FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Sonic Inhabitation: An Interactive Study of Embodied Sound Cognition in Indoor Spaces

Margarida Mesquita Leal



Master in Multimedia: Sound Design and Interactive Music

Supervisor: Gilberto Bernardes de Almeida

Second Supervisor: Alexander Refsum Jensenius

August 31, 2026

Sonic Inhabitation: An Interactive Study of Embodied Sound Cognition in Indoor Spaces

Margarida Mesquita Leal

Master in Multimedia: Sound Design and Interactive Music

August 31, 2026

Abstract

Sound is a constant presence in everyday life, yet indoor environments still tend to treat it as background, a by-product of architecture rather than something that actively shapes how people move and behave. This dissertation investigates how humans experience and inhabit indoor spaces through an interactive 3D audio experience. This work challenges that view by examining the reciprocal interaction between a moving body and a responsive acoustic field, conducted at the fourMs Lab of the RITMO research centre (University of Oslo).

To this end, Sonic Inhabitation was developed, a fully auditory installation in which a participant's own movement is the exclusive means of accessing an otherwise invisible room. The system combines a Qualisys motion-capture array, a 24-speaker ambisonic loudspeaker setup, and a Max/MSP patch that maps the participant's position and velocity onto sound parameters, joining real-time synthesis and sample manipulation. A set of ambisonic recordings of rooms from different countries, the Intercontinental Dataset, is presented to the participant through a three-phase narrative structure, progressing from a single stationary zone to four overlapping rooms and, finally high interactive environment. Thirteen participants completed the experience: their kinematic data were analysed through Quantity of Motion (QoM), and each session was followed by a semi-structured interview on presence, embodiment, and narrative comprehension.

The results are exploratory but point to a common direction: movement in the acoustic overlap zone between rooms was significantly higher than in the rest of the space, independently corroborated by participants, who spontaneously describe boundary-seeking as their dominant exploratory strategy. Movement significantly decreased in the seconds immediately following a sound-trigger event, running against the initial hypothesis that sound would invite more movement rather than a brief pause and the qualitative reading of the interviews, extended by Lena Bauer-Wolf's formal thematic analysis, surfaces recurring themes of cultural resonance, a head-and-neck-centred locus of embodiment, and a tension between acoustic overwhelm and silence as a place of recovery. Together, these results support the claim that the sonic characteristics of an indoor space measurably shape human behaviour, even in the complete absence of visual information.

These findings open a direct path forward. The next steps are to bring the full-body motion-capture data (currently unanalysed) into the picture alongside the head-tracking data already used here, to test whether the drop in movement after a sound trigger holds across all sound types or only some, and, longer term, to move the system off a fixed, expensive motion-capture rig and onto more portable tracking, so that a version of this installation could be deployed and tested in other spaces and configurations.

Acknowledgments

I want to thank friends, colleagues, advisors, and family who helped me all the way through this last year.

To my parents, brother and Bea for believing in me and giving me all the tools to be where I am today.

To my friends, for keeping my heart full and being my biggest fans, even from far away.

To Lena, for accompanying me through every step of this process and teaching me how to trust myself.

Thanks to everyone at RITMO for welcoming me with open arms and teaching me so much.

To my advisors, Prof. Gilberto Bernardes and Prof. Alexander Refsum Jensenius, thank you for guiding me through this journey of discovery. I am fortunate to have had your guidance and to have worked with you this past year.

To all of you, I am grateful.

Margarida

Contents

1 Introduction **1**
- 1.1 Motivation and Context . . . 1
- 1.2 Research questions . . . 2
- 1.3 Project and Methodology . . . 3
- 1.4 Structure of the Dissertation . . . 3
- 1.5 Use of Artificial Intelligence Tools . . . 4

2 State of the Art **5**
- 2.1 Human Perception and Embodiment Cognition . . . 5
  - 2.1.1 Signals . . . 6
  - 2.1.2 Listening Modes and Frameworks . . . 9
  - 2.1.3 Embodied Cognition . . . 9
- 2.2 Space . . . 11
  - 2.2.1 Soundscapes . . . 11
- 2.3 Affective Spaces . . . 12
  - 2.3.1 Perceived Affective Quality of Soundscapes . . . 13
  - 2.3.2 Affective Spaces and Embodied Psychoacoustics . . . 14
  - 2.3.3 Affordances . . . 14
  - 2.3.4 The Schaefferian Sound Object . . . 15
- 2.4 Listening . . . 15
  - 2.4.1 Psychoacoustics . . . 15
  - 2.4.2 Auditory Scene Analysis . . . 16
  - 2.4.3 Acoustemology . . . 16
  - 2.4.4 Spatial Audio and Its Psychoacoustics . . . 17
- 2.5 Interactivity . . . 20
  - 2.5.1 Action–Sound Couplings . . . 21
  - 2.5.2 Agency and Shared Control . . . 21
  - 2.5.3 Sound Actions and Movement Analysis . . . 21
  - 2.5.4 Body-driven Musical Interfaces . . . 22
  - 2.5.5 Performing in the "Air" . . . 22
  - 2.5.6 The Action-Perception Loop . . . 23
- 2.6 Related Works . . . 23
- 2.7 Key Theorists . . . 24
  - 2.7.1 Diana Deutsch . . . 25
  - 2.7.2 Alva Noë . . . 25
  - 2.7.3 Thomas Metzinger . . . 26
- 2.8 Summary and Conclusions . . . 26

3 System architecture **29**
3.1 Audio Framework . . . 29
3.1.1 Intercontinental Dataset . . . 30
3.1.2 Soundmarks . . . 30
3.1.3 Sonic Wall . . . 32
3.1.4 Spatialization and Multi-Channel Configuration . . . 33
3.2 Motion Capture System . . . 34
3.3 System Architecture and Data Integration . . . 35
3.3.1 High-Level Architecture and Data Flow . . . 35
3.4 Max/MSP Patch Architecture . . . 36
3.4.1 Overview . . . 36
3.4.2 QTM2OSCStreaming_NEW.maxpat: The Motion-Capture Bridge . . . 39
3.4.3 fFE_3D1_NEW.maxpat . . . 39
3.4.4 CONTROL.maxpat . . . 40
3.4.5 1phase.maxpat: The Phase 1 Sound Engine . . . 40
3.4.6 2phase.maxpat: The Phase 2 and 3 Sound Engine . . . 41
3.4.7 spatializer.maxpat: Loudspeaker Rendering . . . 42
3.5 Summary . . . 42

4 User Study Protocol **45**
4.1 Physical Space . . . 45
4.2 Instruments . . . 46
4.2.1 Highly Sensitive Person (HSP) Questionnaire . . . 46
4.2.2 Motion Capture Metrics . . . 47
4.2.3 Post-Exposure Interview . . . 48
4.3 3-Phase Narrative Design . . . 49
4.3.1 Three-Phase Approach . . . 49
4.3.2 Phase 1: Sensory Calibration and Immersion (3 Minutes) . . . 49
4.3.3 Phase 2: Spatial Agency (3 Minutes) . . . 50
4.3.4 Phase 3: Complex Architecture and Environmental Anticipation . . . 50
4.4 Step-by-Step Experimental Procedure . . . 50
4.4.1 Pre-Exposure Phase . . . 50
4.4.2 The Briefing and Exposure Phase . . . 51
4.4.3 Post-Exposure Phase . . . 52
4.5 Summary . . . 52

5 Methods for Data Analysis **53**
5.1 Data Sources . . . 53
5.1.1 Behaviour Log . . . 53
5.1.2 QTM Motion Capture . . . 54
5.1.3 Audio Recordings . . . 54
5.2 Participants and Data Completeness . . . 54
5.3 Preprocessing Pipeline . . . 55
5.3.1 Marker Selection for Motion Capture . . . 55
5.3.2 Room Occupancy and the Simultaneous Zone . . . 56
5.3.3 Quantity of Motion . . . 56
5.4 Metrics and Statistical Procedures . . . 57
5.4.1 Phase Duration and Room Dwell Time . . . 57
5.4.2 Sound-Onset Response . . . 58

5.4.3 Statistical Tests . . . 58
5.5 Limitations of the Analytical Pipeline . . . 58

6 Results and Discussion **61**
6.1 Participants . . . 61
6.2 Results . . . 62
6.2.1 Session Duration and Overall Movement . . . 62
6.2.2 Movement Across the Three Phases . . . 63
6.2.3 The Simultaneous / Between-Rooms Zone . . . 65
6.2.4 Movement Response to Sound Onsets . . . 65
6.2.5 Room Dwell Time and Movement . . . 66
6.2.6 Individual Differences in Room Preference . . . 66
6.2.7 A Formal Analysis of the Interview Transcripts . . . 67
6.3 Discussion . . . 71
6.3.1 Faster Movement in the Simultaneous Zone . . . 71
6.3.2 Movement Response to a Sound Onset . . . 72
6.3.3 Room Dwell Time and Movement . . . 73
6.3.4 Room Preference . . . 74
6.3.5 Research Questions . . . 74
6.4 Concluding Remarks . . . 76

7 Conclusions and Future Work **81**
7.1 Finishing the Data Analysis . . . 81
7.1.1 Cleaning the Qualisys Dataset . . . 81
7.1.2 Deeper Analysis of the Cleaned Movement Data . . . 82
7.2 Cross-Analysing Sound and Movement and Publication . . . 82
7.2.1 Aligning the Audio and Motion-Capture Records . . . 82
7.2.2 Soundmark-Specific Analysis . . . 82
7.3 Publication . . . 83
7.3.1 Open-Access Data Publication . . . 83
7.3.2 Publishing the System . . . 83
7.3.3 Exporting to Different Physical Setups . . . 83
7.3.4 Further Directions Suggested by This Study's Findings . . . 84
7.4 Limitations . . . 84
7.5 Concluding Remarks . . . 85

References **87**

A Full Interview Transcripts **97**
A.1 Participant 1-T (17 June 2026, 14:30) . . . 97
A.2 Participant 2-A (18 June 2026, 09:30) . . . 98
A.3 Participant 3-A (18 June 2026, 11:00) . . . 101
A.4 Participant 4-K . . . 102
A.5 Participant 5-B (22 June 2026, 14:00) . . . 103
A.6 Participant 6-A (23 June 2026, 13:00) . . . 105
A.7 Participant 7-C (23 June 2026, 13:30) . . . 107
A.8 Participant 8-E (24 June 2026, 13:00) . . . 108
A.9 Participant 9-S (24 June 2026, 15:00) . . . 109
A.10 Participant 10-B (25 June 2026, 12:30) . . . 110

A.11 Participant 11-L (26 June 2026, 12:00) 112
A.12 Participant 12-H (30 June 2026, 13:30) 113
A.13 Participant 13-N (30 June 2026, 15:00) 116

List of Figures

2.1 The path of a sound event, extended to include the digital mediation loop. A physical excitation produces an acoustic wave, which may either travel directly to the ear or be captured by a microphone, converted into an audio signal, processed digitally, and re-emitted by a loudspeaker as a second, physically distinct acoustic wave before reaching the ear. . . . 7
2.2 Correspondence between Schaeffer's three modes of listening and Gaver's everyday/musical distinction. . . . 9
2.3 R. Murray Schafer's Soundscape Taxonomy . . . 12
2.4 Principal components model of soundscape perception, showing orthogonal dimensions of pleasantness and eventfulness alongside diagonal axes (Axelsson et al., 2010) . . . 13

3.1 R. Murray Schafer's soundscape taxonomy applied to the system's three audio layers. . . . 30
3.2 24-Channel Loudspeaker Array Configuration . . . 35
3.3 RITMO decoder . . . 36
3.4 The headpiece - rigid-body tracker used to control the Max patch through Qualisys 37
3.5 Mocap Suite using Sports marker set . . . 37
3.6 System architecture: motion capture and spatial audio data flow. . . . 38
3.7 Data flow across the six Max/MSP patches. Solid arrows carry positional/audio data. The dashed arrows show the external MTC-based clock (Section 3.3) feeding timestamps to both the QTM side and the Max/MSP side independently. . . . 38
3.8 QTM2OSCStreaming_NEW.maxpat. A UDP bridge, based on the fourMs Lab's public `qualisys-osc` patch, that requests and receives a continuous stream of 6-degree-of-freedom rigid-body data for the head. . . . 39
3.9 fFE_3D1_NEW.maxpat (Nymoen, fourMs Lab). . . . 40
3.10 CONTROL.maxpat. The dashed arrow shows that the phase/room/sound flags being logged are themselves fed back from the active phase patch (Figure 3.7). . 40
3.11 1phase.maxpat . . . 41
3.12 2phase.maxpat . . . 42
3.13 Block diagram of Sonic Wall architecture. . . . 43
3.14 spatializer.maxpat . . . 43

4.1 Technical Control Room. . . . 46
4.2 Main Room . . . 46
4.3 Phase 1 room layout: a single circular zone at the centre of the physical Main Room. . . . 49
4.4 Phase 2 room layout: four overlapping rooms. . . . 50

4.5 Phase 3 room layout: four overlapping rooms with soundmarks. 51

5.1 QTM tracking quality per participant. Bars show the percentage of frames requiring linear interpolation for the best-tracked available marker cluster per participant, with that cluster labelled below each bar. 55
5.2 The four room zones to scale, drawn from the `2phase.maxpat nodes` geometry ($r_x = 1.44$ m, $r_y = 1.10$ m). The shaded bands where adjacent ellipses overlap are the zones that the inclusive assignment and the `simultaneous` flag both pick out. 57

6.1 Session duration across all 13 participants, shown as a box plot with each participant's own duration labelled. 64
6.2 Session duration against total QoM, with each point labelled by participant and a linear fit (shaded band: 95% CI). 64
6.3 QoM rate (total QoM divided by session duration) for all 13 participants, sorted in descending order, with each participant's total session duration labelled below their bar. 2-A (red) uses the QTM chest-marker fallback rather than the head tracker. 65
6.4 Mean per-frame QoM aggregated across participants, by phase, with individual participant means shown as labelled points and error bars showing ± 1 SD. 66
6.5 Cumulative QoM within each phase, per participant. Bars are blank where a participant's clock or mocap data did not allow a phase breakdown (Table 5.1). . . . 67
6.6 Movement rate (QoM per second) in the simultaneous/between-rooms zone versus the rest of the space, per participant. 68
6.7 Time spent in the simultaneous zone as a percentage of each participant's total session duration, sorted descending, with each participant's total duration labelled below their bar. 69
6.8 Change in QoM from a 1-second pre-onset window to a 2-second post-onset window, across all 1,244 deduplicated sound-trigger onset events in Phase 3. 71
6.9 Average share of session time spent in each room across participants, under the exclusive, nearest-room assignment (error bars: ± 1 SD). 72
6.10 Room dwell time against cumulative QoM while in that room, one point per participant per room under the exclusive assignment (n=52), coloured by room, with a Pearson correlation of $r = 0.79$. 73
6.11 Time spent in each room as a percentage of that participant's own session duration, per participant, under the exclusive, nearest-room assignment. 74
6.12 Time spent in each room in seconds, per participant, under the exclusive, nearest-room assignment. 75
6.13 Average share of session time spent in each room across participants, under the inclusive, overlapping-ellipse assignment (error bars: ± 1 SD). Compare Figure 6.9. 76
6.14 Average time in seconds spent in each room across participants, under the exclusive, nearest-room assignment (error bars: ± 1 SD). 77
6.15 Average time in seconds spent in each room across participants, under the inclusive, overlapping-ellipse assignment (error bars: ± 1 SD). 78
6.16 Time spent in each room as a percentage of that participant's own session duration, per participant. Compare Figure 6.11. 79
6.17 Time spent in each room in seconds, per participant. Compare Figure 6.12. . . . 79

List of Tables

2.1 Comparison between lower-order and higher-order Ambisonic microphones . . . 19
2.2 Ambisonic microphones in research and practice . . . 20

3.1 Metadata for the Rooms Selected from the Intercontinental Dataset . . . 31
3.2 Explicit List of Soundmarks and Audio Sample Sources . . . 32
3.3 24-Channel Loudspeaker Array Configuration (Spat AED, Delay, and Gain Correction Values) . . . 34
3.4 Qualisys camera array configuration . . . 34

4.1 HSP Item Responses ($n = 12$; 1–7 scale) . . . 47

5.1 Data completeness per participant . . . 56

6.1 Participant Demographics ($n = 13$ questionnaire respondents) . . . 61
6.2 Participant Musical Background ($n = 13$) . . . 62
6.3 Participant nationality (self-reported, used as a proxy for cultural background) and lighting condition, cross-referenced with session duration and total QoM . . . 62
6.4 One-sentence summary of each transcribed post-exposure interview ($n = 13$ available) . . . 70

Abbreviations

ASA	Auditory Scene Analysis
AEM	Audio Energy Maps
QoM	Quantity of Motion
ITD	Interaural Time Difference
ILD	Interaural Level Difference
HRTF	Head-Related Transfer Function
VBAP	Vector-Based Amplitude Panning
HOA	High Order Ambisonics
DMI	Digital Music Instrument
ITD	Interaural Time Difference
ILD	Interaural Level Difference

Chapter 1

Introduction

In this dissertation, I investigate how humans experience and inhabit indoor spaces through an interactive 3D audio experience. Although sound is a constant presence in our daily lives, its role in indoor environments is often treated as a secondary by-product of architecture rather than a primary influence on human behaviour. My research challenges this perspective by examining the reciprocal interaction between a moving body and a responsive acoustic field. By combining motion capture and narrative, I seek to understand how the invisible acoustic infrastructure of indoor spaces shapes our movement and our perception of the world.

1.1 Motivation and Context

My motivation for this research is rooted in a life of constant enquiry into my surroundings and how music and sound affected me. Between the ages of six and nineteen, my education took place in a music conservatory, where my daily life was defined by total immersion in music. This early foundation shaped my way of perceiving the sonic world around me. A shift in my perspective occurred when I entered the faculty of engineering and started engaging with peers from very different backgrounds and passions; there, I understood how many different ways of perceiving the world exist. The constant attention that I had on my sonic surroundings wasn't the same as my colleagues', and that made me wonder how humans notice sounds, and what the difference in behaviour is when some sounds are noticed versus when they are not. When I started investigating this topic, I noticed how little attention indoor environments have received, rooms such as offices, corridors, and living rooms, in contrast to studies made on urban or natural outdoor soundscapes. Many of us spend much of our daily lives in indoor spaces, offices, schools, and our own home, yet the impact of these sonic environments on our perception, mood, and bodily behaviour remains largely unexplored. This dissertation grew out of these observations. I move beyond personal intuition to investigate the relationship between sound and human agency and perception, driven by the central question: how do the sonic characteristics of indoor spaces that surround us shape our behaviour? The opportunity to work in the AMBIENT team at RITMO, the Centre for Interdisciplinary Studies in Rhythm, Time and Motion, gave me the chance to explore this question

and to develop a multidisciplinary approach joining fields such as musicology, psychology, and technology research. RITMO is a highly multidisciplinary research group, part of the University of Oslo, and the AMBIENT project's main research objective is the study of how indoor spaces affect human behaviour, specifically, indoor non-musical rhythms. The project sets out to examine how everyday auditory and visual elements combine to generate perceptible rhythms within a space, and how those rhythms in turn shape the way people move and feel while inhabiting it (RITMO, 2021). This research was conducted within the fourMs Lab (Music, Mind, Motion, Machines), a research infrastructure of international standing dedicated to the study of human body movement and physiological response within an immersive multimedia setting (RITMO, 2021). RITMO's approach to this kind of research is characteristically interdisciplinary, bringing together qualitative traditions from the arts and humanities, such as interviews and cultural and aesthetic interpretation, with quantitative methods drawn from psychology and neuroscience, including behavioural experimentation, motion capture, and neuroimaging, and further combining these with computational approaches from informatics, such as machine learning and rapid prototyping (RITMO, 2021). This convergence of methods opens up interpretative possibilities that a single disciplinary lens could not offer, and it is this combination that I draw on throughout the experimental work presented in this dissertation.

1.2 Research questions

The project presented in this dissertation seeks to explore how the spaces we occupy shape our bodily actions and emotional well-being. The following sections outline my main research question, along with the three sub-questions that have guided the structure of this work. The core research question of this thesis is: **How do humans experience and act within indoor environments?**

This question is posed at the scale of a general research problem; the empirical work that follows addresses it through a single bespoke installation, one laboratory, and thirteen participants, and the three sub-questions below narrow it to what that design can actually license a claim about. The results in Chapter 6 are accordingly reported as exploratory evidence bearing on this question, not as a general claim about indoor environments as such, a scope discussed further in the limitations in Section 7.4.

To address this, I examine the following sub-questions, each investigated through a specific combination of kinematic and qualitative measures reported in Chapter 6:

1. **How does the body engage with indoor sonic space?** I take this question to concern the moment-to-moment coupling between a participant's body and the sonic environment surrounding it, in the sense developed in the action-perception loop literature (Section 2.5.6). Quantity of Motion, derived from the participant's tracked movement across the experience's three narrative phases and around individual sound-trigger events, is the primary quantitative measure used to address it.

2. **How do participants navigate simulated rooms with different cultural and functional characteristics?** The Intercontinental Dataset simulates rooms drawn from different countries, each with its own acoustic and cultural character, and participants are free to move between them at will.

   Understanding how the social and functional identity a participant constructs for a space affects their physical behaviour, by simulating different environments, is essential to conclude how behaviour changes in response to them. This conclusion is drawn from movement data and from interpreting the participants' own account of their movement.

3. **In what ways do interactive sonic objects alter the subjective perception of presence and place?** Presence is treated here not as binary but as graded, following the participants' own descriptions of moving in and out of a felt sense of being "in" the room. This question is addressed through the post-exposure interview's direct questions on immersion and control, read alongside the measured movement response to individual sound-trigger events, on the reasoning that how a participant's body responds to an interactive sound is itself evidence of how present that sound made the space feel.

   By exploring the role of interactivity, I can assess how a "sense of place" is constructed. I analyse how the responsiveness of virtual objects affects the participant's subjective experience, to understand whether it increases their sense of presence.

Chapter 6 returns to each of these three sub-questions directly.

1.3 Project and Methodology

To answer my research questions, I developed an interactive auditory experience that simulates a variety of indoor rooms from different geographical and cultural contexts. The experience took place in the fourMs Lab at RITMO, using a Qualisys motion-capture system and a 24-speaker setup. The auditory environment is built from Ambisonic recordings of different indoor spaces combined with sound synthesis. The system uses Qualisys motion capture for spatial tracking, which provides input data to a Max/MSP patch. The patch serves as the central engine for interaction, mapping the participant's position and velocity to specific audio parameters; based on the Qualisys data, it adjusts the auditory scene through real-time synthesis and sample manipulation. This interactive loop allowed me to study how participants navigate and perceive simulated environments. A mixed-methods approach let me join physical movement patterns with the participant's lived experience: the Qualisys system provides quantitative data on physical behaviour, while narrative interviews conducted after the experience describe their subjective perception of the space, emotional state, and sense of immersion.

1.4 Structure of the Dissertation

The structure of this dissertation is as follows:

Chapter 1 Introduction introduces the research context, establishes the motivation for the research, presents the research question and objectives, and outlines the structure of the dissertation.

Chapter 2 State of the Art presents the state of the art, reviewing the theoretical and technical literature on sound, listening, space, and embodiment that grounds the project's design, together with the artistic and research precedents that inform it most directly.

Chapter 3 System architecture describes the system built to run the experience: the Qualisys motion-capture setup, the Max/MSP patch architecture that turns tracked movement into sound, the ambisonic spatialisation pipeline, and the sonic-wall mechanism used to bound each room.

Chapter 4 User Study Protocol sets out the user study protocol: the physical space, participant screening and pre-exposure questionnaire, the three-phase experience design, and the post-exposure interview procedure.

Chapter 5 Methods for Data Analysis details the methods used to process and analyse the resulting data, from cleaning the Qualisys and behaviour-log recordings to deriving the Quantity of Motion measure used throughout the results.

Chapter 6 Results and Discussion reports the quantitative and qualitative results, discusses them against the theoretical framework of Chapter 2, and states the study's limitations.

Chapter 7 Conclusions and Future Work closes the dissertation with the future work this project opens up, from finishing the data analysis to publishing the dataset and the system itself.

1.5 Use of Artificial Intelligence Tools

In the course of this dissertation, I used AI-assisted tools to support my work.

Claude (Anthropic) was used to revise text, particularly with language and structure. Claude Code was used during the data-analysis stage to help write and debug the scripts used to process the motion-capture and QoM data reported in Chapter 6. All statistical outputs were checked against the underlying data, and the choice of tests, interpretation of results, and conclusions drawn from them are my own.

NotebookLM (Google) was used during the literature review to help organise and query the set of source documents I had collected.

These tools were used only as writing, coding, and organisational aids, not as sources of data, findings, or scientific claims.

Chapter 2

State of the Art

Introduction

To understand how humans perceive and experience the acoustical world around us, it is necessary to integrate multiple areas of research, such as cognitive science, psychology, acoustics, and music technology. Human perception is inherently multimodal and is based on the continuous interaction between the body and the environment. So, to study human perception and behaviour in indoor spaces, it is essential to have a multimodal approach to study it. This chapter introduces key concepts from embodied cognition and human perception, acoustic ecology, psychoacoustics, and sonic interaction design to establish a theoretical foundation for the project that follows, along with recent research and projects made in this area. Initially, this review explores embodied cognition to establish the body and its physical relation with the surroundings as the central frame for world understanding. This provides context for exploring space and soundscapes, where the focus shifts to the spaces that surround us and the qualitative construction of a room. The discussion continues by comparing the categorical aspect of Auditory scene analysis in contrast with Acoustemology, where the cultural background of the listeners is taken into account in the perception of the auditory environment. In the last section, the current state of spatial audio and ambisonic technology is evaluated in terms of how it facilitates the creation of immersive environments and helps research in perception and movement.

2.1 Human Perception and Embodiment Cognition

Human experience is a complex interconnection of our senses and cognition that work to better understand the world (Jensenius, 2024). In "The Senses Considered As Perceptual Systems", James J. Gibson establishes senses not as passive channels for sensory input but as active interrelated systems. The author distinguishes sensation from perception, where sensation refers to the activation of sensory receptors, whereas perception alludes to the ability of detecting properties and the structure of the world surrounding us. The core challenge of perception is explaining how an observer achieves constant perception of their surroundings in everyday life, even though the observer's

movement and the stimulus energies, such as sound, keep changing as the observer moves (Ernst and Bülthoff, 2004). Barbara Tversky further argues that action and perception are fundamentally intertwined and equally important to our understanding of the world (Tversky, 2019). This supports the view that human understanding of the world is constantly being reassessed, based on our previous experiences and existing preconceptions (Freeman, 1969). Multimodality is a crucial aspect of human perception, where the key to a better understanding of the world is to combine multiple sources of sensory information (Ernst and Bülthoff, 2004). In this view, auditory stimuli deeply affect how we perceive the world and how they make us feel (Watkins, 2018).

2.1.1 Signals

Before addressing perception and meaning, this section sets up a working vocabulary for sound as a physical and technical phenomenon, and looks at the main techniques used to generate or manipulate it.

2.1.1.1 Sound versus Audio

A foundational distinction is that between *sound* and *audio*. Sound, in its physical sense, is a travelling variation in air pressure: a mechanical wave produced by a vibrating source and propagated through a medium (Müller, 2015). Audio is a representation of that phenomenon instead: an electrical or numerical signal, analogue or digital, that stands for sound and can be stored, transmitted, and manipulated independently of any acoustic event. This distinction marks the point where a physical process becomes encoded information and is therefore the point from which synthesis, recording, and digital signal processing (DSP) become possible (Truax, 2001). Alexander Refsum Jensenius draws a related distinction between *sound design*, the practice of producing the sounding wave itself, and *sonic design*, the multimodal fusion of sound with the sensations of motion and embodiment that accompany it (Jensenius, 2023). Only once a sound wave arrives at the ear does it become the object of psychoacoustic study, the point at which physical acoustics gives way to the perceptual and cognitive processing discussed throughout the rest of this chapter (Truax, 2001). Figure 2.1 schematically traces this path, from the physical excitation of a sound-producing event, through its transduction into an audio signal, to its eventual arrival at the point of psychoacoustic interpretation.

2.1.1.2 Sound Synthesis and Sample Manipulation

The soundscapes developed for this project draw on two broad families of sound-generation techniques: synthesis, where sound is generated algorithmically, and sample manipulation, where pre-recorded material is manipulated and reassembled. Both are used in the system described in Chapter 3, and each has different implications for the fidelity, flexibility, and cultural specificity of the resulting soundscape.

Synthesis: Sound design and synthesis strategies can be categorised into two different methodological approaches: signal models and physical models (Wanderley and Depalle, 2004). Physical

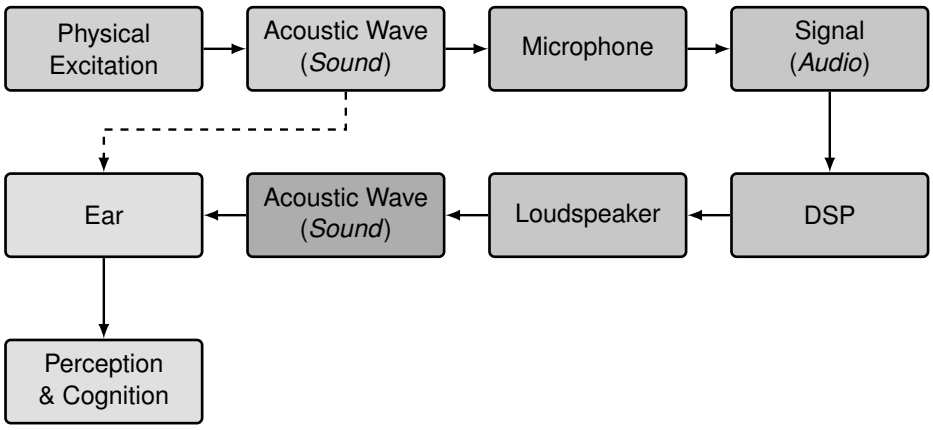


Figure 2.1: The path of a sound event, extended to include the digital mediation loop. A physical excitation produces an acoustic wave, which may either travel directly to the ear or be captured by a microphone, converted into an audio signal, processed digitally, and re-emitted by a loudspeaker as a second, physically distinct acoustic wave before reaching the ear.

modelling describes the mechanical and acoustic behaviour of a sounding instrument through equations for vibrating structures (Välimäki et al., 2006; Wanderley and Depalle, 2004). These models simulate real-world energy relationships and boundary conditions, such as digital waveguides, spring-mass models, and modal synthesis (Jensenius, 2024; Wanderley and Depalle, 2004). Because each model is built around a specific physical structure, it is structurally deterministic and difficult to interpolate between different instruments, since physical models generally lack automated parameter-analysis methods (Holbrook, 2022). Their main advantage is a highly intuitive gestural mapping: input parameters correspond directly to physical actions such as blowing pressure or bow velocity (Wanderley and Depalle, 2004). Signal modelling takes a different approach, using abstract mathematical structures to represent the spectral and temporal properties of a sound without reference to the instrument that produced it (Roads, 1996; Wanderley and Depalle, 2004). Additive synthesis, the phase vocoder, frequency modulation, and the source-filter model all belong to this type of synthesis. Except for frequency modulation, signal models benefit from well-developed analysis-by-synthesis tools, which let a designer extract parameters directly from a recorded sound and continuously morph between different timbres (Holbrook, 2022).

Sample manipulation: While synthesis constructs sound from scratch using mathematical abstraction, sample manipulation is made directly on recorded material. This was developed with Pierre Schaeffer and the founding of musique concrète, a method that starts from a "harvest of found sounds" and develops a musical macrostructure (Schaeffer, 2017). Early sample manipulation was restrained by the technology of phonographic discs. Schaeffer and his colleagues used the closed-groove technique (*sillon fermé*) to isolate short sound fragments, forcing the phonograph stylus to loop continuously and separating the sound from its original temporal context. This repetition created a musical effect on the fragment by shifting the listener's attention away from the sound's origin and toward its acoustic properties. By physically or electronically removing the initial attack of a recorded bell sound in the cut-bell experiment (*cloche coupée*), the sound's

dynamic profile changed, and the perceived timbre was radically changed, to the point where the sustaining body of the bell became indistinguishable from a flute. The post-war introduction of magnetic tape considerably expanded these techniques. Sample manipulation came to include tape splicing, tape looping, speed alteration (which transposes pitch and time together), backward playback, and multi-track superimposition. These techniques are still used today with the same name but using algorithms and digital processing (Schaeffer, 1966). This digital technology extended sample manipulation into the micro-temporal domain. Granular synthesis splits an audio file into small pieces, typically 30–100 ms long. By grouping these at high density and manipulating them, the creation of new sounds is achieved, playing with pitch and time independently to create different compositions (Roads, 2001). Barry Truax's *Riverrun* (1986) is an example application of this technique, employed to build a dense, sonic texture from short vocal phonemes (Truax, 1988). The grain density and temporal distribution simulate a fluid, riverine motion, showing how micro-level changes to digital samples can generate macroscopic sonic form.

2.1.1.3 Procedural Audio

Procedural audio is synthetic sound created in real time from a set of algorithms and live data input (Menexopoulos et al., 2023; Müller, 2015). Procedural audio synthesises sound from scratch in real time; on the other hand, procedural *sound design* maps pre-recorded samples onto dynamic variables rather than generating sound algorithmically. For a sound model to be truly procedural, it must expose meaningful, real-time physical or perceptual parameters, such as walking speed rather than abstract mathematical terms (Menexopoulos et al., 2023). There are three main contributors on the scene of procedural audio: Perry Cook, working on physics-based models of rigid-body sound, such as strings, bars, and tubes (Cook, 1996); Davide Rocchesso, on the physical principles behind everyday interactive "sounding objects" (Rocchesso et al., 2003); and Andy Farnell, who coined the term itself and wrote the field's foundational handbook, *Designing Sound* (Farnell, 2010). Procedural audio has been used for multiple purposes, from game sound design to rehabilitation and sports science. Systems such as *SONI-Gait* sonicating gait in real time from instrumented insoles to support motor coordination Horsak et al. (2016), and accelerometer-driven systems such as *Low-Impact Runner* overlay dynamically adjusted noise onto a runner's music to discourage high-impact footfall Van den Berghe et al. (2019). Assistive-technology applications map the position of tracked obstacles to spatial audio parameters, for example by panning to horizontal position and pitch to height, to support navigation for visually impaired users Ji et al. (2018), and interactive sound installations such as Rawlinson's *w[i]nd* use procedural synthesis to generate environmentally responsive soundscapes in real time Rawlinson (2021). This work establishes procedural audio as an approach to connect continuous physical parameters, including bodily movement, to continuously generated sound, the same design rationale behind the real-time sonic-wall synthesis described in Chapter 3.

2.1.2 Listening Modes and Frameworks

This section looks at how a sound is listened to, rather than what its physical signal contains. This is done through two perceptual frameworks: the classical listening-mode typologies developed by Schaeffer and Gaver, and the literature on how the auditory channel interacts with other sensory modes. The way a listener listens to sound is not fixed, but changes according to what the listener is trying to understand from it. Pierre Schaeffer, in his *Traité des objets musicaux* and later in the *Solfège de l'objet sonore*, introduced three main listening modes: *causal listening* (*écouter*), directed toward identifying the source or cause of a sound; Semantic *listening* (*comprendre*), directed towards the meaning or the sound, as in speech; and *reduced listening* (*écoute réduite*), directed toward the perceptual qualities of the sound itself, deliberately bracketing its source or meaning (Schaeffer, 1966; Chion, 2009). William Gaver's distinction between *everyday listening*, hearing sound as information about events and their sources in the world, and *musical listening*, attending to the perceptual attributes of the sound itself, such as pitch or timbre, parallels Schaeffer's causal and reduced listening. Gaver's approach draws directly on Gibson's ecological theory of perception, (Gaver, 1993), which holds that perceptual systems are attuned not to raw sensory stimulation, but to invariant structures in the ambient energy that specify meaningful properties of the environment directly, without the need for internal reasoning and reconstruction (Gibson, 1979). Applied to hearing, this view treats sound not as an abstract acoustic signal to be decoded, but as a direct carrier of information about events, materials, spatial layout, and the general understanding of the world. Figure 2.2 sets Schaeffer's typology and Gaver's dichotomy side by side.

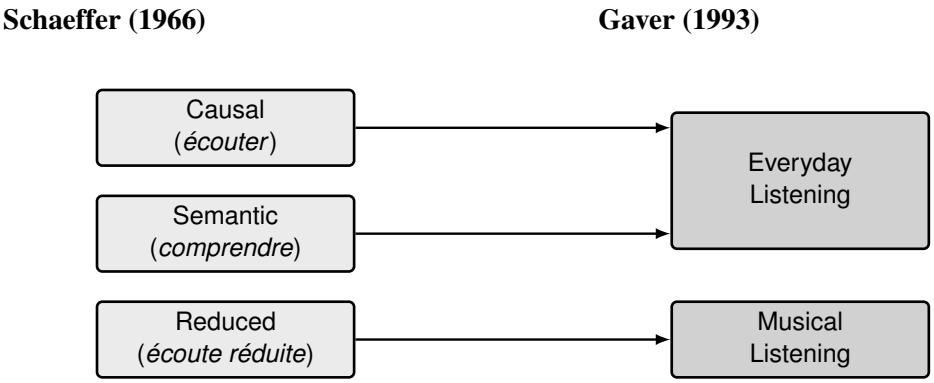


Figure 2.2: Correspondence between Schaeffer's three modes of listening and Gaver's everyday/musical distinction.

2.1.3 Embodied Cognition

Embodied cognition is the view that cognitive processes cannot be fully understood by looking at the brain in isolation, because they are constituted in part by the body's ongoing interaction with its environment. Francisco Varela, Evan Thompson, and Eleanor Rosch set out the foundational statement of this position, arguing that cognition is *enactive*: it does not consist in the recovery of

a pre-given world by an internal representational system, but arises through a history of structural coupling between an active organism and its environment, so that perceiving and acting are two aspects of the same process rather than separate stages joined afterwards (Varela et al., 1991). Marc Leman extends this enactive premise specifically to music and sound, arguing that a listener's engagement with sound is mediated by the body's own action-oriented representations, so that musical meaning is constructed through corporeal simulation of the sound-producing gesture rather than decoded from the acoustic signal alone (Leman, 2008). Rolf Inge Godøy narrows this further to the level of individual gesture, showing that listeners spontaneously and covertly simulate the motor gesture that could have produced a given sound, a mechanism discussed at length in Section 2.3.4 below (Godøy, 2006).

This enactive premise, that the body and its perceptual and motor capabilities are the basis of how we interact with our surroundings, has since been taken up and applied specifically to sound design and spatial audio by researchers affiliated with RITMO. Ana Tajadura-Jiménez frames embodied cognition as the view that cognitive processes can only be fully understood when considering the interaction of the body with space, the opposite of a traditional picture in which the body and mind are two disconnected identities and the brain acts as a passive receptor (Tajadura-Jiménez, 2008; Jensenius, 2017). As Shaun Gallagher puts it, "the body shapes the mind": the body is not solely a physical object but a site of experiencing the world, fundamentally shaping the way we think and understand our surroundings (Gallagher, 2005). As Thomas Görne explains, in order to mentally grasp immaterial phenomena such as sound, humans often project them into space to attribute meaning to them (Görne et al., 2021). This spatial projection allows the mind to interpret a signal, for example, a sound, through concrete physical objects, making it easier to understand by classifying and measuring it. This mechanism implies that perception and action are inseparable and that knowledge is deeply embodied and inseparable from the sensory–motor system (Gallese and Lakoff, 2005; Jensenius, 2024).

The action-perception loop, treated in full in Section 2.5.6, is one of the core elements of embodied cognition put this way. It emerges from the active interaction with the environment (Riaz, 2024) and is characterised by a continuous flow of re-afferents between the subject and the environment, where perception guides action, and action, in turn, shapes perception (Riaz, 2024; Freeman, 1969). Guo et al. (2025) studied the relationship between spatial-temporal auditory stimuli and human micro-motion when standing still and showed that even in stillness the body has an exchange with the environment. The study has shown a relation between involuntary movement and the directionality of environmental sounds, revealing that we are in a "continuous action-perception loop with our surroundings" (Guo et al., 2025).

Existing in a spatial space is inseparable from the body (Riaz, 2024), and in *embodiment spatialization*, the body acts as the central frame of the relation between the observer and the environment (Tajadura-Jiménez, 2008), reflecting on how personal each perspective and experience with the world are. Embodied spatialization views space as an acoustic field that emerges from the continuous interaction between the body and the surroundings (Tajadura-Jiménez, 2008; Riaz, 2024).

2.2 Space

The concept of space is fundamental to human experience and perception, serving as a physical container as well as an emotional and dynamic environment (Lefebvre, 1991). Henri Lefebvre creates a clear distinction between space and place (Lefebvre, 1991); space can be understood as a three-dimensional, boundless site where events occur, linked to lived experience. In contrast, a place is a physical space that is grounded and where social activities occur. A room encompasses physical dimensions (length, width, height), but inside this room there lives a social dimension (how people interact within it) and a psychological dimension (how it feels) that characterise the space within it. The combination of these characteristics defines our experience in space. Architectural aspects in indoor environments have been shown to have psychological and physiological effects on humans. Elements such as light, colour, smell, and sound have a great impact on human experience (St-Jean et al., 2022). The main focus of this thesis lies on how sound impacts human experience in these spaces, specifically inside indoor spaces.Görne et al. (2021) puts it as: "Space is experienced through movement and interaction with the environment"

2.2.1 Soundscapes

The concept of soundscapes is a framework for understanding how the environment is experienced through sound. Soundscape is a widely used term (Mitchell et al., 2021); in acoustics research, the term is widely used as the acoustic environment perceived by people in a context (Hasegawa and Lau, 2021). According to ISO 12913-1, a soundscape is a "perceptually constructed" environment based on a "physical phenomenon", where the physical phenomenon is called the acoustic environment (iso, 2014). This definition is based on the work of R. Murray Schafer, which reinforces the importance of subjectivity as a central aspect of a soundscape (Schafer, 1993). The context of the person's experiences and sociocultural background is essential to determine their experience and perception of a soundscape (Jensenius, 2024; Varoutsos and D'Arcy, 2021). Visi et al. (2024) make a comparison between everyday listening and music listening, stating that the difference between the two is based on experience and not on the characteristics of the sound (Visi et al., 2024); this can be extrapolated to the difference in how the perception of the sound surrounding us changes based on our experiences and predisposition. Early studies in this field mainly concentrated on outdoor soundscapes, but in the last decade the importance of indoor soundscapes and sound environment design of these spaces grew significantly (Kang, 2006). Outdoor soundscape studies focus specifically on features such as geophonic, bio-phonic and anthropogenic sounds, while indoor soundscape studies focus primarily on more controlled sources, such as ventilation systems, machine noise and reverberant effects of the room (Riaz, 2024). Some artistic approaches to soundscapes focus on the relationship between sound and space, with the aim of turning the listener into an active participant (Carson and Gordon, 2024). Hildegard Westerkamp transforms urban and natural sound recordings into auditory narratives, examining how their manipulation can alter the emotional perception of a space (Riaz, 2024). In *Soundscape Speaks*, one of her most recent works, she draws on 40 years of field recordings to explore the relationship between

environmental sound and human presence, weaving these recordings together with her own voice. The piece encourages listeners to reassess their relationship with the environment through reflective listening (Westerkamp, 2021). Adopting a more interactive strategy, *Resonant Landscapes*, developed by Tate Carson and Carter Gordon, is a web application that employs ambisonic audio and GPS technology to layer field recordings from state parks onto an urban university campus. This allows users to experience immersive environmental soundscapes through their smartphones and headphones when placed in a location with different sound features. Spatial audio makes a difference here by creating immersion in nature sounds while physically remaining in an urban setting. The project examines the connection between soundscape studies and ecological awareness, bringing into discussion the lack of green spaces in cities (Carson and Gordon, 2024).

2.2.1.1 Soundscape taxonomy

R. Murray Schafer's soundscape taxonomy divides sounds into three main categories: keynotes (persistent background sounds like traffic hum or ventilation), sound signals (foreground sounds consciously listened to, such as bells or sirens), and soundmarks (unique community sounds that define a place, like local bell melodies) (Schafer, 1993).

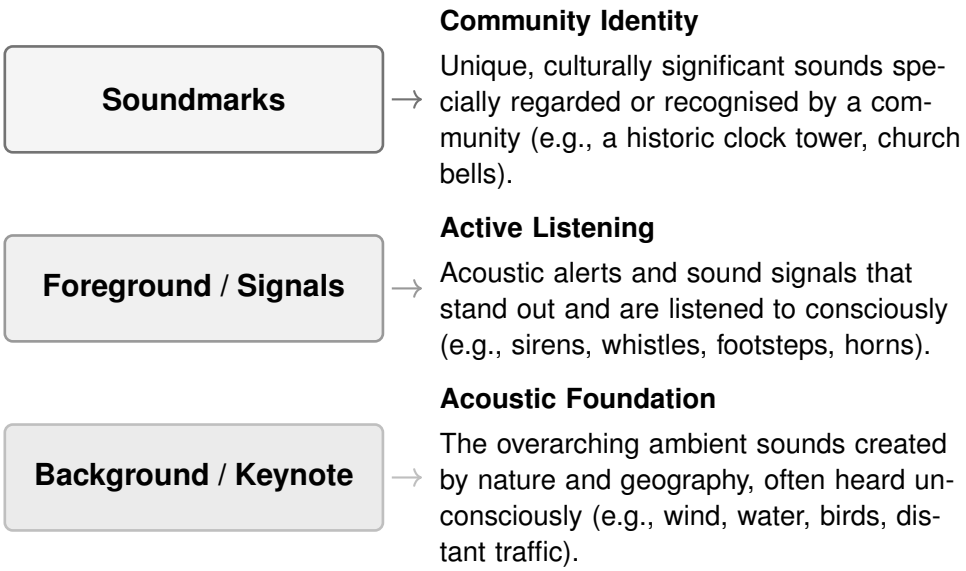


Figure 2.3: R. Murray Schafer's Soundscape Taxonomy

2.3 Affective Spaces

The previous section looked at how soundscapes are structured and categorised. This one looks at how they are experienced affectively: how an acoustic environment feels pleasant, tense, calm, or threatening, and how the body registers sound as occupying space around it.

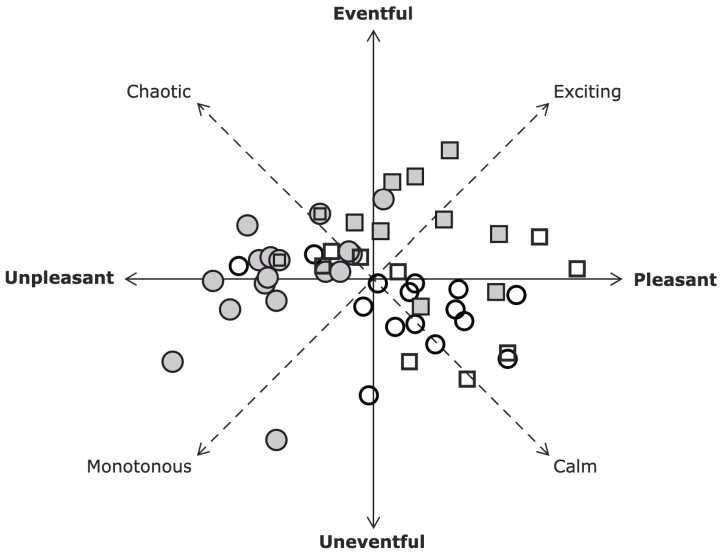


Figure 2.4: Principal components model of soundscape perception, showing orthogonal dimensions of pleasantness and eventfulness alongside diagonal axes (Axelsson et al., 2010)

2.3.1 Perceived Affective Quality of Soundscapes

Beyond identifying what a sound is, listeners routinely evaluate how a soundscape as a whole makes them feel, even if unconsciously. The most widely adopted model is the circumplex model proposed by Axelsson, Nilsson, and Berglund; it was developed from a listening experiment in which 100 participants rated 50 excerpts of urban soundscape recordings across 116 descriptive attribute scales (Axelsson et al., 2010). This model maps human perception into two dimensions: *Pleasantness* and *Eventfulness*, which together account for most of the variance in how a soundscape is perceived (Axelsson et al., 2010). Soundscapes dominated by technological sound tend to be rated unpleasant; those dominated by natural sound, pleasant; and those dominated by human activity, eventful (Axelsson et al., 2010). This two-dimensional space has since become the basis of the ISO 12913 soundscape standard, and is widely used to classify acoustic environments into quadrants such as *calm*, *exciting*, *monotonous*, and *chaotic* (ISO 12913-2:2018; Aletta et al., 2016). The perceived quality of a soundscape is not a linear product of sound pressure level; it's dynamic and influenced by personal, situational and multisensory factors (Tajadura-Jiménez, 2008). Factors like pre-exposure mood heavily bias the subjective experience of an individual to the soundscape. Soundscape understanding is not purely auditory; research shows that visual context also affects how a given sound is perceived, as the presence of greenery, water, or other visual nature elements has been found to reduce reported annoyance and increase perceived pleasantness for an acoustically similar environment (Van Renterghem et al., 2016; De Coensel et al., 2011). Interestingly, this effect is also seen with different thermal conditions: when a room is warm, adding sounds of water in order to mask traffic noise improves pleasantness and calmness, but when the room is cold, the water sounds can enhance annoyance (Torresin et al., 2019).

2.3.2 Affective Spaces and Embodied Psychoacoustics

A big part of research looks not at how a soundscape is judged, but at how the body registers sound as occupying and moving through the space around it. Neuhoff showed that listeners systematically overestimate how close an approaching sound has come, relative to an equivalent receding sound, a bias that fits an evolved warning function in which approach simply matters more than retreat (Neuhoff, 1998). Neuroimaging has since backed this up: looming sounds produce stronger activation than receding ones in auditory and motor-preparatory brain regions (Seifritz et al., 2002; Bach et al., 2009).

Research on peripersonal space adds more: the space immediately around the body is represented differently from space further away, with multisensory neurons coding stimuli, sound included, in a body-part-centred frame that expands or contracts with movement and tool use (Graziano and Cooke, 2006; Serino, 2019). Listeners can also judge how large a room is just from its reverberation: by hearing and understanding its impulse response, Traer and McDermott showed that people, without noticing, understand the size of a room based on the relation between the natural reverberation and the pure sound from the source (Traer and McDermott, 2016). Both findings back up the basic premise this project's design rests on: that an indoor auditory environment can carry a physical room's spatial and behavioural affordances even with no visual information at all.

2.3.3 Affordances

The concept of the affordance, introduced by James J. Gibson, tells what an environment or object offers an organism in terms of possibilities for action, a property that exists relationally, together determined by the physical characteristics of the object and the perceptual-motor capacities of the perceiver (Gibson, 1977, 1979). Crucially, the perception of an affordance depends on the perceiver's experience of manipulating similar objects: a given surface or object does not present a fixed, universal set of action possibilities, but a set calibrated to what a particular organism has learned it can do with things of that kind (Gibson, 1977). Applied to the sonic domain, a *sonic affordance* can be understood as the set of possible sonic outcomes made available by the physical properties of an object and the actions that can be performed on it, the sonic qualities an object affords, given the ways it can be struck, scraped, shaken, or otherwise excited (Jensenius, 2023). Withagen, de Poel, Araújo, and Pepping extend Gibson's original, relatively passive formulation by arguing that affordances are not merely possibilities waiting to be recognised: they can actively invite exploration and behaviour, soliciting action rather than simply permitting it. They propose a mutualist model in which the strength of such invitations depends on the multidimensional relationship between the animal and its environment, rather than on the object's static physical properties alone (Withagen et al., 2012).

2.3.4 The Schaefferian Sound Object

This subsection returns to the enactive premise set out in Section 2.1.3, that perceiving a sound is inseparable from the bodily gesture that could have produced it, and gives it a concrete perceptual unit to work with: the sound object. Schaeffer brings the notion of *sound object* (*objet sonore*): a fragment of sound perceived, under reduced listening, as a single coherent perceptual unit, independent of its source or cause (Schaeffer, 1966). His typology and morphology of sound objects classify sound objects first by type, distinguishing broadly between impulsive sounds, a single, sharp excitation, such as a strike; sustained sounds, continuous excitation, such as a bowed or blown tone; and iterative sounds, a rapid succession of impulses perceived as a single grainy or granular texture. Each object is further characterised along a set of morphological criteria, including mass (the sound's pitched or noise-like quality), dynamic (its intensity over time), grain (its fine internal texture), and gait (e.g. vibrato or tremolo) (Schaeffer, 1966).

This has been substantially extended by Rolf Inge Godøy's notion of the gestural-sonorous object, which reframes Schaeffer's sound object as bound up with the bodily gesture that produces or could produce it, arguing that listeners spontaneously and covertly simulate the motor gesture implied by a sound object even when no visible performer is present (Godøy, 2006), the same simulation mechanism introduced in Section 2.1.3. Godøy further distinguishes the *sound-motion object* as a unit of perception in which the temporal shape of a sound and the physical trajectory of the gesture that could produce it are experienced as fused (Godøy, 2006).

2.4 Listening

In this section, it is important to distinguish Listening from Hearing. Although hearing refers to the physiological process by which the auditory system registers sound, listening is an active mechanism that requires attention and cognitive resources to seek meaning and interpretation (Freeman, 1969; Jensenius, 2024). As Barry Truax puts it: "Both in theory and practice, listening is the crucial interface between the individual and an environment. It is also a set of sophisticated skills which appear to be deteriorating within the technologized urban environment, both because of noise exposure which causes hearing loss and physiological stress, and because of the proliferation of low information, highly redundant, and basically uninteresting sounds which do not encourage sensitive listening" (Truax, 2001, p. 35).

2.4.1 Psychoacoustics

Psychoacoustics studies the relationship between the physical properties of a sound wave, its frequency, amplitude, and spectral and temporal structure, and the subjective, perceptual response that wave produces in a listener (Truax, 2001). Where acoustics describes what a sound wave objectively is, psychoacoustics describes what a listener hears it as: two physically distinct signals can be perceived as identical, and two physically identical signals can be perceived very differently depending on context, prior exposure, and the listener's own physiology. The three subsections

that follow look at this listener-side construction of the acoustic world at three different scales: how a continuous acoustic field is segmented into discrete, trackable objects (Auditory Scene Analysis), how that segmentation is shaped by a listener's cultural and experiential history rather than by acoustics alone (Acoustemology), and how the auditory system locates a segmented sound object in three-dimensional space (Spatial Audio).

2.4.2 Auditory Scene Analysis

Auditory Scene Analysis (ASA) is central to research in cognitive psychology, music cognition, soundscape studies, and spatial audio, creating a theoretical basis for understanding how listeners experience sound environments. Motor theories of perception argue that, to make sense of the sounds registered by our auditory system, we must understand their origin, whether it is an action or a source (Haugen, 2024) and that this understanding is crucial to constructing a coherent representation of the world Bregman (1990). Having the ability to discriminate complex sound events in a soundscape, known as auditory streams, is what Bregman calls "Auditory Scene Analysis" (Bregman, 1990). ASA is a process by which the human auditory system organises sound into distinct mental organisations called streams. Streams serve as centres for description, allowing the brain to then cluster different properties depending on frequency, temporal cues, spatial cues, and timbre (Bregman, 1990). One example of a classification system is Foreground vs Background, which is used mostly in urban noise control and architectural acoustics. In indoor soundscapes, foreground sounds are the ones that capture individual attention, such as a ringing phone or a sudden loud noise; background sounds are the ones that form a continuous acoustic context, such as ventilation noise or refrigerator sound (Riaz, 2024). This distinction raises methodological challenges in how to effectively record and analyse complex soundscapes. Existing research on indoor soundscapes remains limited, with most studies focusing on outdoor environments or controlled laboratory conditions (Schafer, 1993; Kang, 2006). In contrast, applied audio fields such as music production and game audio systematically make this distinction by using ambient microphones to capture background spatial information, while close microphones isolate foreground sound sources (Riaz et al., 2025a).

2.4.3 Acoustemology

Acoustemology is a term defined by Steven Feld in the early 1990s that joins acoustics and epistemology to define sound as a way of knowing the world (Feld, 2025), and presents an ecological approach to ASA, shifting the focus from abstract "sounds" to "events". Although ASA distinguishes foreground and background sounds as perceptual categories, acoustemology sees these layers as relational structures that help produce place and meaning. Acoustemology originates from ethnographic research in the Bosavi rainforest in Papua New Guinea, and its core methodology treats listening as an embodied place-based practice defined as "listening to histories of listening" (Feld, 2025). The sensory engagement of a person with an environment is a product

of their history of listening (Feld, 2025), and from this point of view, the perception of a soundscape is shaped by the person's individual experiences, context, and socio-cultural background (Varoutsos and D'Arcy, 2021; Beadling and Vickers, 2023). This perspective shift defines sound as a mode of cognition and relational agency, where communities cultivate "mastery of locality" through the daily presence in their surroundings (Feld, 2025). In Feld's premises, listening is the primary mechanism that makes a community, existing as the linking point between nature and human, human and non-human, and past and present, where the human is not the dominant part of the environment but the perspectival centre, similar to the view of the body as a central frame of perception in Tajadura-Jiménez (2008).

2.4.4 Spatial Audio and Its Psychoacoustics

Human experience is inherently three-dimensional, and sound strongly shapes how people perceive and understand their surroundings (Rumsey, 2001). Although visual perception often dominates judgments about size and shape, auditory cues still provide important complementary information (Rumsey, 2001). Spatial audio, a set of technologies for creating an auditory environment similar to the one experienced in real life, can be used to build immersive soundscapes (Guo et al., 2025). It lets listeners perceive sound in three dimensions: the spatial relation between listener and source, and the motion of the source itself (Riaz et al., 2025a). This technology enhances immersion in settings such as augmented reality, virtual reality, and any context where a heightened sense of presence is desirable (Guo et al., 2025).

The perception this technology exploits is itself grounded in a small set of psychoacoustic mechanisms, reviewed here before turning to Ambisonics, the specific spatial-audio technology used in this project. The perception of a sound's location in space is based on binaural and monaural cues. Interaural Time Difference (ITD) and Interaural Level Difference (ILD) together form the structure of sound localisation, first developed by John William Strutt, and describe how a sound arriving from an off-centre azimuth reaches the two ears at slightly different times and intensities, with ITD dominant for low-frequency sounds and ILD dominant for high-frequency sounds, owing to the acoustic shadowing effect of the head (Rayleigh, 1907). Head-Related Transfer Functions (HRTFs) capture the direction- and frequency-dependent filtering imposed by the outer ear, head, and torso, and are essential to resolve front-back and elevation ambiguities that ITD and ILD alone cannot (Wightman and Kistler, 1989). Summing localisation describes the perceptual fusion of multiple loudspeakers into a single phantom source when their signals arrive at the ear in close temporal proximity, the psychoacoustic principle underlying amplitude-panning techniques such as the VBAP and HOA decoding used in the system described in Chapter 3 (Blauert, 1997). The precedence effect, or law of the first wavefront, describes how, when a sound and one or more delayed copies of it, as early room reflections, arrive within a short time window, the auditory system suppresses the localisation cues of the later-arriving copies and attributes the sound's location almost entirely to the first-arriving wavefront (Wallach et al., 1949). This suppression is not

absolute, but strongly dependent on context, including the delay interval, the similarity of the competing signals, and prior exposure, and can break down or be overridden under certain conditions (Litovsky et al., 1999).

2.4.4.1 Ambisonics

Ambisonic audio technology has been proven to create soundscapes that accurately reproduce the perceptual dimensions of recorded sound environments (Carson and Gordon, 2024). This audio technology covers microphone design as well as audio encoding and playback processes to capture an immersive 360-degree sound field (Riaz et al., 2025a; Gerzon, 1983). Unlike traditional channel-based systems (5.1 or 7.1), which are designed for a fixed speaker layout, ambisonics is a scene-based audio format (Guo et al., 2025; Riaz, 2024). This means that it captures the entire sound field at a single point in space, representing the auditory scene that can be later decoded for any number of speakers or even for binaural rendering over headphones (Riaz et al., 2025a). Ambisonic surround sound design has been explored in research, in conjunction with movement, creating immersive three-dimensional auditory spaces with which the user can inhabit and interact (Wijnans, 2010). This audio technology is used because it captures a scene-based sound field, which allows high directional precision and dynamic manipulation of the audio in real-time, which is essential for body and sound synchronisation (Guo et al., 2025; Wijnans, 2010). In *ChoreSonic Environments*, Wijnans creates performance spaces where dance movements are transformed into real-time 3D spatial sound compositions, creating an experience where improvised dance meets interactive performance environments (Wijnans, 2010). The author introduces the term *TranSonic Perception* to describe the feeling of manipulating the sound with the body position and movement. The sound becomes a "second skin" for the dancer, going beyond body limitations. In this case, spatialised sound increases emotional perception and realism, creating an intimate feedback loop where the sound environment is fully controlled by the movement of the dancer (Wijnans, 2010). In the study of human micromotion and human behaviour by the *RITMO* research team, Ambisonic technologies are used to analyse spatial-temporal stimuli of the indoor environment. Recent findings indicate that human postural sway patterns are not random; they show a potential spatial relation with the directionality of environmental sounds. The researchers calculated the directional audio energy maps to pinpoint the precise location of sound sources in a 360-degree field and cross-referenced that information with swaying patterns recorded by an Android-based smartphone attached to a string around the neck (Guo et al., 2025). In this research, Ambisonics was used due to its ability to create a full-sphere representation of sound, facilitating post-processing work and analysis (Guo et al., 2025). Although ambisonic research is not a new topic in music technology, it is not so often integrated with motion capture systems. To do so, analysing the spatial accuracy of these recordings often requires a cross-modal comparison with motion capture. Some techniques used for this analysis include:

1. **Anglegrams:** a 2D graphical tool that helps visualise the azimuth and elevation changes of a dominant sound source over time (Riaz et al., 2025a).

2. **Audio Energy Maps (AEM):** maps that provide a quantitative visualisation of the energy distribution across a captured space. By decoding Ambisonics recordings onto a spherical grid, energy maps can be overlaid on equirectangular 360-degree video to identify how sounds interact with the physical environment (Riaz, 2024).

3. **Quantity of Motion (QoM):** A metric used to describe the total energy of a moving body (Jensenius, 2024). It is used to analyse the physical response of listeners to music (Guo et al., 2025) or to establish a baseline for human micromotion during stillness (Jensenius et al., 2017).

Ambisonics is a surround sound format that represents audio from all directions, allowing full-sphere spatial capture. There are different devices to record ambisonic audio that range from high-end laboratory systems to accessible consumer-grade devices. One example is the Zoom H3-VR, an entry-level first-order microphone valued for its portability and how easy it is to use in spontaneous field trips (Riaz, 2024). Higher-order microphones, such as ZYLIA and Eigenmike, are required when researchers need to do more advanced spatial analysis, allowing more accurate localisation and sharper source images. The type of Ambisonic microphone used in a project depends on the level of spatial resolution and directional precision required. The order of the microphone (N) has a $(N+1)^2$ relation with the number of audio channels and, consequently, the size of the generated files. The following table compares low-order and high-order Ambisonic microphones. The following table compares different microphones and their use.

Table 2.1: Comparison between lower-order and higher-order Ambisonic microphones

	Lower Order (1st)	**Higher Order (3rd+)**
Device	Zoom H3-VR	ZYLIA ZM-1, Eigenmike
Channel Count	4 channels	16+ channels
Pros	High portability and low cost (Riaz et al., 2025a)	High analytical and spatial resolution, wide and expanded listening area (Riaz, 2024)
Cons	Spatial blurring, high-frequency aliasing and a small and limited listening area (Riaz, 2024; Wijnans, 2010)	Higher price and large file size (Carson and Gordon, 2024)

Table 2.2: Ambisonic microphones in research and practice

Microphone	Study/Paper	How It Was Used	Why It Was Used
Zoom H3-VR (First-order)	Where to Put That Microphone? (Riaz et al., 2025a)	Arranged in two configurations: a central setup and a four-microphone corner setup to track a moving audio source.	To evaluate the impact of microphone placement on sound localisation accuracy in indoor spaces.
Zoom H3-VR (First-order)	Comparing Spatial Audio Recordings from 360-degree Video Cameras (Guo et al., 2020)	Mounted in the centre of the room, 1 meter away from the floor, to record impulse sounds and continuous sounds.	Used as an Ambisonics benchmark to compare against built-in surround modes of consumer action cameras, due to being a budget-friendly option
Zoom H3-VR (First-order)	Cross-Modal Analysis of Spatial-Temporal Auditory Stimuli (Guo et al., 2025)	Mounted on a tripod at chest level next to a 360-degree camera to document the indoor acoustic environment.	To provide high-quality standardised audio for analysing the relationship between sound directionality and human postural sway.
ZYLIA ZM-1 (Third-order)	Where is That Bird? (Riaz et al., 2025b)	Positioned in the centre of an office environment to capture indoor soundscapes and artificial birdsong.	To provide high spatial resolution and directional precision for producing Audio Energy Maps (AEM).
Core Audio OctoMic (Second-order)	Resonant Landscapes (Carson and Gordon, 2024)	Used to capture 8-channel field recordings of nature soundscapes in state parks.	To preserve immersive 360-degree environments for a locative media web application with body-oriented tracking.
Eigenmike (Sixth-order)	Immersive Audio Concepts (Zotter and Frank, 2012)	Cited as a professional benchmark for high-fidelity spatial capture up to sixth order.	To increase spatial resolution, sharpen source imaging, and expand the listening area.

2.5 Interactivity

This section looks at how agency is distributed between a system and its user, how bodily movement becomes a channel of sonic control, and how these principles have shaped unconventional, movement-driven musical instruments and installations. That question of agency, who or what is in control of a given sound event, drives through every subsection that follows, down to the design decisions behind the sonic-wall mechanism in Chapter 3.

2.5.1 Action–Sound Couplings

An action-sound coupling describes the relationship between the biomechanics of a sound-producing action and the acoustic properties of the resulting sound: the same object, acted upon in the same way, reliably produces the same sound, and this regularity is what allows a listener to predict, from long experience with the physical world, the likely sonic outcome of an observed or imagined interaction with an object (Jensenius, 2023). From an ecological perspective, this predictive capacity is essential to everyday interaction with objects and environments; identifying and characterising a sound in terms of the action and object that produced it is a core function of everyday listening, closely related to Gaver's account of sound-source perception and to Warren and Verbrugge's classic demonstration that listeners can reliably distinguish the higher-order temporal structure of a bouncing object's sound from that of a breaking one (Warren, 1984). Godøy distinguishes two forms of mental imaging involved in the perception of musical and sound-producing action: motor images, corresponding to the imagined bodily movement that could produce a given sound, and material images, corresponding to the imagined physical object or material being acted upon (Godøy, 2006).

2.5.2 Agency and Shared Control

The design of digital musical instruments (DMI) requires understanding the boundary between human and machine agency. In acoustic instruments, the physical coupling between gesture and sound is fixed by the instrument's construction; in a DMI, the gesture-to-sound mapping is arbitrary and must be explicitly designed, opening a much larger, and potentially disorienting, design space (Miranda and Wanderley, 2006). Marc Leman's theory of embodied music cognition places this negotiation within a broader account of how humans relate to sounding environments, drawing an explicit parallel between the way a musical instrument or interactive system responds to a performer's gesture and the earliest forms of human communicative coordination (Leman, 2008; Malloch and Trevarthen, 2009). Leman uses this parallel to argue that musical interaction with technology can be understood as a technologically mediated extension of an innate social and corporeal capacity for entrainment and mutual responsiveness (Leman, 2008). Leman's framework also covers biofeedback, systems that sonify a performer's own physiological signals back to them, and self-augmented interaction, where a system amplifies or extends a performer's existing expressive capacities rather than replacing them with an arbitrary new mapping (Leman, 2008).

2.5.3 Sound Actions and Movement Analysis

Building on the action-sound coupling framework, Jensenius developed a broader *sound actions* model for analysing and designing interactive sonic systems, distinguishing between intentional sound-producing actions, deliberately performed to generate a specific sonic result, as in playing an instrument, and incidental sound and movement feedback, which arises as an unintended by-product of an action performed for some other purpose (Jensenius, 2017). This distinction matters

for Sonic Inhabitation specifically: its participants are not instructed to produce sound intentionally, but instead encounter sound as an incidental, continuously present consequence of ordinary movement through space. After understanding the system, this can change, and sound can become intentional.

2.5.4 Body-driven Musical Interfaces

A recurring theme in the design of unconventional, body-driven musical interfaces is the tension between openness and guidance: too many simultaneous possibilities can overwhelm a new user, while too few foreclose expressive depth. Magnusson argues that electro-acoustic instruments benefit from deliberately constraint-based design, since an unconstrained instrument's near-infinite possibility space is disorienting rather than liberating for the performer (Magnusson, 2009). In recent work, Magnusson proposes that a well-designed musical instrument should "teach, adapt, explain, direct, suggest, entice" its user toward productive interaction, treating the instrument as an active pedagogical agent rather than a neutral tool (Magnusson, 2018). This resonates with Maeda's design principle that simplicity is achieved "by subtracting the obvious, and adding the meaningful" (Maeda, 2006), and in music-technology design, Wessel and Wright's criterion that a computer-based instrument should offer a "low entry fee with no ceiling on virtuosity", accessible enough for a novice to produce a pleasing result immediately, yet in-depth enough to reward years of continued practice (Wessel and Wright, 2002).

Jensenius suggests giving the controller finer spatio-temporal resolution, so it can register and respond to more subtle gestures. A larger possibility space also means a heavier cognitive load for the performer, which brings back the very tension Magnusson's constraint-based approach was meant to resolve. In the unconventional-instrument literature, this tension is sometimes framed as a question of what the interface actually is: a *sound maker*, built for timbral and textural exploration, or a *music maker*, built around the more constrained pitch- and rhythm-organised structures conventionally associated with making music (Jensenius, 2023).

2.5.5 Performing in the "Air"

A different kind of unconventional instrument drops the physical sounding object altogether and maps free bodily gesture, moving through open air, straight to sound. Godøy calls this *playing air instruments*: the spontaneous, unprompted way people mimic playing an instrument that is not actually there. Researchers have used it to study how listeners translate music they hear into imagined motor gesture. Godøy, Haga, and Jensenius found that trained and untrained participants alike produce structured, recognisable air-instrument gestures when asked to mimic a piece of music (Godøy et al., 2006). The literature groups these into three broad categories. Touchless instruments are controlled with no physical contact at all: the theremin, played through the performer's proximity to two antennas, is the classic example. Object-based air instruments instead have the performer hold a motion-sensing device, as with the SoundSaber synthesis architecture

that this dissertation's own sonic-wall mechanism draws on (Chapter 3), where the object's trajectory through space is mapped to sound. Muscle-based air instruments map electromyographic (EMG) signals from an armband such as the Myo sensor directly from the performer's muscles to synthesis parameters (Donnarumma, 2011; Erdem et al., 2020). This line of research at the University of Oslo keeps coming back to the same question: how bodily movement and musical sound relate to each other. Testing these instruments has turned up one especially interesting, recurring finding: formally trained musicians tend to do well, as expected, but the most surprising results come from participants who consider themselves "tone-deaf" and still perform capably on these movement-based instruments. It suggests that people are, on the whole, considerably more musical than they think they are (Godøy et al., 2006).

2.5.6 The Action-Perception Loop

Underlying all of the interaction framework reviewed in this section is the concept of the action-perception loop: a continuous, closed cycle in which a participant's ongoing micro-movements, including postural sway, balance adjustments, and orienting responses, both shape and are shaped by the responsive sonic environment surrounding them, such that perception and action cannot be meaningfully separated into distinct, sequential stages (Jensenius, 2023). Micromotion, the small, often subconscious postural corrections a person makes even while intending to stand still, has itself been shown to correlate systematically with the properties of an ambient soundscape, including the spatial directionality of environmental sound sources (Guo et al., 2025).

2.6 Related Works

Several artistic precedents inform the interaction design of this project developed in this dissertation.

Alvin Lucier's *I Am Sitting in a Room* (1969) repeatedly re-records a spoken text played back into, and re-recorded from, the same physical room, so that the room's own resonant frequencies are progressively reinforced across successive iterations until the original speech dissolves entirely into the resonant identity of the space (Lucier, 1969), an early and direct demonstration that a room possesses its own sonic character independent of, and ultimately capable of overwhelming, any sound introduced into it. David Rokeby's *Very Nervous System* (1986) used computer vision to translate a participant's whole-body movement, in real time, into responsive electronic sound within an open gallery space (Rokeby, 1998), and is widely regarded as a foundational work in the lineage of movement-driven interactive sound installation that *Sonic Inhabitation* extends. Imogen Heap's Mi. Mu Gloves, developed with Thomas Mitchell and Sebastian Madgwick over several iterations from the early 2010s onward, translate hand gesture and orientation into real-time control of live vocal looping and effects processing (Mitchell et al., 2012), a widely visible contemporary example of the object-based air-instrument category discussed in Section 2.5.5. Stephanie Cheng Smith's *Bell Controller*, documented from 2016 onward, drives fifteen vibration motors and LEDs mounted on jingle bells through a custom Arduino routine, letting a performer adjust note timing,

intensity, and motor behaviour in real time; it has been performed at venues including MOCA's Geffen Contemporary (2017) and REDCAT (2018) (Cheng Smith, 2016).

The works discussed throughout this chapter fall into two broad categories with respect to *Sonic Inhabitation*: theoretical or technical literature that grounds the project's methodology (Sections 2.1 to 2.5), and artistic or research configuration that supports its design directly. This section bridges the theoretical review above to the system and experience-design decisions detailed in Chapters 3 and 4.

Rokeby's *Very Nervous System*, introduced above, established the basic proposition that whole-body movement, tracked in real time and without a handheld controller, can drive a responsive electronic soundscape within an inhabited physical space, the same proposition underlying the motion-capture-to-synthesis pipeline described in Chapter 3, though *Sonic Inhabitation* extends it further by removing the visual channel Rokeby's installation retained. Wijnans's *ChoreSonic Environments* (Section 2.4.4.1) is the closest existing precedent for the specific coupling this dissertation relies on between body position, ambisonic spatialisation, and a real-time-controlled 3D sound field; Wijnans's notion of *TranSonic Perception*, sound as a "second skin" extending beyond the body's physical limits, anticipates the experiential claims made about spatial agency in Chapter 4. Carson and Gordon's *Resonant Landscapes* (Section 2.2.1) demonstrates, at a technological level, that ambisonic field recordings can be convincingly relocated into a different physical environment via real-time playback triggered by a participant's position, the same underlying operation performed by the Intercontinental Dataset described in Chapter 3, applied there to recreate specific indoor rooms from around the world rather than outdoor nature recordings. A wider set of works exists in adjacent fields, musical performance and interactive concert practice in particular, that engage similar questions of embodiment, agency, and spatial sound without being direct precedents for an architectural or installation-based experience, and are noted here without further elaboration for that reason: the broader corpus of digital-musical-instrument and NIME-community research discussed in Section 2.5 (Magnusson, 2018) belongs to this category, as does the wider literature on live-electronic and audiovisual concert performance more generally.

2.7 Key Theorists

Three further theorists, while not primarily concerned with sound, offer conceptual resources directly relevant to the experiential claims this dissertation makes about embodiment, agency, and presence in an audio-only environment. They are placed here, after the sound-specific and artistic literature reviewed in Sections 2.1 to 2.6 and before the chapter's Summary, because their role in this dissertation is interpretive rather than foundational to the system's design: none of the three theorists influenced a design decision in Chapter 3 or Chapter 4, but each supplies a specific conceptual tool used when reading the qualitative results in Chapter 6. Deutsch, Noë, and Metzinger were selected, out of a much larger possible field, because each maps onto one distinct interpretive need that arises directly from this study's audio-only, movement-driven design, rather than for general theoretical breadth: Deutsch's auditory illusions are direct empirical evidence that spatial

hearing is actively constructed rather than passively read off the acoustic signal, exactly the kind of construction this installation's participants are asked to perform without visual confirmation; Noë's enactive account extends the embodied-cognition frame of Section 2.1.3 into a full philosophical account of perception as active bodily skill rather than passive receipt; and Metzinger's self-model theory gives a vocabulary for what it might mean for a participant's felt bodily presence within the room to become unsettled or foregrounded, the specific phenomenon the post-exposure interview is designed to probe.

2.7.1 Diana Deutsch

Diana Deutsch's decades of work in music psychology and psychoacoustics have documented a series of robust and striking auditory illusions that show how actively constructive, rather than passive, auditory spatial perception really is. The *octave illusion* presents listeners with two tones an octave apart, alternating between the ears such that each ear receives a different tone at any given moment. Rather than perceiving two independent pitch streams, most listeners perceive a single melodic line, and the perceived spatial location of that line, whether it seems to come from the left or right ear, depends systematically on the listener's handedness, not on the actual physical arrangement of the tones (Deutsch, 1974). The closely related *scale illusion* distributes ascending and descending scale fragments unpredictably between the two ears, yet listeners reconstruct two smooth, spatially coherent melodic contours, one high and one low, that do not correspond to what either ear individually received, showing that the auditory system actively regroups fragmented input by pitch proximity rather than ear of origin (Deutsch, 1974). The *tritone paradox* presents a pair of tones separated by exactly half an octave, an interval that is acoustically ambiguous as to whether it is "rising" or "falling," and shows that listeners' judgments of directional motion are strongly predicted by their native language and dialect, implying that a listener's long-term linguistic exposure calibrates an implicit template against which ambiguous pitch relationships are later judged (Deutsch, 1991). Deutsch has also documented illusions of the auditory-verbal boundary itself, most notably the *speech-to-song illusion*, in which a spoken phrase, looped without alteration, is spontaneously reheard by listeners as sung rather than spoken purely as a function of repetition (Deutsch et al., 2011). This body of work, synthesised in *Musical Illusions and Phantom Words* (Deutsch, 2019), is a sustained practical case that auditory spatial and categorical perception is a construction shaped by the listener's individual physiology, linguistic history, and expectation.

2.7.2 Alva Noë

Alva Noë's enactive account of perception, developed in *Action in Perception* and building on the sensorimotor contingency theory he co-developed with psychologist J. Kevin O'Regan, argues that perceiving is not a process that happens to a passive organism upon receipt of sensory input, but an active skill, a form of exploratory bodily activity guided by an implicit, practical mastery of the way sensory stimulation changes as one moves (Noë, 2004; O'Regan and Noë, 2001). On this view, seeing a fruit, for instance, is not a matter of an internal representation being built from

retinal input, but a matter of practically understanding how the fruit's appearance would change were one to move one's eyes, head, or body, what O'Regan and Noë call the object's "sensorimotor contingencies." Perceptual experience is constituted by, rather than merely accompanied or caused by, the perceiver's sensorimotor engagement with the environment. Noë's framework has been influential because it relocates the site of perceptual experience from the brain alone to the coupled system of brain, body, and environment, and it has drawn both adoption and criticism within philosophy of mind, with critics arguing among other objections that the theory struggles to account for perceptual experience during genuine immobility.

2.7.3 Thomas Metzinger

Thomas Metzinger's self-model theory of subjectivity, set out in *Being No One: The Self-Model Theory of Subjectivity* (Metzinger, 2003), proposes that the sense of being a self is not a fixed, given entity but a dynamically constructed, representational model, the phenomenal self-model (PSM), generated by the organism's own information-processing system. Central to Metzinger's argument is the concept of transparency: the PSM is transparent in the sense that the organism does not, and typically cannot, experience it as a model or representation at all; the model's own representational character is invisible to introspection, so the organism simply takes itself to be a self, directly and without mediation, rather than experiencing itself as the output of an underlying modelling process. It is precisely the transparency of the PSM, on this account, that gives rise to the pre-theoretical conviction of being a genuine, continuously existing self, rather than a model of one, hence the book's title, since on Metzinger's view no one, in the sense of an unmediated, non-modelled subject, ultimately exists. Metzinger's theory has since been read alongside, and partly absorbed into, the predictive-processing framework in cognitive neuroscience, which similarly treats the sense of self as a hierarchically generated prediction rather than a direct perceptual given (Seth and Tsakiris, 2018; Wiese and Metzinger, 2017). Metzinger's framework offers a way of theorising what may be at stake when that reconstruction becomes unstable or altered: not merely a change in what the participant perceives about the room, but a possible modulation of the transparent self-model through which the participant's own bodily presence within the room is represented at all. Where an ordinary, visually grounded environment leaves the transparency of the PSM largely undisturbed, an audio-only environment that destabilises a participant's confidence in their own spatial location arguably exposes, however briefly, the normally invisible modelling process Metzinger's theory describes, a phenomenon this dissertation's qualitative interview data (Chapter 4) is well placed to probe directly, by asking participants to reflect on their felt sense of bodily presence during the experience.

2.8 Summary and Conclusions

This chapter has reviewed six bases of literature bearing on *Sonic Inhabitation*: the physical and technical vocabulary of sound and audio and the synthesis and sampling techniques used to generate it; the perceptual frameworks, Schaefferian, ecological, and multisensory, through which

listeners make sense of an auditory scene; the affective and embodied-spatial literature on how soundscapes are felt rather than merely identified; the sound object and the psychoacoustic mechanisms of spatial hearing and action-sound coupling that underlie it; the negotiation of agency in interactive and unconventional musical instruments, together with the artistic and research precedents that most directly inform the present design; and, finally, three theorists whose work on auditory illusion, enactive perception, and the phenomenal self bears directly on the experiential claims examined in the chapters that follow.

Reviewing these frameworks together is what makes it possible to understand how humans perceive, experience, and interact with sound in space, rather than treating each body of literature as a self-contained topic. The discussion of space and soundscapes has shown that space is not merely a physical construct, but a dynamic and affective construct shaped by sensory input, cultural context, and personal experience. Soundscapes reveal how auditory environments contribute to spatial understanding and emotional experience. Artistic and technological projects demonstrate how spatial sound can reshape perception and foster ecological and reflective listening practices. Spatial audio technologies, particularly Ambisonics, offer powerful tools for capturing, reproducing, and analysing immersive sound environments. The reviewed studies demonstrate how spatial audio can be used to investigate human behaviour, perception, and interaction, especially when combined with motion capture and behavioural metrics. Despite advances, several research gaps persist. Indoor soundscapes remain less explored than in outdoor contexts. Research in how humans experience these indoor spaces is still reduced to a response to annoyance, focusing on minimising noise, instead of on how to use sound as a positive experience in indoor spaces. These studies mainly rely on monaural metrics and acoustically averaged levels of sound intensity in a room, overlooking the three-dimensional nature of the human auditory experience. Finally, there is a notable absence of studies that effectively bridge engineering and artistic research. Engineering approaches provide the quantitative rigour and technical data, such as the quantity of motion and audio energy maps, which provide important results on micromotion and movement. Artistic research provides the qualitative depth necessary to understand the nuances of how humans experience space and interact with it. This research at the RITMO centre directly addresses this gap. By using the fourMs Lab to simulate realistic rooms, I can combine high-precision kinematic measurements with narrative inquiry. This mixed-methods approach allows me to investigate the sound-motion relationship as a holistic, interactive process. In doing so, I aim to provide a more comprehensive framework for how humans inhabit and process the indoor acoustical world.

Chapter 3

System architecture

To address research questions about embodied sound cognition, spatial agency, and how cultural specificity in soundscapes shapes human interaction with space, a custom interactive architecture was required. This chapter presents the architecture and design rationale of the system developed to support this inquiry, structured around three interdependent layers: the audio framework, the motion capture system, and the real-time data pipeline connecting motion capture to spatial audio synthesis. This architecture follows the methods established within RITMO's AMBIENT project and the fourMs Lab more broadly, where ambisonic field recordings of indoor spaces are paired with motion capture to study how people respond bodily to a soundscape (Guo et al., 2025; Riaz et al., 2025a).

3.1 Audio Framework

The system's central purpose is to immerse the participant in an auditory environment where sound, not sight, shapes the experience. Without visual feedback, the auditory design has to do three things: establish spatial boundaries, sustain narrative progression, and deliver interactive feedback. The soundscapes were therefore built from a combination of sample playback, sample manipulation, and real-time sound synthesis. The soundscapes are composed following R. Murray Schafer's soundscape taxonomy, which distinguishes between *keynote* (background) sounds, *signals* (foreground) sounds, and *soundmarks* Schafer (1993). The following subsections detail the auditory assets selected for each layer, the synthesis engine used to define the spatial boundaries between virtual rooms, and the multichannel spatialisation strategy used to render the environment across the loudspeaker array. These recordings and samples have been deposited on Zenodo, an open-access research repository, under a CC-BY 4.0 licence (Mesquita Leal, 2026), so they remain available independently of this thesis.

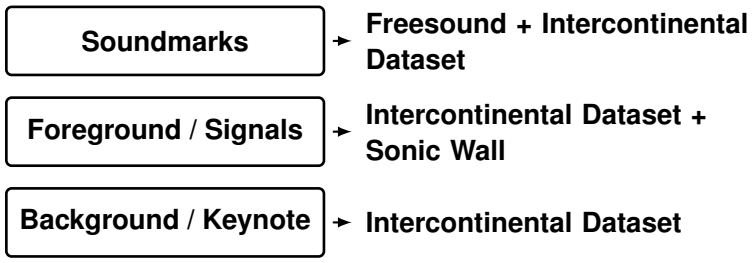


Figure 3.1: R. Murray Schafer's soundscape taxonomy applied to the system's three audio layers.

3.1.1 Intercontinental Dataset

To examine how cultural context shapes spatial perception, the virtual rooms in the experience were constructed from original field recordings, referred to here as the Intercontinental Dataset. The dataset has been developed by the AMBIENT research group since 2023, using a Zoom H3-VR ambisonic recorder to capture indoor soundscapes in countries around the world. Its purpose is to allow systematic comparison of the sonic character of indoor spaces across cultures. First-order ambisonic recordings preserve high-fidelity spatial and textural information with a distinctive cultural, environmental, and acoustic signature, and can evoke physical spaces the participant could not otherwise access. At the time of the project development, the dataset contains 72 first-order ambisonic recordings collected in 11 countries. Each recording is approximately ten minutes in duration and, whenever possible, was captured at the centre of the room, positioned on a table. The recordings chosen for the final experience were based on: audio quality, clear structural separation between background and foreground sound layers, geographic diversity, and auditory salience. The first approach to the dataset was to shortlist one representative recording per country, prioritising material in which a distinct acoustic contrast between continuous background atmosphere and discrete foreground soundmarks could be clearly identified. This contrast was necessary for downstream processing and layering. Selection was also decided by privacy considerations, particularly where recordings contained identifiable speech. Table 3.1 summarises the four rooms drawn from the dataset that were used to build the final experience, together with the descriptive metadata recorded at the time of capture.

3.1.2 Soundmarks

The recordings drawn from the Intercontinental Dataset carry the acoustic fingerprint of the real rooms in which they were captured. Whenever possible, soundmarks were extracted directly from these field recordings, and external libraries were used only to supplement or in cases where a soundmark could not be isolated from the original material through editing. Each of the four environments required a distinct set of soundmarks. Where extraction from the original recording was feasible, individual events were isolated by snipping and filtering the source material, followed by noise reduction using Audacity's built-in algorithm. Where extraction was not possible, supplementary material was sourced from Freesound.org. Freesound assets were selected according to two criteria, applied in order of priority: recordings made in the same city and similar setting as

Table 3.1: Metadata for the Rooms Selected from the Intercontinental Dataset

Room	Location	Near / Foreground Elements	Distant / Background Elements	Notes
Room 0	Zhangjiakou, China — Market	Speech; smartphone	—	Market setting; customers and sellers speaking, content not intelligible. ID 54, mic M3, recorded 09.01.2024, 11:08.
Room 1	Dali, China — Exhibition room	Ventilation; speech; footsteps	Sweeping; speech; background music	Copyright issues: yes. Privacy issues: yes. Clear speech audible; background music present ($\sim$8 min mark). ID 57, mic M3, recorded 21.01.2024, 11:08.
Room 2	Barcelona, Spain — Public library	Indistinct chatter; footsteps on hard floor; book handling; furniture movement; page turning; zips/bags; typing and mouse clicks; occasional coughing; distant speech; door opening/closing; book-cart movement	City traffic noise, car horns, faint exterior voices	Recorded in a study area between bookshelves; historic stone architecture produces natural reverberation. ID 68, mic M6, recorded 18.12.2025, 13:26.
Room 3	Oslo, Norway — Gardermoen Airport	Footsteps (occasional running); rolling suitcase wheels; nearby indistinct speech; chair movement/scraping	Distant unintelligible conversation and crowd murmur; very distant aircraft engine sound from airside	ID, mic, and capture date/time not recorded for this entry.
Room 4	Doha, Qatar — Airport food court	Chatter; music; cutlery; plates	None	Description: many static and non-static sound elements. Copyright issues: yes. Privacy issues: yes. Music audible at approx. 5:05; general chatter unintelligible, with occasional dialogue close to the microphone. ID 3, mic M1, recorded 17.06.2023, 14:18.

the corresponding field recording, and, where city-matched material was unavailable, recordings made in a comparable type of environment. In the case of the Barcelona Library soundscape, for example, several soundmarks were sourced from recordings made in the UPF library in Barcelona, a setting closely similar to the original field recording. Other soundmarks for this room, such as the sound of tearing paper, were selected based on environmental similarity, such as an office or study space rather than geographic matching. Table 3.2 provides the complete inventory of soundmarks used in the experience. This inventory is also shown in Max/MSP patches: the `buffer~` object filenames found in the `soundactions` subpatch of `2phase.maxpat` (Section 3.4.6) match this table exactly, including Freesound IDs.

Table 3.2: Explicit List of Soundmarks and Audio Sample Sources

Location	No.	Soundmark Action	Source / ID
Room 1: Dali Exhibition (Museum Room, China)	1	Whistling	Original Recording
	2	Talking and walking	Original Recording
	3	Talking 2	Original Recording
	4	Talking 1	Original Recording
	5	Cough	Original Recording
	6	Door closing	Original Recording
	7	Door opening	Original Recording
	8	Car honking	Original Recording
	9	Footsteps 3	Original Recording
	10	Footsteps 2	Original Recording
	11	Footsteps 1	Original Recording
	12	Camera shutter	Freesound (485711)
Room 2: Barcelona Library (Public Library, Spain)	1	Metal chair across concrete floor	Freesound (740448)
	2	Ringing phone	Freesound (240187)
	3	Writing on wood	Freesound (849311)
	4	Women speaking	Freesound (257862)
	5	Printer buttons	Freesound (153343)
	6	Copying machine / photocopies	Freesound (233018)
	7	Computer keyboard	Freesound (233020)
	8	Cough	Freesound (508117)
	9	Tearing paper	Freesound (852211)
Room 3: Oslo Airport (Oslo Airport, Gardermoen, Norway)	1	Airport baggage-claim ambience	Freesound (416816)
	2	PA announcement	Original Recording
	3	Metal chair 2	Original Recording
	4	Metal chairs	Original Recording
	5	Suitcase 1	Original Recording
	6	Suitcase 2	Original Recording
Room 4: Doha Food Court (Airport, Qatar)	—	*No soundmarks extracted*	N/A

3.1.3 Sonic Wall

To give the virtual environment physical gravity, the boundaries between rooms are defined through synthesised sound and manipulation of the soundscape. As a participant comes close to the limit

between the virtual rooms, the system generates a localised, synthetic texture that acts as a haptic-auditory boundary. This texture is designed to give the participant a firm perceptual impression of an obstacle, letting them map the width and height of an otherwise invisible room without physical contact with any object. This synthesised barrier is called the *sonic wall*. Its underlying engine was adapted and extended from the open-source architecture of *SoundSaber*, a motion-capture instrument developed within the fourMs Lab framework, and is built around a high-frequency impulse train modulated through resonant feedback systems. Section 3.4.6 explains the algorithm in detail.

3.1.4 Spatialization and Multi-Channel Configuration

The recordings in the Intercontinental Dataset are captured in first-order ambisonic (B-format), so spatial information is already encoded within the source material itself. Within Max/MSP, this information is encoded and decoded in real time using the SPAT5 library and RITMO decoder, adapting the spatialization to the main room of the laboratory. The *soundmarks* and the *sonic wall*, on the other hand, also had to be spatialized using the SPAT library from IRCAM. The three-dimensional auditory experience is rendered through a loudspeaker array of 24 independent speakers distributed throughout the main laboratory room. The primary objective of this configuration is to ensure that, as the participant moves freely through the space, the spatialization of the auditory scene remains coherent and continuously updated relative to the participant's position. Real-time spatialization is managed using the Spatialisateur (*Spat5*) framework developed by IRCAM in the Max/MSP environment. To convert the participant's tracked coordinates into multichannel control signals, the system implements spatial panning algorithms, specifically Higher-Order Ambisonics (HOA) decoding to render virtual point sources across the 24-channel loudspeaker geometry. Table 3.3 presents the calibration configuration per-speaker used by the Spat engine, containing the azimuth, elevation, and distance of each loudspeaker relative to the listening position, together with the delay and gain corrections applied to compensate for the physical geometry of the array.

3.1.4.1 RITMO Decoder

The RITMO decoder was developed alongside Balint Laczko, a doctoral research fellow at RITMO whose work spans 3D audio, sonification, and audiovisual analysis. The decoder was designed specifically for the new monitor setup to facilitate future 3D audio work developed in the laboratory. The decoder stores loudspeaker positions in the Ambisonic Equidistant Decoding (AED) format, specifying each speaker's azimuth, elevation, and distance. The array has twenty-four channels arranged unevenly across the horizontal and upper hemispheres: a lower ring at zero elevation carries the main horizontal image, a second ring sits at roughly 17° to 25° elevation, and a few near-zenith speakers (above 36° elevation) cover the top. This uneven, dome-like layout reflects a common trade-off in Ambisonic playback: horizontal localisation is prioritised in the plane where a standing or moving listener needs it most, while the raised speakers widen the perceived sound field without requiring full spherical symmetry.

Table 3.3: 24-Channel Loudspeaker Array Configuration (Spat AED, Delay, and Gain Correction Values)

Speaker	Azimuth (°)	Elevation (°)	Distance (m)	Delay (ms)	Gain (dB)
1	23.0	0.00	2.924	3.3333	-2.9
2	65.0	0.00	3.387	1.9834	-1.6
3	101.0	0.00	3.378	2.0105	-1.6
4	133.0	0.00	3.391	1.9726	-1.6
5	178.6	0.00	2.478	4.6329	-4.3
6	-143.0	0.00	3.318	2.1853	-1.8
7	-112.0	0.00	3.787	0.8188	-0.6
8	-76.0	0.00	3.741	0.9528	-0.7
9	-42.0	0.00	3.252	2.3776	-1.9
10	30.0	20.45	3.291	2.2640	-1.8
11	62.0	16.96	4.055	0.0379	0.0
12	84.0	18.08	3.780	0.8392	-0.6
13	105.0	17.28	4.049	0.0554	0.0
14	137.0	21.40	3.354	2.0804	-1.7
15	-149.0	20.72	3.330	2.1503	-1.7
16	-117.0	17.16	4.063	0.0146	0.0
17	-91.0	18.18	3.782	0.8333	-0.6
18	-74.0	16.97	4.068	0.0000	0.0
19	-46.0	20.88	3.339	2.1241	-1.7
20	-12.0	25.39	2.752	3.8345	-3.4
21	1.0	47.90	1.849	6.4656	-6.8
22	87.0	37.14	2.211	5.4108	-5.3
23	-174.0	48.62	1.834	6.5099	-6.9
24	-96.0	36.83	2.237	5.3351	-5.2

3.2 Motion Capture System

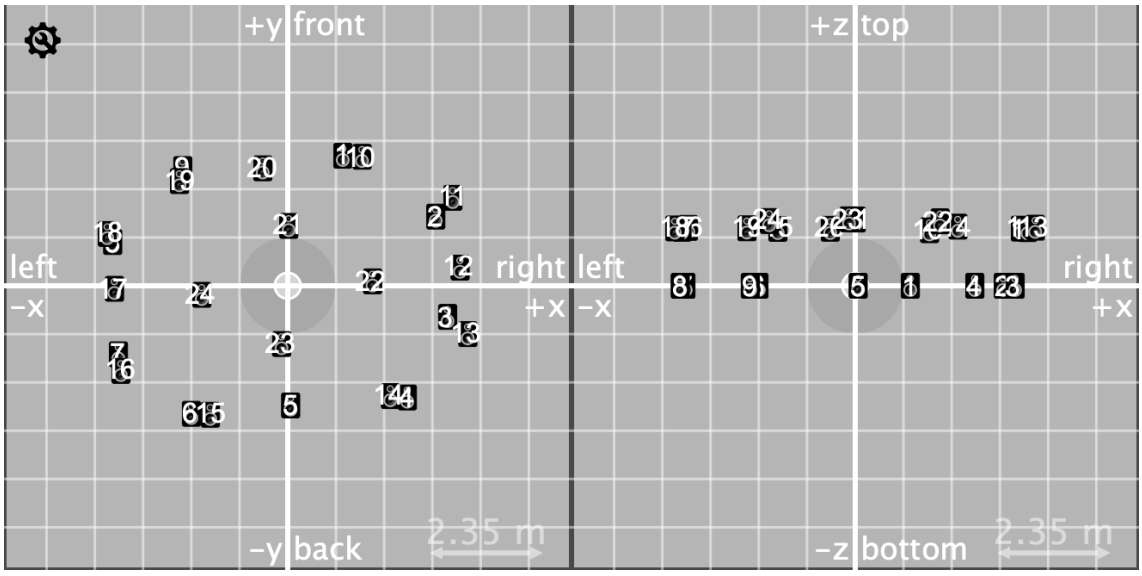
The motion capture was made using 14 cameras from a Qualisys infrared motion-capture array (Qualisys AB, Gothenburg, Sweden), calibrated before every session. The fourteen configured cameras have a fully resolved position in the current calibration, at a height of approximately 2.5 to 2.8 m with a capture frequency of 100 Hz. This aligns with the analysis pipeline described in Chapter 5, which presents a logging rate of 96.57 Hz for the behaviour log.

Table 3.4: Qualisys camera array configuration

Parameter	Value
Configured camera slots	14 (all flagged active)
Cameras with a resolved calibrated position	13
Camera mounting height	≈2.5–2.8 m
Estimated capture-volume footprint (plan)	≈5.3 m × 7.4 m
Capture frequency	100 Hz
Marker set (AIM model)	Qualisys “Sports” (`Sports.qam`)

To close the interaction loop between physical movement and acoustic generation, a real-time data pipeline was designed to connect the motion-tracking hardware with the spatial audio syn-

Figure 3.2: 24-Channel Loudspeaker Array Configuration



thesis engines. Both the motion-capture computer running QTM and the audio computer running the Max/MSP patch were synchronised to a shared external clock, the AMBIENT system clock. QTM's own recordings carry a `TIME_STAMP` field derived from this clock, and the Max/MSP behaviour log's embedded hour/minute/second/millisecond fields are drawn from the same source, so the two recordings can be aligned post hoc on this shared timestamp.

3.3 System Architecture and Data Integration

The following section explains in detail the architecture of the system created, as well as the data flow across all the different Max/MSP patches.

3.3.1 High-Level Architecture and Data Flow

The system operates as a continuous action-reaction loop composed of three layers: Motion Capture (sensing), Network Communication (routing), and Audio Synthesis/Spatialisation (execution).

1. **Sensing:** The participant, equipped with a motion-capture suit with passive reflective markers (Qualisys Sports marker set), moves freely in the Main Room. The 14-camera Qualisys array tracks the physical markers at 100 Hz, computing three-dimensional spatial coordinates (X, Y, Z) and rigid-body skeletons in real time, with both systems' clocks nominally referenced to the shared AMBIENT system clock described above.

2. **Routing:** The position data is compiled by Qualisys Track Manager (QTM) and broadcast to the audio-processing computer. The specific patch responsible for this bridge is described in detail in Section 3.4.2

Figure 3.3: RITMO decoder



3. **Execution:** A custom Max/MSP patch analyses the incoming position data. The raw spatial coordinates are unpacked, normalised, and mapped simultaneously to all the patches.

Figure 3.6 shows the system's structure and the real-time, two-way relationship between the participant, the hardware, and the software components. Section 3.4 below expands this same architecture into a detailed explanation of the Max/MSP implementation.

3.4 Max/MSP Patch Architecture

The interaction system is implemented across six connected Max/MSP patches. This section walks through each in turn, from the raw motion-capture bridge to the final loudspeaker output, and illustrates each with a schematic figure.

3.4.1 Overview

Figure 3.7 outlines the data flow through all six patches, extending the higher-level diagram already given in Figure 3.6 into a more detailed explanation. Positional data is sent from QTM as one real-time-tracked rigid body, placed on the head, and is picked up by `QTM2OSCStreaming_NEW.maxpat`. `CONTROL.maxpat` is the system's nucleus: it extracts movement features from each tracked point and sends them to the `fFE_3D1_NEW` abstraction, forwards the head's data onward to whichever phase patch (`1phase.maxpat` or `2phase.maxpat`) is currently active, and independently logs the full set of position, phase, room, and sound-trigger flags, together with a timestamp, to the behaviour log analysed in Chapter 5. The active phase patch generates the room and boundary audio and passes it to `spatializer.maxpat`, which renders it across the 24-channel array via IRCAM Spat5.



Figure 3.4: The headpiece - rigid-body tracker used to control the Max patch through Qualisys



Figure 3.5: Mocap Suite using Sports marker set

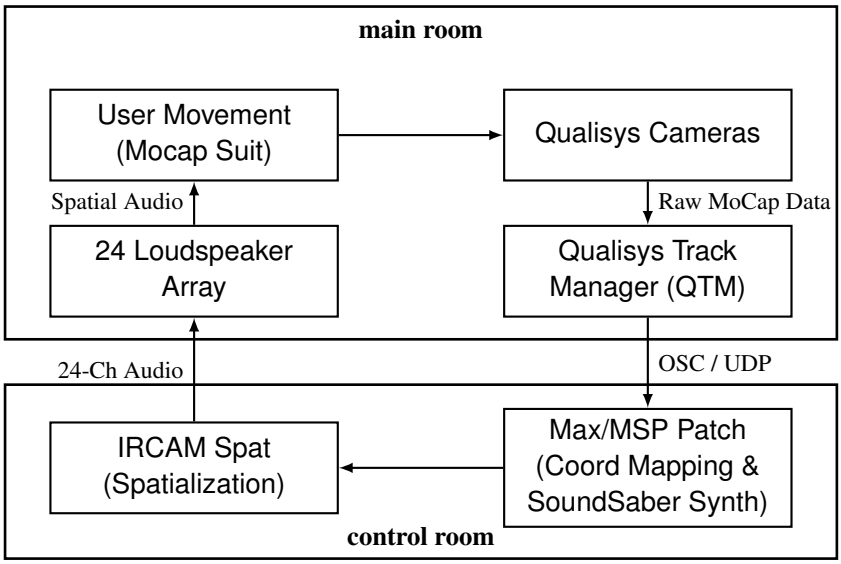


Figure 3.6: System architecture: motion capture and spatial audio data flow.

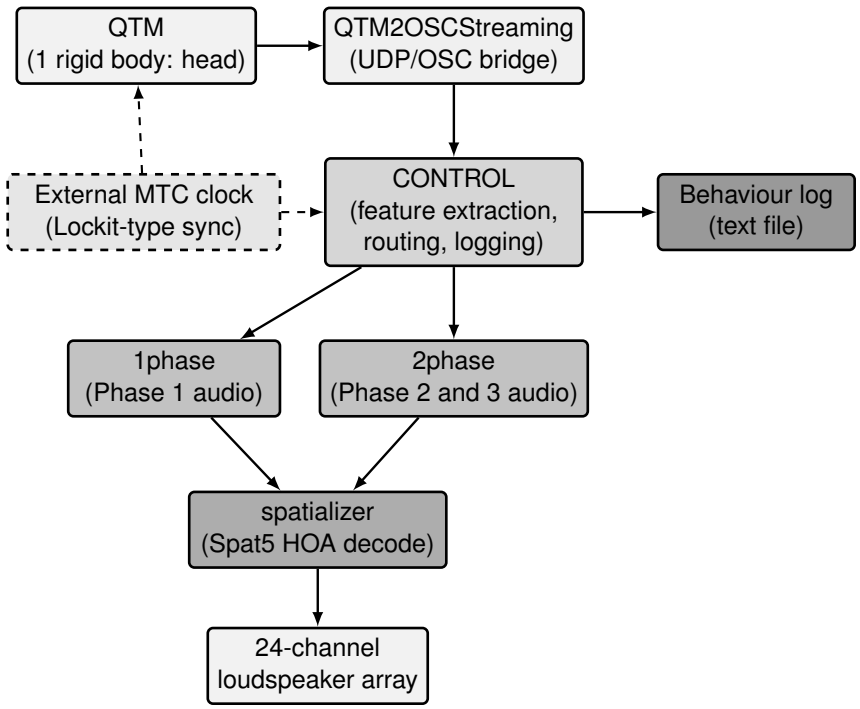


Figure 3.7: Data flow across the six Max/MSP patches. Solid arrows carry positional/audio data. The dashed arrows show the external MTC-based clock (Section 3.3) feeding timestamps to both the QTM side and the Max/MSP side independently.

3.4.2 QTM2OSCStreaming_NEW.maxpat: The Motion-Capture Bridge

This patch is adapted from the fourMs Lab's `qualisys-osc` bridge.[1] It opens a UDP connection to QTM (`udpreceive 45454` to receive, `udpsend 172.22.40.50 22225` to send control messages back), and, once connected via the message `/qtm Connect 45454`, requests a continuous stream of 6-degrees-of-freedom rigid-body data with `/qtm StreamFrames AllFrames 6D`. One rigid body is streamed this way, corresponding to the head, and an extra rigid body (a plastic piece with 3 reflectores) is added to the participant's motion-capture suit. This is a separate data pathway from the labelled-marker `.tsv` export analysed offline in Chapter 5: the real-time pathway described here streams one pre-defined rigid body for live sound control, while the offline export used for movement analysis is built from the full labelled marker set. The two should not be conflated when describing the system's tracking capabilities.

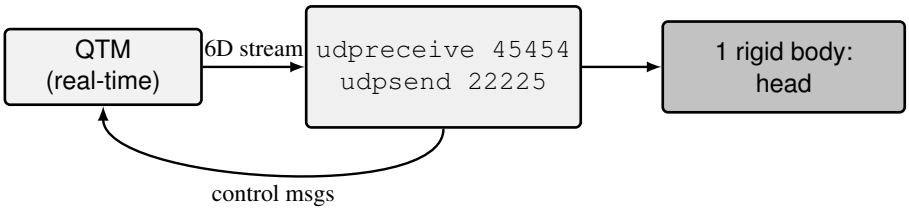


Figure 3.8: QTM2OSCStreaming_NEW.maxpat. A UDP bridge, based on the fourMs Lab's public `qualisys-osc` patch, that requests and receives a continuous stream of 6-degree-of-freedom rigid-body data for the head.

3.4.3 fFE_3D1_NEW.maxpat

Developed by Kristian Nymoen (fourMs Lab), this abstraction converts a raw 3D position stream into movement features used in the rest of the system. The patch computes: absolute velocity and horizontal velocity (via `expr` distance formulas over successive filtered positions), horizontal distance, and the position in polar (azimuth/elevation/distance) form via `cartopol`, alongside the raw filtered x, y, and z. The coordinates are first passed through data objects (such as `slide` or `deltaclip`) that filter out high-frequency tracking jitter without introducing perceptible latency. Once smoothed, low-level motion features, such as absolute velocity ($\|v\|$) and acceleration, are calculated. The absolute Euclidean distance (d) between the participant's head marker (X_u, Y_u, Z_u) and the nearest virtual wall coordinate (X_w, Y_w, Z_w) is computed continuously according to Equation 3.1:

$$d = \sqrt{(X_u - X_w)^2 + (Y_u - Y_w)^2 + (Z_u - Z_w)^2} \tag{3.1}$$

[1] fourMs Lab, *qualisys-osc*, University of Oslo. Retrieved from https://github.com/fourMs/qualisys-osc.

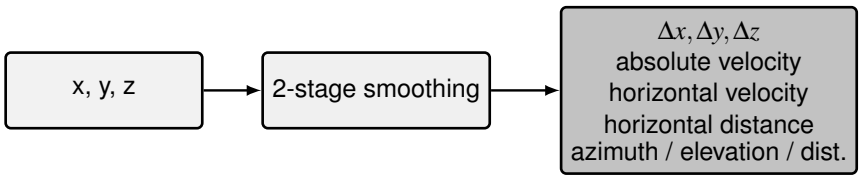


Figure 3.9: fFE_3D1_NEW.maxpat (Nymoen, fourMs Lab).

3.4.4 CONTROL.maxpat

This patch is the system's control, doing routing, feature extraction, and logging of the data. It embeds `QTM2OSCStreaming_NEW.maxpat` directly (Section 3.4.2) inside a subpatch named `your-3D-sensor-data-stream-here`, routes the one incoming rigid-body streams by address (`route /qtm/6d/head` and passes each through its own instance of `fFE_3D1_NEW` (Section 3.4.3). The head's features are forwarded on, via `send data_head`, to whichever phase patch is currently active. Separately, `CONTROL.maxpat` receives every phase flag, room flag, the simultaneous flag, and all 27 sound-trigger flags back from the active phase patch, packs them together with the current head position and a timestamp (built from `date` and `prepend time`/`prepend date`), and writes the resulting row to a `coll` object, which is what is saved as the behaviour log's text file. The patch also contains a subpatch named `p lockit_connection`, which receives MIDI input (`midiin`) and decodes it with bit-shift and mask operations consistent with MIDI Time Code quarter-frame parsing. This is the decoding of the AMBIENT system clock, Lockit hardware timecode synchroniser.

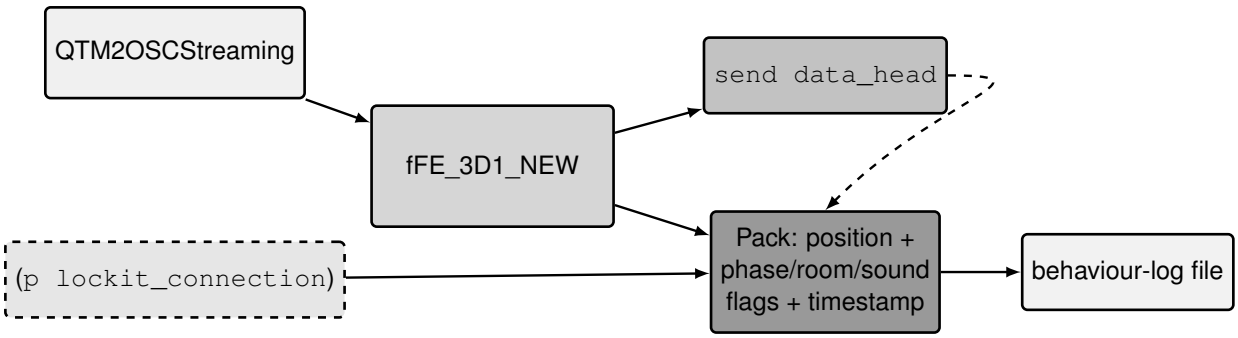


Figure 3.10: CONTROL.maxpat. The dashed arrow shows that the phase/room/sound flags being logged are themselves fed back from the active phase patch (Figure 3.7).

3.4.5 1phase.maxpat: The Phase 1 Sound Engine

This patch renders the one-room soundscape described in Chapter 4 (Phase 1: Sensory Calibration and Immersion). Head position is converted, in the `room0` subpatch, into a measure of proximity to two concentric circular thresholds ("how much inside" the large and small circle), which drives a chain of ramped gain and filter automation (subpatches `filters` and `volume_alterations`) applied to a four-channel ambisonic buffer (`buffer~ phase1 file2_1.WAV`, played with `mc.groove~`). Concretely, crossing the inner threshold ramps the room's gain between roughly

−100 dB (silent) and −20 to −5 dB over 500–1500 ms, while simultaneously sweeping a biquad filter's cutoff frequency, producing the room's graded, distance-dependent audibility. A separate branch, `p z_movement`, derives an additional control signal from vertical movement (`delta_z`) combined with horizontal velocity. The processed four-channel signal is passed to an `ambisonics` subpatch, which decodes it via IRCAM Spat5 (`spat5.spat~ @initwith "/panning/type hoa3d"`) together with the custom `ritmodecoder~` object described in Section 3.1.4.1, and a loaded speaker preset (`ritmo_speaker_config.txt`).

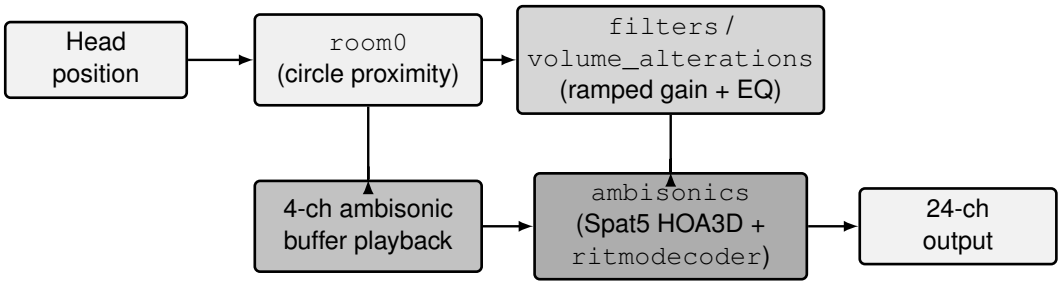


Figure 3.11: 1phase.maxpat

3.4.6 2phase.maxpat: The Phase 2 and 3 Sound Engine

This substantially larger patch extends the same room-proximity logic of Section 3.4.5 to four rooms in parallel (subpatches `room01`–`room04`, each structured like Section 3.4.5's `room0`, and each independently detecting the "simultaneous" overlap state analysed in the Results chapter), and adds three further components not present in Phase 1. The `wall` subpatch implements the boundary synthesis discussed in Section 3.1.3: a bass tone (`cycle~ 30`, i.e. a 30 Hz oscillator) combined with two delay lines (`tapin~`/`tapout~`, roughly 1.46–20 ms, in a `delay and feedback` sub-subpatch with an 0.8/0.2 feedback mix) independently modulated by vertical and horizontal velocity, passed through a lowpass biquad filter whose cutoff tracks the participant's vertical (z) position, and periodically punctuated by a `click~` impulse; the 30 Hz tone and periodic click are the two elements flagged in Section 3.1.3 as missing from that subsection's current description. The `soundactions` subpatch plays the 27 discrete, Freesound-and-original-recording sound triggers catalogued in Table 3.2 (its `buffer~` object filenames, e.g. `room1_Dali_footsteps1.wav`, `room1_Dali_485711__camera-shutter.m4a`, match that table exactly); where multiple recorded variants exist for one soundmark (three footstep recordings, two talking recordings), a `counter` object cycles between them on successive triggers so the same action does not always play an identical recording. Finally, an `underwater` subpatch (`spat5.reverb~`, with separately scaled low/medium/high-frequency decay times) as an additional immersive layer, scaled by vertical position and proximity.

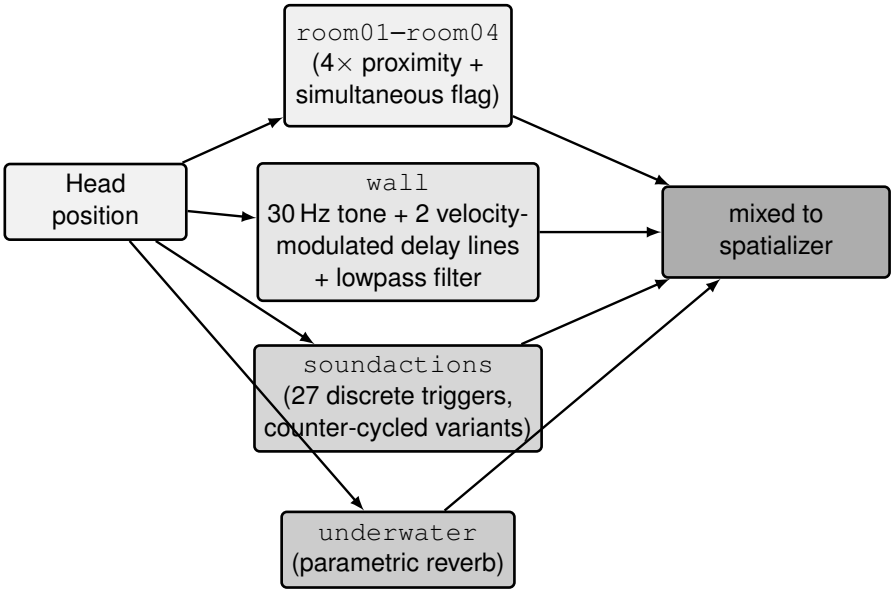


Figure 3.12: 2phase.maxpat

3.4.7 spatializer.maxpat: Loudspeaker Rendering

The final stage converts a participant's position (in millimetres, from QTM) into a normalised source position for IRCAM Spat5, and renders the mixed room/wall/soundaction/reverb signal across the loudspeaker array. Position is scaled and offset (`* 0.001` to convert mm to m, then per-axis offset), packed, and sent to `spat5.oper` to update the source's location; `spat5.spat~ @inputs 1 @outputs 25 @mc 1` performs the panning itself. A message box loaded at startup (`/speakers/aed ...`) sets the azimuth/elevation/distance of all speakers, matching the 24-channel configuration already given in Table 3.3, with a 25th value present in the same message for the system's reference/subwoofer channel; output is finally sent to `mc.dac~` across channels 1–24.

3.5 Summary

This chapter has presented the technical architecture underpinning the experience: the three-layer audio framework built from the Intercontinental Dataset and its extracted soundmarks; the synthesis engine used to render acoustic room boundaries; the 24-channel spatialisation pipeline; the physical configuration of the fourMs Lab; the real-time data pipeline linking motion capture to sound synthesis; and a patch-level walkthrough of the six Max/MSP files that implement it. Throughout, the design choices described, the recording format and placement used to build the Intercontinental Dataset, and the pairing of optical motion capture with ambisonic audio follow methods already established and validated within the AMBIENT project and the fourMs Lab.

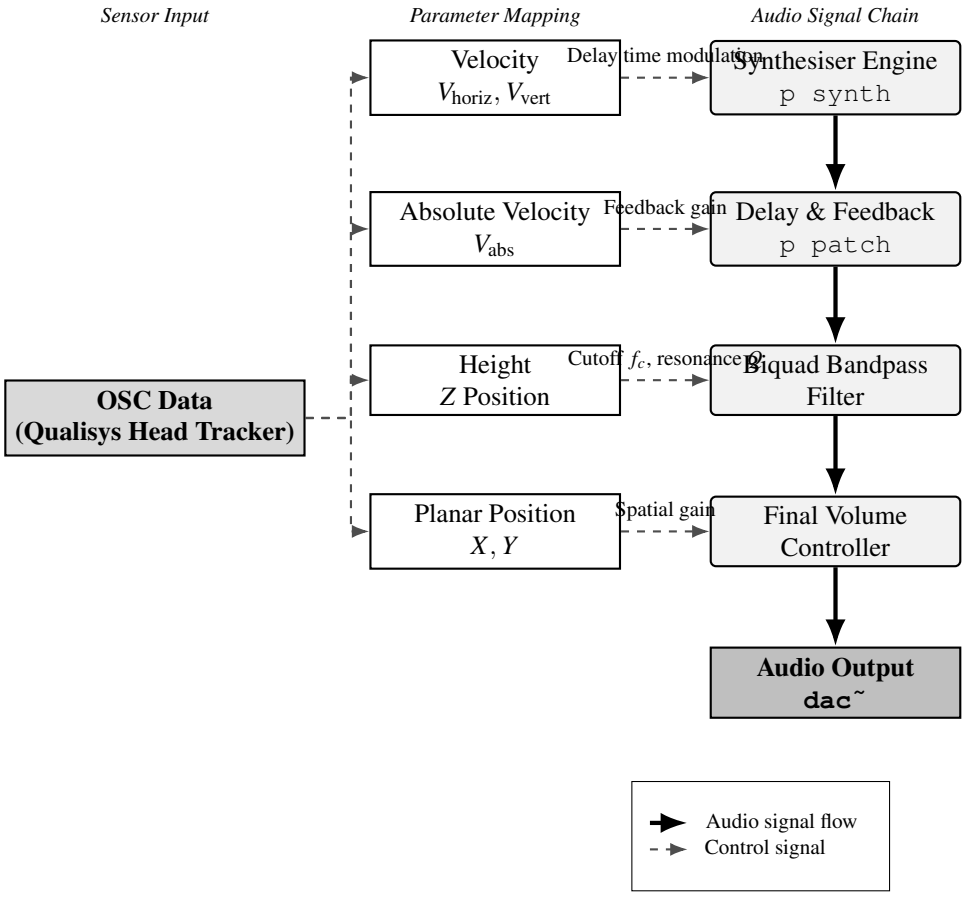


Figure 3.13: Block diagram of Sonic Wall architecture.

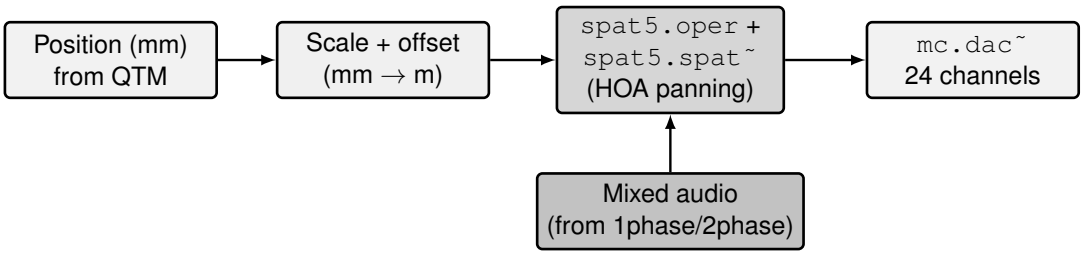


Figure 3.14: spatializer.maxpat

Chapter 4

User Study Protocol

This chapter details the design of the user study conducted within the fourMs Lab, run with thirteen participants: the experiential and narrative logic of the auditory installation, the sample and instruments used to evaluate it, and the standardised procedure followed with every participant. The study uses a mixed-methods design, combining quantitative data, motion-capture kinematics, and a pre-exposure questionnaire with qualitative data from a semi-structured post-exposure interview, so the objective record of how participants moved through the space can be read alongside their own account of what that movement felt like. This follows the same approach as the AMBIENT project uses to study movement in indoor soundscapes: motion or accelerometer data is paired together with participants' own descriptions of the space (Guo et al., 2025).

The central premise of the experience, and of the protocol built around it, is that a participant does not merely walk through the space but that their movement actively constructs it.

4.1 Physical Space

The empirical phase of this research was conducted within the fourMs Lab. The physical layout of the lab is divided into two spaces: the main laboratory room and the technical control room. The main room is where the participant experiences the experiment. Based on the footprint of the calibrated camera array (Table 3.4), the capture volume covers approximately 5.3 m $\times$ 7.4 m in plan. The room is treated with acoustic absorption curtains to minimise natural room reflections and keep natural light from entering. The acoustic control is necessary for the experience, as it prevents the physical acoustic properties of the real room from interfering with the virtual acoustic environment rendered by the 24-speaker array. The floor is left empty to maximise participant safety and to eliminate any tactile or physical boundary. The technical control room, by contrast, works as a monitoring centre and control point for data acquisition. It has the primary computing units running Qualisys Track Manager (QTM) and the Max/MSP patch. The spatial separation

between the Control Room and the Main Room isolates the researcher from the participant, ensuring that the participant can neither see nor hear the researcher during the three-phase exposure protocol. This separation is intended to heighten the participant's sense of isolation, agency, and embodiment within the virtual space.



Figure 4.1: Technical Control Room.

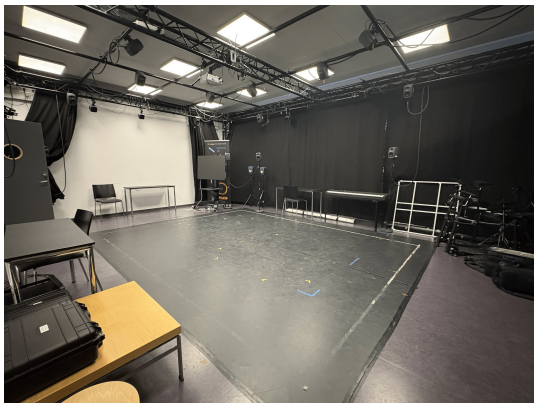


Figure 4.2: Main Room

4.2 Instruments

4.2.1 Highly Sensitive Person (HSP) Questionnaire

The pre-exposure questionnaire included items adapted from Aron's Highly Sensitive Person (HSP) scale (Aron and Aron, 1997), used to evaluate participants' sensory processing sensitivity before exposure. The scale was recommended by Jonna K. Vuoskoski (RITMO Centre, University of Oslo)[1], and was included because sensory processing sensitivity offered a way to relate a participant's own trait-level tendencies to how they were observed exploring the space, read alongside what they later described in the interview. This choice follows the same general logic behind RITMO's own use of individual-difference personality measures to predict movement responses

[1] Personal communication

to sound, seen for instance in work linking empathic concern to spontaneous movement responses to music (Zelechowska et al., 2020), even though that study used a different instrument, the Interpersonal Reactivity Index, rather than the HSP scale itself. Cross-referencing HSP scores with movement behaviour and interview content was not carried out within the scope of the present study and is left for future work. The version administered via Nettskjema included 19 of the 27 items in Aron's original scale. Questions belonging to several published subscales (May et al., 2020) were only partially represented, so canonical subscale scores cannot be reconstructed, and only a group-level summary is reported (Table 4.1).

Table 4.1: HSP Item Responses ($n = 12$; 1–7 scale)

No.	Item	*M*	*SD*
1	Easily overwhelmed by strong sensory input	4.00	1.95
2	Aware of subtleties in the environment	5.33	0.99
3	Affected by other people's moods	5.25	0.97
4	More sensitive to pain	4.25	1.22
5	Needs to withdraw during busy days for relief from stimulation	4.50	1.45
6	Sensitive to the effects of caffeine	3.92	1.93
7	Overwhelmed by bright lights, strong smells, coarse fabrics, or sirens	3.50	1.98
8	Has a rich, complex inner life	5.50	1.24
9	Uncomfortable with loud noises	4.42	1.73
10	Deeply moved by the arts or music	6.33	0.65
11	Nervous system feels frazzled, needs to be alone	4.75	1.77
12	Conscientious	5.58	1.00
13	Startles easily	4.67	1.44
14	Gets rattled with a lot to do in a short time	4.75	1.82
15	Knows how to make an uncomfortable physical environment more comfortable	4.42	1.51
16	Annoyed when asked to do too many things at once	4.83	1.40
17	Tries hard to avoid mistakes or forgetting things	5.17	1.80
18	Avoids violent movies and TV shows	1.92	1.17
19	Unpleasantly aroused when a lot is going on	3.75	1.71
Mean across administered items		**4.57**	—

Item 10 ("deeply moved by the arts or music," $M = 6.33$, $SD = 0.65$) produced the highest and most tightly clustered response; item 18 ("avoids violent movies and TV shows," $M = 1.92$, $SD = 1.17$) the lowest, consistent with a sample self-selected for engagement with sonic/artistic material (Table 6.2).

4.2.2 Motion Capture Metrics

Position, velocity, and Quantity of Motion (QoM) were derived from the Qualisys motion-capture system during exposure; the full processing pipeline for these signals is described in Chapter 5.

4.2.3 Post-Exposure Interview

A semi-structured qualitative interview was conducted with each participant immediately after exposure, designed to capture their felt experience of presence, embodiment, and narrative comprehension. Fully audio-transcribed interviews are available for all participants. Table 6.4 summarises each in one or two sentences, and a fuller thematic reading is given in Section 6.2.7.

The interview guide was developed together with the researcher and psychologist Lena Bauer-Wolf, and its seven questions move from concrete detail toward broader reflection. The first two questions anchor the interview in specific, observable parts of the experience: memorable sounds and what drove the participant to explore. The next three turn to how the experience felt in the body: emotional response, bodily involvement, and the sense of being immersed in the space. The sixth question asks about the participant's sense of control over the auditory environment. The final question is left open for anything that the earlier questions did not cover. This progression, from what happened, to how it felt, to how much agency the participant felt they had, follows directly from the presence, embodiment, and narrative comprehension the interview set out to capture.

The interviews were recorded using a microphone and transcribed automatically, using Autotekst's implementation of OpenAI's Whisper V3. These automatic transcripts may contain recognition errors and do not distinguish who is speaking, since the tool returns one continuous stream of text rather than labelled turns. The full transcripts are reproduced in Appendix A, close to verbatim, with only broken line breaks merged into readable paragraphs. The transcripts were reviewed by listening to each interview separately before making the analysis.

The interview guide was as follows:

1. During the experience, did you experience any specific sounds that felt memorable to you? (Follow-up: where in the room?)

2. What aspects made you want to explore the room, and how did you choose which parts to explore?

3. Did you experience any specific emotions during the experience? What were they?

4. How did you experience your body during the listening experience? Did specific parts of your body feel like they were part of the listening experience?

5. To what extent did you feel like you were one with the auditory space, or immersed in it?

6. To what extent did you feel in control of the auditory space?

7. Is there anything else you would like to add about the experience?

4.3 3-Phase Narrative Design

4.3.1 Three-Phase Approach

The experience's narrative and psychological progression is divided into three phases, guiding the participant from a lower-stimulus environment toward a more complex, interactive one. Because the system presents itself exclusively through sound, participants must learn, in real time, how their movement maps onto sensory feedback. Structuring the experience into different phases both reduces the risk of an overwhelming initial environment and lets this learning curve develop smoothly, limiting cognitive fatigue. The experience takes place in a semi-dark environment to induce sensory re-weighting: suppressing visual input compels the nervous system to rely more heavily on auditory and proprioceptive cues.

4.3.2 Phase 1: Sensory Calibration and Immersion (3 Minutes)

The first phase has the objective of lowering the participant's perceptual baseline and inducing in-depth listening. It promotes stillness and introduces the interaction's fundamental rule: the participant's movement is registered by the system, and they hold control over the sonic experience, establishing their first sense of physical presence in the space while their auditory and proprioceptive senses adjust. The phase is built around a single circular zone at the centre of the physical room, representing one recorded room from the Intercontinental Dataset. Sound is audible only while the participant stands still at the centre of the circle, where positional sensitivity is set very high (Figure 4.3). Building a phase around induced stillness, rather than free movement, follows RITMO's own standstill research: an instructed standstill makes involuntary micromotion easier to observe, and that micromotion can itself become material for an interactive or artistic system (Jensenius et al., 2014).



Figure 4.3: Phase 1 room layout: a single circular zone at the centre of the physical Main Room.

4.3.3 Phase 2: Spatial Agency (3 Minutes)

The second phase has the objective of inducing curiosity and increasing spatial exploration. A clear sonic cue marks the transition from Phase 1, signalling that changing in interaction. The single zone of Phase 1 is replaced by four overlapping elliptical zones, each a distinct soundscape from a different country. In the overlap between two circles, both soundscapes are audible simultaneously, letting the participant experience the acoustic boundary between two cultural environments directly. This overlap region is analysed quantitatively in Section 6.2.3 as the "simultaneous zone" (Figure 4.5) and is independently described by several participants as the most actively investigated part of the space (Section 6.2.7).

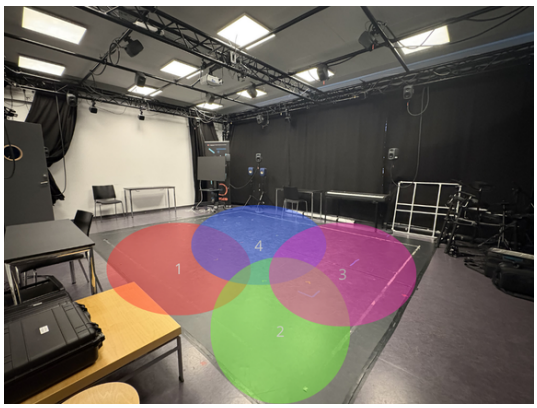


Figure 4.4: Phase 2 room layout: four overlapping rooms.

4.3.4 Phase 3: Complex Architecture and Environmental Anticipation

In the third phase, the participant has maximum agency over the soundscape. Unlike the Phase 1 to Phase 2 transition, no explicit cue marks the boundary into Phase 3. This phase also doesn't have a fixed duration: participants remain as long as they wish, in contrast to the fixed three-minute durations of Phases 1 and 2, which is why measured Phase 3 durations vary so widely across participants (from 7 minutes to over 38 minutes; Chapter 6). The objective of this phase is to push the boundaries of interactivity, contrasting rhythmic and non-rhythmic soundscapes while testing the limits of participant control. By introducing the soundmarks (Figure 4.5), this phase is sonically a lot more complex, inviting more exploration and interaction with the system.

4.4 Step-by-Step Experimental Procedure

4.4.1 Pre-Exposure Phase

Participants completed a Nettskjema questionnaire (University of Oslo) comprising informed consent, demographic items (age, gender, self-reported nationality, and self-reported musical background), and the 19-item HSP measure (Section 4.2.1). Self-reported nationality is the basis for the cultural-background variable reported in Table 6.3. The consent statement described the

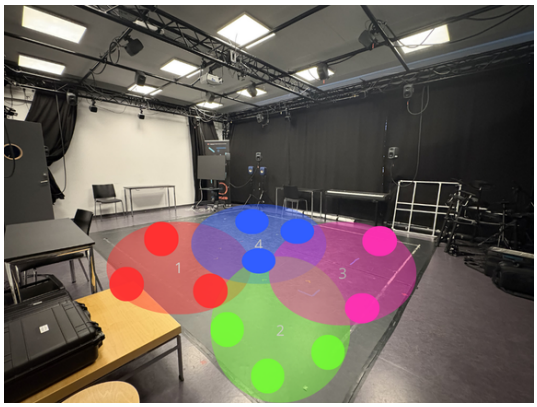


Figure 4.5: Phase 3 room layout: four overlapping rooms with soundmarks.

study's focus on embodiment, agency, and human-environment relationships in a space experienced entirely through sound. Participation was voluntary, restricted to adults 18 or older, with the right to withdraw data up until preprocessing. Participants then underwent onboarding: fitting of the motion-capture suit and placement of a Qualisys sports marker set on the arms, legs, and head, followed by system calibration. Height and weight were collected as calibration parameters for the Qualisys skeleton model; these fields do not appear among the demographic items in the Nettskjema results, so it remains unclear whether they were recorded separately during suit fitting.

4.4.2 The Briefing and Exposure Phase

Every participant received an identical, standardised verbal briefing before exposure:

> *"Thank you so much for taking part in the experience. First, we are going to ask you to fill out a form, then dress you in the mocap suit and place the sensors, then perform a calibration of the system, and then start the experience. In the end, we are going to ask you a few questions about your experience." "You will experience an audio experience in the centre of the room, and after three minutes it will change to the whole room. After another three minutes, a new phase will start. You can stay in the room for as long as you desire, and can leave when you feel ready. There are no right or wrong movements or experiences; the objective is to have fun."*

The ambient lighting condition of the Main Room is treated as a controlled variable. Participants are given the autonomy to choose between fully extinguished lighting and a heavily dimmed setting, a choice designed to suppress dominant visual stimuli and shift the balance of sensory processing toward the auditory pathway.

During exposure, the participant remained alone in the Main Room for the full session while the researcher monitored remotely from the Control Room, until they chose to re-enter the Control Room, at which point exposure was considered complete. This spatial separation removes the researcher's physical presence from the participant's perceptual field, mitigating the observer effect. However, one participant (13-N) reported a residual, self-directed performance pressure

tied to their professional essence despite the researcher's physical absence (Table 6.4), suggesting spatial separation alone may not eliminate every form of observer-effect-like self-consciousness (Section 6.2.7).

4.4.3 Post-Exposure Phase

Immediately after the experience, participants took part in the semi-structured interview described in Section 4.2.3, eliciting their felt experience of presence, embodiment, and narrative comprehension across the three phases; the summarised transcripts are given in Table 6.4, and a fuller thematic reading is given in Section 6.2.7.

4.5 Summary

This chapter described the mixed-methods user study: 13 participant sessions, screened for the absence of movement, hearing, and uncorrected vision impairment; a pre-exposure questionnaire combining demographic items and a 19-item HSP adaptation; a three-phase experiential structure moving from single-room stillness to multi-room exploratory agency to open-ended anticipatory interaction; a standardised procedure of briefing, isolated exposure, and semi-structured debriefing, whose guide has now been reconstructed from the transcripts themselves (Section 4.2.3); and full or partial interview data for 12 of the 13 participants, summarised in Table 6.4 and read against the movement data in Chapter 6.

Chapter 5

Methods for Data Analysis

This chapter describes how the quantitative data collected during the exposure phase (Chapter 4) were processed, cleaned, and analysed.

5.1 Data Sources

Three independent data streams were recorded per participant: a behaviour log written directly by the Max/MSP patch, containing a head-mounted tracker position alongside the participant's phase, room, and sound-trigger state on every row; a Qualisys Track Manager (QTM) motion-capture export of the full sports marker; and a pipeline that processes these files, which was built specifically for this dissertation. This chapter summarises its logic, the data-quality problems it was built to handle, and the statistical procedures applied to its output.

The guiding metric throughout is Quantity of Motion (QoM), the frame-to-frame Euclidean displacement of a tracked body point, following the usage established in Section 2.4.4.1 (Jensenius, 2024). Framing movement this way keeps the analysis consistent with the action-perception loop described in Section 2.5.6: displacement is not treated here as an incidental by-product of the interaction, but as the directly observable half of the loop, the continuously produced action that both responds to, and in turn changes, the sonic environment (Section 2.1.3).

5.1.1 Behaviour Log

The behaviour log is a plain-text file with 52 columns per row, written once per simulation frame by the Max/MSP patch. Each row contains the position (x, y, z) of a head-mounted tracker, logged independently and in real time by the interaction patch; the participant's current phase (three flag columns); their occupancy of each of the four virtual rooms (four flag columns); a fifth flag marking the "simultaneous" state, active when two rooms are triggered at once; the state of 27 sound-trigger columns; and an embedded hour/minute/second/millisecond timestamp. Because the head-tracker position and the phase/room/sound state are written to the same row by the same

process, they are intrinsically time-aligned and require no cross-system synchronisation, unlike the QTM export below.

5.1.2 QTM Motion Capture

Each participant's full-body Qualisys export was also retained, in the labelled-marker `.tsv` format described in Chapter 3. This export records every marker's 3D position at the system's native frequency, on a separate internal clock from the behaviour log, with data quality that varies by participant depending on marker occlusion. The full-body Sports marker set (Qualisys AB, b) was used, rather than tracking only the head, because a complete labelled marker set is what Qualisys's own Calqulus pipeline needs to compute joint angles and other biomechanical variables from motion-capture data (Qualisys AB, a). This full-body data was not analysed within the scope of the present study, being left for future work.

5.1.3 Audio Recordings

The audio output was recorded for each session, separated by phase (Phase 1 and Phase 2). Recordings were made in four channels to preserve the spatialisation of the sound. The audio was captured exactly as it was delivered to the participant, with no post-processing applied. These recordings exist so the 3D audio can later be cross-referenced with the movement data, comparing qualities such as frequency content and loudness against a participant's direction and amount of movement, the same kind of cross-modal analysis RITMO researchers have used elsewhere (Guo et al., 2025). This cross-referencing was not carried out within the scope of this thesis and is left for future work.

5.2 Participants and Data Completeness

Thirteen participant sessions were recorded, matching the total reported in Chapter 4. Data completeness varies across participants (Table 5.1, reproduced from Chapter 4). One participant (1-T) has no QTM file; the behaviour log's own head-tracker signal is used instead, as it is for every participant but one. Three participants (2-A, 11-L, 13-N) have an unreliable behaviour-log clock, though the failure differs in severity between them and was handled differently as a result. For 11-L and 13-N, the log contains a full run of rows but its timestamp column never advances, so their elapsed time was approximated from row index using the empirical median logging rate (96.57 Hz) observed across participants with a working clock; this approximation is flagged throughout and is a coarser measurement than the directly timed sessions. For 2-A, the failure is more severe: her behaviour log is truncated to only six logged rows, too few for the row-index approximation to be meaningful, so her session duration was instead taken from the separately recorded session audio file and her QoM from the QTM chest-marker fallback rather than from the (effectively absent) headpiece log, as Table 5.1 reflects. QTM tracking gaps range from 4.9% to 43.3% across the twelve participants with a mocap file; none exceeds the 50% threshold used elsewhere to flag a severely reconstructed trajectory. One participant, 9-S, lost a motion-capture sensor during the

session (Chapter 4); her session timing was unaffected, since it is derived from the behaviour log rather than QTM, but her full-body marker export should be treated as less reliable than other participants' for any analysis that uses it directly rather than as a secondary cross-check.

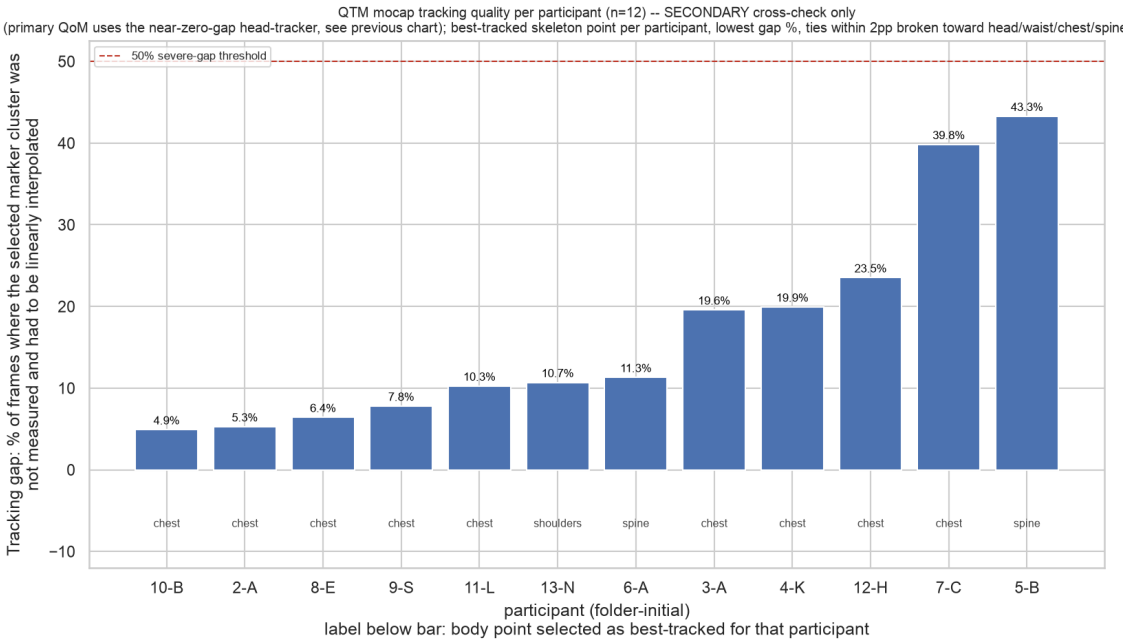


Figure 5.1: QTM tracking quality per participant. Bars show the percentage of frames requiring linear interpolation for the best-tracked available marker cluster per participant, with that cluster labelled below each bar.

5.3 Preprocessing Pipeline

5.3.1 Marker Selection for Motion Capture

Instead of assuming the head marker is always the best tracker, every anatomically plausible marker group (head, waist, chest, spine, shoulders) was evaluated for every participant, and the group with the lowest interpolation-gap percentage was selected, with ties within 2 percentage points resolved toward the theoretically preferred group (head, then waist, chest, spine, shoulders). In practice, the chest marker, a single, centrally placed sternum marker, was selected for eight of the twelve QTM participants, the shoulders for one, and the spine for two; a fixed head preference, as an earlier version of the pipeline used, would have kept a measurably worse signal in most participants (one participant's head cluster carried an 11.7% gap against 4.9% for chest). This QTM-derived signal remains a secondary cross-check against the near-gap-free head-tracker signal, not the primary movement measure.

Table 5.1: Data completeness per participant

Folder	Participant	QoM source	QTM gap %	Duration	Notes
1	T	headpiece	—	14:46	no QTM file
2	A	QTM (chest, fallback)	5.3	17:54	log truncated (6 rows)
3	A	headpiece	19.6	16:49	—
4	K	headpiece	19.9	18:28	—
5	B	headpiece	43.3	7:36	shortest session
6	A	headpiece	11.3	9:28	—
7	C	headpiece	39.8	21:01	—
8	E	headpiece	6.4	16:30	—
9	S	headpiece	7.8	16:13	lost a sensor during session
10	B	headpiece	4.9	26:18	best QTM tracking
11	L	headpiece	10.3	12:56	clock never advances; approximated
12	H	headpiece	23.5	45:01	longest session
13	N	headpiece	10.7	14:02	clock never advances; approximated

5.3.2 Room Occupancy and the Simultaneous Zone

The behaviour log's own room1–room4 occupancy columns were found to be zero in every row for all 13 participants: room location was never written to these columns during data collection due to an error in the system. Room occupancy is instead derived geometrically, from the head-tracker's x/y position against four elliptical zones (semi-axes $r_x = 1.44$ m, $r_y = 1.10$ m) centred on the physical room coordinates supplied by the researcher, matching the room-trigger geometry defined directly in the `2phase.maxpat` patch (`nodes` object, Chapter 3) rather than an independently chosen radius. Because adjacent ellipses overlap along a shared band (Figure 5.2), two room-membership rules are used depending on the question being asked. An *exclusive* assignment gives each frame to its single nearest room centre and is used for any comparison that should count a frame toward one room only. An *inclusive* assignment instead gives a frame to every room whose ellipse contains it, and is used to capture time spent in the overlapping band itself; because a frame in that band is counted toward both adjacent rooms, inclusive per-participant room times can sum to more than the session's total duration. The behaviour log's own `simultaneous` flag is used to identify the "between rooms" state in which two rooms are jointly active.

5.3.3 Quantity of Motion

QoM is computed as the frame-to-frame Euclidean displacement of the selected tracking point, $\sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}$, following Jensenius's usage of the term as a summary measure of a moving body's total kinematic energy (Jensenius, 2024). The head-tracker signal is the primary source for all but one participant, both for its near-zero gap rate and because it requires no cross-system time alignment with the phase/room/sound data it is compared against, being written to the same row by the same process. Short gaps in either signal (rows where a marker or the head tracker returns exactly zero on all three axes) are linearly interpolated before QoM is computed, and the resulting gap percentage is reported alongside every participant's QoM figures (Table 5.1, Figure 5.1).

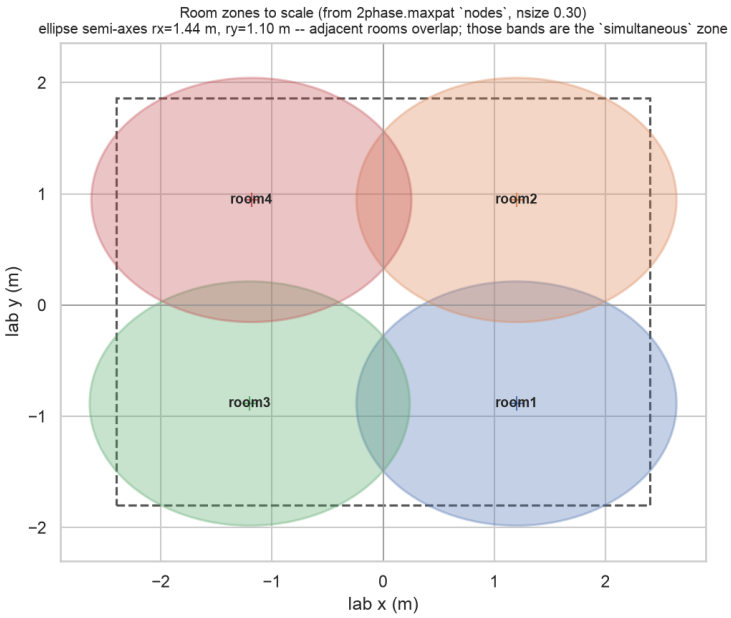


Figure 5.2: The four room zones to scale, drawn from the `2phase.maxpat` `nodes` geometry ($r_x = 1.44$ m, $r_y = 1.10$ m). The shaded bands where adjacent ellipses overlap are the zones that the inclusive assignment and the `simultaneous` flag both pick out.

QoM was chosen instead of a discrete gesture classifier or a raw acceleration trace. A gesture classifier needs a fixed set of named gestures, and the movement studied here is different: small, continuous postural adjustment rather than distinct actions. A raw trace keeps every fluctuation in the signal, but gives no single number to compare across participants and phases. QoM solves both problems by reducing continuous movement to one comparable value per frame. This also matches how the AMBIENT team has used the metric before: Guo, Tørresen, and Jensenius compute a similar frame-to-frame QoM from accelerometer data to study how indoor soundscapes relate to involuntary micromotion during stillness (Guo et al., 2025), and Gonzalez-Sanchez, Zelechowska, and Jensenius use the same continuous-displacement approach for music and involuntary micromotion during standstill (Gonzalez-Sanchez et al., 2018).

5.4 Metrics and Statistical Procedures

5.4.1 Phase Duration and Room Dwell Time

The duration of a phase is the accumulated time its flag column is active, the sum of per-row time deltas while the flag equals 1, rather than the span between its first and last occurrence. This distinction matters because several participants' phase flags toggle off and back on, or briefly overlap with an adjacent phase, rather than form one clean contiguous block; participants 7-C and 9-S, among others, showed a first-to-last span noticeably longer than their accumulated active time. The room dwell time and the simultaneous-zone dwell time are computed in the same way.

5.4.2 Sound-Onset Response

Each rising edge of one of the 27 sound-trigger columns during Phase 3 was treated as an onset. For every onset, mean QoM in the one second immediately preceding it was compared with mean QoM in the two seconds immediately following it. Onsets without a full pre- and post-window available were excluded. The event table contained a small number of exact duplicate rows (identical participant, sound, and onset time), removed before analysis; the deduplicated table used here contains 1,244 onset events across 12 participants.

5.4.3 Statistical Tests

Descriptive statistics (mean, SD, range) are reported for all metrics. For paired comparisons of theoretical interest, given the small and non-independent sample, both a paired t-test and its non-parametric equivalent, the Wilcoxon signed-rank test, were run, with the more conservative result emphasised in the text and Cohen's d reported as an effect size. Three such comparisons were performed: per-frame QoM in Phase 1 versus Phase 3, QoM rate in the simultaneous zone versus the rest of the space, and QoM immediately before versus after sound onset. Two correlations were also computed, session duration against total QoM and room dwell time against cumulative in-room QoM, using Pearson's r and Spearman's ρ given the plausibly non-normal distribution of movement data. No correction for multiple comparisons was applied, a limitation discussed further below and in Chapter 6.

5.5 Limitations of the Analytical Pipeline

Four limitations should be kept in mind when interpreting Chapter 6. First, the sample available for any given test is smaller than the full 13 participants, since the tests require both a working clock and, for QoM comparisons, a mocap or head-tracker signal. The Phase 1 versus Phase 3 comparison, for instance, draws on 11 participants. Second, the approximated timeline used for participants 2-A, 11-L, and 13-N assumes a constant logging rate, or in 2-A's case an audio-derived duration, and cannot correct for any recording pause, so their timing is coarser than the other ten participants'. Third, total QoM is strongly correlated with session duration (Section 6.2.1), so raw totals conflate how much a participant moved with how long they stayed; QoM *rate* is used wherever a fair between-participant comparison is intended. Fourth, no multiple-comparisons correction was applied across the five statistical tests reported in the next chapter, so the reported p-values are indicative rather than confirmatory, consistent with the exploratory character of this analysis. This choice, rather than a simple oversight, follows from what a correction would do at this sample size: a Bonferroni or Holm correction across five tests would require a per-test threshold near $p < 0.01$ for the family-wise error rate to stay at the conventional 0.05, and with only 11–13 participants per comparison, a test already needs a fairly large effect to reach significance at the uncorrected $p < 0.05$ threshold, so a corrected threshold would leave the analysis powered to detect only very large effects, in effect discarding the exploratory, hypothesis-generating results

this chapter reports without the sample size to properly confirm or reject any of them at a stricter threshold. The single significant, uncorrected result reported here (the simultaneous-zone comparison, Section 6.2.3) is accordingly presented throughout as motivating a pre-registered, adequately powered replication, not as a confirmed effect, and every reported p-value should be read with this in mind. Room occupancy, finally, is derived geometrically from a single head-tracker point rather than logged directly, since the system's own room columns were never populated.

Chapter 6

Results and Discussion

This chapter documents the findings from the analysis pipeline described in Chapter 5, discusses them against the theoretical framework in Chapter 2 and the qualitative interview data introduced in Chapter 4, and concludes with limitations and reflections. The results are exploratory throughout: the sample is small (11–13 participants depending on the comparison), no multiple-comparison correction was applied, and several participants' timing is approximated rather than measured (Section 5.5). They are reported in full nonetheless, including where a finding runs against the intuitive expectation, because a mixed-methods study of this kind is only useful if its quantitative arm is reported honestly alongside its qualitative one.

6.1 Participants

Thirteen participant sessions were recorded in total (Chapter 5). The pre-exposure Nettskjema questionnaire was reported below. Inclusion required the absence of any movement impediment and the absence of hearing or vision impairment, with corrected vision (e.g. glasses) permitted.

Table 6.1: Participant Demographics ($n = 13$ questionnaire respondents)

Variable	Statistic	Detail
Age	$M = 31.8$, $SD = 3.5$	Range: 27–38 years
Gender	Male: 11 (84.6%) Female: 2 (15.4%) Non-binary / Prefer not to say / Other: 0 (0%)	

The sample skewed toward participants with some degree of musical training: seven of 13 (53.8%) identified as Amateur musician or above, relevant given that musical training is independently associated with auditory attentional and localisation skills. Table 6.3 adds a second layer of description not captured by the two tables above: each participant's self-reported nationality, collected as a demographic item on the pre-exposure Nettskjema alongside age, gender, and musical

Table 6.2: Participant Musical Background ($n = 13$)

Self-Reported Category	Count	%
Nonmusician	2	15.4
Music-loving nonmusician	4	30.8
Amateur musician	0	0.0
Serious amateur musician	3	23.1
Semiprofessional musician	2	15.4
Professional musician	2	15.4

background (Section 4.4.1), and the lighting condition they chose, cross-referenced with session duration and total movement (Chapters 5–6). Nationality is used here as a practical, self-reported proxy for cultural background; it necessarily collapses a person's actual cultural formation, which may draw on places lived, studied, or raised rather than citizenship alone, into a single label, a simplification made explicit here rather than left implicit in the table.

Table 6.3: Participant nationality (self-reported, used as a proxy for cultural background) and lighting condition, cross-referenced with session duration and total QoM

ID	Nationality	Lighting	Duration	Total QoM (m)	QoM rate (m/s)
1-T	Italian	Off	14:46	378.1	0.427
2-A	US-American	Off	17:54	269.3	0.251
3-A	Chinese	Dimmed (on)	16:49	319.9	0.317
4-K	Japanese	Off	18:28	364.6	0.329
5-B	Norwegian	Off	7:36	141.8	0.311
6-A	Norwegian	Off	9:28	190.4	0.335
7-C	German	Off	21:01	171.4	0.136
8-E	Norwegian	On	16:30	283.9	0.287
9-S	Italian	Off	16:13	214.4	0.220
10-B	Hungarian	Off	26:18	251.3	0.159
11-L	Norwegian	Off	12:56	149.8	0.193
12-H	British (Scottish)	Off	45:01	636.3	0.236
13-N	Norwegian	Off	14:02	209.5	0.249

The sample is accordingly weighted toward Northern European (six of 13 participants were Norwegian) and, more broadly, European backgrounds, with single Chinese, Japanese, and US-American participants. This imbalance is worth bearing in mind when reading the cultural-resonance theme in Section 6.2.7, since it is drawn from a sample with limited cultural range rather than a deliberately stratified cross-cultural one.

6.2 Results

6.2.1 Session Duration and Overall Movement

Session duration varied across the 13 participants, from 7 minutes 36 seconds (5-B) to 45 minutes 1 second (12-H), mean 18.2 minutes, SD 9.3 minutes, consistent with the open-ended design of Phase 3, in which participants remain as long as they wish (Chapter 4). Total QoM ranged from

141.8 m to 636.3 m (mean 275.4 m, SD 132.6 m) and was strongly correlated with session duration (Pearson $r = 0.81$, $p < 0.001$; Spearman $\rho = 0.58$, $p = 0.039$; Figure 6.2). This is expected, since a longer session mechanically accumulates more displacement, and is why QoM *rate* (total QoM divided by duration) rather than raw total QoM is used for the between-participant comparisons that follow, including the per-room comparisons in Section 6.2.6. QoM rate itself varies more than threefold across the sample, from 0.136 m/s (7-C) to 0.427 m/s (1-T; Table 6.3, Figure 6.3), a first indication that participants engaged with the installation at very different overall tempos before any room-by-room breakdown is considered.

This more than sixfold spread in duration is itself worth pausing on, since nothing in the protocol forced it: Phase 3 has no fixed length, so the 7:36-to-45:01 range is a direct behavioural record of how long each participant chose to keep exploring. Two candidate explanations are worth setting out, though the present data can only partly adjudicate between them. The first is a trait-level one: since the pre-exposure questionnaire included Aron's HSP scale precisely to relate a participant's own sensory-processing tendencies to how they were later observed exploring the space (Section 4.2.1), sensory processing sensitivity, or the personality trait of openness more broadly, is a natural candidate predictor of how long someone stayed. That cross-reference was explicitly left for future work rather than carried out here (Section 4.2.1), since only a group-level HSP summary, not a reconstructable per-participant score linkable to duration, is reported in this dissertation; testing it directly is accordingly listed among the priorities in Chapter 7 rather than answered here. The second candidate explanation is a property of the soundscapes themselves rather than of the participant: the four rooms are not acoustically equivalent, and two of the four, the Dali exhibition room and the Doha food court, contain audible background music at specific points in the field recording (Table 3.1), while the Zhangjiakou market and Oslo airport recordings do not. A soundscape with a recognisable musical element plausibly invites a different, more sustained kind of engagement than one built entirely from environmental and speech sound, though this dissertation did not test session duration against room-level musical content directly, and the per-participant room-preference data (Section 6.2.6) shows attachment to a specific room tracking personal and cultural associations at least as often as any single acoustic feature. Both remain open, empirically tractable questions for the replication proposed in Chapter 7 rather than ones this exploratory sample can settle.

6.2.2 Movement Across the Three Phases

Mean per-frame QoM rose from Phase 1 to Phase 3 for 7 of the 11 participants with usable data in both phases (Phase 1 mean 0.00175 m/frame; Phase 3 mean 0.00257 m/frame), a moderate effect (Cohen's $d = 0.51$) that did not reach significance at this sample size (paired $t(10) = 1.70$, $p = 0.121$; Wilcoxon $W = 17$, $p = 0.175$; Figure 6.4). The direction is consistent with the intended progression from Phase 1's stillness-promoting design to Phase 3's higher-agency, faster-responding environment (Chapter 4).

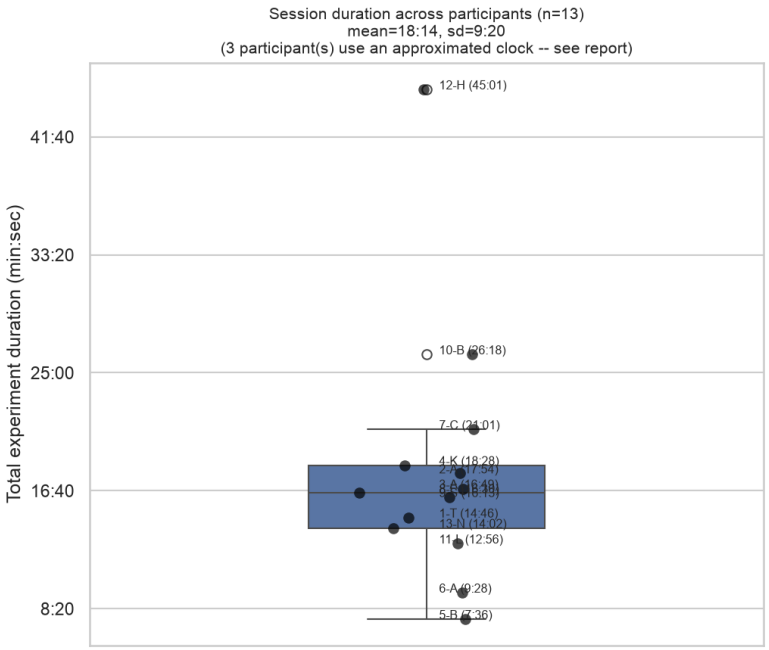


Figure 6.1: Session duration across all 13 participants, shown as a box plot with each participant's own duration labelled.

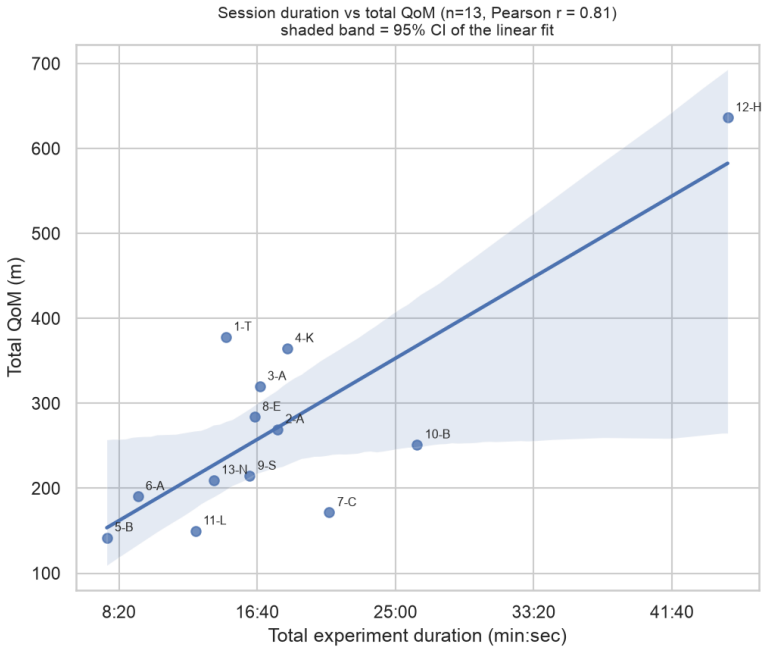


Figure 6.2: Session duration against total QoM, with each point labelled by participant and a linear fit (shaded band: 95% CI).

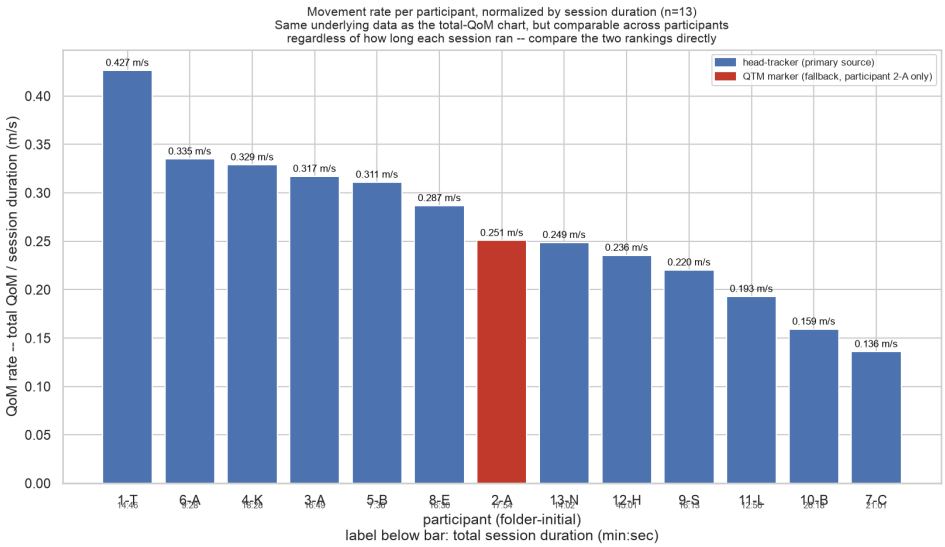


Figure 6.3: QoM rate (total QoM divided by session duration) for all 13 participants, sorted in descending order, with each participant's total session duration labelled below their bar. 2-A (red) uses the QTM chest-marker fallback rather than the head tracker.

6.2.3 The Simultaneous / Between-Rooms Zone

QoM rate was higher while the `simultaneous` flag was active, while a participant occupied the acoustic overlap between two rooms (Figure 4.4, Chapter 4), than in the rest of the space; participants moved on average around 0.12 m/s faster there Figure 6.6). Nine of thirteen participants moved faster in the simultaneous zone. Across the sample, it is the one part of the installation that reliably draws extra engagement: the four rooms themselves turn out to appeal very differently to different people (Section 6.2.6), but the overlap band shows up in the movement data for a clear majority of participants regardless. They also spent a modest but fairly variable proportion of each session there (mean 8.2%, coefficient of variation 91% Figure 6.7). This quantitative pattern is corroborated by the interview data (Section 6.2.7).

6.2.4 Movement Response to Sound Onsets

Across 1,244 deduplicated sound-onset events (Phase 3 only, 12 participants; Section 5.4.2), the change in QoM from the second immediately preceding a sound onset to the two seconds following it did not resolve into a single consistent direction: 44% of events showed higher QoM after the onset against 56% showing lower QoM, close enough to an even split that no reliable population-level increase or decrease emerged (Figure 6.8). At the participant level, the mean post-onset delta was similarly mixed, with some participants moving more after an onset and others less, rather than concentrated in one direction, unlike what would be expected if a sound onset reliably drew participants either into more active exploration or into a brief stillness response.

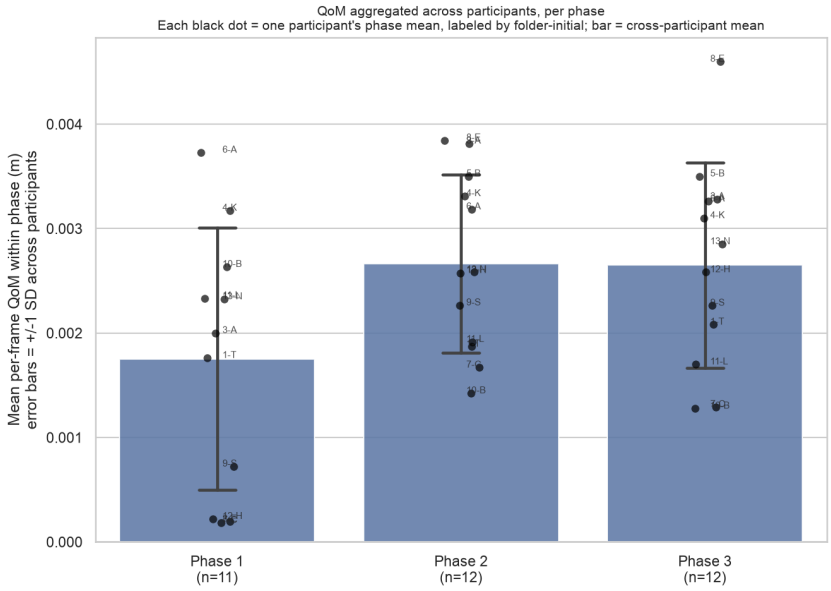


Figure 6.4: Mean per-frame QoM aggregated across participants, by phase, with individual participant means shown as labelled points and error bars showing ± 1 SD.

6.2.5 Room Dwell Time and Movement

Time spent in a room's zone and cumulative QoM while in that room were strongly correlated across the 52 room-by-participant observations available under the exclusive, nearest-room assignment (Pearson $r = 0.79$, $p < 0.001$; Figure 6.10). Expressed as a rate (QoM per second of dwell), the four rooms again show broadly similar movement rates with overlapping standard deviations across all four. The share of session time spent in each room is somewhat less even: room 3 accounts for a higher average share than the other three under the exclusive assignment, a pattern also visible under the inclusive assignment (Figure 6.9).

6.2.6 Individual Differences in Room Preference

Figure 6.11 shows the share of each participant's session spent in each room individually, and the spread within a single participant is far wider than the spread between the four group averages. 1-T spent roughly a third of his session in a single room while falling below 10% in several others. 9-S and 4-K each show a comparable single-room peak above 30%; and 5-B, 7-C, and 12-H each have at least one room occupying under 10% of their time. No participants share the same profile. A roughly even population average built out of these widely divergent individual distributions is what a personalised room preference looks like: rather than converging on a shared favourite room, or a shared least-favoured one, each participant appears to have formed their own, largely personal relationship to the four soundscapes.

Because adjacent rooms overlap (Section 5.3.2), every result in this section is reported under both room-membership rules introduced in Chapter 5: the exclusive assignment used above, which gives each frame to its single nearest room and so never double-counts a frame, and an inclusive

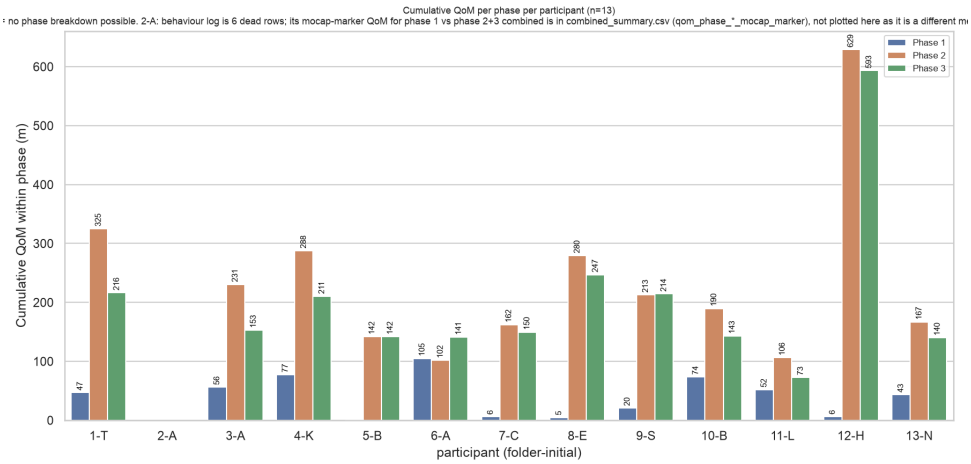


Figure 6.5: Cumulative QoM within each phase, per participant. Bars are blank where a participant's clock or mocap data did not allow a phase breakdown (Table 5.1).

assignment, which credits a frame to every room whose zone contains it and so can sum to more than a session's duration. The two rules disagree in absolute terms: inclusive room times are uniformly larger, since time spent in an overlap band is credited to both adjacent rooms at once, but every conclusion drawn from them here holds under both. Room 3 keeps the highest average share of session time under both the exclusive (Figure 6.9) and the inclusive assignment (Figure 6.13), and in both seconds (Figures 6.14 and 6.15) and percentage terms. More importantly, the per-participant picture of wide, individually specific variation looks the same under either rule (Figure 6.16 against Figure 6.11; Figure 6.17 against Figure 6.12): a participant who spends an unusually large or small share of their session in a given room under one rule does so under the other as well.

6.2.7 A Formal Analysis of the Interview Transcripts

The review and analysis of the qualitative data was conducted by researcher and psychologist Lena Bauer-Wolf, with a specialisation in experimental psychology.

She coded the transcripts according to Braun and Clarke's six-phase framework for reflexive thematic analysis (Braun and Clarke, 2006), moving from familiarisation and line-by-line coding through candidate-theme generation, review, and definition, to a final narrative synthesis. Consistent with that framework's reflexive premise, her own position as an analyst is treated as an interpretive resource rather than a source of bias to be eliminated, and the coding moves deliberately from the semantic content of what participants said, the specific sounds and movements they described, to the latent psychological states that content points to, such as a need for agency or a felt loss of temporal orientation. Four themes emerged from this analysis. This subsection reports Bauer-Wolf's own reading of what each theme means, since that reading is itself the result of her formal coded analysis rather than an inference added afterwards; Section 6.3 then draws these four themes together with the quantitative results reported above.

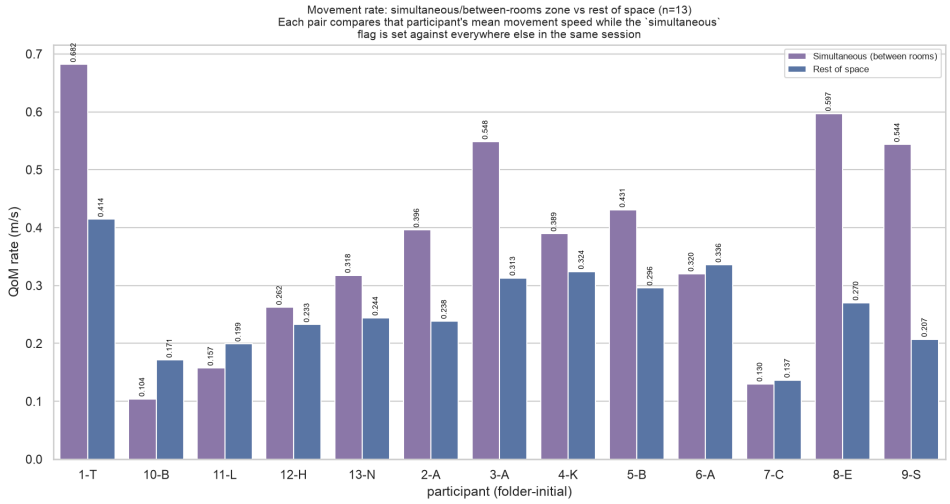


Figure 6.6: Movement rate (QoM per second) in the simultaneous/between-rooms zone versus the rest of the space, per participant.

The first theme, which Bauer-Wolf terms *interactive play and rhythmic agency*, describes participants' shift from passive observation to active experimentation once they had identified how the system responded to them. 3-A described a striking sensation of *teleportation*, in which moving between zones felt like abrupt displacement into a different sonic reality rather than a continuous walk between them; 12-H discovered a form of *vertical agency* by testing whether crouching would trigger sound, finding that verticality offered its own fine-grained control beyond horizontal movement. Several participants turned this agency into something closer to a rhythmic practice: 5-B synchronised his footsteps to a recorded footstep sound, effectively dancing with the audio, while 10-B moved from an analytical, exploratory stance to a playful one, deliberately toggling between sound loops to assemble an improvised beat. Bauer-Wolf reads this progression, from passive listening to treating the room as a rhythmic instrument, as evidence that geolocated sound triggers invite participants to build a mental map of the space using their own bodies as a reference.

The second theme, *nostalgic transport and anthropomorphic imagination*, concerns the way abstract or realistic sounds were transformed into personal narrative. 2-A's Phase 1 soundscape of waves and clatter reconstructed a specific, vividly recalled lunch in Mallorca; 8-E, a Norwegian participant, described feeling physically transported to Beijing by scrambled Chinese voices and traffic noise, an unexpected sense of belonging to an unfamiliar, imagined place. Distinct cultural markers, such as the Chinese mobile-payment notification sound recognised by 3-A, acted as strong anchors grounding the listener in a specific, recognisable social reality. Where a soundscape was more ambiguous, participants instead constructed spatial theories to make sense of it: 13-N hypothesised in turn a formal office, an industrial hallway, and a kebab shop, reasoning from the reverberation and surface quality of the recorded sound alone.

The third theme, the *embodied disconnect*, addresses the friction created by a system that tracks head position but is felt, by many participants, to call for a whole-body response. 12-H again

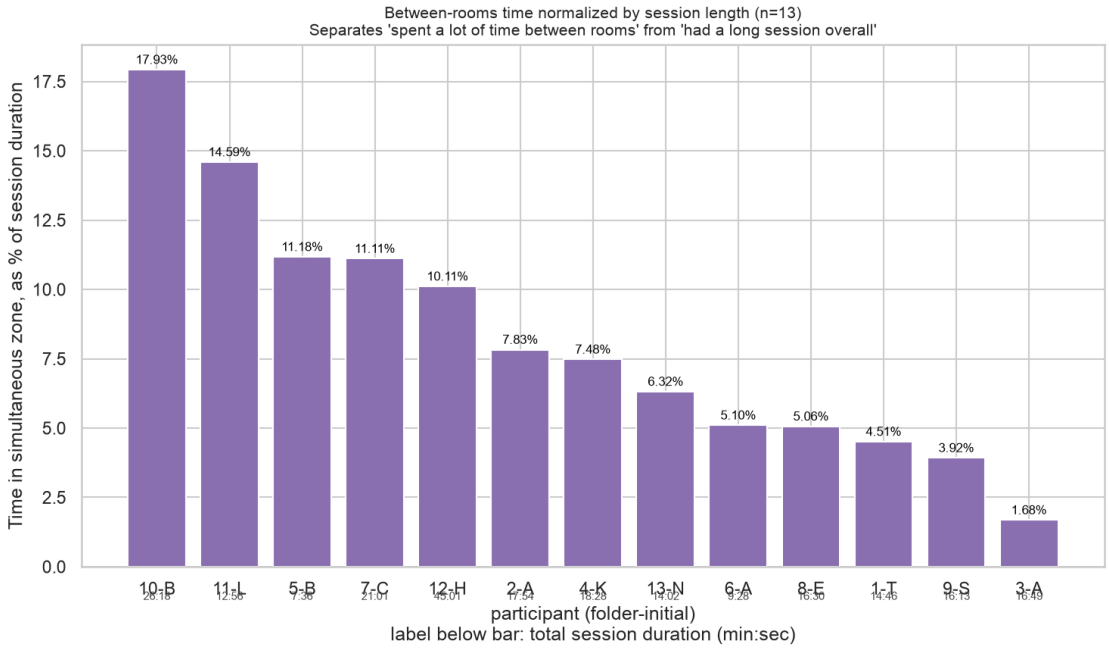


Figure 6.7: Time spent in the simultaneous zone as a percentage of each participant's total session duration, sorted descending, with each participant's total duration labelled below their bar.

supplied the analysis's clearest image, describing the neck as a "magnifying glass" or a "torch" that did the work of exploration while the rest of the body played a supporting role; 10-B, by contrast, felt that the torso, or centre of mass, ought to have been the primary driver of the interaction. 6-A and 7-C both reported disappointment at the absence of full-body tracking, describing themselves as outside observers rather than integrated participants, and several participants developed physical workarounds to compensate: 7-C had to lift one foot to reach a threshold of silence when the torso alone was not enough, 13-N ran in tight circles to test how quickly the audio would respond, and 8-E leaned the torso to trigger sounds without moving the legs at all. This theme sharpens, with formally coded support, the same head-versus-whole-body inconsistency flagged as an open question in the preliminary reading of the results above (Sections 6.2.6–6.2.7): on Bauer-Wolf's account it is better understood not as inconsistent system behaviour but as a mismatch between a head-tracked interaction design and participants' own, more whole-body expectations of how an interactive room should respond to them.

The fourth theme, *acoustic tension*, sets sensory overwhelm against silence as a place of recovery. 6-A and 7-C both reported a form of repetitive sound stress, in which the looping structure of the audio produced a loss of what Bauer-Wolf calls temporal autonomy, participants losing track of time and feeling captured by the noise; 11-L independently described the experience in similarly forceful terms as "drastic." Against this, the zone of silence at the centre of the room functioned for several participants as a deliberate recovery mechanism rather than a mere absence of sound: 12-H described it as a physical place for "sensory resetting," while 8-E and 12-H both used explicitly restorative language, calling it calm, peaceful, and a place to "breathe out."

Table 6.4: One-sentence summary of each transcribed post-exposure interview ($n = 13$ available)

ID	Summary
1-T	Reported losing his sense of time; described actively assigning a "personality" to each room, explaining this in terms of the mind being predisposed to look for stories, and finding the ambiguity of unidentified sounds engaging rather than frustrating.
2-A	Reported a vivid autobiographical memory triggered by Phase 1 ambient sound; distinguished soundscapes that felt socially "included"/homey (e.g. a train station) from ones that felt awkward or exclusionary (near-silent rooms with a single distinct sound).
3-A	Reported losing his sense of time and a sense of familiarity with the sounds; explicitly estimated feeling "in the room" only around 20% of the time, the rest spent "playing with" or "manipulating" it. Had some prior knowledge of the experience before taking part (see the caveat below).
4-K	Confused by the initial instructions; did not understand he could leave and remained until the researcher ended the session, by his account, after roughly 30 minutes
5-B	Reported an extensively exploratory, playful session, describing it at one point as "trippy" (mapping the whole space, synchronising his steps to a footstep sound, composing rhythmic patterns between trigger zones); felt in control throughout except one unannounced phase transition, and wished the experience were longer.
6-A	Reported curiosity giving way to anxiety, stress, and a feeling of being "captured" by increasing sound intensity in the later phases; explicitly disappointed that only head position, not the whole body, appeared to register.
7-C	Was, like 4-K, a little confused by the initial instructions; reported initial insecurity about how to use his body, later resolving into strong immersion, and found the loops' repetition and the loss of time-awareness stressful, but ultimately felt "very present."
8-E	Reported losing his sense of time and unexpected nostalgia for Beijing triggered by the soundscape despite a Norwegian background; described trying arm, lean, and tilt movements and found several sufficient without leg involvement.
9-S	Compared her experience to viewing art at the Venice Biennale, an observer rather than a participant, except during a footstep sound that alone felt embodied and co-present. One motion-capture sensor was lost during her session (see the data-quality caveat below), though her session timing itself remained accurate.
10-B	Initial self-conscious, "analytic" stance eased into playful exploration; reported the system responding mainly to hip/torso position rather than hand gestures, and described a shift from identifying sound sources to attending to their timbral evolution as loops repeated.
11-L	Reported comfort shifting to unease as sound transitions became less predictable; repeatedly and emphatically stated she never felt in control of the space.
12-H	Reported the neck as the primary locus of exploration, describing it as a "torch" pointing toward areas to investigate; oscillated between curiosity-driven exploration and calm during silent zones; longest session in the dataset.
13-N	Constructed an elaborate spatial theory of what each room represented (airport, market, office, industrial space); reported a self-conscious "need to perform" during the session despite the researcher's physical absence.

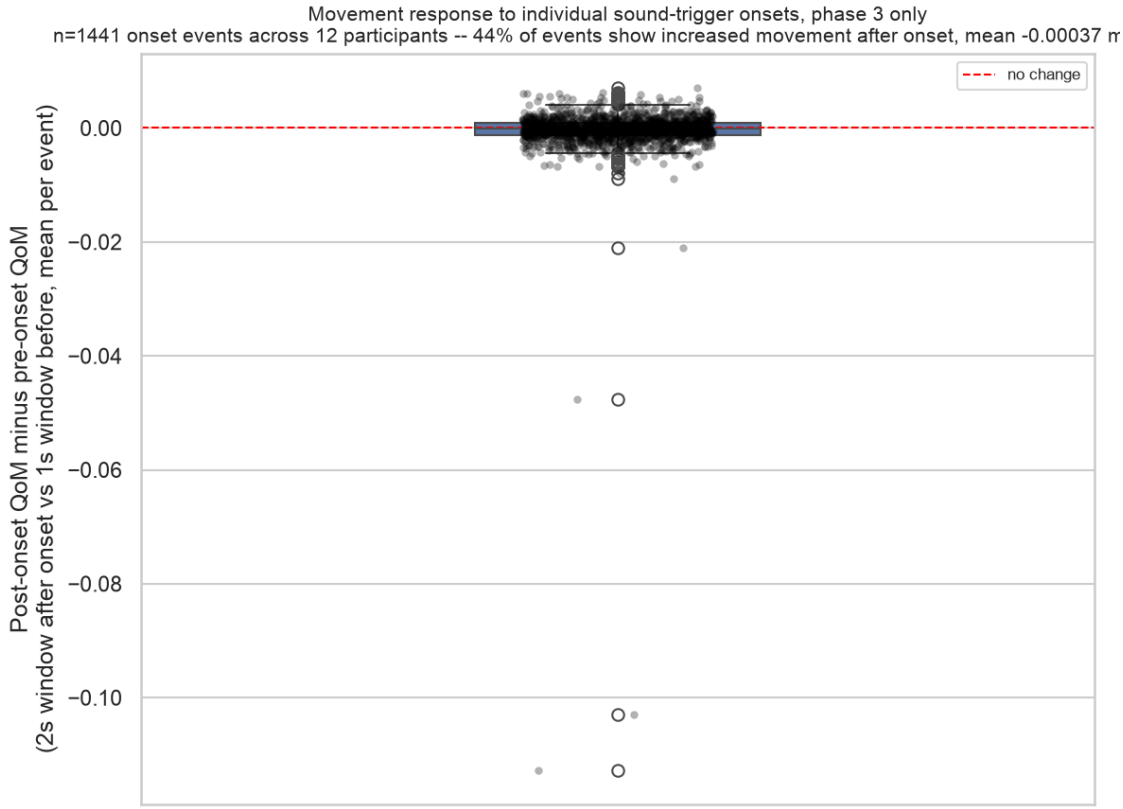


Figure 6.8: Change in QoM from a 1-second pre-onset window to a 2-second post-onset window, across all 1,244 deduplicated sound-trigger onset events in Phase 3.

6.3 Discussion

6.3.1 Faster Movement in the Simultaneous Zone

The significantly higher movement rate in the simultaneous, between-rooms zone is one of the clearer findings in this dataset, qualitatively corroborated by the interviews (Section 6.2.7): seven participants independently described boundary-seeking as their central exploratory strategy, several in strikingly similar terms ("video game dynamic," "mapping," a "grid"). Several elements of the theoretical framework offer converging explanations for why this zone in particular draws sustained attention. First, the zone is, by construction, an acoustic transition: a participant standing there hears two soundscapes simultaneously (Chapter 3), a more attention-demanding state than being settled within a single room's soundscape, consistent with everyday listening's orientation toward locating and resolving an ambiguous auditory scene (Section 2.1.2). Second, the zone sits at, or near, the sonic-wall boundary mechanism (Chapter 3), whose synthesised drone is designed to intensify with proximity to a boundary, exploiting the looming bias documented by Neuhoff (Section 2.3.2). Faster movement through this region is consistent with participants passing through an actively intensifying acoustic zone rather than lingering in it. Third, from an action-perception loop perspective (Section 2.5.6), a zone defined by ambiguity and change is,

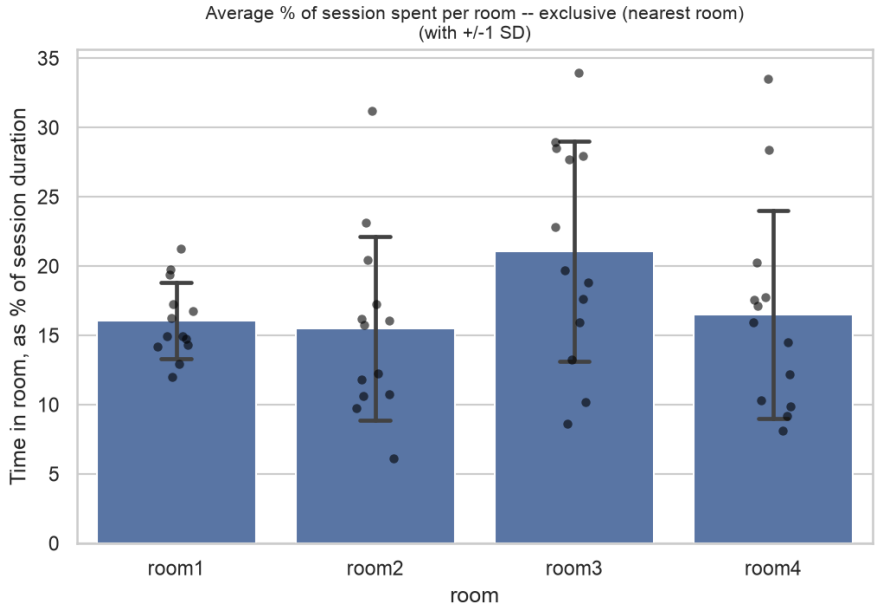


Figure 6.9: Average share of session time spent in each room across participants, under the exclusive, nearest-room assignment (error bars: ± 1 SD).

almost definitionally, one in which the coupling between movement and sonic feedback is most informative, and so most likely to sustain continued exploratory movement rather than settling. This can be proven by 10-B's own account of "settling on four-ish areas" and then deliberately hunting for further ones, and 12-H's description of "these shifts" as what made the experience "memorable," which describe exactly this kind of information-seeking behaviour in the participants' own words.

6.3.2 Movement Response to a Sound Onset

Unlike the phase and simultaneous-zone comparisons, the sound-onset comparison does not resolve into a single group-level direction: of 1,244 deduplicated onset events, 44% show higher QoM in the two seconds after the onset and 56% show lower QoM, close enough to an even split that no reliable population-level increase or decrease emerges, and the same mixed picture holds at the participant level. Three considerations are worth setting out, none of which the current data can fully adjudicate between. The first is methodological: the 1-second pre-onset window is short and may not fully separate a participant's baseline movement from movement already changing for unrelated reasons, which would add noise that a near-even split cannot easily be distinguished from a genuinely absent effect. The second is perceptual, and the interview data suggest the underlying picture is a genuine mixture of both directions rather than a single, hidden effect obscured by noise: some participants describe a momentary halt consistent with an orienting response, 2-A recalled a sneeze or cough that "took me off guard... it just surprised me," and 7-C described a bell-like sound that "made me laugh" after catching him by surprise, both accounts of a brief

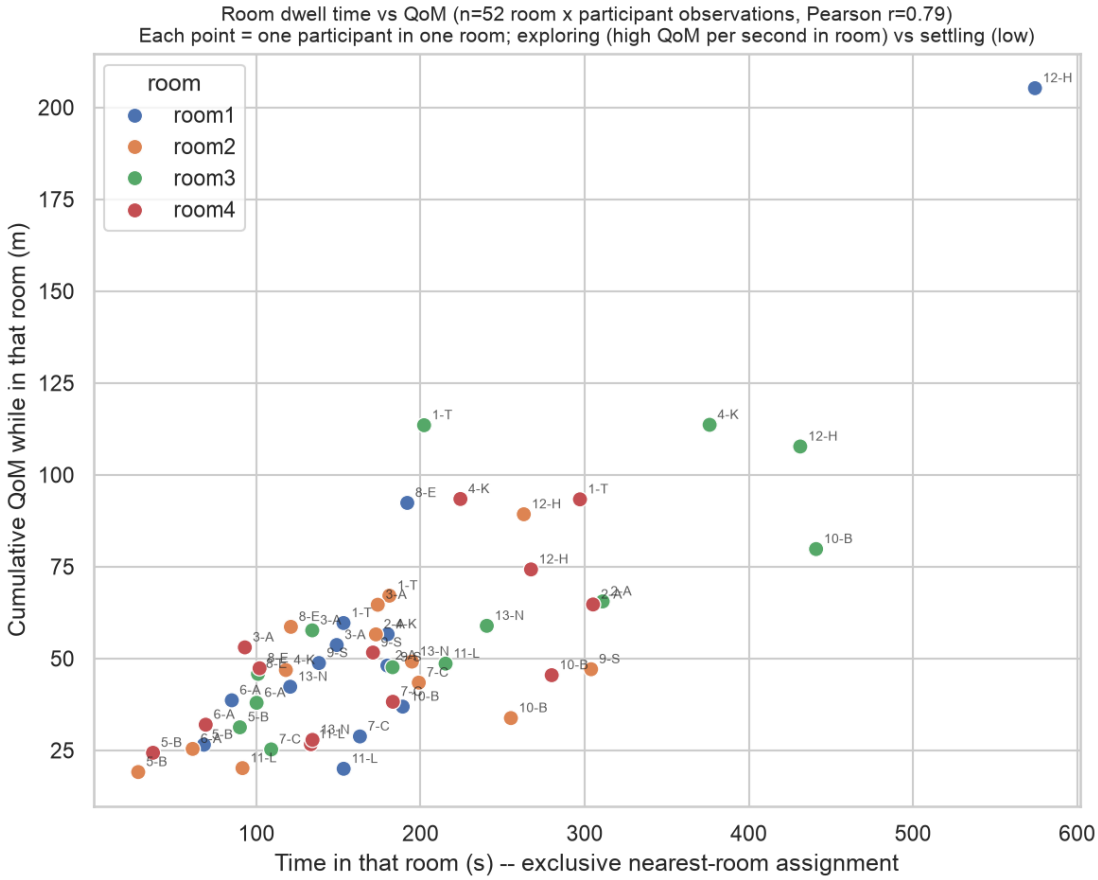


Figure 6.10: Room dwell time against cumulative QoM while in that room, one point per participant per room under the exclusive assignment (n=52), coloured by room, with a Pearson correlation of $r = 0.79$.

interruption to ongoing movement, while other participants' more exploratory, play-oriented accounts of the interaction (Section 6.2.7) are consistent with treating a sound onset as a cue to keep moving rather than to pause. The third interpretation is specific to this dissertation's design: since Phase 3 was built to introduce anticipatory behaviour that "challenges" the participant with sudden structural change (Chapter 4), participants may simply differ in whether a challenge event prompts them to pause and reassess or to press on and explore further, a difference an aggregate comparison cannot capture. Checking whether the direction of an individual's post-onset response depends on the specific sound-trigger type, or on a participant's more cautious or more playful interview-reported style, is a priority for further analysis (Chapter 7).

6.3.3 Room Dwell Time and Movement

The correlation between room dwell time and cumulative room QoM confirms that no specific room was associated with participants "freezing" in place; once expressed as a rate, movement was broadly comparable across all four rooms, even though the rooms are not visited for equal shares

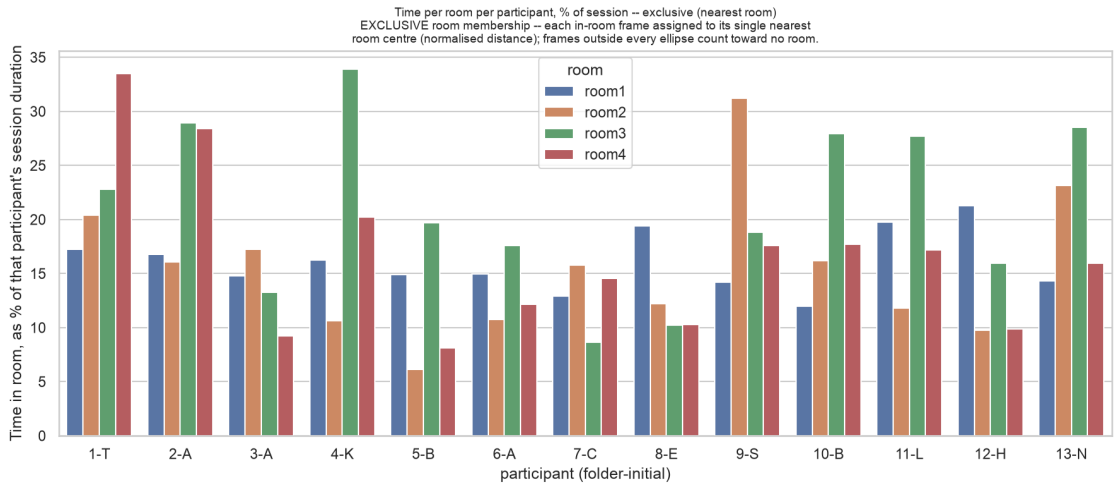


Figure 6.11: Time spent in each room as a percentage of that participant's own session duration, per participant, under the exclusive, nearest-room assignment.

of a session (Figure 6.9). This is a useful negative result: the differences reported in Sections 6.2.2 and 6.2.3 (across phases, and between the simultaneous zone and the rest of the space) do not simply reflect participants spending more time in one particular room's soundscape; they reflect genuine differences tied to phase and to the specific perceptual demands of the transition zone.

6.3.4 Room Preference

This chapter contains two findings about spatial preference, and they behave quite differently. The simultaneous zone (Section 6.3.1) is preferred by something close to the whole sample: a clear majority of participants move faster there, and describe doing so in convergent language, "borders," "boundaries," "video game dynamics," regardless of how they otherwise felt about the installation's content. Preference for a specific room does not work this way. The per-participant room-time data (Section 6.2.6) show every participant with a different, often strongly peaked, distribution across the four rooms, and the interviews suggest why: 2-A was pulled toward a specific soundscape for an autobiographical reason, tied to a remembered lunch. 8-E's pull was cultural, an unanticipated resonance with a place he had never lived. 11-L's response to a third soundscape was aversive rather than attractive, an active avoidance of one jarring sound. Three participants, three different rooms, three different reasons: this looks less like a shared preference than like three separate, individual ones. The boundary-seeking finding supports a general claim about this installation, that acoustic ambiguity itself draws sustained attention.

6.3.5 Research Questions

Chapter 1 posed one core research question and three sub-questions. In this section, each is revisited against the results reported above.

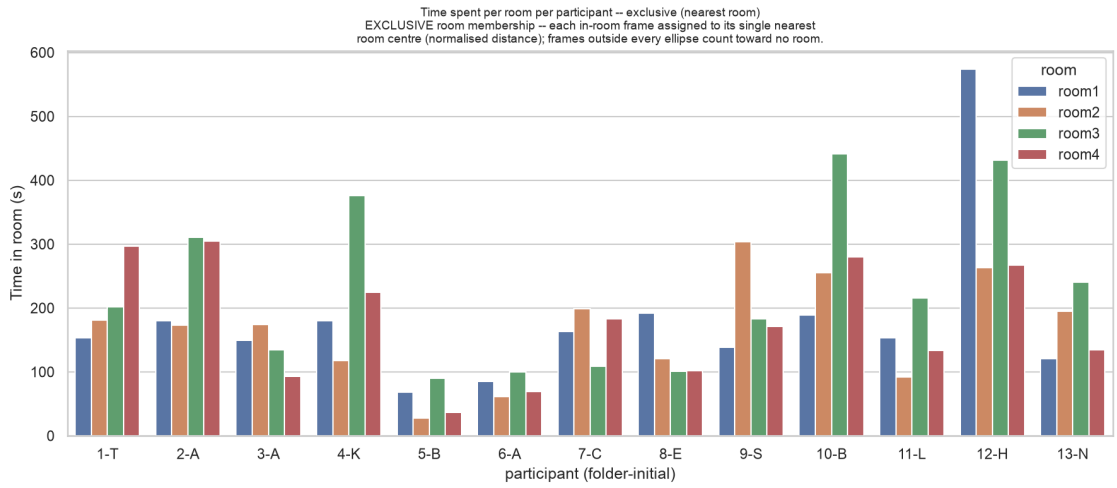


Figure 6.12: Time spent in each room in seconds, per participant, under the exclusive, nearest-room assignment.

The first sub-question asked how indoor sonic spaces impact the embodied body, addressed through Quantity of Motion across the three narrative phases and around individual sound-trigger events. The answer this dissertation offers is qualified but real: movement rose from Phase 1 to Phase 3 in the direction the narrative design intended, though not to a statistically significant degree at this sample size (Section 6.2.2), and individual sound-trigger events prompted a real but heterogeneous response, some participants briefly stilling and others moving more, rather than a single uniform reaction (Section 6.3.2). Taken together, these results support treating indoor sonic space as something a participant's body actively responds to, moment to moment, rather than a passive backdrop to movement that would have occurred anyway.

The second sub-question asked how participants navigate simulated rooms with different cultural and functional characteristics, addressed through movement at the boundaries between rooms, through room-by-room dwell time, and through the interview theme of participants constructing a functional identity for each one. The answer has two parts, and they point in different directions. At the level of the whole sample, participants did not experience the four rooms as interchangeable containers but as places actively worth comparing, moving significantly faster through the acoustic overlap between rooms than through the rest of the space (Section 6.3.1), and independently narrating that same boundary-seeking behaviour as their dominant strategy for making sense of the installation (Section 6.2.7); this part of the answer generalises reasonably well across participants. At the level of any individual room, however, the answer does not generalise: which room a given participant gravitated toward, or avoided, varied from person to person and tracked personal, cultural, or autobiographical associations rather than any shared ranking of the four soundscapes (Sections 6.2.6 and 6.3.4). The anthropomorphic and cultural-resonance findings (Section 6.2.7) extend this second part of the answer qualitatively: at least some participants read a room's cultural specificity, not only its acoustic novelty, directly off the soundscape itself, and that specificity is personal enough that it never adds up to a shared room preference across the sample.

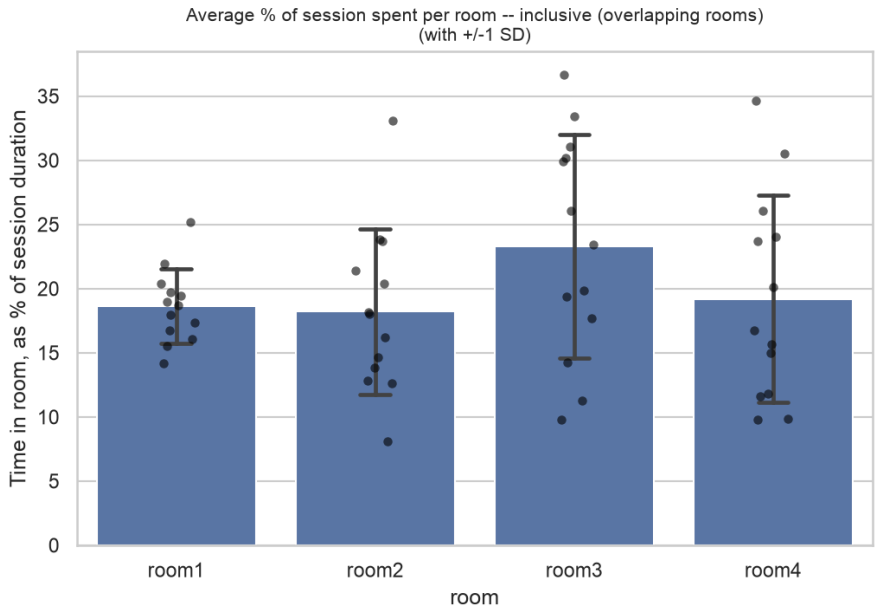


Figure 6.13: Average share of session time spent in each room across participants, under the inclusive, overlapping-ellipse assignment (error bars: ± 1 SD). Compare Figure 6.9.

The third sub-question asked in what ways interactive sonic objects alter the subjective perception of presence and place, addressed through the interview's direct questions on immersion and control and through the measured movement response to sound onsets. The interview data show presence to be graded rather than binary, from 9-S's explicit estimate of feeling "in the room" only twenty per cent of the time to 5-B's sustained absorption (Section 6.2.7), and show that interactivity did not uniformly increase that sense of presence: several participants (6-A, 7-C, 11-L) describe a system whose responsiveness felt inconsistent or unpredictable as reducing rather than deepening their sense of place, while others (5-B, 10-B, 12-H) describe the opposite. The residual, professional-identity-linked self-consciousness reported by 13-N despite the researcher's physical absence (Section 6.2.7) further suggests that an interactive object's capacity to alter subjective presence is bounded by factors, such as a participant's own self-monitoring, that the sonic design itself does not control. The mixed, participant-dependent movement response to individual sound onsets (Section 6.3.2) is itself informative here: it suggests that an interactive object does not engage every participant's body in the same way, some responding with a brief pause and others with renewed exploration, so that presence and engagement may be altered along genuinely different, individually variable pathways rather than a single shared response this project's design had anticipated.

6.4 Concluding Remarks

Taken together, these results support the central premise of this dissertation: that a participant's movement through *Sonic Inhabitation* is not a neutral means of exploring a passive sound field, but

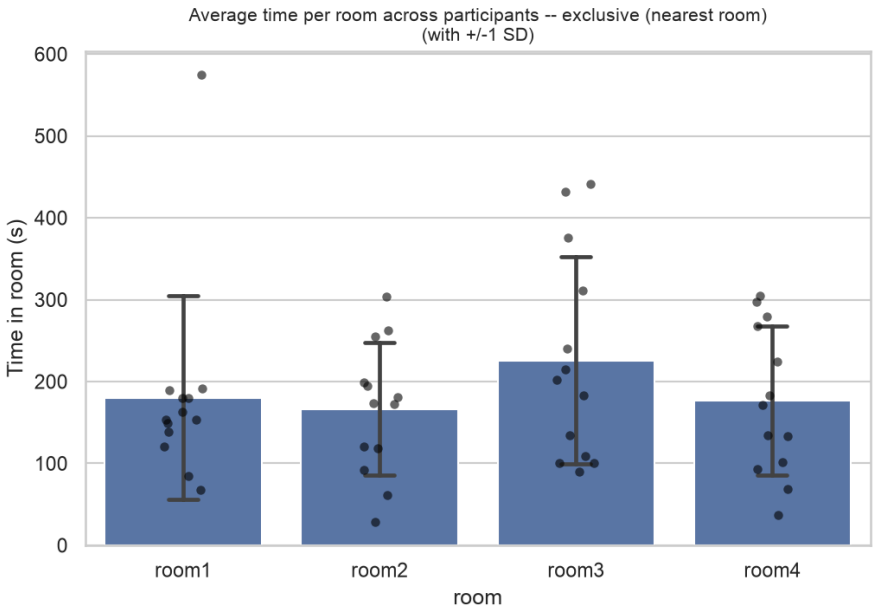


Figure 6.14: Average time in seconds spent in each room across participants, under the exclusive, nearest-room assignment (error bars: ± 1 SD).

an active process in which the acoustic environment measurably shapes behaviour, and behaviour in turn resolves the environment's ambiguity. The clearest quantitative evidence for this is the elevated movement rate in the simultaneous, between-rooms zone, a finding that the interview data independently and strongly confirms through participants' own accounts of boundary-seeking as their dominant exploratory strategy. Room-level preference did not generalise at all. Behind the roughly even average time spent in each room sits a different pattern of attachment and avoidance for every participant, the same under both room-membership conventions used in the analysis, and each traceable to that participant's own reasons for gravitating toward, or away from, a particular soundscape. The mixed, individually variable movement response immediately after a sound onset complicates a simpler "more sound, more movement" model of the interaction, pointing instead toward at least two coexisting responses, a brief listening pause for some participants and renewed exploratory movement for others, both of which fit comfortably within the action-perception loop framework used throughout this dissertation. Bauer-Wolf's thematic analysis adds a layer to the movement data: a room experienced through sound only can encourage a playful, rhythmic engagement, autobiographical and cultural association, and a felt tension between sensory overwhelm and the restorative pull of silence, alongside real friction between how the system tracks the body, through the head, and how participants feel they ought to be tracked, through the whole body.

These findings remain exploratory throughout, for the reasons stated at the opening of this chapter: a small sample, no correction for multiple comparisons, and some approximated timing data. Even so, the convergence of the quantitative and qualitative results, particularly around the between-rooms zone and the tension between sonic engagement and stillness, gives reasonable

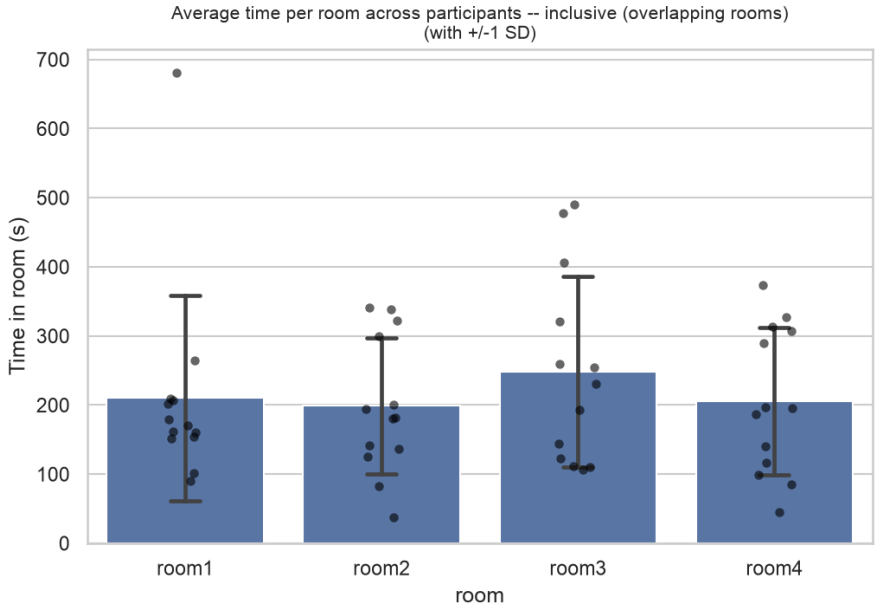


Figure 6.15: Average time in seconds spent in each room across participants, under the inclusive, overlapping-ellipse assignment (error bars: ± 1 SD).

assurance that the sonic characteristics of an indoor space measurably shape how people move through it, even in the complete absence of visual information.

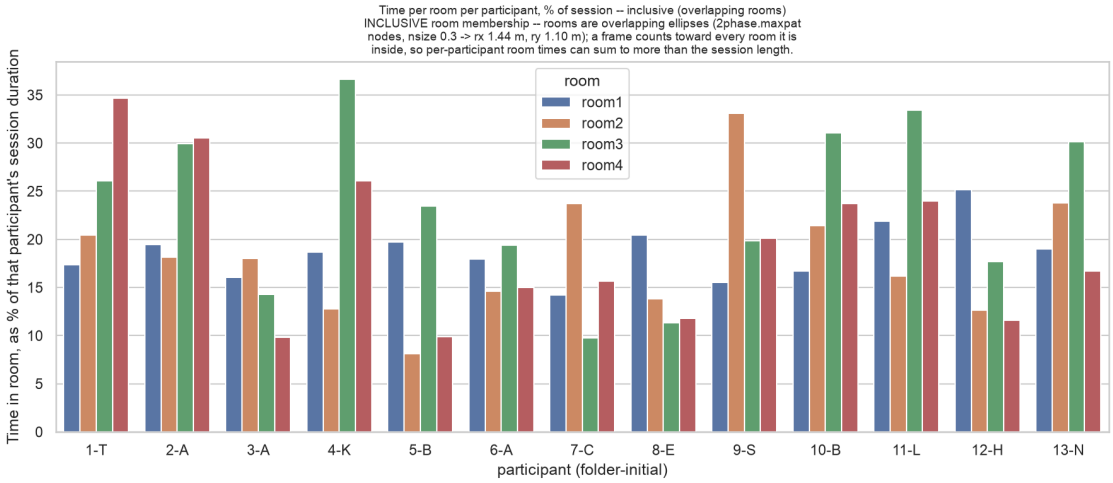


Figure 6.16: Time spent in each room as a percentage of that participant's own session duration, per participant. Compare Figure 6.11.

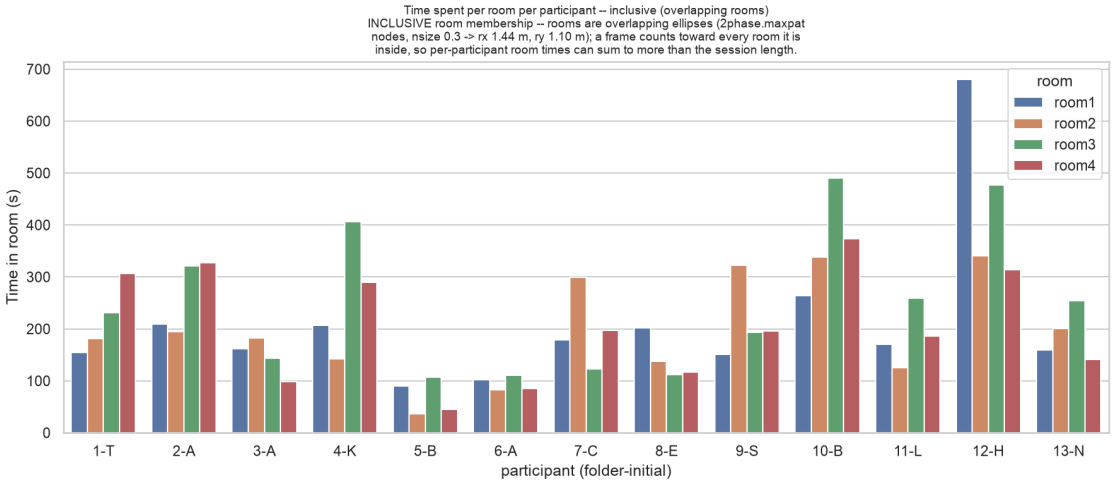


Figure 6.17: Time spent in each room in seconds, per participant. Compare Figure 6.12.

Chapter 7

Conclusions and Future Work

This dissertation reported an exploratory, mixed-methods study: thirteen participants moved through an auditory experience, while their kinematics were captured by a Qualisys motion-capture system and their felt experience was recorded in a post-exposure interview. Chapter 6 already set out three immediate analytical priorities arising directly from the reported results (Section 6.3): testing the head/neck-versus-whole-body inconsistency directly against the full-body Qualisys data once it is cleaned, breaking down the sound-onset response by individual trigger type, and running a pre-registered replication with a larger sample. Rather than repeat those points, this chapter widens the lens: it sets out the data work still needed before this project's raw material can be considered fully analysed, most pressingly the full-body dataset flagged as the fifth limitation in Section 7.4, the further analyses that data work would make possible, and the steps required to turn a single-installation study into a reusable system and a public research resource.

7.1 Finishing the Data Analysis

7.1.1 Cleaning the Qualisys Dataset

The results in Chapter 6 rely, for twelve of the thirteen participants, on the head-tracker rigid-body signal rather than the full-body marker set, precisely because the rigid-body signal has an effectively zero tracking gap while the labelled-marker export is subject to occlusion (Chapter 5). That full-body export has, as a result, gone largely unanalysed: it has been used only as a secondary cross-check on the head-tracker signal for a single participant (Section 5.2). Before it can support any further analysis, it needs a proper cleaning pass, gap-filling and, where necessary, manual relabelling of mislabelled or swapped markers, cluster-by-cluster quality auditing of the kind already applied to the head-tracker signal, and a documented, reproducible pipeline for going from Qualisys's native export to an analysis-ready dataset. This is a precondition for essentially everything else in this chapter: the cross-analysis with the audio record, any full-body kinematic analysis, and open publication of the dataset all depend on this cleaned export existing.

7.1.2 Deeper Analysis of the Cleaned Movement Data

Once cleaned, the full-body dataset opens analyses that a single head-tracker point cannot support. Quantity of Motion in this dissertation was computed from head velocity alone; a cleaned marker set would allow QoM to be decomposed by body segment, testing directly whether the "head-versus-whole-body" inconsistency that emerged independently from both my own preliminary reading and Bauer-Wolf's formally coded analysis (Section 6.2.7) reflects a genuine difference in how the system registered different participants' movement, or a difference in how each participant interpreted a signal that was in fact consistent. A full skeleton also supports postural and gestural analysis that head position cannot, for instance, identifying the specific gestures, leaning, crouching, arm movement, that different participants substituted for head movement when trying to trigger sound (Section 6.2.7), rather than relying on their own retrospective description of what they did. Finally, a full-body dataset would let Quantity of Motion be validated against an independent movement measure, checking whether the head-tracker-derived QoM used throughout this dissertation systematically over- or under-estimates whole-body movement for particular kinds of exploratory behaviour.

7.2 Cross-Analysing Sound and Movement and Publication

7.2.1 Aligning the Audio and Motion-Capture Records

The behaviour log analysed in this dissertation records which sound events fired and when, but not the audio signal itself in a form suitable for acoustic analysis. The session audio recordings exist separately (Chapter 5 already uses them as a fallback timing source for participant 2-A) and, once time-aligned with the Qualisys and behaviour-log clocks, would let each recorded sound event be paired directly with the kinematic trace that produced or followed it. This is a natural extension of the sound-onset analysis in Section 6.2.4, which already shows that individual sound onsets provoke a real but individually variable movement response; a proper cross-analysis against the audio record would make it possible to ask which specific sounds, rather than sound onsets in general, are associated with a pause versus renewed exploration.

7.2.2 Soundmark-Specific Analysis

Chapter 2 distinguishes, following Schafer, between a soundscape's background keynotes, its foreground signals, and its soundmarks, the unique, culturally specific sounds that a community recognises as belonging to a particular place. The Intercontinental Dataset underlying *Sonic Inhabitation* was built from recordings that plausibly differ along exactly this dimension, and the qualitative data already gives a direct empirical foothold for testing that difference: the autobiographical and cultural-resonance theme identified in both the preliminary reading and Bauer-Wolf's formal analysis (Section 6.2.7), the Chinese mobile-payment notification recognised by 3-A, the Beijing association reported by 8-E, the specific memory of Mallorca triggered in 2-A, shows that at least

some sounds in this dataset function as soundmarks for at least some listeners. A dedicated soundmark analysis would first classify the sound events in the Intercontinental Dataset along Schafer's keynote/signal/soundmark distinction, then test whether movement and dwell-time patterns (Sections 6.2.3–6.2.6) differ systematically around soundmark-classified events compared to keynote or generic signal events, and whether a soundmark's cultural specificity predicts the strength of the affective or memory response it elicits in the interview data.

7.3 Publication

7.3.1 Open-Access Data Publication

Once the Qualisys export is cleaned and properly documented, and once the necessary consent and anonymisation steps are confirmed with participants, the resulting dataset, cleaned motion-capture data, the behaviour log, and the anonymised interview transcripts, should be published in an open-access repository. An indoor, movement-and-sound dataset of this kind, combining high-resolution kinematics with matched qualitative interview data across a culturally varied set of simulated rooms, is a resource other researchers working on embodied sound cognition, soundscape perception, or human-computer interaction could reuse directly, rather than each research group collecting a comparable dataset from scratch.

7.3.2 Publishing the System

The Max/MSP patch architecture described in Chapter 3, the QTM-to-OSC bridge, the feature-extraction and routing patch, the phase-specific sound engines, and the ambisonic spatialisation layer, was built as a working research instrument rather than as distributable software, and currently depends on lab-specific paths, hardware addresses, and manual configuration steps that are not documented for an external user. Publishing it as a usable open-source system would mean documenting its dependencies (IRCAM Spat5, the specific Max/MSP externals used), removing hard-coded configuration in favour of a settings file, and writing setup instructions for a researcher without prior knowledge of this specific installation. Doing so would let other spatial-audio or embodied-interaction researchers build on this system directly rather than reimplementing an equivalent pipeline.

7.3.3 Exporting to Different Physical Setups

The system as built assumes the specific geometry of the fourMs Lab: a 24-speaker ambisonic array, a fixed capture volume, and a Qualisys installation calibrated to that particular room. Exporting the experience to a different physical space would require the spatialisation layer to be reconfigured for a different speaker count and layout (the RITMO decoder's elevation-band structure, Chapter 3, was itself tuned to the specific dome shape of the current array) and the sonic-wall and room-boundary logic to be re-parametrised for a different room footprint. Doing this for two or three physically distinct rooms, rather than one, would also let a future study test directly

whether the effects reported in this dissertation, the simultaneous-zone movement increase and the individually variable sound-onset response, are properties of this specific installation or generalise across different physical spaces.

7.3.4 Further Directions Suggested by This Study's Findings

A smaller set of additional directions follows more speculatively from specific results reported in Chapter 6. The residual, professional-identity-linked observer effect reported by 13-N despite the researcher's physical absence (Section 6.2.7) suggests that spatial separation alone may not fully control for self-consciousness in a study of this kind, and a future protocol could test additional measures, such as an explicitly anonymous framing of the task, against this residual effect directly. The near-unanimous preference for total darkness (Table 6.3) was treated in this dissertation as a background methodological choice rather than a variable of interest; a dedicated comparison between the fully dark and dimmed lighting conditions, with a sample large enough to power that specific comparison, could test whether the small amount of retained vision in the dimmed condition measurably changes movement or reported presence. Finally, since immersion emerged from the qualitative data as graded rather than binary, ranging from 9-S's explicit "20 per cent of the time" estimate to 5-B's sustained absorption (Section 6.2.7), a future study could probe this gradient directly and quantitatively, for instance through repeated in-session presence ratings, rather than inferring it only after the fact from a single retrospective interview.

7.4 Limitations

Beyond the pipeline-level limitations already discussed in Section 5.5, five further limitations apply to the results reported here. First, the small sample size (11–13 participants depending on the comparison) limits statistical power considerably, and several of the non-significant findings above (notably the Phase 1 to Phase 3 comparison) may reflect an underpowered test of a real effect rather than the absence of one. Second, no correction for multiple comparisons was applied across the statistical tests reported in this chapter; at an uncorrected threshold, roughly one in twenty tests would be expected to reach $p < 0.05$ by chance alone, so the significant finding reported here (the simultaneous-zone comparison) should be treated as motivating further, pre-registered confirmation rather than as a final result. Third, even with Bauer-Wolf's formally coded analysis now incorporated (Section 6.2.7), it remains an analysis of the interview data alone; it has not been independently inter-rater checked, and my own preliminary reading and her coded analysis were not produced by fully independent analysts working blind to each other's prior notes, so the agreement between them (Section 6.2.7) should be read as a useful but imperfect form of triangulation, not as inter-rater reliability in the formal sense. Fourth, the timing discrepancy for participant 4-K (Section 6.2.7) is now better understood qualitatively, the researcher's session notes and Bauer-Wolf's coded analysis agree that he did not realise he could leave the room, but the specific numbers (an estimated 30 minutes against a logged 18 minutes 28 seconds) remain unreconciled, a reminder that the behaviour-log clock, however carefully reconstructed (Section 5.2),

measures what the logging system recorded, not necessarily the participant's full physical presence in the room. Fifth, every movement result in this chapter is derived from a single tracked point, the head-mounted rigid body, rather than from the full-body marker set collected for every participant (Chapter 5); this choice is well motivated by data quality: the head-tracker signal has an effectively zero tracking gap while the full-body export is subject to occlusion, but it means the whole-body movement patterns several participants describe in their interviews (Section 6.2.7) have not yet been tested directly against the corresponding kinematic data. Closing this gap is the priority set out in Chapter 7.

7.5 Concluding Remarks

This dissertation set out to ask how the sonic characteristics of an indoor space shape human behaviour, using *Sonic Inhabitation*, a fully auditory installation in which a participant's own movement is the sole means of accessing an otherwise invisible room, as the experimental instrument for investigating that question directly. The quantitative results in this chapter offer a first, exploratory answer: participants' movement was measurably, if not always statistically decisively, shaped by the structure of the experience, moving faster through acoustically ambiguous transition zones between rooms, showing a modest and theoretically consistent (if not significant) rise in movement across the three narrative phases, and responding to individual sound events in a genuinely mixed way, some participants briefly stilling and others moving more, rather than with the single uniform increase the intuitive "sound invites movement" hypothesis this project began with would have predicted. Read against the full interview transcripts, these patterns gain both support and nuance: the boundary-seeking behind the simultaneous-zone finding is confirmed directly in participants' own words, while the tentative link this dissertation initially drew between negative affect and shorter sessions turned out, on fuller data, to hold for some participants and not others, a correction offered here in the interest of reporting this mixed-methods design honestly rather than letting an early, partial reading stand uncorrected.

Taken together with the theoretical grounding of Chapter 2 and the system and protocol described in Chapters 3 and 4, this work demonstrates that combining high-precision kinematic measurement with a designed, narrative auditory environment is a workable methodology for studying embodied sound cognition in indoor space, and one whose most useful findings, including its counter-intuitive ones, only become fully interpretable once the quantitative and qualitative strands of the mixed-methods design are read together, and are revised as more of that qualitative data becomes available, as it did in the course of writing this very chapter.

References

Acoustics — soundscape — part 1: Definition and conceptual framework, 2014. URL https://www.iso.org/standard/52161.html. Cited on page 11.

Francesco Aletta, Jian Kang, and Östen Axelsson. Soundscape descriptors and a conceptual framework for soundscape assessment. *Frontiers in Psychology*, 7:182, 2016. doi: 10.3389/fpsyg.2016.00182. Cited on page 13.

Elaine N. Aron and Arthur Aron. Sensory-processing sensitivity and its relation to introductory psychology students' trait variables. *Journal of Personality and Social Psychology*, 73(2):345–368, 1997. doi: 10.1037/0022-3514.73.2.345. Cited on page 46.

Östen Axelsson, Mats E. Nilsson, and Birgitta Berglund. A principal components model of soundscape perception. *The Journal of the Acoustical Society of America*, 128(5):2836–2846, 2010. doi: 10.1121/1.3493436. Cited on pages ix and 13.

Dominik R. Bach, John G. Neuhoff, Walter Perrig, and Erich Seifritz. Looming sounds as warning signals: The function of motion cues. *International Journal of Psychophysiology*, 74(1):28–33, 2009. doi: 10.1016/j.ijpsycho.2009.06.004. Cited on page 14.

Andrew Beadling and Paul Vickers. Listener Perception of Spatialised Audio for Embodied Interaction in Sonification. In *Proceedings of the 28th International Conference on Auditory Display (ICAD2023)*, pages 15–22. International Community for Auditory Display, 2023. ISBN 978-0-9670904-9-8. doi: 10.21785/icad2023.3814. URL http://hdl.handle.net/1853/72870. Cited on page 17.

Jens Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, Cambridge, MA, revised edition, 1997. ISBN 978-0-262-02413-6. Cited on page 17.

Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006. doi: 10.1191/1478088706qp063oa. Cited on page 67.

Albert S. Bregman. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, 1990. ISBN 9780262022970. Cited on page 16.

Tate Carson and Carter Gordon. Resonant Landscapes. In *Audio Mostly 2024 - Explorations in Sonic Cultures*, pages 525–532. ACM, 2024. ISBN 979-8-4007-0968-5. doi: 10.1145/3678299.3678354. URL https://dl.acm.org/doi/10.1145/3678299.3678354. Cited on pages 11, 12, 18, 19, and 20.

Stephanie Cheng Smith. Bell controller. Artist website, 2016. URL https://music.stephiescastle.com/works/bell-controller. Fifteen-motor, LED-and-bell interactive instrument; documented in the Experimental Music Yearbook 2016; performed at MOCA's Geffen Contemporary (2017) and REDCAT (2018). Cited on page 24.

Michel Chion. *Guide to Sound Objects: Pierre Schaeffer and Musical Research*. 2009. URL http://ears.huma-num.fr/onlinePublications.html. English translation of *Guide des objets sonores* (1983). Cited on page 9.

Perry R. Cook. Physically Informed Sonic Modeling (PhISM): Percussive Synthesis. In *Proceedings of the International Computer Music Conference (ICMC 1996)*, Hong Kong, 1996. International Computer Music Association. Cited on page 8.

Bert De Coensel, Timothy Van Renterghem, and Dick Botteldooren. The influence of natural sounds on the perception of road traffic noise. *The Journal of the Acoustical Society of America*, 129(3):1480–1490, 2011. doi: 10.1121/1.3543975. Cited on page 13.

Diana Deutsch. An auditory illusion. *Nature*, 251(5473):307–309, 1974. doi: 10.1038/251307a0. Cited on page 25.

Diana Deutsch. The tritone paradox: An influence of language on music perception. *Music Perception*, 8(4):335–347, 1991. doi: 10.2307/40285517. Cited on page 25.

Diana Deutsch. *Musical Illusions and Phantom Words: How Music and Speech Unlock Mysteries of the Brain*. Oxford University Press, New York, NY, 2019. ISBN 978-0-19-020683-3. doi: 10.1093/oso/9780190206833.001.0001. Cited on page 25.

Diana Deutsch, Trevor Henthorn, and Rachael Lapidis. Illusory transformation from speech to song. *The Journal of the Acoustical Society of America*, 129(4):2245–2252, 2011. doi: 10.1121/1.3562562. Cited on page 25.

Marco Donnarumma. Xth Sense: A biophysical model for gesture interaction and musical performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2011)*, pages 501–504, Oslo, Norway, 2011. NIME. Cited on page 23.

Çağrı Erdem, Alexander Refsum Jensenius, and Rolf Inge Godøy. Controlling sound with muscle tensegrity: A framework for muscle-sensing interfaces in interactive music. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2020)*, pages 420–425, Birmingham, UK, 2020. NIME. Cited on page 23.

Marc O. Ernst and Heinrich H. Bülthoff. Merging the senses into a robust percept. 8(4):162–169, 2004. ISSN 13646613. doi: 10.1016/j.tics.2004.02.002. URL https://linkinghub.elsevier.com/retrieve/pii/S1364661304000385. Cited on page 6.

Andy Farnell. *Designing Sound*. MIT Press, Cambridge, MA, 2010. ISBN 978-0-262-01441-0. Cited on page 8.

Steven Feld. *Acoustemology: Four Lectures*. Steven Feld / VoxLox, 2025. URL https://www.stevenfeld.net/exposures. Open-access multimedia eBook. Cited on pages 16 and 17.

Robert B. Freeman. The senses considered as perceptual systems. James J. Gibson. 44(1):104–105, 1969. ISSN 0033-5770, 1539-7718. doi: 10.1086/406033. URL https://www.journals.uchicago.edu/doi/10.1086/406033. Cited on pages 6, 10, and 15.

Shaun Gallagher. *How the Body Shapes the Mind*. Oxford University Press, 2005. ISBN 9780199271948. Cited on page 10.

Vittorio Gallese and George Lakoff. The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22(3-4):455–479, 2005. doi: 10.1080/02643290442000310. URL https://www.neurohumanitiestudies.eu/archivio/Gallese-Lakoff_2005.pdf. Accessed: 2026-02-10. Cited on page 10.

William W. Gaver. What in the world do we hear? an ecological approach to auditory event perception. *Ecological Psychology*, 5(1):1–29, 1993. doi: 10.1207/s15326969eco0501_1. Cited on page 9.

Michael A. Gerzon. Ambisonics in multichannel broadcasting and video. In *Audio Engineering Society Convention 74*. Audio Engineering Society, 1983. Cited on page 18.

James J. Gibson. The theory of affordances. In Robert Shaw and John Bransford, editors, *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, pages 67–82. Lawrence Erlbaum Associates, Hillsdale, NJ, 1977. Cited on page 14.

James J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, MA, 1979. ISBN 978-0-395-27049-3. Cited on pages 9 and 14.

Rolf Inge Godøy. Gestural-sonorous objects: embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound*, 11(2):149–157, 2006. doi: 10.1017/S1355771806001439. Cited on pages 10, 15, and 21.

Rolf Inge Godøy, Egil Haga, and Alexander Refsum Jensenius. Playing "air instruments": Mimicry of sound-producing gestures by novices and experts. In Sylvie Gibet, Nicolas Courty, and Jean-François Kamp, editors, *Gesture in Human-Computer Interaction and Simulation*, volume 3881 of *Lecture Notes in Computer Science*, pages 256–267. Springer, Berlin, Heidelberg, 2006. doi: 10.1007/11678816_29. Cited on pages 22 and 23.

Victor Evaristo Gonzalez-Sanchez, Agata Zelechowska, and Alexander Refsum Jensenius. Correspondences between music and involuntary human micromotion during standstill. *Frontiers in Psychology*, 9:1382, 2018. doi: 10.3389/fpsyg.2018.01382. Cited on page 57.

Michael S. A. Graziano and Dylan F. Cooke. Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia*, 44(6):845–859, 2006. doi: 10.1016/j.neuropsychologia.2005.09.009. Cited on page 14.

Jinyue Guo, Maham Riaz, and Alexander R. Jensenius. Comparing four 360-degree cameras for spatial video recording and analysis. In *RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion*, 2020. Cited on page 20.

Jinyue Guo, Jim Tørresen, and Alexander Refsum Jensenius. Cross-modal analysis of spatial-temporal auditory stimuli and human micromotion when standing still in indoor environments. In *Proceedings of the International Symposium on Computer Music Multidisciplinary Research (CMMR 2025)*, 2025. URL https://zenodo.org/records/17502603. RITMO Centre / AMBIENT project, University of Oslo. Cited on pages 10, 17, 18, 19, 20, 23, 29, 45, 54, and 57.

Thomas Görne, Kristin Kuldkepp, and Stefan Troschka. Implications of crossmodal effects and spatial cognition on producing in spatial audio. 2021. Cited on pages 10 and 11.

Yoshimi Hasegawa and Siu-Kit Lau. Audiovisual Bimodal and Interactive Effects for Soundscape Design of the Indoor Environments: A Systematic Review. 13(1):339, 2021. ISSN 2071-1050. doi: 10.3390/su13010339. URL https://www.mdpi.com/2071-1050/13/1/339. Cited on page 11.

Mari Romarheim Haugen. Musical meter as shape: An embodied perspective on metrical trajectories and curves. In Alexander Refsum Jensenius, editor, *Sonic Design: Current Research in Systematic Musicology*. Springer Nature Switzerland, Cham, Switzerland, 2024. ISBN 978-3-031-57892-2. doi: 10.1007/978-3-031-57892-2_6. Cited on page 16.

Ulf A. S. Holbrook. *Objects and Structures: Aesthetical inquiry and artistic experimentation into the relationships between sound objects and spatial audio*. Ph.d. thesis, University of Oslo, Oslo, Norway, 2022. Cited on page 7.

Brian Horsak, Ronald Dlapka, Michael Iber, Anna-Maria Gorgas, Anita Kiselka, Christian Gradl, Tarique Siragy, and Jakob Doppler. SONIGait: a wireless instrumented insole device for real-time sonification of gait. *Journal on Multimodal User Interfaces*, 10(3):195–206, 2016. doi: 10.1007/s12193-016-0216-9. Cited on page 8.

ISO 12913-2:2018. Acoustics – Soundscape – Part 2: Data collection and reporting requirements. Standard, International Organization for Standardization, Geneva, CH, December 2018. Cited on page 13.

Alexander Refsum Jensenius. *Sound Actions: Conceptualizing Musical Instruments*. MIT Press, Cambridge, MA, 2017. ISBN 9780262037028. Cited on pages 10 and 21.

Alexander Refsum Jensenius. *Sound Actions: Conceptualizing Musical Gestures*. MIT Press, Cambridge, MA, 2023. ISBN 978-0-262-54477-1. doi: 10.7551/mitpress/14060.001.0001. Cited on pages 6, 14, 21, 22, and 23.

Alexander Refsum Jensenius, editor. *Sonic Design: Explorations Between Art and Science*, volume 12 of *Current Research in Systematic Musicology*. Springer Nature Switzerland, 2024. ISBN 978-3-031-57891-5 978-3-031-57892-2. doi: 10.1007/978-3-031-57892-2. URL https://link.springer.com/10.1007/978-3-031-57892-2. Cited on pages 5, 7, 10, 11, 15, 19, 53, and 56.

Alexander Refsum Jensenius, Kari Anne Vadstensvik Bjerkestrand, and Victoria Johnson. How still is still? exploring human standstill for artistic applications. *International Journal of Arts and Technology*, 7(2/3):207, 2014. doi: 10.1504/IJART.2014.060943. Cited on page 49.

Alexander Refsum Jensenius, Agata Zelechowska, and Victor Gonzalez Sanchez. The musical influence on people's micromotion when standing still in groups. 2017. Cited on page 19.

Chang-ho Ji, Young-ji Cho, Young-guk Kim, and Won-hyung Cho. Obstacle detection and spatial audio feedback system for visually impaired orientation and mobility. *Applied Sciences*, 8(11): 2175, 2018. doi: 10.3390/app8112175. Cited on page 8.

Jian Kang. *Urban Sound Environment*. CRC Press, Boca Raton, FL, 2006. ISBN 9780415385368. Cited on pages 11 and 16.

Henri Lefebvre. *The Production of Space*. Blackwell, Cambridge, MA, 1991. ISBN 9780631181774. URL https://philpapers.org/archive/LEFTPO-4.pdf. Accessed: 2026-02-10. Cited on page 11.

Marc Leman. *Embodied Music Cognition and Mediation Technology*. The MIT Press, Cambridge, MA, 2008. ISBN 978-0-262-12293-1. doi: 10.7551/mitpress/7476.001.0001. Cited on pages 10 and 21.

Ruth Y. Litovsky, H. Steven Colburn, William A. Yost, and Sandra J. Guzman. The precedence effect. *The Journal of the Acoustical Society of America*, 106(4):1633–1654, 1999. doi: 10.1121/1.427914. Cited on page 18.

Alvin Lucier. I am sitting in a room. Sound work for voice and electromagnetic tape, 1969. Composed 1969 at Brandeis University's Electronic Music Studio; premiered 1970 at the Solomon R. Guggenheim Museum, New York; recording released by Lovely Music, Ltd. (LCD 1013, 1981). Cited on page 23.

John Maeda. *The Laws of Simplicity: Design, Technology, Business, Life*. The MIT Press, Cambridge, MA, 2006. ISBN 978-0-262-13472-9. Cited on page 22.

Thor Magnusson. An epistemology of digital instruments. In *Proceedings of the International Computer Music Conference (ICMC 2009)*, pages 311–314, Montreal, Canada, 2009. Michigan Publishing. Cited on page 22.

Thor Magnusson. *Sonic Writing: Technologies of Material, Symbolic, and Signal Inscriptions*. Bloomsbury Academic, London, UK, 2018. ISBN 978-1-5013-1386-8. doi: 10.5040/9781501313899. Cited on pages 22 and 24.

Stephen Malloch and Colwyn Trevarthen, editors. *Communicative Musicality: Exploring the Basis of Human Companionship*. Oxford University Press, Oxford, UK, 2009. ISBN 978-0-19-856628-1. Cited on page 21.

Anna C. May, Bianca P. Acevedo, Arthur Aron, and Elaine N. Aron. Sensory processing sensitivity: A review in the context of other personality traits, clinical disorders, and environmental sensitivity. *Frontiers in Psychology*, 11:588656, 2020. doi: 10.3389/fpsyg.2020.588656. Cited on page 47.

Dimitris Menexopoulos, Pedro Pestana, and Joshua D. Reiss. The state of the art in procedural audio. *Journal of the Audio Engineering Society*, 71(12):826–848, December 2023. URL https://www.aes.org/e-lib/browse.cfm?elib=22346. Cited on page 8.

Margarida Mesquita Leal. Audio recordings — sonic inhabitation, 2026. URL https://zenodo.org/records/22306873. CC-BY 4.0. Cited on page 29.

Thomas Metzinger. *Being No One: The Self-Model Theory of Subjectivity*. The MIT Press, Cambridge, MA, 2003. ISBN 978-0-262-13417-0. doi: 10.7551/mitpress/1551.001.0001. Cited on page 26.

Eduardo Reck Miranda and Marcelo M. Wanderley. *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*, volume 21 of *Computer Music and Digital Audio Series*. A-R Editions, Inc., Middleton, WI, 2006. ISBN 978-0-89579-585-4. Cited on page 21.

Andrew Mitchell, Tin Oberman, Francesco Aletta, and Jian Kang. Searching for a common understanding of 'soundscape': A critical look at the definitions and uses of the term. *The Journal of the Acoustical Society of America*, 149(6):4299–4308, 2021. doi: 10.1121/10.0004989. Cited on page 11.

Thomas J. Mitchell, Sebastian Madgwick, and Imogen Heap. Musical interaction with hand posture and orientation: A toolbox of gestural control mechanisms. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, Ann Arbor, MI, 2012. doi: 10.5281/zenodo.1180543. Cited on page 23.

Meinard Müller. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer, Erlangen, Germany, 2015. ISBN 978-3-319-21944-8. doi: 10.1007/978-3-319-21945-5. Cited on pages 6 and 8.

John G. Neuhoff. Perceptual bias for rising tones. *Nature*, 395(6698):123–124, 1998. doi: 10.1038/25862. Cited on page 14.

Alva Noë. *Action in Perception*. The MIT Press, Cambridge, MA, 2004. ISBN 978-0-262-14088-1. doi: 10.7551/mitpress/1056.001.0001. Cited on page 25.

J. Kevin O'Regan and Alva Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5):939–973, 2001. doi: 10.1017/S0140525X01000115. Cited on page 25.

Qualisys AB. Calqulus. QTM Documentation, a. URL https://docs.qualisys.com/qtm/content/applications/calqulus.htm. Cited on page 54.

Qualisys AB. Qualisys sports marker set. QTM Documentation, b. URL https://docs.qualisys.com/qtm/content/processing_measurement/skeleton_marker_set_sports.htm. Cited on page 54.

Joseph Rawlinson. w[i]nd: An interactive and generative audio-visual composition/installation in procedural and orchestrated space. In Miguel Carvalhais, Mario Verdicchio, Luísa Ribas, and André Rangel, editors, *Proceedings of the Ninth Conference on Computation, Communication, Aesthetics & X (xCoAx 2021)*, pages 483–489. i2ADS, 2021. Cited on page 8.

Lord Rayleigh. On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 13(74):214–232, 1907. doi: 10.1080/14786440709463595. Cited on page 17.

Maham Riaz. *Embodied Spaces: Explorations in Spatial Audio–Video Recording and Ambient Interventions*. Ph.d. thesis, University of Oslo, Oslo, Norway, 2024. Cited on pages 10, 11, 16, 18, and 19.

Maham Riaz, Jinyue Guo, Çağrı Erdem, and Alexander Refsum Jensenius. Where to put that microphone? a study of sound localization in ambisonics recordings. In *Proceedings of the International Symposium on Computer Music Multidisciplinary Research (CMMR 2025)*, 2025a. RITMO Centre / AMBIENT project, University of Oslo. Cited on pages 16, 17, 18, 19, 20, and 29.

Maham Riaz, Jinyue Guo, Bilge Serdar Göksülük, and Alexander Refsum Jensenius. Where is that bird? the impact of artificial birdsong in public indoor environments. In *Proceedings of the 20th International Audio Mostly Conference*. ACM, 2025b. Cited on page 20.

RITMO. Bodily entrainment to audiovisual rhythms (AMBIENT), 2021. URL https://www.uio.no/ritmo/english/projects/ambient/. Cited on page 2.

Curtis Roads. *The Computer Music Tutorial*. MIT Press, Cambridge, MA, 1996. ISBN 978-0-262-68082-0. Cited on page 7.

Curtis Roads. *Microsound*. MIT Press, Cambridge, MA, 2001. ISBN 978-0-262-18215-7. Cited on page 8.

Davide Rocchesso, Roberto Bresin, and Mikael Fernström. Sound design for interactive applications. *IEEE MultiMedia*, 10(2):52–63, 2003. doi: 10.1109/MMUL.2003.1195160. Cited on page 8.

David Rokeby. The construction of experience: Interface as content. In Clark Dodsworth, editor, *Digital Illusion: Entertaining the Future with High Technology*, pages 27–47. ACM Press / Addison-Wesley, 1998. Cited on page 23.

Francis Rumsey. *Spatial Audio*. Routledge, New York, 1 edition, 2001. ISBN 9780080498195. doi: 10.4324/9780080498195. Cited on page 17.

Pierre Schaeffer. *Traité des objets meśicaux: Essai interdisciplinaire*. Éditions du Seuil, Paris, 1966. Cited on pages 8, 9, and 15.

Pierre Schaeffer. *Treatise on Musical Objects: An Essay Across Disciplines*. University of California Press, Berkeley, CA, 2017. ISBN 978-0-520-28429-6. doi: 10.1525/j.ctt1qv5pqb. Cited on page 7.

R. Murray Schafer. *The Soundscape: Our Sonic Environment and the Tuning of the World*. Destiny Books, Rochester, VT, 2 edition, 1993. ISBN 9780892814558. Cited on pages 11, 12, 16, and 29.

Erich Seifritz, John G. Neuhoff, Deniz Bilecen, Klaus Scheffler, Henrietta Mustovic, Hansjörg Schächinger, Rita Elefante, and Francesco Di Salle. Neural processing of auditory looming in the human brain. *Current Biology*, 12(21):2147–2151, 2002. doi: 10.1016/S0960-9822(02)01356-8. Cited on page 14.

Andrea Serino. Peripersonal space (pps) as a multisensory interface between the individual and the environment, defining the space of the self. *Neuroscience & Biobehavioral Reviews*, 99:138–159, 2019. doi: 10.1016/j.neubiorev.2019.01.016. Cited on page 14.

Anil K. Seth and Manos Tsakiris. Being a beast machine: The somatic basis of selfhood. *Trends in Cognitive Sciences*, 22(11):969–981, 2018. doi: 10.1016/j.tics.2018.08.008. Cited on page 26.

P. St-Jean, O. G. Clark, and M. Jemtrud. A review of the effects of architectural stimuli on human psychology and physiology. 219:109182, 2022. doi: 10.1016/j.buildenv.2022.109182. URL https://doi.org/10.1016/j.buildenv.2022.109182. Cited on page 11.

Ana Tajadura-Jiménez. *Embodied Psychoacoustics: Spatial and Multisensory Determinants of Auditory-Induced Emotion*. Chalmers University of Technology, 2008. ISBN 978-91-7385-217-3. Cited on pages 10, 13, and 17.

Simone Torresin, Rossano Albatici, Francesco Aletta, Francesco Babich, and Jian Kang. Assessment Methods and Factors Determining Positive Indoor Soundscapes in Residential Buildings: A Systematic Review. 11(19):5290, 2019. ISSN 2071-1050. doi: 10.3390/su11195290. URL https://www.mdpi.com/2071-1050/11/19/5290. Cited on page 13.

James Traer and Josh H. McDermott. Statistics of natural reverberation enable perceptual separation of sound and space. *Proceedings of the National Academy of Sciences*, 113(48):E7856–E7865, 2016. doi: 10.1073/pnas.1612524113. Cited on page 14.

Barry Truax. Real-time granular synthesis for sound design and composition. *Computer Music Journal*, 12(2):14–26, 1988. doi: 10.2307/3679938. Cited on page 8.

Barry Truax. *Acoustic Communication*. Ablex Publishing, 2nd edition, 2001. Cited on pages 6 and 15.

Barbara Tversky. *Mind in Motion: How Action Shapes Thought*. Basic Books, New York, NY, 2019. ISBN 9780465093069. Cited on page 6.

Vesa Välimäki, Jyri Pakarinen, Cumhur Erkut, and Matti Karjalainen. Discrete-time modelling of musical instruments. *Reports on Progress in Physics*, 69(1):1–78, 2006. doi: 10.1088/0034-4885/69/1/R01. Cited on page 7.

Pieter Van den Berghe, Valerio Lorenzoni, Rud Derie, Joren Six, Joeri Gerlo, Dirk De Clercq, and Marc Leman. Music-based biofeedback reduces tibial shock in over-ground running. *Medicine & Science in Sports & Exercise*, 51(11):2284–2290, 2019. doi: 10.1249/MSS.0000000000002060. Cited on page 8.

Timothy Van Renterghem, Kris Vanhecke, Paul Devos, Pieter Thomas, Bert De Coensel, and Dick Botteldooren. Interactive effect of vegetation and noise level on human appreciation of urban outdoor spaces. *Landscape and Urban Planning*, 154:48–58, 2016. doi: 10.1016/j.landurbplan.2016.03.013. Cited on page 13.

Francisco J. Varela, Evan Thompson, and Eleanor Rosch. *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press, Cambridge, MA, 1991. ISBN 978-0-262-72021-2. Cited on page 10.

Georgios Varoutsos and John D'Arcy. The sonic imagery of the covid-19 pandemic. In *Proceedings of the SARC: Centre for Interdisciplinary Research in Sound and Music*, Queen's University Belfast, Belfast, Northern Ireland, 2021. Accessed via SARC research archive. Cited on pages 11 and 17.

Federico Visi, Rodrigo Schramm, Kerstin Frödin, Åsa Unander-Scharin, and Stefan Østersjö. Empirical analysis of gestural sonic objects combining qualitative and quantitative methods. In Alexander Refsum Jensenius, editor, *Sonic Design: Current Research in Systematic Musicology*, pages 115–137. Springer Nature Switzerland, Cham, Switzerland, 2024. ISBN 978-3-031-57892-2. doi: 10.1007/978-3-031-57892-2_7. Cited on page 11.

Hans Wallach, Edwin B. Newman, and Mark R. Rosenzweig. The precedence effect in sound localization. *The American Journal of Psychology*, 62(3):315–336, 1949. doi: 10.2307/1418275. Cited on page 17.

Marcelo M. Wanderley and Philippe Depalle. Gestural control of sound synthesis. *Proceedings of the IEEE*, 92(4):632–644, 2004. doi: 10.1109/JPROC.2004.825882. Cited on pages 6 and 7.

Richard M. Warren. Auditory induction: Perceptual synthesis of missing sounds. *Perception & Psychophysics*, 35(4):371–377, 1984. doi: 10.3758/BF03206341. Cited on page 21.

Julie Watkins. Composing Visual Music: Visual Music Practice at the Intersection of Technology, Audio-visual Rhythms and Human Traces. 17(1):51, 2018. ISSN 1470-9120. doi: 10.16995/bst.296. URL https://www.bstjournal.com/article/10.16995/bst.296/. Cited on page 6.

David Wessel and Matthew Wright. Problems and prospects for intimate musical control of computers. *Computer Music Journal*, 26(3):11–22, 2002. doi: 10.1162/014892602320582945. Cited on page 22.

Hildegard Westerkamp. The soundscape speaks, 2021. URL https://hildegardwesterkamp.ca/sound/comp/1/soundscapespeaks/. Accessed February 5, 2026. Cited on page 12.

Wanja Wiese and Thomas Metzinger. Vanilla PP for philosophers: A primer on predictive processing. In Thomas Metzinger and Wanja Wiese, editors, *Philosophy and Predictive Processing*. MIND Group, Frankfurt am Main, Germany, 2017. doi: 10.15502/9783958573024. Cited on page 26.

Frederic L. Wightman and Doris J. Kistler. Headphone simulation of free-field listening. I: Stimulus synthesis. *The Journal of the Acoustical Society of America*, 85(2):858–867, 1989. doi: 10.1121/1.397557. Cited on page 17.

Hortense (Stan) Wijnans. The body as a spatial sound generating instrument: Defining the three dimensional data interpreting methodology (3dim), 2010. Submitted in partial fulfilment of the requirements of the University of the West of England and Bath Spa University. Cited on pages 18 and 19.

Rob Withagen, Harjo J. de Poel, Duarte Araújo, and Gert-Jan Pepping. Affordances can invite behavior: Reconsidering the relationship between affordances and agency. *New Ideas in Psychology*, 30(2):250–258, 2012. doi: 10.1016/j.newideapsych.2011.12.003. Cited on page 14.

Agata Zelechowska, Victor Evaristo Gonzalez-Sanchez, Bruno Laeng, Jonna K. Vuoskoski, and Alexander Refsum Jensenius. Who moves to music? empathic concern predicts spontaneous movement responses to rhythm and music. *Music & Science*, 3, 2020. doi: 10.1177/2059204320974216. Cited on page 47.

Franz Zotter and Matthias Frank. All-Round Ambisonic Panning and Decoding. 2012. Cited on page 20.

Appendix A

Full Interview Transcripts

This appendix contains the automatically transcribed post-exposure interviews described in Section 4.2.3, referenced by participant ID (Table 6.4). Each interview followed the seven-question guide given in that section. Transcripts were generated automatically (Autotekst, using OpenAI Whisper V3) and may contain recognition errors; they are reproduced here without correction beyond merging broken line breaks into readable paragraphs.

A.1 Participant 1-T (17 June 2026, 14:30)

Okay, then I'm starting the recording now. I will now ask you a few questions about your experience during the experiment and feel free to take your time to think back about the experiments and just let me know when you're ready. Yes, I'm ready.

During the experience, did you experience any specific sounds that felt memorable to you? Yes, a bunch of them. So what I was trying to do was to try to figure out more or less the general vibes of the different rooms, so I was looking for something specific, and for example in the one back there I was able to hear footsteps, which meant that the room was kind of quieter compared to the other ones probably, and also at some point I heard a car honk, like something like that, and then in the other one over there at some point I heard like some machine going like beep beep, like an alarm of some kind. Um, and the other one that comes to mind was in this one here, I was able to hear like, um, children voices, I guess, yeah, like not only adults speaking in the kind of background but also children.

Okay, okay, thank you! Um, what aspects made you want to explore the room and how did you choose which parts to explore? Um, so I would say, at first glance, the rooms feel quite similar to each other. So I wanted to kind of try to decide on something that made them feel different, to try to assign them kind of a personality, I guess, or some kind of idea. Sorry, what was the second part? How did you choose which parts to explore? Okay, so at first I just went around, and then I stayed in a room for a certain time until I heard something that was interesting, like the sounds I told you before, and then another thing is that I tended to go into the corners because I felt that

was closest to the source, I don't know if that's true, but like that's what it felt like to me. So yeah, okay.

Thank you! During the experience, did you experience any specific emotions? It's probably not part of the experiment, but when I was in this room here, I think there's kind of this low-pitched hum that the speakers make, that made the whole metal structure kind of vibrate, probably. [laughs] That kind of creeped me out, so I didn't want to spend a lot of time there, and another thing is that when you go into the middle and like everything disappears, it's quite sharp, so it's kind of surprising in a way but not unpleasantly so, it's like, right, very peaceful. Okay. Did you experience any other kind of emotions? Hmm. Do you consider curiosity an emotion? Like just wanting to explore and going around and see what's there? Yeah, I would say this.

Okay, thank you. How did you experience your body while you were listening to the sounds? Hmm, mostly just relaxed, but every now and then I would hear something kind of unexpected, like the beep from the machine, and then I would be like, I would turn around after a while, but still very relaxed, not super fast. Apart from that, I'm not sure what you really want me to say. Did specific parts of your body feel like they were part of the listening experience? Well, apart from my head, in which you kind of change the orientation to try to, you know, get your ear closer to where the source of the sound is, I wouldn't say so, like just yet probably, but that's because that's where the sound is supposed to go, yeah.

Okay, thank you. To what extent did you feel like you were one with the auditory space? One with the auditory space? You mean as if I felt like being there? Being immersed. Immersed. To some of them, to quite some extent, yes, like for this one, for example, this room here with the low-pitched hum, not really, also I wasn't really able to figure out what it was or to give it a personality or an idea, a touch to it. With the other ones a little more, especially those two, because I felt they were more special, I don't know, like one is quieter and the other one felt more like a workshop of some kind or a kitchen or something like that, so I would say just because it was easier to figure out what, like, their nature, what they were, or at least something closer to that, I felt more immersed compared to this one, for example, yeah. Okay, thank you. So yeah, like it's kind of different, a little more for those and a little less for this one, for example, yeah, that's it.

Well, thank you, thank you so much for participating, that was it, and I will stop the recording now, maybe.

A.2 Participant 2-A (18 June 2026, 09:30)

Perfect. Okay, so I'm going to ask you some, thank you for participating. I'm very happy. And before we can fully end the experience, I would like to ask you a little bit of questions about how you felt. And then I'm recording for then I can go through the information. Is that okay? Yeah, perfect. Okay, so it's recording now.

And I'd like to ask you if during the experience, if there were any sounds that were memorable to you? Um, yes, for a couple different reasons, I guess. So in the first phase, well, in the first phase, I guess it was memorable in that it brought back a memory, or yeah, it actually reminded

me of a trip that I took with my ex. In our last relationship, we went to Majorca and we got lunch in this cafe restaurant, and we were sat outside, and it was just a beautiful day. So we could hear the waves, and there were a lot of people around because it was like a kind of peak touristy season, or at the start of it. So there were some sounds that sounded like waves, sounded like maybe some women working in the restaurant, like waitresses, but then like other kind of clatter, so I don't know why, but for whatever reason it just brought me back to that lunch that we had there, and it wasn't even a particular, it was just pleasant, like I don't know why, but it kind of reminded me of that. So that brought back a memory, so it was maybe memorable in that way. Then there were, I don't know if this was in the second or third phase, but then after I was kind of wandering around to the different corners of the space, and the corner that's like here, the closest. Yeah. It was one of the more quieter spaces, but then there was randomly somebody who sneezed or coughed. The first time it happened, it actually kind of took me off guard, because I had thought this is a quieter space. So it just surprised me. And then going back to it, I wanted to see if it surprised me the second time. And it still kind of drew in my attention, but it didn't startle me. And then I guess this was maybe entering into the third phase. There was a fun transition space. And I don't even know what it would sound like, like a bell, or I don't know, it was a sharper higher pitch sound, and I was just kind of playing around with that transition, trying to figure out what part of my body would trigger the transition, and then I think I figured out that it was my head, so yeah, so I was trying to move my body but keep my head still, but it was a lot of fun. So I guess those were the three most memorable noises.

And you were mentioning that you were trying to explore the space with different parts of your body. How did you choose which parts to explore of the space, and which part of the body? Did you have any... At first, like in the beginning of the first phase, I was just wandering around because trying to figure out how it works. You get a new toy, so you're like, okay, what do I do with this? And then I realized that, was I standing in the middle? Or I realized that if I moved outside of the center at the very start, then the noise would stop? Would it stop or would it start? So then I realized I was just testing the boundaries. Of course, yeah. So in that part of the session, I was just walking around and seeing, okay, where are the limits, or where are the boundaries that I can move? Or that, I knew I could move wherever, but like I wanted to see, yeah, what caused them exactly, so at that point it was just moving around, just walking, but then as the sessions progressed, then I guess I, yeah, felt like I was understanding it a bit more, or it was changing, and then felt, yeah, more comfortable to maybe move around, not that I felt uncomfortable before, but before I was, more in control? Yeah, yeah, and more experimental, yeah, yeah, maybe.

And did you experience any type of emotions, specific emotions that you can recall? Maybe in the first phase, just because I was having the memories of that day. And it was kind of surprising to recall because I was also trying to remember, like nothing significant happened on that day, it was literally just lunch. The space was nice, but we were there for a week or so, and we got lunch at different places, so it was funny that, why did those sounds bring back that particular meal on that particular day? So I guess, yeah, there were emotions that were tied to that memory, so the experience kind of brought that memory back. And then there were also, I love ambient sounds. So

whenever I sleep, I always have to have either a fan on or like brown noise, or something that just like kind of a low hum. And so some of the spaces I was able to find something that sounded kind of similar. So it was quite peaceful. But there's a balance between having ambient background noise and then just, I guess, one space, it almost felt a little awkward because it was just dead silence and somebody sweeping or something. So I was like, oh, I feel like I'm in a museum, and, like, trying, I felt like it required me to be more respectful of the space or something, so I was like, oh, I should leave, leave this room, leave this custodian to do their job and finish sweeping or mopping the floor, so I felt, I guess, more comfortable and more invited into the spaces where there was a bit more happening. Yeah, there was one that it felt like I was in a train station, and it reminded me of being in the UK, because I would travel by train a lot when I lived there, and that was again kind of nice, of thinking, like, oh, I can just pop into this coffee shop and grab a coffee, or, yeah, so you feel more included, like, yeah, just a part of the crowd, rather than just one person who's peeking in on a room with paintings and somebody sweeping the floor. Very interesting. Thank you so much. Yeah, I was going to say, maybe these are more images than emotions, but it's more like the emotions that they kind of conjure up. The memory that these bring is very, very important.

And how did you experience your body during the listening experience? Which part of the body felt like it was listening? It was listening. That's an interesting question. I guess because so much of this was very mental or very cognitive, then I'm inclined to say my head. But then, so that's more in a mental, emotional sense of the word, but then, with the last phase, what I was saying, or towards the end of the experiment, then I started to use my body a bit more and see, like, oh, if I sway, like how does that change the transition of the sounds, and so, yeah, I guess because I had moved from this sound brings back this memory, or this sound kind of conjures up this image in my mind, then I put that to the side and then was more physically experiencing the space, yeah, which I wonder if that's kind of how, I think maybe that's how I am in general, like maybe that's a personality trait of mine, that once you have all the thoughts, then you're kind of in your body more, maybe.

Interesting, would you say that you felt one with the auditory space? That I felt one, part of it, connected? Yes. For some, yeah. Like for the first phase where I was having this memory, because it felt like I was there in that moment. So definitely for that. And then for, I guess, the back corner, like, I don't know, the back right corner by the drums, that was the sound space that reminded me of a train station the most, so then I also felt more included in that, and then a little bit of this back, but then the opposite of that, this front left corner, I guess I was trying to figure out if it was traffic or like a bus station, or I don't know, so it felt not as homey or comfortable as the train station did. I did spend a lot of time in the front left, but I think I'm trying to remember why, I guess I felt more excluded in this front corner where there was nothing really happening, because again I guess I felt like I was walking into a room where there weren't many people around, and kind of like, oh, I've just stumbled into this random space and feel slightly awkward, again just with the sound of the sweeping and then one person sneezing, I'm just going to backtrack out of here and go back to the train station where I can grab a cup of coffee and read or do something else, but I

guess, yeah, maybe, yeah, it's interesting to think, like, why I felt more connected or included in these other spaces, it's fun to think about, yeah.

That's my interview. Thank you so much for participating in my experience. I will stop recording now. Thank you once more.

A.3 Participant 3-A (18 June 2026, 11:00)

So I'm starting to record now. I will now ask you a few questions about the experiment. Feel free to take some time if you want to first think back to it, and just let me know when you're ready. Ready? Yes. Okay.

During the experience, did specific sounds feel memorable to you? Memorable. I think there were two. One was a footstep in a large hall, well, it's not particularly reminds me of any event, but just something that sounds very familiar. And another was this announcement sound in Chinese, it's saying like a payment or something, because it's everywhere in the small shops in China, like this mobile phone payment thing, like whenever you give a payment, they will have this notification sound. And that was very, that was something like clearly memorable, yeah. Uh, maybe also some car horns, uh, yeah, I think that's all I can recall now.

Okay, thank you, um, what aspects, um, did you, do you make him want to explore the room, and how did you choose to explore the space? Well, in the first phase, because there's only sound in the middle, so I was mostly just listening to what events are happening there. And then in the second and the third, because I can move around in the different rooms, I mean, I knew there are different rooms before, so, but like the changing between the rooms, it's something fun to explore with. The sound that it makes? Well, I'm thinking, yeah. I think the most interesting part for me was the ability to jump in different rooms. It's almost like a teleporting experience, like teleporting to a different space. So then in the second phase, I wasn't specifically listening to the details of the sounds in each space, but just the ability to switch in different spaces makes me want to explore it. And then in the third phase, where there were sound events at specific points, then I think the goal changed to, my goal changed to listening to specific sounds in each room and then trying to find the point in each room where I can trigger the special sound, so yeah, second phase it was changing in the rooms, and then in the third it was looking for specific points, yeah.

Okay, cool, thank you, um, did you experience any specific emotions during the experience, um? I think there were some deja vus that made me recall some places I've been to. I'm not sure if that counts as an emotion, but yeah. And when I heard the notification sound in Chinese, it definitely made me feel a bit nostalgic. And then in the second phase, when I, so I realized I can change between the rooms, and then I start to run to explore the different rooms, that feels sort of exciting, I guess, because it's a, I think it's a new experience. I think that was three emotions.

Okay, thank you. How did you experience your body during the listening experience? Did you feel like any specific parts of it were present in the listening experience? I wasn't specifically aware of my body, so to speak. But in the third phase, I knew it's the head that triggers the sound

events. So I was sort of trying to move my head back and forth, try to re-trigger the sound events. Yeah.

Okay, um, and to what extent would you say you felt like you were one with the auditory space? I were one? Or like, how, how immersed did you feel in the auditory space? How immersed, like in a scale, or how should I? Generally, how would you say, like, you experience the feeling of, like, you felt part of the auditory space, like you were present, yeah. I think in a few moments I felt embarrassed, but most of the time I do realize that I'm manipulating the space. I mean, they sounded like rooms for sure, but yeah. What was it? Probably like 20% of the time that I felt I was in the room. Other times I feel like I'm just playing around with the rooms. Yeah, maybe that's how I will describe it.

Do you remember the specific moments where you felt immersed? Do you remember, kind of, why, what, like, the frame was, why you felt immersed? Uh, there was one moment when I was standing in the corner and I heard the footstep, as I mentioned before, the footstep was a loud reverb, in that moment I sort of actually felt like there's someone passing by there, so that was an immersive experience, yes, and the car horn, I can't remember if it was in the same, it's probably in the same room, but it did sound like it's coming from a distance, as normal horns on the road, so, yeah, I think it's, but, um, so I guess the room creates a base of it, and then when some very realistic event happens, that made me feel immersed, yes, yeah.

Okay, cool, thank you so much, that was it, thank you so much for participating, and I will stop this now.

A.4 Participant 4-K

[Verify: session date/time]

Okay, then I will start the recording now. I will now ask you a few questions about your experience during the experiment. Feel free to take your time and to think back to the experience and just let me know whenever you're ready. You ready? Yes, I'm ready.

During the experience, did specific sounds feel memorable to you? Well, there were, uh, it's like a surprising for me. Yeah. So it's not, it's not what I mean. Yeah. Yeah. Did any specific sounds, like what sounds? It's like a, oh, how do you say it? Like a knife. Like a knife. *[makes swishing sounds (sweeee)]* Like a scraping sound? Yeah, or shwee. Yeah, that sound surprised me. Okay. And like this. Wait, I can't hear. Yeah, yeah, sorry. Some sound like, like the sound loudness increased. In this room? Yeah, yeah, yeah. Around this point, yeah. *[points to middle of the front]*

Okay. Okay, thank you. What aspects made you want to explore the room and how did you choose which parts to explore? Okay, so first I tried to move to listen to the sound first. After that, it was the first period, and after a while I heard sound anytime, right? So I moved around in this room and I found at some point sound doesn't happen at some point. After a while I tried to, like, oh, this motion makes sound and this motion didn't make sound, like, yeah, so I found, like, a head position, yeah, which position makes sound and which position didn't make sound, yeah,

yeah, and there I found several types of sound happen depending on where I stand. So I try to categorize this point, I can hear like in a station, and this point is like in a restaurant, yeah.

Did you experience any specific emotions during the experience? First I felt like scary, that sound is like a horror video game, so probably first five minutes or ten minutes I felt scared, the feeling, it's got reduced after a while, so after that, just like a look around what kind of sound happened, yeah, okay.

Um, you mentioned that you were experiencing how motion and sound kind of went together. Um, how did you experience your body in the listening experience? Oh, body, okay. So you mean, uh, when I feel like, uh, when I feel scary, then? Generally, just how did you experience it? How did you feel your body while you were doing the experiment? Body? Yes, body, yeah. Or did any specific parts feel like they were very present or part of the experience? Okay, so when I was surprised, I feel like around this is like, around your chest? Yeah, yeah, yeah, around chest, my chest is like, I felt like my heart rate increased when I was surprised, but I didn't feel any other feeling about my body.

Did you feel like you were in control of the sound? So you mean, like you could control it? Yes, yeah, yeah, yeah. So probably last 5 minutes or 10 minutes, I found like which sound I can, I could figure out what sound comes from in each specific point. So this point is like a station, this point is like a restaurant. I could feel like I could control the sound, yeah, and when I wanted to keep silent, then I moved to like this point, or center, or a point, and then the sound stopped, so I feel like I can control the sound, yeah.

And did you feel immersed in the listening experience? Immersed? Ah, yes, yeah, yeah, yeah, yeah, yeah, yeah. The sound is like a, I can, I heard sound anywhere, so yeah, yeah, yes, yeah. Did you feel like you were really there? Did you feel like you were somewhere else? Yeah, yeah, yeah, yeah. But it's not like a, I didn't feel like, I didn't feel like that, I didn't feel like, I stay specific in real world space. Yeah, I feel like I stay in, like a horror video game, yeah, it's like in a strange world, yeah.

Okay. Well, that was it. Thank you so much. I will now stop the recording.

A.5 Participant 5-B (22 June 2026, 14:00)

Then I will start the recording now. I will now ask you a few questions about your experience during the experiment. Please feel free to always take time and remember back and just let me know whenever you're ready. Okay, during the experience did any specific sounds feel memorable to you?

So yeah, a few of them. I think, like, the footsteps, especially during the last part, the footsteps that were in this corner were kind of memorable, maybe because they remind me of a specific track, a song I listened to lately, and also they were fun to play with, so I kind of tried to synchronize my own steps with the footsteps, but also the bell from the, I think it was the train station at the airport, and that was also a memorable sound, but it took a while until I actually recognized it was from the airport, because I didn't listen long enough, I didn't stay in one position to listen long

enough to hear what's being said. But it's, of course, easy to remember something that's spoken in a language you understand. But during the first part, I think in some ways they were memorable, but they were spoken words, but I couldn't understand them, maybe in a foreign language to me, or because they were blurred or just in the background. But I kind of still remember the experience of a lot of the sounds. And I was first surprised when I experienced the busy street sound, I think that was towards the second part, yeah. Okay, so the footsteps in the front left corner? Yeah, front right, if I'm looking at the screens, yeah. And the back left was the bell? No, the bell was the back right, for the train station. Okay, thank you. Those specifically stood out, I think. There were also some sounds all the way towards the back here, which were a bit more calming, I found, it was like, especially during the last part, it was like, it felt like even behind the train station with this, it almost felt like, I think it was still a city recording, but it felt more like standing near a waterfall or something, which I kind of enjoyed, but I didn't stay there for too long because I wanted to play, not chill out.

You said you wanted to play. What kind of, what aspects made you want to explore the room and how did you choose which parts to explore? So, just the fact that I realized that, I tried to figure out, of course, what is initiating the sounds. I realized they're all recordings and they start anew once you get into a certain, within a certain area, and it's mainly the head that got tracked, so, um, at first I think I just wanted to examine what was the active area, so I started playing around and I figured out during the first part there was a circle in the middle of the room, so I kind of mapped out the circle, but then it changed the second part, I didn't fully get what kind of pattern it was arranged in, but it was much more complex, but I think I also got more busy trying to make like a rhythm to it, and then I had this period where I just looked for the samples that had an interesting start, because then you could make an interesting sound by moving between them, so finding the edges and moving between them and doing small dances, and towards the last one I think I mapped out the whole space, and basically found the intersections between the different zones, and in some places I found these interesting sounds, for example the bell that I mentioned, it was quite hard to find the exact spot to make it play, but it was fun to look for it, and fun to look for it in a specific pattern so that it would make an interesting sound with the surrounding sounds, to make like a rhythmic pattern. And also with one of the sounds in the front right, towards the last part, there was some, like a, sounding like voices, but not really. And I think if I spent more time I could make something more interesting, but I wanted to make an interesting sound with it, so both mapping out the space but also making something which could sound like a track. But the last section I figured out there were different, maybe there were circles, or I imagined triangles, but there were these intersections where it was completely silent towards the end, but if I stood in the middle of those two, between the two silent ones, I was hitting three different zones, and just moving the head I can play between them, and I found that nice, to just get some feedback on how large my movements were.

Cool, thank you. I got a bit aware of the body, because more aware of the body movements and how large they are, because normally you don't really get any sort of feedback on where you really are placed, it felt like I got the ground truth on my placement in the room, which was interesting

in itself, in addition to that, so there were layers of the experience, I think.

How did you experience your body in the listening experience? What parts did you feel like were really part of the experience? So I quickly noticed that my head was the important tracker for playing the sounds. When I said I got conscious of my position in space, I was also really unconscious of my body at times because I was so immersed in the experience, so I was just playing around, doing things naturally, so I think a bit of both, some parts I got really conscious of, am I moving now, am I standing still, especially when I mapped out the boundaries, because then small movements can initiate the sound or not, and then I got really aware of, or I tried to be aware of, even small movements, but other times it was more like what's happening in the bigger picture, if I just move like this, and it felt more like dancing, yeah, and parts of me wanted to be able to synchronize with the sounds even more to make it more immersive, so for example with the footsteps I kind of wanted to find a way of stepping so that the steps matched my, like the sound matched my own steps, yeah.

Cool, um, you mentioned you felt immersed, um, to what extent did you feel like you were being one with the space, and to what extent did you feel like you were really presently there? I felt, to a large extent, that I was present there, but also, it was, I had a change of, like when I exit the room I comment on how much time I spent, because it felt like the time went fast, but also afterwards it felt like a lot of time had passed, that I spent a lot of time in there, but it was fun, I think it was mainly because it felt like play, like when you just go and play and then you forget about the time and what happened, I kind of forgot at one point that it was an experiment, it just felt like a nice new way of playing.

To what extent did you feel in control of the auditory space? Pretty much, yeah. Except when it changed between the different parts, I didn't catch the first change until a bit after that it changed, so I got a bit surprised, so then I felt a loss of control from having control, but most of the part, except for that transition, I felt like I was in control.

And during the experience, did you experience any specific kind of emotions, or what emotions did you experience? Mostly just this playfulness, this easy joy, like a bit of curiosity and easy enjoyment, yeah, I felt it was intriguing, I wanted to, if there were more parts to it, it would be fun for it to last even longer, but it was nice with this long, more complex landscape towards the end, so I could really map it out, but I think, just, an even more complex, larger space, I think I could play for hours. Thank you so much, that was it, I will stop the recording, and thank you again so much for participating. Thank you.

A.6 Participant 6-A (23 June 2026, 13:00)

Then I will start the recording now. Okay, I will now ask you questions about your experience during the experiment. Please always feel free to take your time to think back and just let me know whenever you're ready. Okay, during the experience, did any specific sounds feel memorable to you? There was something happening in this corner. Which corner? That's the southeast corner. This one? Yes. So back right. Yeah. Okay. Back right. Okay. What kind of sound? I think

this is kind of sharp, kind of a sharp sound, maybe, like something metallic, yeah, if I remember correctly, but, yeah, yeah, that's mostly it. I think the silence was more memorable than the sound. The silence when? The silence in this last condition, when there was noise all around the room except for this island of silence in the middle. I'm not sure if that's correct, but it felt like the audio was a bit louder in this last condition, so it was a bit loud, then being in the middle felt very nice, where it was quiet, yeah, so that was the most memorable part, I think.

Okay, thank you. Which parts made you want to explore the room? How did you choose what sounds to explore? I think I was mostly interested in the borders more than the rooms in themselves, so, like, going back and trying to find the border and to see if it's only my head or if it's my body that's triggering the entrance into a space, so I played a lot with that, more than, I thought the spaces were a bit noisy. What was it about the borders that was interesting to you? It was like a video game dynamic, it felt like going into a new space and there's like this new world, I thought that could be cool, and also, like, just doing something and having an immediate response, that sensation was nice, and just trying to map out the space, I guess, and see where the borders are at, or, yeah, to have a grid of it, okay, yeah.

You mentioned that you were trying to explore the space with your body, how did you experience your body while you were doing this listening experience? It felt like my head was the only part that mattered, because it was the only one that gave an effect, so I felt that was a bit disappointing, because I would like my whole body to give some feedback to the system, so that it would matter, because it felt like the only important feature was where I was standing, not what I was doing, so yeah.

To what extent did you feel like you were in control of the experience and of the room, or let's say the auditory room? More in the first part than in the second part, I don't know, I felt a bit captured by all the noise, so then I felt like I was detained to the silent corner or center, so I didn't feel in control of the room, just like if I had the choice to interact with it or not, but I think I would wish for more feedback or control, perhaps.

Okay, thank you. You said that it was very noisy, what kind of emotions did you experience during the experience? I think I got a bit anxious at the end, it was a bit overwhelming, I got a bit stressed, almost irritated, but in the start I was more curious and open, yeah, so it was a shift, and I think it had to, it felt like it had to do with the audio level, yeah, I'm not sure. And it wasn't specific sounds that led to those feelings of anxiety? I'm not sure, because I then tried to see if there were some of the rooms that were better than the other, but they all felt at that point like a bit too overstimulating, I think, but it might have been one room that triggered it, but I can't remember, so yeah, might be the case.

Do I have anything else for you? Yeah. Did you feel like you were one with the auditory space? Did you feel like you were immersed? No, I don't think so, or to an extent. Can you elaborate on that? I'm not sure, I think, I imagined that it would be really cool to have audio, to have some visual features as well to interact with, that would make me more immersed, I think, than the audio alone, because I'm not sure, it goes a bit in my head, then not so much in my body, I also think that my physical head was the only part that mattered, it didn't afford so much being

in my body, I think, maybe that would have led to a larger degree of immersion, yeah.

Okay, that was already it, everything we wanted to know, thank you so much, I will stop the recording now.

A.7 Participant 7-C (23 June 2026, 13:30)

I will switch to English now. Tell me if you don't understand anything. I will now ask you a few questions about your experience. Feel free to always take time and think back to your experience, and just always let me know when you're ready or when you're uncomfortable with anything. During the experience, did any specific sounds feel memorable to you? Yeah, there was this one sound that, it reminded me of somebody cleaning the floor with a broom or something, and I always came back to it to figure out what it is, but I couldn't really figure out what it was, but it sounded a bit like cleaning noise, but it was too fast, it was some fast cleaner, maybe, I don't know, but it just made me curious, I guess. Do you remember in what space it was, like where in the room? Yeah, in the front right. So over there? From my point of view, front right, not really corner, but maybe in the middle, a bit right from the center. Did any other sounds stand out to you? I don't know what it was, but it was like a bell or something. A what? A bell or an alarm going off or something, but not very intense, but it made me laugh in the end, because I somehow expected some unexpected sounds to occur, but it was mainly like a white noise with some noises that you could then tell apart, but that was the only thing that caught me by surprise a bit, so I had to laugh.

Nice. Okay, what aspects made you want to explore the room, or how did you choose how to explore this auditory space? Yeah, I tried whether there's some interaction possible, and I was a bit disappointed because there was not too much, I thought it's more like, when I do this move maybe something happens or something, so in the beginning I was just listening, and then I tried to discover whether things happen if I move, and the only real thing I found out was that when I got on the central line to the left, because there were like three points in the space, or places where you could shut off the noises, where it was totally still, and the center, then the central line or the right corner in the back maybe, or maybe it was just on the central line to the right, I don't know, but there you could just stand still or not moving and it was quiet, but on the left side of the room in the center there I had always had to lift my right foot to make it quiet, that was the only interaction I found, like real, something related to my movements other than standing still or moving around, yeah, and because I found this out, then I started to, maybe there's another sound level if I go lower, if I go sit down, if I crawl, but then not too much to discover.

Did you feel like you were in control of the room? Yeah, I guess, because, yeah, if I wanted it to be silent I knew how to get it silent pretty quickly, so if I had to spend a longer time I would just have chosen the silent spots at one point, because in the end I, like the repeating noises, that made me a bit stressed, because I was losing my time consciousness, I didn't know how long I was already in there, because I was just told it's three minutes and then as long as I want, and I was like, okay, for the second three minutes I don't even know how long this middle part was,

and, yeah, and then I was like, how long is the last part now, am I here 15 minutes, am I here 10, I don't know, it felt super long in the end, and then the noises made me stressed because they were repeating, okay, yeah.

Did you experience any other emotions, or what kind of emotions during the experience? Not so many, actually, curiosity, um, it felt a bit uncomfortable, because it's a very unusual environment, so I didn't, well, I did not know really what to expect, will there be like rhythms, will there be like scary noises coming at some point, like surprises or something, or would it just be, in the end it was quite relaxed, I felt like, from the soundscape I expected something different, yeah, so not too many emotions.

Okay, how did you experience your body during the listening experience, did you feel like your body was part of it, or did you feel like any specific parts of your body were part? Yeah, as I said before, I thought my body would interact more, maybe, with the soundscape, so in the end it felt a bit, uh, I felt a bit distance maybe from my body, because I was like, how should I use my body, I was a bit maybe, how do you say, a bit insecure about how to move, and I'm usually not used to being insecure with my body, so it was a bit weird feeling.

Okay. And last question, did you feel like you were really there, or did you feel like you were immersed in the sound? Yeah, after a while, definitely, like in the end part, because I said, like, I wasn't aware how long I was there, and then I got familiar with where and which noises came from, and I figured out which parts I liked more and which not, so I could really dive into the experience, I guess, so I was very present, I felt. Okay, is there anything else you would like to add about the experience? Not so much, I guess. Okay, well then thank you very much for participating, and I will stop the recording now.

A.8 Participant 8-E (24 June 2026, 13:00)

So I will now ask you a few questions about your experience during the experiment. Always feel free to think back to how you felt in the moment and always let me know when you're ready. During the experience, did specific sounds feel memorable to you? Yeah, I guess it was this horn in the corner that I felt like I didn't always manage to trigger, so I think I might have gone back and forth there a couple of times to try and trigger it, and then there was some sound, like announcement sound for like a train station or something, there were a lot of just talking, can't remember much, like many other distinct sounds, actually. Okay. Do you remember where in the space the sounds were? Yeah, like, the horn was in around that area, so front left, front left, and then the announcement sound was front right, probably some other sounds that I have forgotten.

Okay, yeah, you said that you tried to trigger a specific sound several times, what other aspects made you want to explore, and how did you choose to explore the room? Uh, I think I stood still for a long time, and then I moved a bit, and then I realized, oh, I'm supposed to move now to have sound, and then I started just to move around a bit and see what happened, the different parts, and I even tried waving my arms and see if that would have any like effect, can't say it did, not that I

noticed at least, um, and just walking to and from and to and from, backwards running, uh, yeah, okay.

Okay, um, during the experience, did you experience specific emotions, or what emotions did you have? A bit nostalgia, because I felt like I was back in Beijing, because it sounded a bit like scrambled Chinese or something, the voices, so maybe that, and with all the noise, people talking and honking and stuff like that, announcement sounds, it felt like I was in a big city with a lot of people, yeah, so a bit nostalgia was probably one of the feelings, other than that, just curiosity, see what would happen where, and, yeah, not much, I wasn't annoyed or sad or angry or any other like emotions like that, it was just curious and nostalgic, yeah, I think, yeah, if that's feelings enough for you.

Yes, of course that's enough. Did you feel like you were in control of the room, or to what extent did you feel in control? Yeah, I mean, when I moved into different zones, if you will, I guess, then the sounds triggered and started, and if I moved away and came back then it started over again, so I felt like I was the one who started the sounds, they weren't like active all the time, that would probably be a bit too much noise if every, yeah, and, yeah, that was that, I guess, I was in control of starting and stopping, and I found the points where it got quiet, so that was also like, okay, if I want to kind of reset or figure out my next steps in silence, then I can do that, yeah, I think, yeah.

Okay. How did you experience your body during the experiment? What parts of the body do you feel like were involved in the experience? I tried different things, like with just arms or just tilting, and so I felt like it was the whole body, and it didn't, or I didn't need the whole body to trigger the sound, so I could lean, I could stand in the middle and lean to the left or right, and then it would be triggered, they didn't need my legs to follow with it, so that was interesting, that, yeah, so the body decided what sounds were going to come, in a way.

Okay. And the last question that I have is, to what extent did you feel like you were one with the space, or how immersed did you feel with the sound space? Yeah, pretty immersed, I think, it felt like sounds and noises that could be in a space like with a lot of people out and about, so, yeah, felt immersive, uh, yeah, probably what kicked off the nostalgia feeling anyway, so, yeah, I would say pretty immersive, yeah. Cool. Thank you, that was it, thank you so much, I will stop the recording now, maybe.

A.9 Participant 9-S (24 June 2026, 15:00)

Okay, then I will start the recording now. I will now ask you a few questions about your experience during the experiment. Always, if you want, take your time to think back and just let me know whenever you're ready. During the experience, did any specific sounds feel memorable to you? Yeah, I think the first one is the footsteps, I think that I imagine that there was footsteps on the stair, and then a music in the opposite part, I hear something, I think these are the two main sounds that I noticed. Then I tried to understand the voices, our school is about people, but I didn't understand

anything, and another sound, and I didn't understand what it was, but I imagine like a station or something like that.

Okay, um, how did you choose to explore the room, and what aspects would you want to explore it? I explored at the beginning, I just walked around to hear the different sounds, and then I tried to see if it was changing, or was always the same in the same point, and if it changes the composition, so if I go in a point and then in another, if it's going to change, it will remain the same. Okay, cool. You said, yeah, but walking around, how did you feel your body during the listening experience? Do you feel like your body was part of the listening experience? I don't know too much, not much, because maybe the footsteps was the only one that I understand, and so that I could feel, yeah, that could be there, and that was someone working in the sand, but in the other sounds I was there as outside, because people were not working with me, and I didn't understand what they said, the music I can't hear exactly what was, so not so part of the environment, like I was exploring, but I was listening, like viewing something, like I think an experience, it remembers me something like art, when you are at the Biennale of Venice, and you are exploring something, but I don't feel part of this something.

So you wouldn't say that you felt immersed in the sound? At some point a little bit, but when I changed the position, I, not part of, um, I think at some point I was part of this sound but not altogether. Okay, okay, okay, thank you, yeah, um, okay, did you feel like you were in control of the auditory space? Yeah, this seems to me like I'm in control, because it seems to me that the sounds was reacting to my moves, yeah, so in some way I was in connection with, okay. Okay. Would you say you experienced any specific kind of emotions, or what emotions did you experience during the experiment? I don't, curiosity, I think it was the one, yeah, yeah, the curiosity, would, to hear better, to imagine what is happening, what the ambience was, what was going on, yeah. Okay, well, that was it, thank you so much, thank you, I'll stop recording now, okay, thank you so much for participating.

A.10 Participant 10-B (25 June 2026, 12:30)

Okay, so first I'd like to ask you a lot of some questions about your experience during the experiment, and feel free to take your time to answer, and I want to ask you about how you feel, and there's no right or wrong, these questions are about the experience. So, during your experience, did you remember any specific sounds, do you remember any specific sounds from the experience that were memorable? Yes, I counted five, maybe six recordings, well, the one that was here behind you had a very distinct kind of humming in it, uh, there was of course the Oslo airport ding with the metro, or T-bane, no, sorry, that's the RJ train, um, and the other three, or I'm not sure if there was another one somewhere there, so the others were a bit less distinct, although there was this kind of glassy shooshing sound in that one that is in the background in that corner, yes, around the drum sets, um, yeah, no, I mean, of course all of them, or many of them, had this kind of ambient recording thing, so people speaking, some noisy textures of objects moving, or people being busy, vehicles, things like that, so, um, yeah, but I guess I said all the distinct ones.

And what makes you want to explore the rooms, different spaces? Well, I was just curious, um, of first of all how many there are, so in this first phase I kind of learned, or at least I got an intuition, that, like, what the system reacts to, so there was maybe a more like a childish means or aspiration to just explore the boundary levels or the boundary lines, and then I kind of got, for a long time, settled on four-ish areas, and then quite a lot into the thing, I accidentally stumbled upon this fifth one, the Oslo airport, so then I kind of started making rounds and see, oh, did I kind of miss, is there like another one, and I wasn't quite sure about that area, but maybe I should have paid more attention about what they are actually playing. Yeah, what was the question? I'm sorry. No, it was like, what drove you to explore these rooms? Ah, yeah, just curiosity, as you mentioned. I think it was also largely motivated by me realizing that the system reacts to my position, and I kind of tried some other things, I tried like crouching down or tilting myself in various ways or doing stuff with the hands, but to me it felt like it mostly reacts to either hips or some kind of like my kind of core or torso or something, so then I kind of just kept, how should I say, using that, or like I kind of at some point I just gave up on the hands, and maybe I shouldn't have, I don't know, but yeah, I was just kind of moving around, and, kind of, yeah, the interactivity was, I think, the biggest motivation, yeah.

And which kind of emotion did you feel during the experience? Yeah, first it was of course a bit awkward or artificial at the start, because of course you don't often experience this lab situation, like it was my second time, you know, having markers on me, so that was kind of awkward, and I felt a sense of, so at first I couldn't really let myself go, and I felt like, you know, a professional participating in another professional's experiment, so I was like having all my technical thoughts, and I had this kind of, it's not really a negative feeling, but like it's this kind of, not really judgmental either, but, you know, like analytic, I think, so I had this kind of like a detective kind of feeling, and then I kind of slowly eased into it, and, these kind of ambient sounds always make me kind of zoned out, so it was hard for me to focus on details, so when a distinct element in one of them came up, it was, um, kind of rewarding, so I guess, yeah, midway through I kind of forgot that it was an experiment and started kind of playing with the system as a child, because of course, and I was kind of feeling playful after a while, not necessarily, so I kind of, to be honest, started to ignore the original content of the sounds, I recognized that they were loops, and then when you restart loops a lot, you start to listen to it in a different way, so it's not necessarily about what is in there, like the source references, but it's more like just a spectrum morphology, or how should I say, like how the timbre evolves over time, etc., so I don't know, at some point I was silly enough to start trying to make beats with it or something, so, yeah, I think in the end it was very playful, and there were some occasional, you know, when you have this weird feeling of a busy place being silent, like when you go to, I don't know, a train station super late, and it's this, like, I don't know, there is this like the silence you can cut, we say, so it was, I had it twice, this feeling that I entered like an in-between zone, or like a silent zone, and I was just like, you know, that, I think this, this was all, yeah.

And so did you feel in control of the space then? Yeah, I mean, I had an impression that it's all up to me where, what is playing really, I mean, I cannot change the content, but if I want, I can

just make it all silent, if I want it can be this or that, like I can kind of mix it, I had the feeling that I was totally in control.

Do you feel immersed in the space, like you felt like one with the other, the space, um? Time to time, but not generally, because I felt that there was very little directionality in it, so I heard it as like a general kind of ambience, and sometimes I feel like I was somehow standing in the right position, and then I felt some directionality, but it wasn't ever very strong, those sounds that I mentioned earlier, those helped, so it's not like just a general noisy environment, but it's like some distinct sounds, but I also think, to be honest, I was probably a bit distracted by the interactive elements, maybe that also made me listen less to the, you know, fine details in the recordings, I don't know, sometimes, yeah, I got immersed, but generally I would say less, yeah.

And for the last question, which part did you feel that it was listening? Ah, yeah, uh, so I felt like I tried my hands several times, I could snap into a position sometimes where I could just kind of move my head, so I tried to make a stand in between two areas, exactly in the middle line, as far as I could tell, and so if I just tilt my head a little bit, then I can turn on that side, and if I go a little bit the other, like tilt my head a little bit, that I can turn on the other side, so based on that, I would say definitely my, either it was my head all along, but I also felt like sometimes my head alone didn't make the switch, but if I just like slowly pushed my hip towards it, then suddenly at some point it turned on, so it was, I feel like it was something with the hip and the head, or maybe torso or something like that, to be honest, the impression was, after spending some time with it, that it's kind of like a center of mass kind of thing, except it ignores the arms, because even if I tried to do things like this, it never really reacted to the arms, yeah.

Thank you so much for being part of the experience, for the interview, I'll stop recording now, and, yeah, well it was a pleasure.

A.11 Participant 11-L (26 June 2026, 12:00)

So I'd like to know how you felt during the experience, and what did you feel during it, and there's no right and wrong answers, I just want to know your experience, and first I would like to ask you if you remember any specific sounds during the experience, any of the sounds were memorable to you? Yeah, the most memorable sound was by far the train, well, it was two sounds, it was the train brake sound, and then there was this like a bell, like, or, you know, like signal, like when you're in the train hall and there's like a boom sound. I also think in the beginning I was experiencing this more natural watery wave sound combined with sounds of something that sounds like a cafeteria or a cafe. In the first phase? In the first phase, and that's also quite memorable, and I guess what also defined the experience to me was, was also the, just because it felt like it was, it was a lot of it was, um, the sound of people, just like, like the wash of people, um, yeah.

What are any emotions associated to it? I think at first I was like, um, pretty just curious and comfortable, and kind of like the further into it, it got a little bit more and more uncomfortable, I guess, it felt like it became more and more drastic, also with the cuts, it was very unpredictable.

And which part of your body felt like it was listening? Oh, good question, I don't know, actually I don't know, I don't have a clue about that.

Did your body feel immersed in the experience? Yeah. I definitely jumped back a few times, I was like, whoa, not, not like always, I think, like, I think, like, sometimes I was in it and sometimes I wasn't, but I think that there's many reasons, but for instance, like the cutting in and out, you know, made me like, what's going on now? And then like, oh, maybe it's over, and then I'm back into it, but I guess it's sort of immersive as well, maybe, I don't know.

And what made you want to explore the space? I guess sometimes you just, I don't know, it just felt natural to walk the way I did. Yeah. What drove you to explore, what were you looking for? I was, I guess, I don't know, you know, just following the impulse, not really sure, just like, if, also the awareness that it was like a 3D experience, I guess, maybe, made me curious, sometimes I moved because it felt like the sound was loud from an angle, you know, and I was like, oh, maybe it's more quiet the other way. And you wanted it to be quieter then, with the move? Yeah, I mean, for instance, with the train scream, or train sound, I think so, yeah, I think that was a reaction, my body was trying to get away from that sound.

Interesting. And it's the last question, to which extent did you feel one with the auditory space? Felt one like, felt like I was, in control of it. In control, oh, never. Okay. Never, never felt like I was in control of it. Was I? Oh my God, was I? Ah, that's so insane, what a waste of opportunity, I was so careful, wait, that's so crazy, it definitely did not feel like I was in control of it at all. Whoa. Oh, my God. What the hell?

That's it for my interview. Thank you so much for experiencing it, and I will stop the recording now.

A.12 Participant 12-H (30 June 2026, 13:30)

And the recording is starting now. I will now ask you a few questions about your experience, and at all times feel free to think back and take your time, ask me if you don't understand the question that I'm asking. Okay. During the experience, did any specific sounds feel memorable to you? Do you mean sort of external to the experience itself, or sort of in reference to? Both, whatever you noticed. Um, there were, I think what was, um, it was, what I think was memorable to me was sort of the sudden shifts, so when there was a sudden shift in the sound, and being able to, there's this one little area here where, um, it's sort of to one side, this sort of almost gate swinging sound, and then this kind of car horn sound, and, um, I think what was memorable to me in that sense was that it was, that I'd found some other areas, some other sounds, and I got a bit of an idea of, okay, how I can move my head here to sort of interact or to trigger them, but there I was still, I was trying to figure out, okay, what's the relationship between these, how are they, and the relationship, the space, I couldn't quite figure it out, and so I think that's sort of how it was memorable to me, was the sense that it was sort of, okay, how do these sounds interact with the spaces, there's something interesting in the way that I'm moving to trigger these, it was what made things memorable in that sense.

Okay, very cool, did you remember any specific sounds that stood out? Well, there was that part with the car horns and the gate, but there was also the train station kind of announcer sound that was also interesting to play around with a bit, there was also sort of these, again, sort of very linked to the space, I think, in the sense that there was, it's sort of, there were those kind of very prominent sounds where we're kind of relating it to sort of an external sound source, but there was a lot of sort of interesting sounds, which maybe I'm not kind of remembering, in the sense of kind of the sounds themselves, but how I was interacting with the space, I think, um, so there was one area I found where after being here a while I thought, well, I've just been standing the entire time, what happens if I duck, and it was an area where it was silent when I was standing, and ducked, and it triggered the sound, so that was sort of memorable, in that sense that, okay, right, I've got this, okay, I can play with verticality, with that moment when I kind of, that clicked, that was then memorable in that way, so it wasn't sort of directly linked to the sound itself, but rather how I was moving around in the space and how that related to the sounds.

That is so cool, thank you, you said that you tried to explore the room a bit, how did you choose which parts to explore? There wasn't really anything systematic, I just kind of started off, I think right at the beginning I was still standing in the middle, which was silent, and then sort of my head just kind of triggered a little bit, sort of, okay, right, there's stuff going on, and then sort of I just took my time trying to move around the room in different ways, sort of also on different scales, sort of thinking, okay, if I move around the room sort of broadly, just back and forward, or what happens if I kind of find an area that's interesting and then really sort of try and explore the nuances, um, sort of I felt very aware of, I suppose I'd say sort of the locus of the exploration was almost sort of in my neck, there was a lot of sort of feeling very present in my neck, especially on the kind of micro scale of how, at what point does the sound trigger when I sort of move my head just this tiny little bit, and again I think what I found interesting were the boundaries again, sort of the moments when there were sort of these shifts, and that kind of stuck out to me, okay, right, this is a little area that I want to explore for a while here, see what's happening here, so it was very much kind of exploration-led, but then when I found an interesting area, I kind of wanted to interact with that sort of for a while, and kind of test its boundaries and things like that.

Very cool, and you said you experienced your neck very present, yeah, um, how did you experience the rest of your body during the listening experience? Um, there was, um, um, it was sort of in a sense, kind of, I was, because I kind of felt that, okay, most of the exploration work is coming here from the neck, I was very focused on how is that then supporting this, in a way, so I didn't end up actually doing very much with my arms, I tried at one point doing a little bit of this, but it didn't really do much, apart from cause my upper body to also then shake a bit, so in that sense it triggered it, so my arms then felt a little bit less prominent in that sense, I think I just held them at my side for most of the time, there was some parts where I was trying to then also sort of, if I just slide my leg slowly, that kind of moves my upper body, so there, in that sense, I was kind of trying to think of it kind of as a chain up into my neck, that, okay, my leg can also sort of be the thing supporting this, but it was very much sort of this supporting role of, this area is sort of the magnifying glass, or the torch, shall we say, the head area is sort of the thing pointing out where

to explore.

Okay, cool, would you say that you experienced any kind of emotion, what kind of emotions did you experience during the experiment? Um, so the most prominent emotion I'd say would be the sort of the sense of exploration, okay, I've got this area here that I can explore and play around with, so this kind of curiosity behind it was the main driving force, but there were those moments when I found sort of areas of silence and narratives, I'm like, okay, this is actually sort of a very calming feeling, after being sort of in a very strong soundscape for such a long time, sort of coming back into the silence, almost very sort of peaceful and relaxing, and I enjoyed kind of those moments as well, kind of just playing with that area of silence, and just being, okay, if I just stand here, can I stay still, can I just breathe out and relax here, so yeah, I'd say it was sort of these oscillations between kind of curiosity and wanting to go out into the space, this kind of excitement, but then also with those intermittent moments of calm, and sort of wanting to just relax into the silence for a bit.

Did you feel in control of the room and the sound? I would say, I don't want to say sort of I was completely in control of it, but I was more towards that end of the spectrum, I would say, I felt very much that, okay, I can decide how much sound there is, and maybe kind of there were moments where it was sort of trying to figure out, okay, sometimes when I moved through the space, kind of the space there where the station was, for example, in, in the middle you mean, or, this area where there was sort of the announcement from the train station, um, that I could sometimes get it to trigger and sometimes not, but that was much more on sort of the micro scale, and sort of the more macro scale it was sort of, okay, I'm the one who's deciding if there's sound, how much sound, and there was also, for example, there was one knife, and there's like a line of silence there, where, like, if I stood very still I could be there, and then if I just slightly move my head there would be this sound, if I just slightly move my head that way there would be that, um, and that was very much this like, again, thinking about the motion on very kind of a very micro scale of, okay, it's very much I'm the locus here of if the sound is coming or not, and I mean, just in that sense, how connected I feel to my body, I can then talk about that in terms of control, like how much control do I have over my tiny movements, but it was still, I felt, stemming from me in that sense. Okay. You said the line of silence, was that in the back of the room there? It was here in the middle, in the middle between the back two, okay, sort of along where that yellow and green crosses.

Cool, and the one last question that I have for you is, did you feel immersed in the sound, did you feel one with the sound? I did find it very immersive, it was, I think, especially with the lights off, and there were moments I found myself also closing my eyes, especially the moments when I was standing still, I would just close my eyes, but it was still, for me, it always still went back to how I was moving in that sense, that it was the, I sort of didn't feel much of a pull to, okay, just stop and let myself be immersed, but kind of be very immersed in the space as somewhere to move through, as somewhere to explore, in that sense I felt very, very immersed, as sort of very, trying to map out spatially, okay, what is sort of the sonic map of the room in a sense, and so it got this very kind of also spatial quality to it as well, and the immersion, that it wasn't sort of, I

would say that there was sort of this, at times this inner immersion, sort of the moments when I was calm, but it was more sort of, I'd say, immersed in the room as a physical space, so it wasn't sort of an internally oriented immersion in that sense.

Cool, thank you, do you have anything else you would like to add about the experience? No, just that I really enjoyed it, I thought it was really good fun, I'm very glad. Cool, then I'll stop the recording.

A.13 Participant 13-N (30 June 2026, 15:00)

Then it's recording now, and if you at any point don't understand what I'm asking of you, please let me know, and always take your time, think back, just let me know when you're ready. Okay, so I'm very interested in your experience during it, like your emotional experience, and I wanted to know if there were any sounds that were particularly memorable to you during the experience. Yeah, there was one definitely, which was the airport sound, I could recognize that one, and there was like a clatter or something in this one, which is kind of. Which one? Yeah, the one in that corner. So left front? Right front, oh yeah, okay, this is front, yeah, then left, yeah, and then I can remember here you could hear like the steps in a hallway, yeah, and then the back left I could hear like sound, there's a music of some sorts, yeah, those are the ones I can remember. That one, I have a theory of what the different spaces are. Okay, tell us. I mean, I know what this one is. Okay. It's either, I think it's the airport or the central station, I imagine that this is like a hallway with hard surfaces. Okay, so left front, yeah. Yeah, you can hear the reverb a bit, and then this one is tricky to pin down. Right back. Right back, yeah, sorry, I'm so bad at right and left, so I'm terrible, this one, right back, I struggled to pin down, I think it was like, I think it was some sort of, I imagine a kebab or something, but it wasn't noisy enough to be that either, but it was like some sort of public space, this front right was more like a more formal space, like an office or office hallway, and then front left was a larger interior space, but it felt more industrial in its sounds, so it's more like, I remember it had like a beep sound or something, something that was weird that I couldn't really figure out what it was.

Okay, cool, thank you, how did you choose to explore the space, how did you choose to explore it? At first I didn't, I was a bit, I didn't understand how it worked, right, so I just messed around a bit, and then I just heard the noise, and I'm like, okay, sure, and then I tried to go in close to it, and then it stopped, and then I stood still, and then it came back again, so I think that if I moved it kind of stopped, that was at least what I thought first, what happened, and then it changed, I think, to being like geolocated to the different corners, and then, so at first I was like, okay, I can quickly move, then stop, and then like mess with it, so I can see if I could get it like to start, stop or something, I'm not sure if I got that working properly, but then it switched to the corners, and then I tried to explore the boundaries between the sounds, and of course going as close as possible to the corner and the source of the sound, and then I tried to run in a circle to see what happened if I switched super quickly, and I also went into the center and I tried to find the transitional zone between where it would trigger the sound, and if I could find that location somehow.

Cool, how did you experience your body during the experience, what parts did you feel like were important? I mean, I felt like my location was the one that meant the most, but I could use my extremities to kind of reach into some space and also try to span two of them, but I wouldn't say that I felt there was a difference between if I use my legs or my foot, it felt more like my location rather than my specific body parts, I think.

What did you experience your body like when you were moving, did you feel present in your body, or was it more like not relevant? More, I was more exploring, so I was a little bit in my body maybe when I was exploring, like, the thresholds and that, but otherwise I was mostly in the space, if that makes sense, yeah, cool.

What kind of emotions did you experience? Curiosity mostly, and intrigue, but that's maybe the same, and a little bit fun as well, when I did the silly run of the circle, I did laugh at myself because it was a bit stupid, but it's, so yeah, it was a bit fun as well, yeah.

That's so cool, what else do I have? Intriguing, but maybe did I say that, I did, um. To what extent would you say you felt like you were being one with the auditory space, or like you were immersed in it? I'd say pretty, like I tried to really figure out what the spaces were, and then I feel pretty immersed in that, yeah.

Nice, that's cool, did you feel in control of the space? When I figured out how the thing worked, then I could, yeah, I felt like I could work with it, and like I had some control, yeah.

Cool, do you have anything else that you would like to tell us about the experience? There's one thing that I do feel a sense of, that I have to perform, because of what I work with, so I know that I have the want to do something well, if that makes sense, so I can't do it fully neutral, that's the only thing I notice, I want to, subconsciously, I want to impress, you know, and that might have something to say, I don't know, I wouldn't say I was completely natural, but I don't, maybe nobody is, yeah, that's it.

That was it, thank you so much, I hope I didn't leave anything out, oh, thank you so much. Fun. Yeah.