-user experience is shaped by the interplay of hardware, software, and content design. Brimijoin et al. (2013) demonstrated that when sounds remain fixed in space relative to the environment—rather than moving with the listener's head—externalization is significantly enhanced. This highlights why loudspeaker arrays, which provide natural spatial cues, often yield stronger externalization than headphones. By contrast, headphone-based systems dominate consumer markets but are inherently susceptible to the "head-lock" effect unless augmented with real-time head-tracking and personalization strategies.

The Intrusiveness of Vision and the Economics of Immersion

Another insight emerging from this study relates to the overlooked sensory hierarchy and its implications for creating immersive experiences. The literature on multisensory integration has historically emphasized the **dominance of vision**, a faculty that offers the highest bandwidth and spatial precision for navigating the physical world (Gallace et al., 2012). The classic **"ventriloquism effect,"** where the brain resolves sensory conflict by weighting cues according to their reliability, serves as a powerful example of this visual capture; an ambiguous sound is automatically perceived at the location of a precise visual event (Alais & Burr, 2004). However, this dominance is not absolute. As research demonstrates, this hierarchy can be contested when auditory information is rendered with sufficient clarity and realism. **High-fidelity spatial audio**, by providing congruent spatial cues, can challenge visual input for perceptual authority, particularly in virtual environments (Jiménez-Navarro, Serrano, & Malpica, 2025). Conversely, in experiences saturated with overly complex graphics, vision can become a form of sensory intrusion, imposing an **extraneous cognitive load** that splits attention and can lead to user fatigue (Mayer & Moreno, 2003).

The design of this study's experimental setting speaks directly to this dynamic. By conducting the experiment in the Habitat Sonore—a specialized room facilitating intentional sensory attenuation—we minimized visual input to isolate the auditory channel. This setup effectively neutralized the intrusive potential of vision, allowing for an unfiltered assessment of audio's immersive capabilities. The results were clear: a crafted spatial audio soundscape was sufficient on its own to induce a strong sense of **physical presence** (the sensation of "being there"), a profound disconnection from the immediate physical world, and significant **temporal dissociation** (a distorted sense of time characteristic of deep engagement, or "flow") (Skarbez, Brooks, & Whitton, 2021; Csikszentmihalyi, 1990). This suggests that sound, processed rapidly by the brain's emotional and memory centers like the amygdala and hippocampus (Koelsch, 2014), can generate a deeply embodied sense of immersion without requiring a corresponding visual world.

What is particularly striking is how sound, often framed as a "secondary" modality, revealed itself as not only sufficient but, in some respects, better for sustaining immersion. According to **cognitive load theory**, when visual systems are oversaturated, adding more visual complexity does not proportionally enhance immersion; rather, it can diminish it by overwhelming the user's finite processing capacity (Sweller, van Merriënboer, & Paas, 2020). By contrast, auditory immersion effectively leverages the brain’s **predictive coding mechanisms** (Friston, 2005). Instead of rendering every detail, a coherent soundscape provides a continuous stream of environmental cues, allowing the brain to generate a robust predictive model of the space. This guides orientation and affective engagement without demanding the same conscious allocation of cognitive resources, explaining why participants described "being wrapped in sound" as the auditory environment established its own perceptual authority.

These findings carry significant implications for the production pipeline and **economics of immersive media**. The creation of high-fidelity visual content-requiring resource-intensive processes like complex 3D modeling and real-time ray tracing- is typically the most costly aspect of development. Our research provides empirical support for the **compensatory function of audio**, where high-quality sound leverages the brain's tendency toward **perceptual completion** to "fill in" for missing or degraded visual detail (Shams & Beierholm, 2010). In fact, studies confirm that congruent, high-quality audio can substantially elevate the perceived realism of even abstract or lower-fidelity visuals (Rojas et al., 2013). An enveloping, three-dimensional soundscape can thus carry the immersive weight of an experience by implying a rich world rather than explicitly rendering it.

For content creators and UX designers, this points toward a shift from a visual-centric model to an **"audio-forward" design philosophy**. Investing in high-quality spatial audio is not merely a cost-saving measure but a strategically potent method to enhance immersion in a neurologically efficient manner. As a major meta-analysis confirms, **audio fidelity is a key technological factor** in increasing a user's sense of presence (Potter, T.; Cvetković, Z.; De Sena, E., 2022). Moreover, this approach opens new possibilities for accessibility, suggesting that immersive experiences could be effectively tailored for visually impaired audiences by leveraging their intact spatial hearing, without sacrificing depth or engagement (Rosin, J., 2023). This reframing of sensory priorities could redefine the economics and aesthetics of immersion, shifting focus from an arms race in graphical fidelity toward a more balanced, multi-sensory approach where sound plays a central, and perhaps even leading, role.

Chapter 6: Limits, Implications and Conclusion

This study is an exploratory study that seeks to explore how spatial audio, in contrast to traditional stereo sound, influences the user's perception of immersion, presence, and emotional engagement within auditory experiences. The findings highlight the higher immersive qualities inherent in spatial audio. This discussion will systematically address the limitations of the current research, delve into the implications of the key findings, and connect these insights to the broader significance of these discoveries for the evolving fields of User Experience (UX) research and design, as well as their potential impact on industry practices and interactions.

Limitations and Future Research Directions

While the present study discusses the benefits of spatial audio, it is crucial to acknowledge several inherent limitations that constrain the generalizability of its conclusions and simultaneously illuminate fertile ground for future research.

Firstly, the participant pool, comprising 38 individuals, while adequate for an exploratory study and to detect statistically significant differences in a within-subjects experimental design, inherently limits the external validity of the findings. Future research endeavors would significantly benefit from recruiting larger and more heterogeneous participant cohorts, encompassing a wider range of demographics, listening habits, and cultural backgrounds. Such an expansion would bolster the statistical power of the study and enhance the confidence with which its conclusions can be generalized across different user segments (Faul et al., 2007).

Secondly, the experimental setting (Centre PHI's Habitat Sonore) represents a highly specialized and optimized listening environment. This room, equipped with 16 strategically placed speakers designed for three-dimensional acoustic diffusion and augmented by a low-light environment to minimize visual distractions, was intentionally chosen to showcase the maximal potential of spatial audio. However, this controlled and high-fidelity setup diverges significantly from the typical listening environments encountered by most consumers, such as museums and public spaces where the audio experience could strongly impact the nature of the experience. Consequently, the observed results may not directly translate to these more common, less controlled scenarios.

Thirdly, the relatively brief duration of the listening sessions, with each audio clip lasting approximately 4.5 to 5 minutes, precludes an examination of the long-term effects of spatial audio on user experience. The initial "wow" factor or novelty effect often associated with new technologies might inflate short-term engagement and immersion scores (Norman, 2013). It remains an open question whether the heightened sense of presence and emotional engagement observed in this study persists with prolonged or repeated exposure. Future longitudinal studies, of sustained spatial audio use, are necessary to ascertain whether the benefits endure, if users experience hedonic adaptation (Frederick & Loewenstein, 1999), or if prolonged exposure leads to sensory fatigue or desensitization. Such research would be critical for understanding the sustainability of spatial audio's impact in real-world applications.

Fourthly, the study's reliance on a limited selection of five distinct audio clips, while diverse in genre (familiar pop, unfamiliar contemporary, Arabic cultural, nature soundscapes), represents only a fraction of the vast auditory landscape. The effectiveness of spatial audio may vary across different content types, such as spoken word (podcasts, audiobooks), interactive soundscapes in virtual reality, or highly complex musical compositions. Future research should investigate the influence of spatial audio across a broader spectrum of auditory content, including those designed for specific cognitive tasks or therapy through sounds, to build a wider understanding of its applications.

Finally, the current study primarily relied on self-reported measures of immersion, presence, and emotional engagement. While these subjective accounts provide rich qualitative data, integrating objective physiological measures would offer a more comprehensive and triangulated understanding of user responses. Incorporating metrics that can provide empirical validation for the reported emotional and cognitive states (Solovey, McDuff, & Czerwinski, 2014). Lastly, a critical area for future inquiry involves cross-cultural studies to understand how different cultural backgrounds, listening habits, and linguistic contexts shape individuals' perceptions and emotional responses to spatial audio (Jacoby et al. 2020). As highlighted by the distinct reactions to the Arabic song "Merrouh" versus the familiar "Get Lucky," cultural context profoundly influences auditory interpretation, suggesting that future designs must be more personal and adaptable to diverse user needs and preferences.

What Does This Study Brings to UX Research and Design

The findings of this study carry implications for the evolving landscape of User Experience (UX) research and design, offering both new avenues for investigation and practical guidelines for application.

First, the research illustrates **how audio fits into multi-sensory UX**, moving beyond a predominantly visual-centric design paradigm. For too long, audio has been relegated to a secondary role in UX design, often treated as mere background or notification. This study unequivocally shows that sound design, particularly spatial audio, can profoundly affect feelings, emotional engagement, and even cognitive states like attention and memory. This necessitates a paradigm shift in UX research, urging a more integrated approach that considers auditory elements as fundamental components of the overall user experience, on par with visual and haptic feedback **(**Gibbs, J. K., Gillies, M., & Pan, X. 2022)

Also, the findings **add to emotional design theory** by illustrating the powerful role of sound in shaping affective experiences. Emotional design, which focuses on creating products that evoke positive emotions and meaningful experiences (Norman, 2013), can significantly leverage spatial audio. The study's demonstration of spatial audio's capacity to elicit a wide range of intense emotions, from relaxation to spiritual connection, and even discomfort that still signifies engagement, highlights its potential as a potent tool for designers aiming to craft emotionally resonant interactions.

What This Means for UX Design: A Practical Roadmap

The study's findings translate into tangible and actionable design guidelines for incorporating spatial audio into various applications, offering a simple roadmap for designers: The growing understanding and application of spatial audio extend beyond individual user experiences, promising significant shifts in industry practices and broader societal implications.

Industry Shifts: The insights gleaned from studies like this are instrumental in **pushing audio technology forward**. As the demand for immersive experiences grows, it will drive innovation in audio recording, mixing, rendering, and playback technologies. This research could inspire **new UX evaluation standards** that move beyond traditional metrics, incorporating the multi-dimensional nature of immersive experiences and the unique contributions of spatial audio. Furthermore, it promotes **inclusive design**, particularly for accessibility needs. Spatial audio offers a powerful tool for individuals with visual impairments, enabling them to navigate complex digital and physical environments with greater independence and confidence (Massiceti D, Hicks SL, van Rheede JJ 2018). This focus on accessibility can foster a more equitable digital landscape.

Social Considerations: Spatial audio holds significant potential for improving digital wellness. By creating calming, soothing, or stimulating auditory spaces, it can serve as a tool for stress reduction, focus enhancement, and overall mental well-being in an increasingly noisy and demanding digital world. This could lead to new applications in mindfulness, meditation, and therapeutic interventions. Moreover, spatial audio may profoundly affect how we socialize digitally, making online spaces feel less distant and fostering a greater sense of connection and presence in virtual interactions. Imagine virtual meeting rooms where participants' voices originate from their virtual positions, creating a more natural and engaging conversational environment. This could mitigate feelings of isolation often associated with purely visual online interactions. Finally, there is big potential for education, where spatial audio could lead to better engagement and stronger learning experiences if well integrated. For example, historical narratives could be brought to life with spatialized sounds of an era, or scientific concepts could be illustrated through auditory simulations that enhance understanding and memory retention.

The Tech Side: Realistic Challenges

Despite its immense potential, the widespread adoption and optimal implementation of spatial audio face several realistic technological challenges that designers and developers must address.

Firstly, most people won’t have fancy multi-speaker setups. The immersive experience of Habitat Sonore, with its 16 precisely placed speakers, is a luxury. For mass consumption, headphones are key, relying heavily on binaural audio technologies that simulate 3D sound over two channels. While binaural audio has made significant strides, achieving truly convincing and consistent spatialization across all users and content remains a technical challenge, as individual head-related transfer functions (HRTFs) vary (Wang, L., Zeng, X., & Ma, X. 2019). Designers must optimize for headphone experiences, understanding their limitations compared to a dedicated multi-speaker room.

Secondly, spatial sound needs enough device power. Rendering complex 3D audio environments in real-time is computationally intensive. This can be heavy to process, particularly for mobile devices with limited resources. This computational demand raises concerns about battery life, device performance, and even environmental impact ("not very green friendly") due to increased energy consumption. Optimization techniques and efficient rendering algorithms are crucial for making spatial audio viable across a wide range of devices.

Finally, there is a delicate balance to strike: too much spatial audio can overwhelm users. While immersion is desirable, excessive or poorly implemented spatialization can lead to sensory overload, cognitive fatigue, or even discomfort. The brain constantly processes auditory information, and an overly dense or chaotic spatial soundscape can become distracting rather than enhancing. Designers must exercise restraint and precision, ensuring that spatial audio cues are meaningful, well-timed, and contribute positively to the user's experience without becoming a source of irritation or confusion. This necessitates careful calibration and extensive user testing to find the optimal balance for different applications and user preferences.

Conclusion

In conclusion, this study provides empirical evidence for the higher immersive qualities of spatial audio, highlighting its impact on perceived presence, emotional engagement, and temporal awareness. Despite its inherent limitations, the findings offer valuable insights for UX researchers and designers, emphasizing the importance of multi-sensory design, innovative measurement techniques, and the intentional, user-centered application of spatial audio. As technology continues its relentless march forward, spatial audio stands poised to reshape how we interact with digital content, fostering deeper connections, richer engagements, and ultimately, more meaningful user experiences across a multitude of applications. The journey towards fully realizing its potential requires continued interdisciplinary research, thoughtful design, and a commitment to addressing the practical challenges of its implementation.

References:

Agarwal, D., Naaman, M., & Vashistha, A. (2025). AI suggestions homogenize writing toward Western styles and diminish cultural nuances. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–21). Yokohama, Japan: ACM. https://doi.org/10.1145/3706598.3713564

Agrawal, S., Bech, S., de Moor, K., & Forchhammer, S. (2022). Influence of changes in audio spatialization on immersion in audiovisual experiences. *AES: Journal of the Audio Engineering Society*, *70*(10), 810–823. https://doi.org/10.17743/jaes.2022.0034

Agrawal, S., Simon, A., Bech, S., Bærentsen, K., & Forchhammer, S. (2019). Defining immersion: Literature review and implications for research on immersive audiovisual experiences.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization*. MIT Press. https://doi.org/10.7551/mitpress/6391.001.0001

Bleidt, R. (2014). Object-based audio opportunities for improved listening experiences. In *Proceedings of the SMPTE Conference*. https://www.iis.fraunhofer.de/content/dam/iis/en/doc/ame/Conference-Paper/BleidtR_SMPTE2014_Object-Based_Audio.pdf

Bob G. Witmer, & Michael J. Singer. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240. https://doi.org/10.1162/105474698565686

Brimijoin, W., Archer-Boyd, A., & Akeroyd, M. (2013). The contribution of head movement to the externalization and internalization of sounds. *PLoS ONE, 8*, e83068. https://doi.org/10.1371/journal.pone.0083068

Brown, E., & Cairns, P. (2004). A grounded investigation of game immersion. In *ACM SIGCHI Conference on Human Factors in Computing Systems* (pp. 1297–1300). https://doi.org/10.1145/985921.986048

Chion, M. (1994). *Audio-vision: Sound on screen*. Columbia University Press.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 25*(5), 975–979.

Conetta, R., Brookes, T., Rumsey, F., Zieliński, S., Dewhirst, M., Jackson, P., Bech, S., Meares, D., & Alumkal, S. G. (2014). Spatial audio quality perception (Part 1): Impact of commonly encountered processes. *Journal of the Audio Engineering Society, 62,* 831–846. https://doi.org/10.17743/jaes.2014.0048

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.

Delve. (2025). 6 proven ways to use AI in qualitative research. https://delvetool.com/blog/how-use-ai-qualitative-research

Denham, S. L., & Winkler, I. (2020). Predictive coding in auditory perception: Challenges and unresolved questions. *European Journal of Neuroscience*, *51*(5), 1151–1160. https://doi.org/10.1111/ejn.13802

Derksen, M. (2019, May 27). Spatial audio: The continuing evolution. Abbey Road Institute Amsterdam. https://abbeyroadinstitute.nl/blog/spatial-audio-continuing-evolution/

Dourish, P. (2001). *Where the action is: The foundations of embodied interaction*. MIT Press.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Ferreri, L., Mas-Herrero, E., Zatorre, R. J., Ripollés, P., Gomez-Andres, A., Alicart, H., Olivé, G., Marco-Pallarés, J., Antonijoan, R. M., Valle, M., Riba, J., & Rodriguez-Fornells, A. (2019). Dopamine modulates the reward experiences elicited by music. *Proceedings of the National Academy of Sciences, 116*(9), 3793–3798. https://doi.org/10.1073/pnas.1811878116

Franco Rojas, M. L., & Rivera Pinzón, D. M. (2022). Application in virtual reality for the learning of concepts associated with spatial sound. *ICERI2022 Proceedings,* 2176–2184. https://doi.org/10.21125/iceri.2022.0548

Francombe, J., Brookes, T., & Mason, R. (2017). Evaluation of spatial audio reproduction methods (Part 1): Elicitation of perceptual differences. *Journal of the Audio Engineering Society, 65,* 198–211. https://doi.org/10.17743/jaes.2016.0070

Francombe, J., Zieliński, S., & Brookes, T. (2023). Towards predicting immersion in surround sound music: A listener-centric case study on object-based and ambisonic audio processing. *Acta Acustica united with Acustica, 109*(1), 1–9. https://doi.org/10.1051/aacus/2023040

Gao, Z., Wang, H., Feng, G., & Lv, H. (2022). Exploring sonification mapping strategies for spatial auditory guidance in immersive virtual environments. *ACM Transactions on Applied Perception, 19*(3), Article 9. https://doi.org/10.1145/3528171

Gallace, A., Ngo, M., Sulaitis, J., & Spence, C. (2012). Multisensory presence in virtual reality: Possibilities and limitations. In *Multiple sensorial media advances and applications: New developments in MulSeMedia* (pp. 1–40). IGI Global.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology, 79*(5), 701–721. https://doi.org/10.1037/0022-3514.79.5.701

Greenberg, D. M., Bodner, E., Shrira, A., & Fricke, K. R. (2021). Decreasing stress through a spatial audio and immersive 3D environment: A pilot study with implications for clinical and medical settings. *Music & Science, 4*, 2059204321993992. https://doi.org/10.1177/2059204321993992

Hameed, A., Möller, S., & Perkis, A. (2024). A holistic quality taxonomy for virtual reality experiences. *Frontiers in Virtual Reality, 5*, Article 1434016. https://doi.org/10.3389/frvir.2024.1434016

Hayes, A. S. (2025). "Conversing" with qualitative data: Enhancing qualitative research through large language models (LLMs). *International Journal of Qualitative Methods, 24*. https://doi.org/10.1177/16094069251322346

He Zhang, Wu, C., Xie, J., Lyu, Y., Cai, J., & Carroll, J. M. (2025). Harnessing the power of AI in qualitative research: Exploring, using and redesigning ChatGPT. *Computers in Human Behavior: Artificial Humans, 4*, 100144. https://doi.org/10.1016/j.chbah.2025.100144

Ivanov, A. Z., King, A. J., Willmore, B. D. B., Walker, K. M. M., & Harper, N. S. (2022). Cortical adaptation to sound reverberation. *eLife, 11*, e75090. https://doi.org/10.7554/eLife.75090

Jacoby, N., Margulis, E. H., Clayton, M., Hannon, E., Honing, H., Iversen, J., Klein, T. R., Mehr, S. A., Pearson, L., Peretz, I., Perlman, M., Polak, R., Ravignani, A., Savage, P. E., Steingo, G., Stevens, C. J., Trainor, L., Trehub, S., Veal, M., & Wald-Fuhrmann, M. (2020). Cross-cultural work in music cognition: Challenges, insights, and recommendations. *Music Perception, 37*(3), 185–195. https://doi.org/10.1525/mp.2020.37.3.185

Kaplan Neeman, R., Roziner, I., & Muchnik, C. (2022). A clinical paradigm for listening effort assessment in middle-aged listeners. *Frontiers in Psychology, 13*, 820227. https://doi.org/10.3389/fpsyg.2022.820227

Kapralos, B., Jenkin, M. R., & Milios, E. (2008). Virtual audio systems. *Presence: Teleoperators & Virtual Environments, 17*(6), 527–549. https://doi.org/10.1162/pres.17.6.527

Koyama, S., De Sena, E., Samarasinghe, P., Thomas, M. R. P., & Antonacci, F. (2025). Past, present, and future of spatial audio and room acoustics. *arXiv*. https://arxiv.org/abs/2503.12948

Kosmenda, K., & Mickiewicz, W. (2024). Immersive two-channel recordings based on personalized BRIRs and their applications in industry. *Applied Sciences*, *14*, 11724. https://doi.org/10.3390/app142411724

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, *15*(3), 170–180. https://doi.org/10.1038/nrn3666

Kunasegaran, K., Ismail, A. M. H., Ramasamy, S., Gnanou, J. V., Caszo, B. A., & Chen, P. L. (2023). Understanding mental fatigue and its detection: A comparative analysis of assessments and tools. *PeerJ*, *11*, e15744. https://doi.org/10.7717/peerj.15744

Langiulli, N., Calbi, M., Sbravatti, V., Umiltà, M. A., & Gallese, V. (2023). The effect of surround sound on embodiment and sense of presence in cinematic experience: A behavioral and HD-EEG study. *Frontiers in Neuroscience*, *17*, 1222472. https://doi.org/10.3389/fnins.2023.1222472

Lee, G. W., & Kim, H. K. (2018). Personalized HRTF modeling based on deep neural network using anthropometric measurements and images of the ear. *Applied Sciences*, *8*(11), 2180. https://doi.org/10.3390/app8112180

Lee, K. (2004). Presence, explicated. *Communication Theory*, *14*, 27–50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Latoschik, M. E., & Wienrich, C. (2022). Presence and immersion in virtual environments: Theoretical foundations and recent advances. *Frontiers in Virtual Reality, 3*, 879567. https://doi.org/10.3389/frvir.2022.694433

Li, K. D., Fernandez, A. M., Schwartz, R., Rios, N., Carlisle, M. N., Amend, G. M., Patel, H. V., & Breyer, B. N. (2024). Comparing GPT-4 and human researchers in health care data analysis: Qualitative description study. *Journal of Medical Internet Research, 26,* e56500. https://doi.org/10.2196/56500

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication, 3*(2). https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Lokki, T., Pätynen, J., Kuusinen, A., & Tervo, S. (2016). Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences. *Journal of the Acoustical Society of America, 140*(1), 551. https://doi.org/10.1121/1.4958686

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). *Effect of the number of response categories on the reliability and validity of rating scales. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 4*(2), 73–79. https://doi.org/10.1027/1614-2241.4.2.73

Malik, F. S., & Terzidis, O. (2025). A hybrid framework for creating artificial intelligence-augmented systematic literature reviews. *Management Review Quarterly*. https://doi.org/10.1007/s11301-025-00522-8

Massiceti, D., Hicks, S. L., & van Rheede, J. J. (2018). Stereosonic vision: Exploring visual-to-auditory sensory substitution mappings in an immersive virtual reality navigation paradigm. *PLOS ONE*, *13*(7), e0199389. https://doi.org/10.1371/journal.pone.0199389

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, *485*(7397), 233–236. https://doi.org/10.1038/nature11020

Norman, D. A. (2013). *The design of everyday things: Revised and expanded edition*. Basic Books.

Paas, F., & Van Merriënboer, J. J. G. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science*, *29*(4), 394–398. https://doi.org/10.1177/0963721420922183

Potter, T., Cvetković, Z., & De Sena, E. (2022). On the relative importance of visual and spatial audio rendering on VR immersion. *Frontiers in Signal Processing, 2,* Article 904866. https://doi.org/10.3389/frsip.2022.904866

Pulkki, V., & Karjalainen, M. (2001, July). Directional quality of 3D amplitude-panned virtual sources. In *Proceedings of the 7th International Conference on Auditory Display* (pp. 203–206).

Rosin, J. (2023, August 2). How extended reality tech could help visually impaired people. *The Parliament Magazine*. https://www.theparliamentmagazine.eu/news/article/virtual-reality-technology-visually-impaired-people-help

Rojas, D., Kapralos, B., Hogue, A., Collins, K. C., Nacke, L., Cristancho, S., Conati, C., & Dubrowski, A. (2013). The effect of sound on visual fidelity perception in stereoscopic 3-D. *IEEE Transactions on Cybernetics, 43*(6), 1561–1571. https://doi.org/10.1109/TCYB.2013.2269712

Rumsey, F. (2002). *Spatial audio*. Focal Press.

Shams, L., & Beierholm, U. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, *14*(9), 425–432. https://doi.org/10.1016/j.tics.2010.07.001

Shernoff, D. J., Csikszentmihalyi, M., Shneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly, 18*(2), 158–176. https://doi.org/10.1521/scpq.18.2.158.21860

Skarbez, R., Brooks, F. P., Jr., & Whitton, M. C. (2017). A survey of presence and related concepts. *ACM Computing Surveys, 50*(6), 96:1–96:39. https://doi.org/10.1145/3134301

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI, 3,* 74. https://doi.org/10.3389/frobt.2016.00074

Stefanowska, A., & Zieliński, S. K. (2024). Spatial sound and emotions: A literature survey on the relationship between spatially rendered audio and listeners' affective responses. *International Journal of Electronics and Telecommunications, 70*(2), 293–300. https://doi.org/10.24425/ijet.2024.149544

Sun, X. (2021). Immersive audio, capture, transport, and rendering: A review. *APSIPA Transactions on Signal and Information Processing, 10,* e13. https://doi.org/10.1017/ATSIP.2021.12

Turner, D., Murphy, D., Pike, C., & Baume, C. (2022). Spatial audio production for immersive media experiences: Perspectives on practice-led approaches to designing immersive audio content. *The Soundtrack, 13*(1–2), 73–94. https://doi.org/10.1386/ts_00017_1

Verhulst, I., Hemming, R., Ganz, A., Bennett, J., Donnelly, R., Watling, D., & Dalton, P. (2024). Predictors of the sense of presence in an immersive audio storytelling experience: A mixed-methods study. *arXiv*. https://doi.org/10.48550/arXiv.2406.05856

Vicente, P.-N., & Pérez-Seijo, S. (2022). Spatial audio and immersive journalism: Production, narrative design, and sense of presence. *El Profesional de la Información*. https://doi.org/10.3145/epi.2022.oct.19

Wang, J., Ye, Q., Shuai, Z., Wang, P., Wang, Y., & Lin, C. (2024). The influence of transportation, social norms, cultural identity, and affective disposition in transnational media enjoyment. *Frontiers in Psychology, 15*, 1377898. https://doi.org/10.3389/fpsyg.2024.1377898

Wang, L., Zeng, X., & Ma, X. (2019). Advancement of individualized head-related transfer functions (HRTFs) in perceiving the spatialization cues: Case study for an integrated HRTF individualization method. *Applied Sciences, 9*(9), 1867. https://doi.org/10.3390/app9091867

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240. https://doi.org/10.1162/105474698565686

Zhu, Z., Zhang, Y., Guo, W., Pan, C., & Zhao, Z. (2025). ASAudio: A survey of advanced spatial audio research. *arXiv*. https://arxiv.org/abs/2508.10924

Appendix A:

HEC MONTRÉAL

Comité d'éthique de la recherche

January 30, 2025

To the attention of: Farah Abou Jaoude

Re: Ethics approval of your research project

Project No.: 2025-6304
Title of research project: Effect of sound quality on immersive experiences

---

Dear Farah Abou Jaoude,

Your research project has been evaluated in accordance with ethical conduct for research involving human subjects by the Research Ethics Board (REB) of HEC Montréal.

A Certificate of Ethics Approval attesting that your research complies with HEC Montréal's *Policy on Ethical Conduct for Research Involving Humans* has been issued, effective January 30, 2025. This certificate is **valid until January 01, 2026.**

Please note that you are required to renew your ethics approval before your certificate expires using Form *F7 – Annual Renewal*. You will receive an automatic reminder by email a few weeks before your certificate expires.

If any changes are made to your project before the certificate expires, you must complete *F8 – Project Modification* and obtain REB approval before implementing those changes. If your project is completed before the certificate expires, you must complete Form *F9 – Termination of Project* or *F9a – Termination of Student Project*, as applicable.

Under the *Policy on Ethical Conduct for Research Involving Humans*, researchers are responsible for ensuring that their research projects maintain ethics approval for the entire duration of the research work, and for informing the REB of its completion. In addition, any significant changes to the project must be submitted to the REB for approval before they are implemented.

You may now begin the data collection for which you obtained this certificate.

We wish you every success in your research.

REB of HEC Montréal

HEC MONTRĒAL

Comité d'éthique de la recherche

CERTIFICATE OF ETHICS APPROVAL

This is to confirm that the research project described below has been evaluated in accordance with ethical conduct for research involving human subjects, and that it meets the requirements of our policy on that subject.

Project No.: 2025-6304

Title of research project: Effect of sound quality on immersive experiences

Principal investigator: Farah Abou Jaoude

Date of project approval: January 30, 2025
Effective date of certificate: January 30, 2025
Expiry date of certificate: January 01, 2026

Maurice Lemelin
Président
CER de HEC Montréal

Signé le 2025-01-30 à 16:40

Appendix B: Audio clips links:

EMEL

1. Rescuer (anglais)
Spotify: https://open.spotify.com/track/6cU78tvYRUn1hf83FbPUzv?si=a012ce88ebd54927
Apple: https://music.apple.com/ca/album/rescuer/1467689493?i=1467689497

2. Merrouh
Spotify: https://open.spotify.com/track/79mMn1BeJCExSK7WfEkixl?si=36cc25a7490d4696
Apple: https://music.apple.com/ca/album/merrouh/1467689493?i=1467689510

JONATHAN KAWCHUCK

1. Solar Plexus
Spotify: https://open.spotify.com/track/5vSYIcEr61rkQyiJjgQvQZ?si=f41742e1dbab43ac
Apple: https://music.apple.com/ca/album/solar-plexus/1593457178?i=1593457347

DAFT PUNK

1. Get Lucky
Spotify: https://open.spotify.com/track/69kOkLUCkxIZYexIgSG8rq?si=647de74a02944a86
Apple: https://music.apple.com/ca/album/get-lucky/1673536063?i=1673536443

SOFIE BIRCH

1. Lolupupi_Kutiku_Unmastered_2.0.wav
https://drive.google.com/file/d/1aKLoUuT_ef_UjWvk5PC6e7RqXJxiy7pp/view?usp=sharing

ELA MINUS

1. Broken
Spotify:https://open.spotify.com/track/3n4FkCfKlscDqYPMcTBdXm?si=7b9b9687303e4d4e
Apple: https://music.apple.com/ca/album/broken/1743438405?i=1743438648

Appendix C: Researcher's Tags

Sense of space	*Storytelling and imagination*	*Quality and nature of the sound*	*Emotional connection and appropriation of the sound*
Felt like the sound wasn't fully around me	The sense of, like, I was on the top of the mountain and, like, I had achieved something	I felt so many layers of the song	The sound was a little annoying to me; the song felt very weird
I felt like I was right there; you're not where you actually are	They were singing for my ears	Quand il y a plus de voix, je trouve que ça transporte énormément (When there are more voices, I find it's incredibly transporting)	I felt bored because it was very monotone
Complètement laisser emporter et j'étais dans une bulle (Completely let go and I was in a bubble)	Si on ferme les yeux en écoutant le son, j'ai l'impression que quand je vais les rouvrir, tout aura changé autour de moi (If you close your eyes while listening to the sound, I feel like when I open them again, everything around me will have changed)	Je trouve que c'étaient les basses qui faisaient des vibrations (I find it was the bass that made vibrations)	Qu'on me criait dessus, ça m'a un peu oppressé (That I was being yelled at, it oppressed me a little)
M'ont laissé me sentir dans un univers particulier (Let me feel like I was in a particular universe)	I felt like I was losing track of time, I didn't even feel the minutes passing	More plural sounds and more, you know, like, ecosystem sounds	Je pense que c'est le 360, le fait de se sentir un peu cerné par le truc, de se dire "je ne sais pas d'où ça vient, d'où ça va venir", mais j'avais l'impression que ça arrivait et que ça me parlait à moi (I think it's the 360, the feeling of being a bit surrounded by the thing, of telling yourself "I don't know where it's coming from, where it's going to come from", but I had the impression that it was happening and speaking to me)
I feel like I am drowning in the environment	J'avais presque l'impression qu'il y avait de l'eau sur moi (I almost felt like there was water on me)	Were different in terms of intensity	Quand je m'assois à côté de ma famille, je m'endors et je les entends parler (When I sit next to my family, I fall asleep and hear them talking)
J'avais l'impression que ça venait plus de la droite ou de la gauche (I had the impression it was coming more from the right or the left)	Je me suis beaucoup moins senti envolé dans le truc (I felt much less like I was flying in the thing)	Very, like, textured	Elle commence par un audio relaxant, et puis il y avait un bruit qui me gênait (It starts with a relaxing audio, and then there was a noise that bothered me)

Il y avait beaucoup moins cette dimension d'espace (There was much less of this spatial dimension)	Ça me donnait vraiment l'espace d'imaginer comme si j'étais dans un rêve (It really gave me the space to imagine as if I were in a dream)	It's a combination of multiple things, which are the volume, like, the different kinds of volume, the different kinds of sounds	Disconnected from your, you know, present state
Wrapped in this sound	On dirait que tu es vraiment entré dans un rêve (It's like you've really entered a dream)	La qualité du son faisait qu'on était dans un environnement sans écho (The sound quality made it so we were in an environment without echo)	Being able to not care about being perceived, about being, you know, seen
Je l'entends à gauche, parfois à droite (I hear it on the left, sometimes on the right)	Vraiment un univers comme avec des oiseaux (Truly a universe as if with birds)	Comme des voix humaines qui paraissaient, entre guillemets, plus naturelles (Like human voices that seemed, quote, unquote, more natural)	J'étais un peu plus en vigilance (I was a little more alert)
Comme si on était à un concert ou ce genre de choses. Donc tu te sens un peu dans l'espace (As if you were at a concert or that kind of thing. So you feel a bit like you're in the space)	Sentir la musique pénétrer un peu la peau, son esprit (To feel the music penetrate the skin a little, one's mind)	Il me semble que les sons ne sont pas naturels (It seems to me that the sounds are not natural)	Il y avait les bruits aigus, enfin, j'ai trouvé que ça irritait un peu les oreilles et que ça n'allait pas vraiment avec le mot "broken" (There were high-pitched noises, well, I found it irritated the ears a bit and it didn't really go with the word "broken")
J'avais l'impression que ça tournait autour (I had the impression it was turning around me)	There's not vibration no diversity	Des sons un peu plus artificiels (Sounds that are a bit more artificial)	J'étais relaxée en entendant ça (I was relaxed hearing that)
C'est être plongé dans un univers, puis c'est avoir l'impression de ne plus être à l'endroit où tu es, d'être un peu dans un autre monde (It's being plunged into a universe, and then having the impression of no longer being where you are, of being in another world a little)	But then it's like it watches you, and it turns around; like, wherever you go, it'll turn its neck. Three-sixty?	Des sons assez naturels et des sons assez différents les uns des autres (Quite natural sounds and sounds that are quite different from each other)	Weird in a good way, but it's, like, interestingly weird
J'avais l'impression d'être dans la nature, un peu dans une jungle (I felt like I was in nature, a bit in a jungle)	Yeah. Kind of. Okay. And it's like, I feel like, yeah, it's subtle, like an owl, because, you know, it flies in the middle of the night, it just sneaks by. And it's like, but at the same time, it's like you feel like the little mice are	Ils étaient un peu surchargés, un peu comme avec des grésillements presque (They were a bit overloaded, almost like with static)	I would definitely pay money to, like, get some sad songs in there and cry my eyes out

	terrified, running in the forest		
Sentir la musique pénétrer un peu la peau, son esprit (To feel the music penetrate the skin a little, one's mind)	Like, they're terrified of it	Il y avait des bruits aigus qui me faisaient un peu penser à des animaux, après il y avait des bruits un peu plus graves où là c'était un peu comme des frottements de feuilles, des choses comme ça (There were high-pitched noises that made me think of animals, then there were deeper noises that were a bit like leaves rubbing, things like that)	What actually gave me goosebumps
Like they were actually around you	Like music because every instrument tells a story, kind of	Il y avait les bruits aigus, enfin, j'ai trouvé que ça irritait un peu (There were the high-pitched noises, well, I found them a bit irritating)	The first one was so calming
The one with the fairy tale, like, in a green land, you know, standing next to a river or something	It felt more like it was genuinely inside your head	La femme qui chantait était très... enfin, j'ai dit le mot qui décrit, c'est "poignant" (The woman who was singing was very... well, I said the word that describes it is "poignant")	It actually made me feel like it also kind of touched my soul
Like, I was in a club or something, and the sound was directly like I was standing next to the singer or something	The films where we have, like, chips inside your brain	Je sentais un peu plus que c'était le haut-parleur, mais sinon, les autres musiques, je sentais vraiment toute la musique autour de moi et pas juste vers un seul endroit (I felt a bit more that it was the speaker, but otherwise, with the other music, I really felt all the music around me and not just from one direction)	It was like someone was humming exactly next to my ear
The sense of, like, I was on top of the mountain and had achieved something	Vu que j'ai repris pas mal le terme de "voler", en gros, entre guillemets, je dirais l'oiseau parce que tu as une sensation de liberté, on va dire, de bien-être, donc ce qui peut être le plus proche, c'est peut-être l'oiseau, je dirais (Since I used the term "flying" quite a bit,	It's only 70 decibels, and I thought it was like 90 decibels at least, you know	D'un point de vue, c'est envoûtant, on peut dire (From one point of view, it's captivating, you could say)

	basically, I'd say the bird because you have a feeling of freedom, let's say, of well-being, so the closest thing might be the bird, I'd say)		
I felt like there was something in between that sounded like an ocean wave or the sea	L'impression vraiment que ça vient te toucher (The impression that it really comes and touches you)	70 decibels, and it felt like it was the perfect amount of sound	Il y a aussi l'aspect émotionnel de connaître le son (There is also the emotional aspect of knowing the sound)
And I felt like I was actually there, and I felt the wind	Un truc un peu "serpentesque", du serpent, que c'est partout, ça laisse des traces, tu vois, ça mue. Un serpent, ça laisse sa vieille peau, il y a un truc un peu de parcours, d'initiation, d'expérience (Something a bit "serpentine," like a snake, that it's everywhere, it leaves traces, you see, it molts. A snake leaves its old skin; there's something of a journey, an initiation, an experience to it)	The scraping was really realistic	Plus tu te détends, mais c'est vrai que c'est très intéressant. Ça fait des frissons quand tu sens que le son se déplace autour de toi (The more you relax, but it's true that it's very interesting. It gives you chills when you feel the sound moving around you)
J'ai vraiment l'impression d'aller en concert (I really feel like I'm going to a concert)	C'était relaxant et j'avais même envie de m'endormir (It was relaxing, and I even felt like falling asleep)	Being able to hear each instrument like it was actually there is very important	J’ai eu des surprises, parce que je ne m'attendais pas à sentir un son venir d'un côté ou de l'autre (I had some surprises because I didn't expect to feel a sound coming from one side or the other)
Tu t'es vraiment emporté; en fait, tu fermes les yeux, tu es vraiment dans un autre environnement, enfin, tu es ailleurs (You really get carried away; in fact, you close your eyes, you're really in another environment, well, you're somewhere else)	Tu t'amuses un peu à savoir comment, juste le fait de fermer les yeux, de t'imaginer dans tel ou tel endroit, c'était quand même assez fun. Et ça montre aussi à quel point même les petites choses peuvent vraiment complètement changer la direction d'une chanson et d'un effet sonore (You have a bit of fun figuring out how, just by closing your eyes, imagining yourself in this or that place, it was still quite fun. And it also shows	You could really separate everything	Qui te guérit aussi (That heals you too)

	how much even little things can completely change the direction of a song and a sound effect)		
Tu fermes les yeux, tu es vraiment ailleurs, en fait (You close your eyes, you're really somewhere else, in fact)	Thème un peu mystique, forêt, tout ça... du coup, c'est le loup qui me vient à l'esprit (A somewhat mystical theme, forest, all that... so it's the wolf that comes to mind)	I could even hear the sound of some animals or insects	Il y a cette sensation d'expérience physique où on a l'impression que c'est fait pour nous (There's this sensation of a physical experience where you feel like it's made for you)
J'étais vraiment ailleurs (I was really somewhere else)	It's the voice of the woman, and also, it makes it feel very real	Des voix très synthétiques (Very synthetic voices)	And a bit of goosebumps
L'impression que ça vient te toucher (The impression that it comes and touches you)	I was thinking about the tiger or the lion, but for that animal, it's kind of terrifying or can be damaging	C'est des voix synthétiques globalement, même si les trois premières voix sont des voix vraiment naturelles qui sont perçantes, on va dire. En fait, oui, c'est des voix fortes (They're synthetic voices overall, even if the first three voices are really natural voices that are piercing, let's say. In fact, yes, they are strong voices)	Not focusing, but focusing at the same time
Vraiment imprégné par le son, où on a l'impression à des moments d'être dans la musique, presque. Où vraiment tu fermes les yeux, tu les rouvres, tu ne sais plus où tu es. Sur des sons, tu peux vraiment te dire : "je suis dans cette musique" (Really impregnated by the sound, where you sometimes have the impression of being inside the music, almost. Where you really close your eyes, you open them, you don't know where you are anymore. With some sounds, you can really tell yourself: "I am in this music")	Tu le ressens un peu plus à l'intérieur de toi (You feel it a little more inside you)	Parce qu'il y avait des basses, des infrabasses, et là, les voix qui vraiment voyageaient d'un côté à l'autre autour de moi (Because there was bass, sub-bass, and then the voices that really traveled from one side to the other around me)	Like sleeping but still awake. Like, I'm relaxed, but at the same time, I'm able to follow the sound
Et là, les voix qui voyageaient vraiment d'un côté à l'autre autour	Un caméléon, parce que c'est un peu capable de représenter plusieurs	C'était des genres de fréquences un peu qui te traversent (They were	C'était vraiment comme si j'avais autour de moi des femmes qui chantaient pour

de moi (And there, the voices that were really traveling from one side to the other around me)	"peaux" à la fois, qui peut se camoufler, qui... en fait, pour moi, ça peut représenter les différentes façons (A chameleon, because it's kind of able to represent several "skins" at once, that can camouflage itself, that... in fact, for me, it can represent the different ways)	kinds of frequencies that sort of pass through you)	moi (It was really as if I had women around me singing for me)
Pour moi, la voix ne bougeait pas; enfin, c'était comme s'il y avait un haut-parleur à ma gauche, quoi, par rapport à là où j'étais placé (For me, the voice didn't move; well, it was as if there was a speaker to my left, you know, relative to where I was placed)	I was able to feel the surroundings and, yeah, if I close my eyes, I'm there	C'était des fréquences qui aidaient aussi à s'immerger dans un autre état de conscience un peu plus doux (They were frequencies that also helped to immerse oneself in another, gentler state of consciousness)	Ça me faisait aussi sortir certaines émotions, peut-être de la peur, du bonheur, des choses comme ça (It also brought out certain emotions in me, maybe fear, happiness, things like that)
On est très enveloppé (We are very enveloped)	I'm in a forest	Il y avait des sons où ce n'était pas travaillé aussi bien que d'autres (There were sounds where it wasn't as well-crafted as others)	I felt like it's kind of terrifying for me; I feel like it's like a scary movie or something like that
Comme quand on est immergé dans l'eau, là on est immergé dans la musique (Like when you're immersed in water, here you're immersed in the music)	Je ne sais pas pourquoi, le son d'une baleine dans l'eau (I don't know why, the sound of a whale in the water)	La qualité du son, l'ambiance, les lumières et le fait d'être assis, d'être détendu, ça m'a permis de me relaxer et de me concentrer d'une manière fluide (The quality of the sound, the atmosphere, the lights, and the fact of being seated, being relaxed, allowed me to relax and concentrate in a fluid way)	Some person or somebody is just, like, talking to me
On the left and on the right, even, like, you can pinpoint where the sound is	Ca me donner une sensation de liberté, comme si je volais comme un oiseau	Même les petits bruits de fond qu'on pouvait entendre, que ce soit peut-être, par exemple, dans la première chanson, c'était les femmes qui chantaient, et cetera, et c'était le fait qu'on les entendait un peu respirer, tout ça. Toute chose qui se rapporte au naturel, un	Somebody is just talking to me from a very specific direction, and I feel it's very real

		peu, c'est vraiment ça qui aide beaucoup (Even the little background noises you could hear, whether it was, for example, in the first song, it was the women singing, et cetera, and it was the fact that you could hear them breathe a little, all that. Anything that relates to what's natural, a little, that's really what helps a lot)	
I can feel like sometimes the voice is from the left and sometimes it's from the right		Qu'il n'y avait pas assez de bruits de fond qui aidaient un peu à être un peu plus immergé dedans (That there weren't enough background noises that helped a bit to be more immersed in it)	Tu fais juste un avec la musique (You just become one with the music)
Je pense qu'on entendait les sons qui venaient d'un peu plus partout avec les voix, peut-être que ça rendait l'expérience un peu meilleure (I think we heard the sounds coming from all over with the voices; maybe that's what made the experience a bit better)		It's really connecting or relates to the sounds of the birds or the peacock or the forest or something like that	I kind of felt hopeful
Everything that you hear from different sides of the room		Not too many artificial voices or human voices inside. So this makes me feel like it's more real	J'ai senti des frissons. Ça, c'est par le corps, tu peux sentir dans ton cœur, tu peux sentir une joie en écoutant des musiques. C'est vraiment le fait de sentir (I felt chills. That comes through the body; you can feel it in your heart, you can feel joy listening to music. It's really the act of feeling)
		There are no vibrations, no diversity within the whole song	
		The beats are so hard, but the music is soft	

Appendix D: Chat GPT Tags and key findings

Tag (code)	Core meaning	Illustrative quotes
#360_Surround	Being acoustically *enveloped*; sounds arrive from every direction and move in space	"the sound was coming from every single angle … the singer was spinning around me" / "le 360, le fait de se sentir un peu cernée par le truc"
#Time_Loss	Losing track of time and place; forgetting the real room	"I was losing the track of time" / "perdre la notion du temps … tu n'existes plus ici"
#Affective_Response	Strong emotional or affective impact (goose-bumps, being "touched", fear)	"it … touched my soul" / "Elle m'a *hyper touchée*, trop profond"
#Body_Vibration	Bodily sensations (bass-induced vibration, tremble) that reinforce presence	"les basses … ça vibrait partout autour de moi, même le siège" / "the basses … making you *tremble* a little bit more"
#Natural_Sounds	Use of environmental/nature sounds that feel "real" and draw the listener in	"les plus immersifs, c'est ceux avec des sons naturels, bruits d'oiseaux" / "I like natural ambiance … birds chirping, waterfalls"
#Imagery	Evoking vivid mental images or narratives; "being somewhere else"	"j'avais l'impression d'être dans la jungle" / "if I close my eyes, I'm in a forest"
#Relaxation_Flow	A state of relaxation, focus or even drowsiness; effortless attention	"c'était *vraiment relaxant*, je voulais m'endormir" / "the first one was so *calming*"
#Sound_Layers	Perceiving fine detail, multiple layers or textures in the mix	"I felt *so many layers* of the song" / "on se rend compte de toutes les petites choses dans la musique"
#Contextual_Factors	Non-audio elements (lighting, seating, smells) that modulate immersion	"the lighting *enhanced* the experience" / "qualité du son, ambiance, lumières – ça m'a permis de me *relaxer*"

Key Findings of chatGPT

1. **Spatial qualities are primary.** Over two-thirds of participants spontaneously defined immersion through 360° placement, movement and wrap-around sound (#360_Surround). Stereo clips perceived as "coming from one speaker" were consistently ranked least immersive.

2. **Presence involves both cognitive displacement and bodily anchoring.** Participants described *simultaneous* loss of time/place (#Time_Loss) and heightened bodily awareness of vibration or proximity (#Body_Vibration), suggesting that effective immersion blends disembodiment with intensified embodiment.
3. **Emotion is not optional.** Emotional resonance (#Affective_Response) – whether calming, thrilling or frightening – was repeatedly framed as evidence that immersion "worked". Goose-bumps were cited as a diagnostic marker.
4. **Naturalistic textures beat familiar pop mixes.** Clips rich in *non-musical* nature cues (#Natural_Sounds) were widely preferred; well-known commercial tracks were often judged "concert-like, not immersive".
5. **Imagery bridges sound and place.** Vivid mental scenes (#Imagery) ranged from rivers and jungles to mystical rituals; participants who struggled to "attach an image" reported lower immersion.
6. **Relaxation and flow are outcome states.** Many equated immersion with meditative relaxation or flow (#Relaxation_Flow), sometimes to the point of drowsiness.
7. **Fine granularity matters.** The ability to discern "layers", micro-details and spatialized objects (#Sound_Layers) was treated as proof of high audio quality and, by extension, immersion.
8. **Room design can amplify or break the spell.** Soft lighting, comfortable seating and absence of distracting odours enhanced immersion, whereas fan noise or foot-odour (#Contextual_Factors) broke it for some.

Appendix E: Use of AI in This Thesis

During the preparation of this thesis, I used AI tools, like OpenAI's ChatGPT (GPT-5 mini), Google Gemini and Claude, to assist with certain tasks such as drafting text, organizing ideas, and checking reference formatting. The AI was helpful in suggesting ways to phrase ideas more clearly, highlighting points I might have overlooked, and flagging potential formatting issues in references.

At the same time, using AI came with challenges. Occasionally, it generated references that did not exist or provided incomplete or inaccurate information, requiring careful verification of every source. It also sometimes failed to locate the references I needed, which meant manual searches in databases were still necessary.

Overall, AI served as a useful but limited assistant. All substantive content, data interpretation, and final decisions remained under my control, ensuring the academic integrity and originality of this work