Politecnico di Torino

Corso di Laurea Magistrale di Ingegneria del Cinema e dei Media Digitali
A.a. 2025/2026
Sessione di laurea Marzo 2026

A Virtual-Reality and Soundscape-Based Approach for Assessment of Urban Sound Design

Relatrice:
prof.ssa Louena Shtrepi

Candidato:
Giorgio Spegis

Table of Contents

List of Tables V

List of Figures VII

1 Introduction 1
1.1 Background and Context . . . 1
1.1.1 The Origins of Soundscapes . . . 1
1.1.2 Soundscapes and Human Health . . . 3
1.1.3 Soundscapes in Urban Design and Environmental Evaluation 3
1.2 Research Questions . . . 4

2 Literature Review 5
2.1 Methods for Perceptual Assessments of Soundscapes . . . 5
2.1.1 Soundwalks . . . 6
2.1.2 Questionnaires and Descriptors . . . 8
2.1.3 Psychoacoustic Metrics . . . 9
2.2 Virtual Environments in Soundscape Studies . . . 13
2.2.1 Recent Applications in Research . . . 13
2.2.2 Methodologies and Approaches . . . 14
2.3 Ecological Validity in Current Research . . . 19

3 Application of Soundscape approach 22
3.1 Soundwalk Case Study: Turin . . . 22
3.1.1 Path Analysis . . . 23
3.1.2 Audio Acquisition Setup . . . 26
3.1.3 Video Acquisition Setup . . . 28
3.1.4 Questionnaire . . . 30
3.2 Data Analysis . . . 30
3.2.1 Questionnaires Evaluations in field . . . 30
3.2.2 Psychoacoustic parameters measurements in field . . . 35
3.2.3 Data Correlation in field . . . 45

3.3 Real and Virtual Environment Models 49
  3.3.1 Real model building 49
  3.3.2 Virtual Model Construction 49
3.4 Laboratory Experiment 58
  3.4.1 Experimental Design and Signal Flow 60
  3.4.2 Questionnaire 62

4 Results 64
4.1 Psychoacoustic parameters comparison 64
4.2 Subjective parameters comparison 68
4.3 Comparison between Objective and Subjective Parameters 75
4.4 Immersivity 76

5 Conclusion 80

Bibliography 85

A Soundscape Assessment Questionnaire 93

List of Tables

2.1 Review Table for virtual soundscape simulation tools. 19

3.1 Soundwalk design for each spot. 23
3.2 Soundwalk spots summary. 25
3.3 Audio acquisition setup. 27
3.4 Video acquisition setup. 28
3.5 Sample size for each spot. 32
3.6 Complete SPL data for each location. 39
3.7 Complete loudness data for each location. 40
3.8 Complete roughness data for each location. 41
3.9 Complete sharpness data for each location. 41
3.10 Complete fluctuation strength data for each location. 42
3.11 Descriptive statistics and ISO 12913-2 attributes for each location. . 47
3.12 Classification of sound sources. 54
3.13 SPL target for each sound source. 56
3.14 SPL headphone calibration gain matching values. 63

4.10 Descriptive statistics and ISO 12913-2 attributes for each condition in RB01. 69
4.11 Descriptive statistics and ISO 12913-2 attributes for each condition in RB02. 69
4.12 Descriptive statistics and ISO 12913-2 attributes for each condition in BL. 71
4.13 Intra-Laboratory Comparison (HP360 vs SP360): Mann-Whitney U test results and descriptive statistics. 72
4.14 In field-In lab Comparison (IF vs HP360): Mann-Whitney U test results and descriptive statistics. 73
4.15 In field-In lab Comparison (IF vs SP360): Mann-Whitney U test results and descriptive statistics. 74
4.16 In lab Comparison (HP360 vs VR): Mann-Whitney U test results and descriptive statistics (RB02 stop). 74

4.17 In lab Comparison (VR vs AVR): Mann-Whitney U test results and descriptive statistics (RB02 stop). 75
4.18 Spearman's significance correlations (HP360). 76
4.19 Spearman's significance correlations (IF). 78
4.20 Spearman's significance correlations (SP360). 78

A.1 In field questionnaire. 95
A.2 In lab questionnaire. 97

List of Figures

2.1 Soundscape definition graph [1] 5
2.2 Example of a soundwalk [1] 6
2.3 SSID Protocol procedure [**mitch**]. 8
2.4 PAQ's scatter plot [12]. 9
2.5 Weighting curves 11
2.6 Fletcher-Munson curves 11
2.7 Workflow and experimental setup [3]. 13
2.8 Pictures from *Wessels et al.* experiment [15]. 14
2.9 City Ditty [22]. 15
2.10 IHTA Park experiment [24]. 16
2.11 TASCAR scene representation. 17
2.12 Virtual Acoustic RAVEN tracing calculation. 17
2.13 Real world frame and ODEON model [27]. 18
2.14 Industry professionals concerns about VR games side effects from *Lavoie et al.* research [35]. 20

3.1 Soundwalk path. 24
3.2 Rebaudengo Station view. 25
3.3 Spazio 211 picture during soudwalk. 26
3.4 Spazio 211 view. 27
3.5 Bologna park picture during soundwalk. 28
3.6 Audio video capture setup during soundwalk with Zylia M-1 microphone and Insta360 One X2 camera. 29
3.7 Simcenter SCADAS XS binuaural headphones. 29
3.8 Location shots taken with Insta360 One X2 at different points. . . . 31
3.9 Age distribution from soundwalk. 32
3.10 Gender distribution. 33
3.11 Sound source domination distribution. 34
3.12 Architectural and urban design rating. 34
3.13 Sound environment appropriateness rating. 35
3.14 Sound environment perception. 35

3.15 Scattering plots from soundwalk spots. 36
3.16 Density plots from soundwalk spots. 37
3.17 Join plots from soundwalk spots. 38
3.18 Comparison of Sound Pressure Level (SPL) Indicators per Stops. . . 40
3.19 Loudness values from every spot. 42
3.20 Roughness values from every spot. 43
3.21 Sharpness values from every spot. 44
3.22 Fluctuation Strength values from every spot. 45
3.23 Pearson field corelogram. 46
3.24 Spearman field corelogram. 47
3.25 Comparison of the soundscapes of the three urban spaces. 48
3.26 Virtual environment model workflow. 50
3.27 Blender's scene pictures during environment building. 51
3.28 *Unity game window* 52
3.29 Rebaudengo noise map. 53
3.30 FMOD Studio screen 55
3.31 Procedural Acoustic System 57
3.32 Fibonacci's sphere logic [29]. 58
3.33 Image from Audio Space Lab (ASL). 59
3.34 Plogue Bidule screen. 60
3.35 Picture from calibration process. 61
3.36 Artificial head by HEADAcoustic with Meta Quest 2 visor and Senheiser HD 600 headphones. 62
3.37 Experiment audio reproduction chain. 63

4.1 Frequency response differences between SP360 and HP360 conditions in RB02 station. 65
4.2 Correlation tables for psychoacoustic parameters divided in simulation methods. 67
4.3 Subjective comparisons between setups and environments plots. . . 70
4.4 Correlation tables between psychoacoustic and subjective parameters for each condition. 77
4.5 Subjective comparisons between setups and environments plots. . . 79

Chapter 1

Introduction

1.1 Background and Context

The soundscape is a fundamental component of the human experience. It evolves alongside societies, technologies, and environments, continuously reshaping the way individuals perceive and inhabit the world. Consequently, the study of soundscapes spans multiple disciplines — from psychology and acoustics to anthropology and medical research — each contributing to an understanding of how sound influences human behavior, cognition, and well-being.

In *The Soundscape: Our Sonic Environment and the Tuning of the World* (1977), Canadian composer and scholar Raymond Murray Schafer poses a pivotal question:

> *"Is the soundscape of the world an indeterminate composition over which we have no control, or are we its composers and performers, responsible for giving it form and beauty?"*

This reflection outlines two distinct attitudes toward the sonic environment: passive submission to an uncontrolled acoustic reality, or active engagement in shaping the world we inhabit. This research is grounded in the latter perspective — the conviction that sound is not merely an accidental byproduct of modern life, but a designable, meaningful, and influential dimension of human experience.

1.1.1 The Origins of Soundscapes

The term "soundscape" stems from the concept of the landscape and was first introduced by urban planner Michael Southworth in *The Sonic Environment of Cities* (1969). Southworth defined the soundscape as "*the quality and type of sounds and their arrangements in space and time.*"

This conceptualization aimed to establish a standardized terminology capable of serving as a metric framework for the field. However, Southworth lacked access to the advanced methodological and technological tools necessary for an in-depth investigation, due to the technical constraints of the era. Consequently, his analyses were primarily restricted to qualitative data and subjective responses regarding sound perception within specific contexts. Nevertheless, his work provided a crucial foundation, fostering the expectation that future technological advancements would facilitate more precise terminology and systematic descriptors.

The term was further developed by Raymond Murray Schafer in 1969 with the publication of *The New Soundscape* and *The Book of Noise*. Schafer established a profound connection between soundscapes and human experience, analyzing them across contexts ranging from music to philosophy and acoustics.

From these seminal works, several key definitions emerged. Schafer and his students categorized soundscape components into three classes: **natural sounds**, **human sounds**, and **sounds produced by tools and technology**. His research tracked the evolution of these classes throughout history, demonstrating that sounds generated by technology have become increasingly dominant. This classification remains so influential that it serves as a core parameter in contemporary regulations, such as **ISO 12913-2:2014**.

In *The Soundscape: Our Sonic Environment and the Tuning of the World* (1977), Schafer identifies a major concern: the increasing exposure of human beings to loud sounds in all aspects of life. While noise pollution is a concrete problem, Schafer argues that noise abatement alone is insufficient. Instead, the acoustic environment should be approached positively by studying every element in relation to human perception. Thus, soundscape studies emerge as a multidisciplinary "middle ground" between science, society, and art.

Other essential definitions from his research, still relevant for evaluation today, include:

- **Keynote**: The fundamental tone around which other sounds modulate, analogous to the musical key or tonality of a composition.
- **Signals**: Foreground sounds that are consciously perceived, such as bells, whistles, horns, and sirens.
- **Soundmark**: A community sound that is unique or possesses qualities that make it specially regarded or noticed by a community, deserving of protection.

Reflecting on the evolution of human environments, Schafer expresses concern that industrial and digital noise, alongside transportation systems, progressively masks natural and local soundscapes. This process leads to a homogenization of acoustic environments, resulting in a loss of cultural traditions and a weakening of the individual’s sense of belonging.

1.1.2 Soundscapes and Human Health

Over the past decades, extensive research has demonstrated a strong correlation between soundscape perception and human health. In *"Ten questions on the soundscapes of the built environment," Kang et al.* [1] emphasize how carefully designed acoustic environments enhance well-being, reporting benefits ranging from improved sleep quality to positive effects on blood pressure regulation.

Further evidence is provided by *Van Kamp et al.* [2], who highlight a significant correlation between environmental sound perception and annoyance levels, particularly in urban settings. Although many studies still overlook natural settings, these findings underline the critical role of soundscapes in shaping everyday quality of life.

Complementary results are presented by *Latini et al.* [3], who investigated the effects of different soundscapes within identical virtual office environments using **Virtual Reality (VR)**. By monitoring neurological responses and blood pressure, their study demonstrates how soundscape variations can substantially influence both mental and physical health. Collectively, these studies reinforce the necessity of integrating soundscape assessment into environmental design.

1.1.3 Soundscapes in Urban Design and Environmental Evaluation

Recent research has led to a growing awareness of soundscape evaluation in urban planning. A holistic, interdisciplinary approach is now recognized as essential for addressing sound-related phenomena.

In this context, *Kurkose Cal et al.* [4] demonstrated how soundscape perception varies according to social and contextual conditions. Their study revealed significant differences in perception between academic staff and students, underlining the need to consider demographic factors to properly interpret evaluation data.

Consequently, soundscape research has moved to the forefront of evaluating both outdoor urban spaces and indoor environments. The introduction of the **ISO 12913-2** standard marked a milestone, promoting methodologies such as *soundwalks* and *in situ* assessments. Additional approaches, such as the **SSID Protocol** developed by *Mitchell et al.* [5], combine audio-visual recordings with subjective questionnaires, showing how emerging technologies can enhance assessment practices.

The COVID-19 pandemic represented a further turning point. As discussed by *Rehman et al.* [6], the need to analyze urban environments during lockdowns accelerated research in Virtual Reality. This technology has shifted the focus toward new design approaches, enabling the evaluation of environments under multiple controlled conditions—a significant advancement for simulation frameworks and customization strategies.

1.2 Research Questions

Virtual Reality offers new opportunities for the advanced investigation of environmental soundscapes, enabling immersive exploration while introducing novel methodological and technological challenges. Within this framework, this thesis aims to address two specific research questions:

RQ1: *Is it possible to define a methodology for the development of a physically based virtual reality soundscape environment?*

RQ2: *What differences emerge between soundscape evaluations conducted in virtual environments and those performed in real-world conditions?*

Based on these questions, the following literature review provides the foundations for framing the state of the art, outlining both the potential and the limitations of contemporary VR applications in soundscape research.

Chapter 2

Literature Review

2.1 Methods for Perceptual Assessments of Soundscapes

Over the years, numerous studies have sought unified methodologies to assess soundscapes, aiming for consistency across diverse measurement contexts. Consequently, the International Organization for Standardization (ISO) published the first edition of the ISO 12913 standard in 2014, defining a soundscape as the "*acoustic environment as perceived or experienced and/or understood by a person or people, in context*" [7].

This standardization aims to define the soundscape by focusing on the distinction between the acoustic environment and the soundscape itself, considering the latter as the multidimensional result of various intersecting factors, such as the social, health, and acoustic domains (**Figure 2.1**).

Figure 2.1: Soundscape definition graph [1]

.

The ISO 12913-2 standard [8], published in 2018, outlines a comprehensive framework for data collection, ranging from specific descriptors to established methodologies such as soundwalks (**Figure 2.2**) and guided interviews.

Figure 2.2: Example of a soundwalk [1]

The last version published, the ISO 12913-3 [9], in 2025 explores the data interpretations and the ways to interpretate and analyze all parameters collected during measurements.

2.1.1 Soundwalks

A soundwalk is an empirical method used to identify a soundscape and its constitutive components across various locations [10]. The term was originally coined by R. Murray Schafer during the late 1960s and early 1970s.
During a soundwalk, a group of participants follows a pre-defined path featuring specific stops, where the soundscape is assessed through standardized questionnaires [11]. Typically, participants utilize binaural microphones or headphones to record the acoustic environment, while documenting the experience with videos, photographs, and field notes. Within this framework, every piece of information is

potentially significant, given the interdisciplinary nature of soundscape research.

With the emergence of immersive technologies such as Head-Mounted Displays (HMD) and Virtual Reality (VR), the Soundscape Indices (SSID) Protocol was developed. This protocol aims to collect the high-fidelity audio and video data during soundwalks necessary to replicate those scenarios in a laboratory setting [5]. The methodology requires a rigorous timeline to avoid interfering with the soundwalk experience and the environmental assessment, as illustrated in **Figure 2.3**.

Furthermore, when adopting the SSID protocol during soundwalks, it is essential to implement a specialized setup capable of capturing several distinct data typologies:

- **Questionnaire** (QUE) for subjective people assessment;

- **Video** (VID) to reproduce the same visual scenario in laboratory;

- **Pictures** (PIC) to take material from field;

- **Binaural** (BIN) recordings to collect useful data for psychoacoustic parameters analysis;

- **Ambisonic** (AMB) recordings to accourately reproduce in laboratory the audio scenario;

- **Sound levels** (SLM) to take note of the power from sources useful for laboratory calibration;

- **Environmental data** (ENV) to take in count other extra data regarding the environment as weather or temperature.

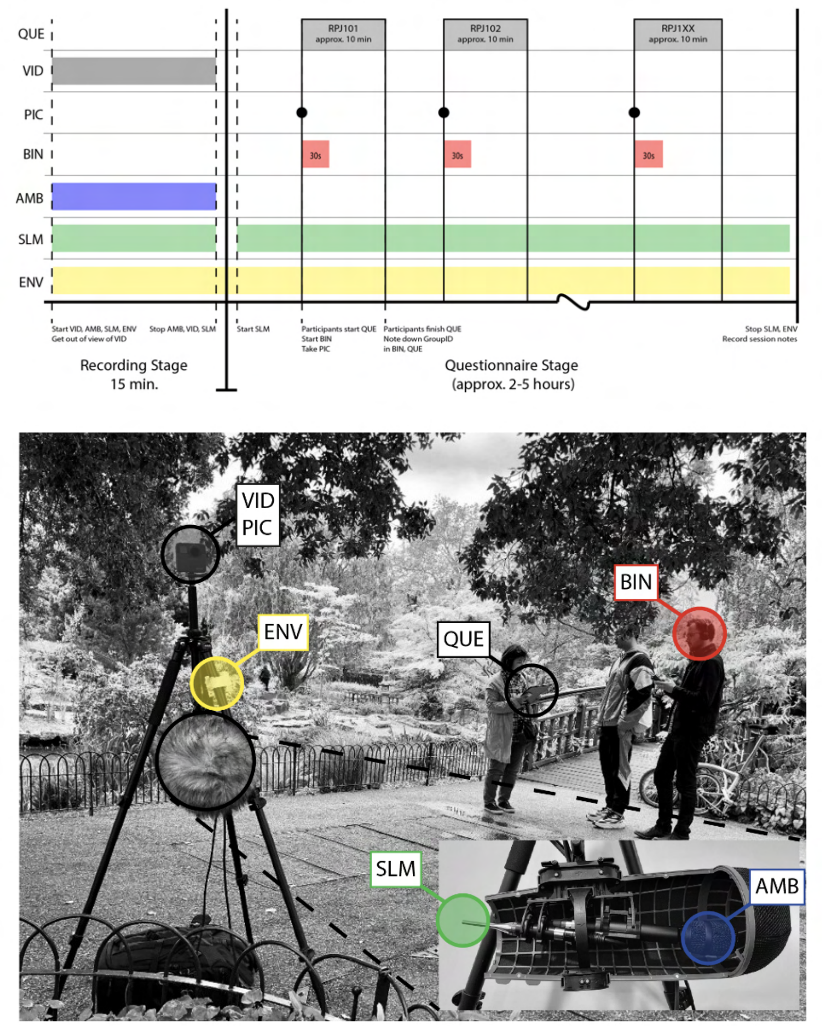


Figure 2.3: SSID Protocol procedure [**mitch**].

2.1.2 Questionnaires and Descriptors

Questionnaires serve as the primary tools for collecting data in subjective soundscape assessments, providing a comprehensive overview of the participants' perceptions during the soundwalk. To be effective, these questionnaires must include specific sections that balance general information about the listener with their immediate sensations during the assessment, following the methodology proposed by *Mitchell et al.* [5].

Within the assessment questionnaire, eight specific descriptors—known as Perceived Affective Qualities (PAQs) [12]—were utilized to evaluate subjective responses, as illustrated in **Figure 2.4**. These descriptors are distributed along four axes:

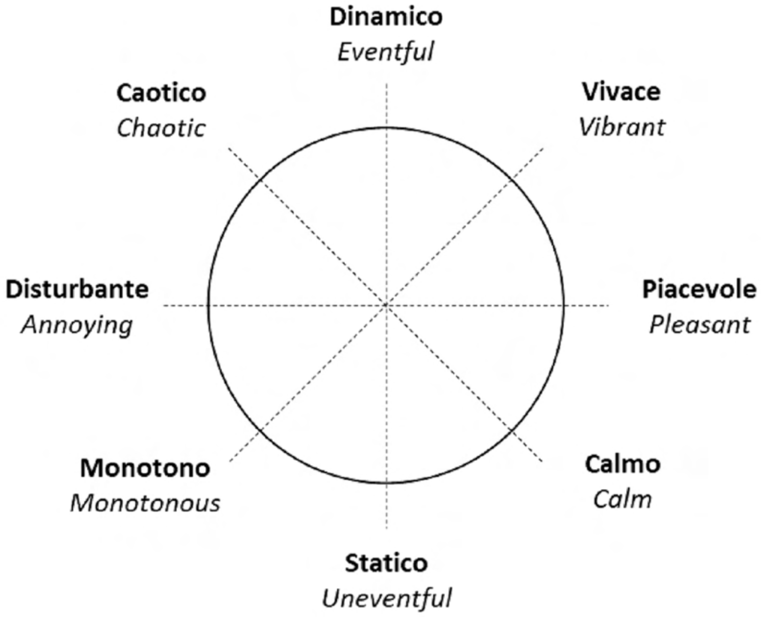


Figure 2.4: PAQ's scatter plot [12].

- Eventful and Uneventful;
- Pleasant and Annoying;
- Calm and Chaotic;
- Vibrant and Monotonous.

The resulting scatter plot is structured according to the *pleasantness/eventfulness* model. These descriptors are integrated as responses within the questionnaire provided to the participants. Once the data are collected, it is possible to analyze the overall subjective assessments by plotting them on a two-dimensional circumplex model.

2.1.3 Psychoacoustic Metrics

To evaluate the recorded data from soundscape assessments using a standardized methodology, it is essential to employ psychoacoustic metrics. These metrics serve as a set of indicators that characterize the acoustic environment by accounting for both spectral content and sound pressure levels.

The **Sound Pressure Level** (SPL) is defined as the local pressure deviation from the ambient atmospheric pressure caused by a sound wave [13]. It is expressed according to the following equation:

$$L_p = 10 \log_{10} \left(\frac{p}{p_{ref}} \right)^2 = 20 \log_{10} \left(\frac{p}{p_{ref}} \right) \text{ [dB]} \tag{2.1}$$

In situations where it is necessary to evaluate sound pressure levels that vary over time, it is standard practice to use the **equivalent continuous sound pressure level** (L_{eq}), defined by the following equation:

$$L_{eq} = 10 \log_{10} \left[\frac{1}{T} \int_0^T \frac{p^2(t)}{p_0^2} dt \right] \text{ [dB]} \tag{2.2}$$

where $p(t)$ represents the instantaneous sound pressure and p_0 is the reference pressure (20 μPa in air).

However, the human ear perceives sound pressure levels differently depending on the frequency. This frequency-dependent sensitivity is accounted for using weighting curves, as shown in **Figure 2.5**, which are derived from the Fletcher-Munson equal-loudness contours (**Figure 2.6**). Generally, the most common is the A-weighting curve, which approximates the human ear's response at low-to-moderate sound levels (specifically, the 40-phon curve).

For a given time interval T, the A-weighted equivalent continuous sound pressure level is calculated as:

$$L_{Aeq,T} = 10 \log_{10} \left[\frac{\frac{1}{T} \int_{t_1}^{t_2} p_A^2(t) dt}{p_0^2} \right] \text{ [dB]} \tag{2.3}$$

where $p_A(t)$ is the instantaneous A-weighted sound pressure at time t and $p_0 =$ 20 μPa.

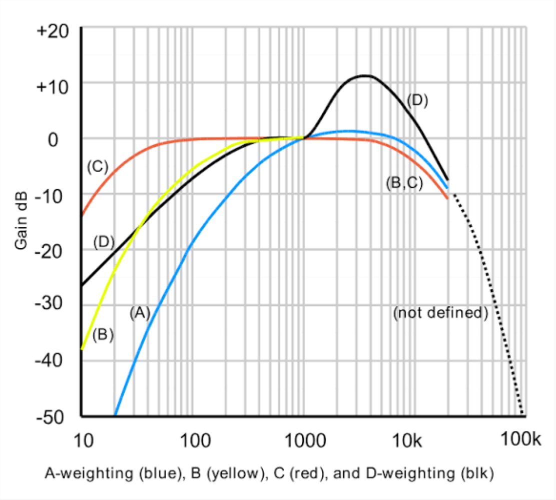


Figure 2.5: Weighting curves

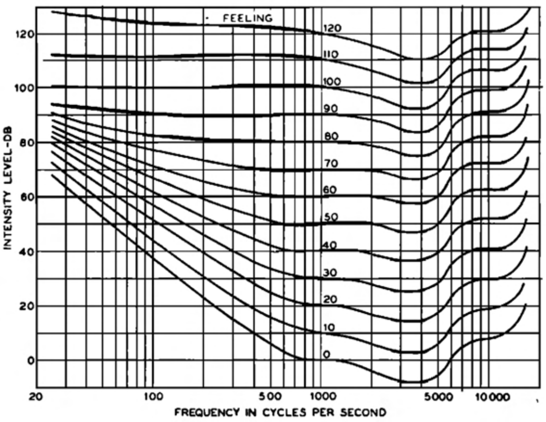


Figure 2.6: Fletcher-Munson curves

Loudness is defined as the "*attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud*" [14]. It is a multidimensional quantity that depends on physiological, physical, and psychological components, representing the subjective perception of sound pressure levels. Loudness is measured in *Sones*; by definition, 1 Sone is equal to 40 Phons, which corresponds to the perceived loudness of a 1 kHz sinusoidal tone at 40 dB SPL.

The relationship between the loudness level in Phons (N_{Phon}) and the loudness in Sones (N_{Sone}) is expressed by the following equation:

$$N_{Sone} = 2^{\frac{N_{Phon}-40}{10}} \tag{2.4}$$

For soundscape standards there are considered the average value ($N_{average}$), the rms value (N_{rmc}), the N_5, N_{95} and their ratio ($\frac{N_5}{N_{95}}$).

Sharpness is the psychoacoustic parameter which indicates how much high frequencies are present into an audio sample. The amount of high frequencies is annoying and make the sound more acid. The unit of measure is *acum*, defined as the noise for a bandwidth with critical band size, 160 Hz ,with 1 kHz as central frequency at 60 dB SPL.
As the frequencies spectrum human audible are subdivived in 24 critical bands, defined as the frequency interval where two pure tones simultaneousy played are not indistinguishable, these 24 bands are defined as Bark's scale. So, with this method, the spetrum content is splitted from 1 Bark to 24 Bark.
The Zwicker and Fastl model sharpness model is proposed as:

$$S[acum] = 0{,}11 \sum_{0}^{24Barks} \frac{N'(z) \times z \times g(z)dz}{N} \tag{2.5}$$

with N' as the specific number for each of 24 barks intervals, z as bark band number and N as total loudness. The $G(z)$ function represents a ponderation factor and its values are the following:

$$g(z) = \begin{cases} 1 & \text{if } z \leq 15.8 \text{ Bark} \\ 0.15e^{0.42(z-15.8)} + 0.85 & \text{if } z > 15.8 \text{ Bark} \end{cases} \tag{2.6}$$

In soundscape evaluation, the most significant indicators for sharpness are the average value (S_{avg}) and the percentile values S_5 and S_{95}.

Tonality is a psychoacoustic parameter that quantifies the prominence of tonal components relative to the broadband noise within a soundscape. According to the ECMA-418-2 standard, the *partial loudness* concept is employed to account for the loudness of a masked sound in relation to its original loudness when propagating in a quiet environment.
The unit of measurement is the tu_{HMS} (Tonality Unit, Hearing Model of Sottek). One tu_{HMS} corresponds to the tonality of a 1 kHz pure tone at 60 dB SPL in a quiet environment.

Roughness (R) quantifies the perception of rapid amplitude modulations within a critical band. This sensation reaches its maximum intensity at a modulation frequency of 70 Hz, beyond which it decreases.
Its unit of measurement is the *asper*. One asper is defined as the roughness produced

by a 1 kHz tone at 60 dB SPL, 100% amplitude-modulated at a frequency of 70 Hz.

Similarly, **Fluctuation Strength** (F) is related to amplitude modulation phenomena. However, it characterizes slower modulations, reaching its peak at a 4 Hz modulation frequency; the fluctuation effect typically vanishes when modulation frequencies exceed 20–30 Hz.
Its unit of measurement is the *vacil*. One vacil is equivalent to a 1 kHz tone at 60 dB SPL, 100% amplitude-modulated at 4 Hz.
In soundscape analysis, research typically accounts for the statistical descriptors of these last two metrics, specifically R_{10}, R_{50}, F_{10}, and F_{50}.

2.2 Virtual Environments in Soundscape Studies

2.2.1 Recent Applications in Research

Virtual Reality (VR) provides researchers with advanced tools to investigate environmental challenges from alternative and highly controlled perspectives. The ability to repeatedly experience the same scenario under varying, calibrated conditions represents a significant methodological advantage, particularly in urban soundscape design and related interdisciplinary fields. For instance, *Latini et al.* [3] investigated human physiological and psychological responses to diverse audiovisual conditions within a simulated work environment. Such technologies enable the systematic creation of scenarios that would be complex, costly, or impossible to reproduce in real-world settings (**Figure 2.7**). Similarly, *Wessels et al.* [15] employed a

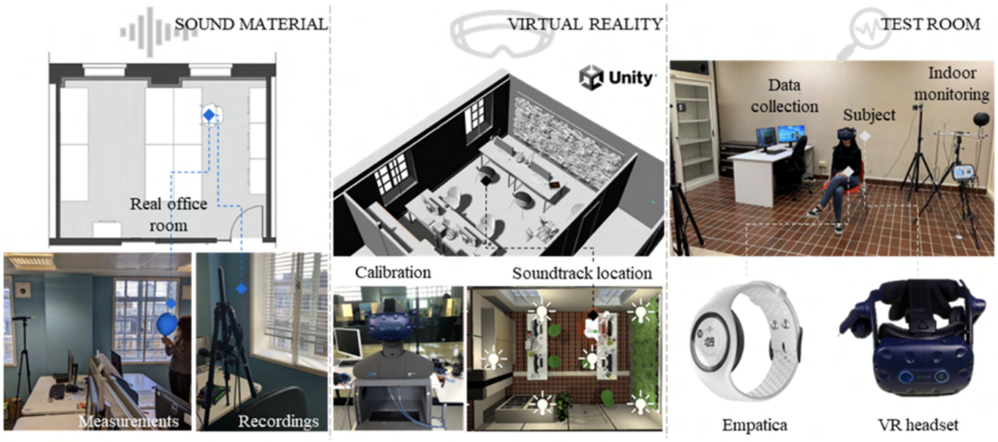


Figure 2.7: Workflow and experimental setup [3].

Virtual Reality framework to analyze perception differences in time-to-collision between internal combustion engine vehicles (ICEVs) and electric vehicles (EVs) (**Figure 2.8**). This approach facilitated the simulation of diverse traffic scenarios while addressing a critical challenge in road safety research. More broadly, recent literature has increasingly adopted Virtual Reality as a robust tool for soundscape research, extending the analytical capabilities of traditional evaluation methods. These technologies enable highly flexible approaches for investigating acoustic environments within specific, reproducible spatial contexts.

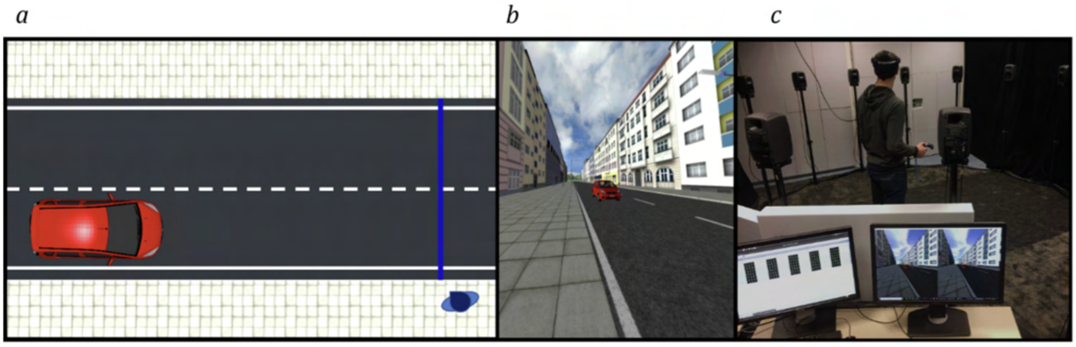


Figure 2.8: Pictures from *Wessels et al.* experiment [15].

2.2.2 Methodologies and Approaches

As a rapidly evolving field, Virtual Reality-based soundscape research currently lacks a unified or standardized workflow. Nevertheless, diverse methodological approaches have emerged, reflecting varying research objectives and technical constraints. A comprehensive systematic review by *Tokgöz et al.* [16] examined the application of VR in soundscape evaluation, identifying a wide range of experimental configurations. Among these, the paradigms most aligned with the objectives of this thesis were further analyzed.

Virtual experiences are currently implemented through distinct interaction paradigms. Some studies adopt desktop-based solutions, where participants navigate virtual environments using standard displays and headphones [17]. Other investigations employ Head-Mounted Displays (HMDs) to provide a more immersive experience, often combined with headphones or multi-screen visualization systems [18].

Visual representation strategies also vary considerably. The use of 360° video playback enables the replication of specific environments with high visual fidelity, capturing real-world events as they occurred [19]. However, such approaches inherently limit the capacity for environmental manipulation. Conversely, three-dimensional computer-generated (CG) environments offer a highly customizable framework, albeit at the cost of increased modeling complexity. This is particularly relevant for auralized environments, where acoustic simulation requires the rigorous modeling of physical phenomena such as sound propagation, reflection, absorption, diffusion, and scattering. Compared to Ambisonic recordings, these auralization-based approaches provide a higher degree of controllability, allowing for the detailed manipulation of environmental parameters.

In this context, auralization requires specialized software capable of physically based acoustic simulation integrated with complex urban geometries. Several tools have been evaluated for accuracy and computational efficiency, as summarized in **Table 2.1**. Regarding visual design platforms, Unity [20] remains one of the most widely adopted engines for soundscape simulation. Its flexible scripting environment and integrated audio framework make it highly suitable for the standardized representation of real-world scenarios. For example, *Jiang et al.* [21] proposed a web-based virtual soundscape assessment platform for Piazza Vittoria in Naples. Furthermore, recent studies have emphasized interactive approaches grounded in Human–Computer Interaction (HCI), promoting active user engagement during the soundscape evaluation process [22] (**Figure 2.9**). Unreal Engine [23] represents another prominent visual platform, offering advanced graphical capabilities that enable the creation of highly realistic environments with sophisticated lighting and material rendering. In the study conducted by *Rehman et al.* [6] (**Figure 2.10**), all scenarios were developed using Unreal Engine 5. This engine facilitated the seamless integration of multiple environmental variables within a single, coherent

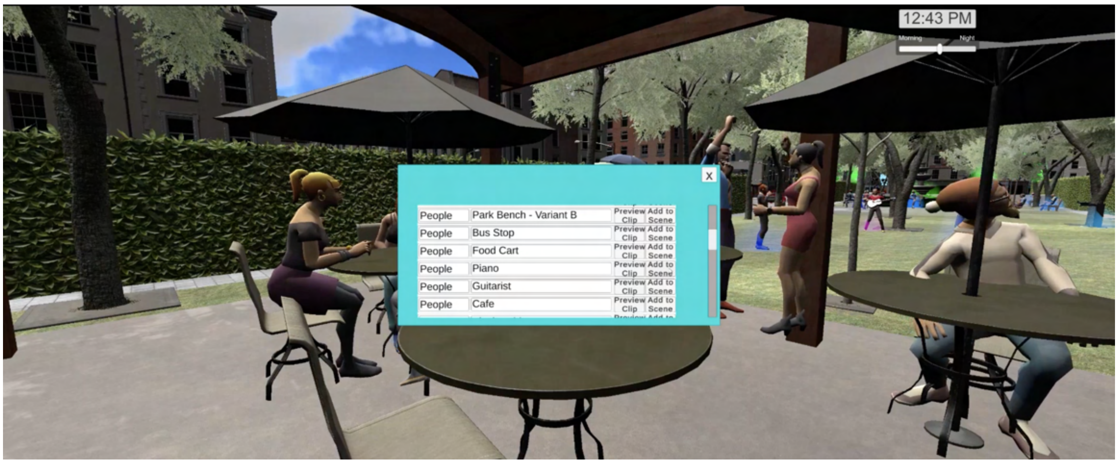
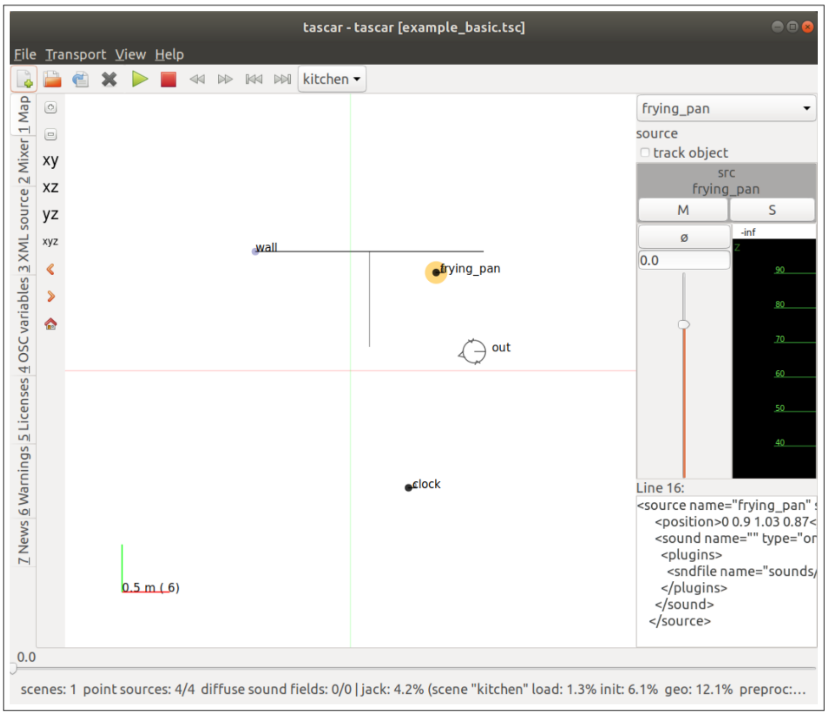


Figure 2.9: City Ditty [22].

virtual space, providing a robust framework for high-fidelity simulations. From an acoustic simulation perspective, various tools with different levels of complexity have been proposed. The native Unity Audio Framework provides a lightweight solution for basic simulations, including sound source directivity, occlusion, and distance-based attenuation modeled through logarithmic approximations of physical laws [21]. However, the inherent lack of accurate reflection and absorption modeling often necessitates the integration of external acoustic engines to achieve higher realism. For instance, *Wessels et al.* [15] utilized the open-source software TASCAR (Toolbox for Acoustic Scene Creation and Rendering) [25]. This tool enables the simulation of complex acoustic scenes and can be synchronized with graphical engines, provided that a robust temporal and spatial alignment between audio and visual components is maintained. Virtual Acoustic [26] (also known as RAVEN) represents another advanced solution, offering high-fidelity acoustic simulation (**Figure 2.12**) and seamless compatibility with external engines such as Unreal Engine [24]. The simulation process begins with pre-recorded source data, which are used to reconstruct the acoustic field within a virtual environment. Through its modular architecture, the software enables the simulation of complex acoustic phenomena, while its output modules allow for the simultaneous routing of sound to multiple devices. Furthermore, the system is highly adaptable, supporting various clients and programming languages. A notable limitation of this software, however, is its exclusive optimization for Windows-based operating systems. In contrast, *Taghipour et al.* [27] employed ODEON [28] to evaluate the soundscapes of inner courtyards (**Figure 2.13**). Although ODEON is widely recognized as

Figure 2.10: IHTA Park experiment [24].



Line 16:

```
<source name="frying_pan" s
    <position>0 0.9 1.03 0.87<
    <sound name="" type="or
     <plugins>
       <sndfile name="sounds/
     </plugins>
    </sound>
  </source>
```

```
scenes: 1  point sources: 4/4  diffuse sound fields: 0/0 | jack: 4.2% (scene "kitchen" load: 1.3% init: 6.1%  geo: 12.1%  preproc:...
```

Figure 2.11: TASCAR scene representation.

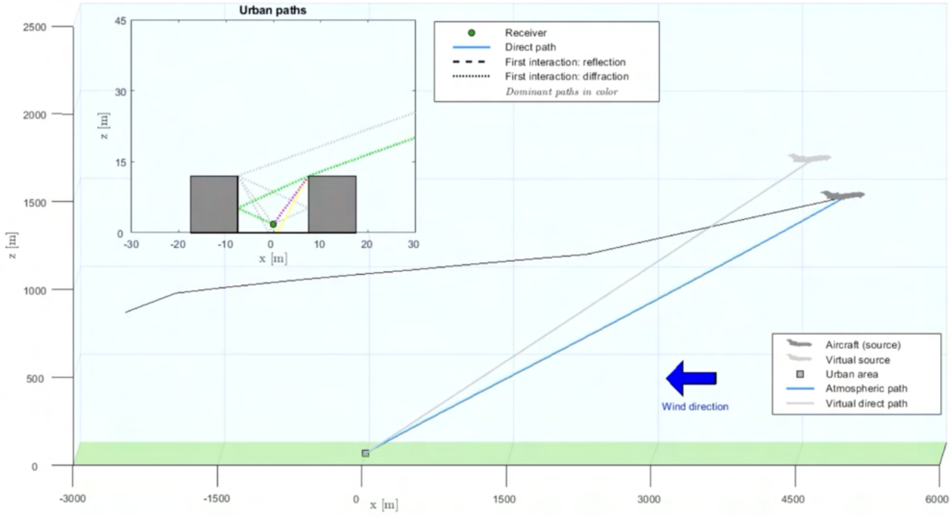


Figure 2.12: Virtual Acoustic RAVEN tracing calculation.

a benchmark for indoor room acoustics and enclosed spaces, its applicability to large-scale, open urban environments can be more constrained. This is primarily due to the increased complexity of sound propagation over long distances and the computational challenges associated with unconstrained outdoor geometries.

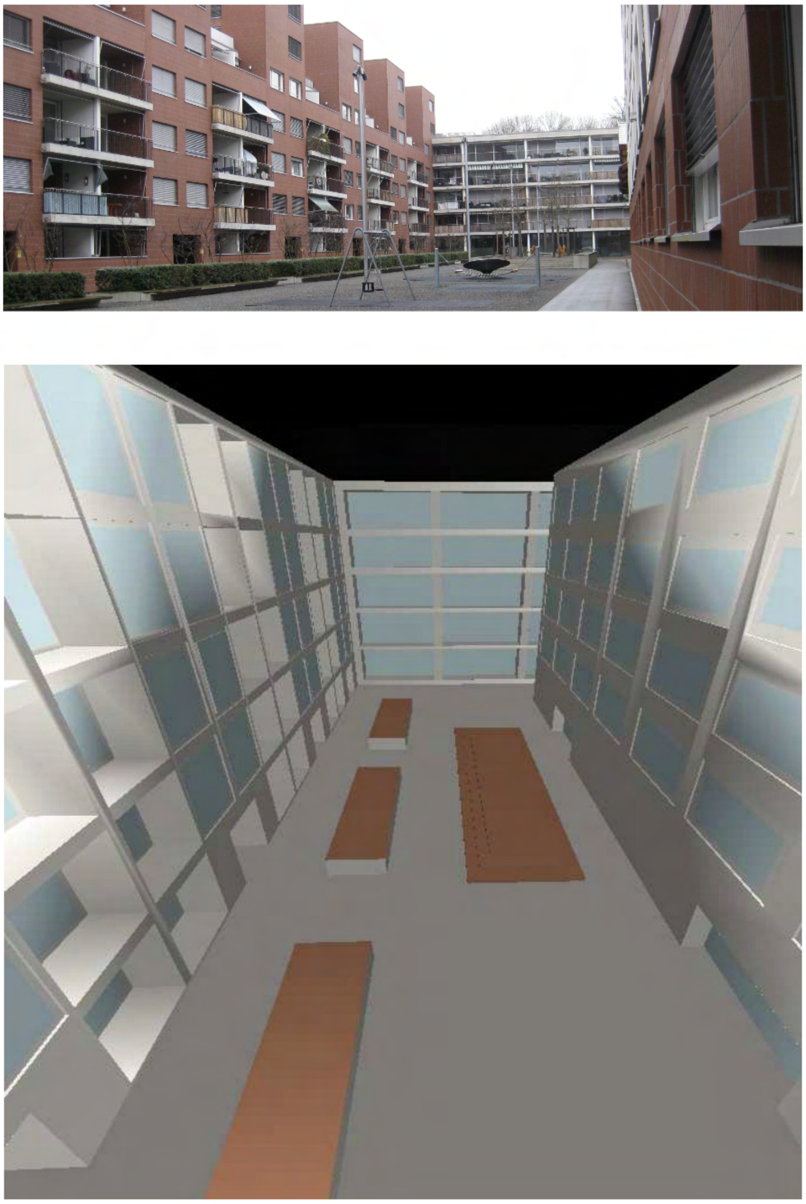


Figure 2.13: Real world frame and ODEON model [27].

Finally, FMOD Studio represents a distinct approach rooted in adaptive game audio design. As a widely utilized middleware supported by an active development community, FMOD Studio can be seamlessly integrated with both Unity and Unreal Engine, providing sophisticated and flexible audio management capabilities. Through custom scripting and API extensions, specific implementations have introduced physically inspired sound behaviors, achieving high levels of realism while maintaining relatively low computational complexity [29]. Overall, the diversity of available methodologies highlights the necessity of selecting soundscape simulation workflows based on specific research objectives, environmental contexts, and technical constraints.

Table 2.1: Review Table for virtual soundscape simulation tools.

Visual Design	**Sound Simulations**
Unreal Engine [24] [6] [17]	Unity Audio Framework [21] [22] [30] [31] [32]
Unity [21] [22] [30] [31] [32]	TASCAR [15]
WorldViz Wizard [15]	Virtual Audio [24] [6]
	ODEON [27]
	FMOD Studio [29]

2.3 Ecological Validity in Current Research

Over the years, the fidelity of soundscape assessments in virtual environments has increased significantly, driven by the rapid development of realistic simulation technologies that bridge the gap between the physical and virtual worlds.

However, the **ecological validity** of these systems must contend with various methodological challenges. A primary issue is that the ISO 12913 standard does not yet incorporate a standardized framework for VR-based assessments. This research field is vast and involves numerous factors that must be meticulously managed during experimental design. A major obstacle remains the medium itself: while Head-Mounted Displays (HMDs) provide highly immersive experiences, they introduce variables that may compromise the authenticity of the soundscape assessment.

As highlighted by *Xu et al.* [33], the cognitive load and memory factors in virtual experiences can make it difficult for users to process the high density of stimuli compared to the information provided by traditional 2D media.

Another critical aspect is the participants' **psychological background**. In their research, *Tokgöz et al.* [34] observed that individuals may become so emotionally involved with certain environments that their subjective assessments are biased. This bias is particularly relevant in immersive reproductions, such as 360-degree videos, where participants may perceive a strong sense of "presence" or even see themselves within the scene.

Furthermore, the demographic composition of participant samples in VR studies presents a significant limitation. There is a scarcity of research involving older populations—one notable exception being the study by *Long et al.* [19]. The majority of studies rely on young individuals, likely due to their greater familiarity with the medium.

Finally, the physiological impact of VR must be considered. Research indicates that users may suffer from **cybersickness**, motion sickness, or visual fatigue. In a study focused on the side effects of VR, *Lavoie et al.* [35] explain that beyond physical discomfort, prolonged VR exposure can lead to emotional distress, alienation, or general unease.

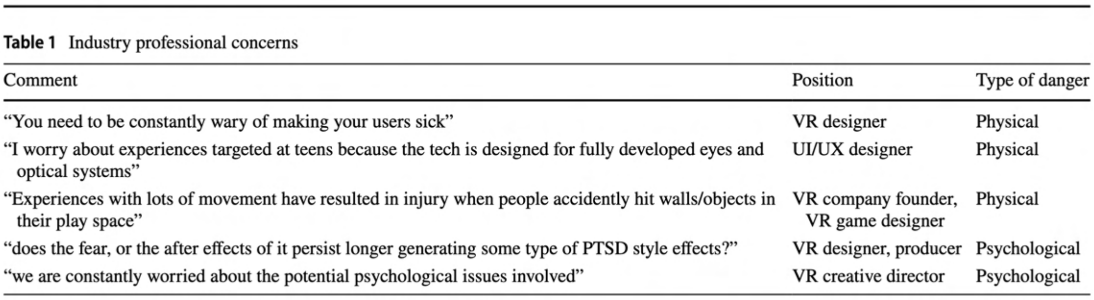


Table 1 Industry professional concerns

Comment	Position	Type of danger
"You need to be constantly wary of making your users sick"	VR designer	Physical
"I worry about experiences targeted at teens because the tech is designed for fully developed eyes and optical systems"	UI/UX designer	Physical
"Experiences with lots of movement have resulted in injury when people accidently hit walls/objects in their play space"	VR company founder, VR game designer	Physical
"does the fear, or the after effects of it persist longer generating some type of PTSD style effects?"	VR designer, producer	Psychological
"we are constantly worried about the potential psychological issues involved"	VR creative director	Psychological

Figure 2.14: Industry professionals concerns about VR games side effects from *Lavoie et al.* research [35].

As illustrated in **Figure 2.14**, safety concerns and environmental awareness represent additional challenges. When a testing area is not adequately prepared for immersive sessions, users may collide with physical objects, leading to what *Tseng et al.* define as "*physical breakdowns*" [36]. This risk obliges experimental designers to meticulously plan every movement required of the participants to prevent potential injuries.

Beyond safety, the ecological validity of both audio and video remains an open question. While the Ambisonic format is currently the industry standard for spatial audio, emerging digital audio frameworks—such as procedural and object-based audio—are offering more flexible and editable alternatives to traditional field recordings [37].

Simultaneously, Computer-Generated (CG) content is redefining soundscape assessments by introducing User Interface (UI) elements and interactive tools that expand the utility of virtual environment evaluations [22].

Looking forward, the upcoming challenges in the field focus on integrating highly interactive components and refined procedures to improve the participant's experience. This evolution aims to facilitate the assessment process, transforming the user into an active "role player" within the research framework [22] [38] .

Chapter 3

Application of Soundscape approach

The experiment aimed to evaluate soundscape assessment within virtual environments to collect and analyze data regarding potential differences and insights compared to real-world soundscape assessments, such as those conducted through soundwalks.
The study was structured into two parts:

- **Fieldwork phase**: data acquisition and analysis derived from *in situ* assessments conducted via soundwalks, specifically focusing on subjective evaluations and psychoacoustic measurements as specified in ISO 12913;
- **Laboratory phase**: development of eight virtual scenarios based on the real-world locations previously evaluated, followed by subjective assessments involving both the original soundwalk participants and a new sample group.

The entire process required diverse approaches and strategies, particularly due to the challenge of replicating real-world scenarios within virtual environments. This involved comparing various solutions regarding audio outputs from both digital and analogue devices, as detailed in the following sections.

3.1 Soundwalk Case Study: Turin

The soundwalk was conducted on October 22, 2025, from early afternoon until the final hours before dusk. The procedure adhered to the SSID protocol [5] and involved a group of students from the "Adaptive Public Spaces" (APS) course at Politecnico di Torino. The students were tasked with assessing and analyzing the soundscape to develop design concepts centered on both acoustic environment

evaluation and urban design for the investigated sites.
The process followed a standardized workflow to ensure the acquisition of comparable data from each location.
The structure of the soundwalk is detailed in **Table 3.1**.

Table 3.1: Soundwalk design for each spot.

Asset	**Listening**	**Assessment**
Questionnaire		2 min
Ambisonic Microphone	2 min	
Binaural Headphones	2 min	
SLM	2 min	
360º Camera	2 min	
Photos	a.t.	a.t.
Environment metrics	a.t	a.t

3.1.1 Path Analysis

The selected path was aligned with the APS course projects, which focused on designing urban solutions for specific city sites designated as future stations for the Metro 2 line in Turin. These locations are situated in the Barriera di Milano district, an area characterized by a strong community and social presence, yet also marked by numerous abandoned buildings requiring urban redevelopment. Within their projects, the students were tasked, among other objectives, with integrating the soundscape as a key thematic element.

Figure 3.1: Soundwalk path.

The three selected sites required an extensive walking route, starting from Rebaudengo Station and proceeding toward Via Bologna, passing by the Giovanni Bosco Hospital. In soundwalk methodology, transition moments are crucial for participants who, through active listening, can achieve a deeper comprehension of the entire acoustic scenario.

Rebaudengo Station (RB01) served as the initial assessment point. The location is a railway station that appears semi-abandoned, characterized by a decadent architectural state. The adjacent road is heavily trafficked by high-speed vehicles due to the nearby entrance of the Corso Venezia highway. Another significant feature is the presence of cycling paths, where the passage of bicycles and scooters is audible. In the background, a large green area serves as a frequented space for walking and recreational activities.
Behind the trees lies the second site.
Spazio 211 (RB02) is a small cultural hub that stands in stark contrast to the large abandoned building in its vicinity. The area experiences intermittent traffic flow. Numerous trees and a pedestrian path connect the site to the neighboring park. The presence of a bus stop, a school, and a library further strengthens the cultural identity of the location.
The final spot is **Via Bologna Park** (BL01), a small "green island" emerging amidst several high-traffic roads. The site leverages its elevated position in an attempt to mitigate road traffic noise. A small fountain is located nearby, and birdsong can be heard through the canopy.

A summarized analysis of each soundwalk location is reported in **Tab 3.2**.

Table 3.2: Soundwalk spots summary.

ID	Location	Acoustic & Environmental Keypoints
RB01	Rebaudengo Station	Heavy traffic flow; semi-abandoned station echoes; accessible cycling paths.
RB02	Spazio 211	Urban green space; rhythmic bus stop sounds; proximity to school and library.
BL01	Via Bologna Park	Isolated "green island"; trickling water from a small fountain; birdsong.

Figure 3.2: Rebaudengo Station view.

Figure 3.3: Spazio 211 picture during soudwalk.

3.1.2 Audio Acquisition Setup

The audio setup selected for the soundwalk conforms to the ISO 12913 standard. The **ambisonic microphone** utilized was a Zylia ZM-1, a third-order ambisonic microphone (3rd order HOA) featuring 19 capsules. This configuration requires remapping to 16 channels to meet the third-order ambisonic standard. This setup ensures high-fidelity reproduction of acoustic environments in terms of spatial clarity and directivity, which are essential factors for subsequent laboratory-based soundscape comparisons.

Audio files were recorded using the **recording software** Bidule by Plogue, with a sampling rate of 48 kHz and a bit depth of 32-bit. This tool, directly interfaced with a PC via cable, enables the simultaneous recording of all 19 raw channels from the ambisonic microphone.

The **binaural microphones**, positioned at the entrance of the ear canal, captured recordings for psychoacoustic parameter analysis. This instrumentation is specifically necessary to emulate human auditory perception through left and right channel recording. Tracks were exported as stereo .wav files.

Figure 3.4: Spazio 211 view.

Lastly, a **sound level meter** (phonometer) recorded sound pressure level (SPL) data, which is critical for subsequent laboratory calibration and further acoustic analysis. The device was calibrated using a 1 kHz sine tone at 94 dB SPL (or 114 dB). Data were collected as arrays and stored in text files. A complete summary of the audio acquisition setup is presented in **Table 3.3**.

Table 3.3: Audio acquisition setup.

Categories	**Model**
Ambisonic Microphone	Zylia M-1
Binaural Headphones	ScadasXS Binaural Headset
Sound Level Meter	NTI Audio XL2
Recording Software	Plogue Bidule

Figure 3.5: Bologna park picture during soundwalk.

3.1.3 Video Acquisition Setup

Videos were captured using an **Insta360 camera**, capable of recording at 5.7K resolution. The footage was shot in a log color space (LOG) to allow for greater flexibility during the color correction process within the virtual experiment design.

Photographs were taken with an iPhone 16 camera to document specific details essential for the virtual reconstruction of the soundscape environments. This was necessary to accurately reproduce the primary features and characteristics of each site scenario.

Categories	**Model**
360º Camera	Insta360 One X2
Camera	iPhone 16

Table 3.4: Video acquisition setup.

Figure 3.6: Audio video capture setup during soundwalk with Zylia M-1 microphone and Insta360 One X2 camera.

Figure 3.7: Simcenter SCADAS XS binuaural headphones.

3.1.4 Questionnaire

The questionnaire was developed in accordance with the ISO 12913 standard and was divided into seven distinct sections:

- **Sound source dominance**: this section evaluates the perceived presence, relative intensity, and dominance of various sound sources within the scene. Sources are categorized into traffic noise, human sounds, natural sounds, and other noises, such as construction or industrial sounds;
- **Perceived Affective Quality (PAQ)**: this section requires an evaluation based on the eight attributes (descriptors) defined by the ISO standard for soundscape assessment;
- **Overall soundscape assessment**: this section focuses on the general quality and pleasantness of the soundscape at each location;
- **Architectural environment**: this section involves an architectural evaluation, specifically addressing the strict correlation between acoustic perception and visual context;

The remaining three sections covered the informed consent required to begin the assessment, the identification of the specific site being evaluated, and demographic information.
The complete questionnaire is provided in the **Appendix A**.

3.2 Data Analysis (in field)

3.2.1 Questionnaires Evaluations (in field)

The questionnaire analysis yielded significant information regarding the participants' subjective evaluations of the soundscape. These data, complemented by the Perceived Affective Quality (PAQ) scores, constitute a fundamental basis for investigating the correlations between acoustic environments and human perception.

As represented in **Figure 3.9**, the mean age of the distribution is 24.44 years, with a standard deviation (SD) of ±2.15. These results reflect a relatively heterogeneous sample in terms of demographic data, with participants ranging from 20 to over 30 years old. The mode of the distribution is 23 years.

The participant density during the soundwalk experienced a decrease across the different locations; the specific sample size for each spot is detailed in **Table 3.5**.

Regarding the gender distribution among participants, a higher prevalence of female individuals (60%) was observed compared to male participants (40%), as illustrated in **Figure 3.10**.

(a) RB01 location shot with Insta360 One X2.

(b) RB02 location shot with Insta360 One X2.

(c) BL location shot with Insta360 One X2.

Figure 3.8: Location shots taken with Insta360 One X2 at different points.

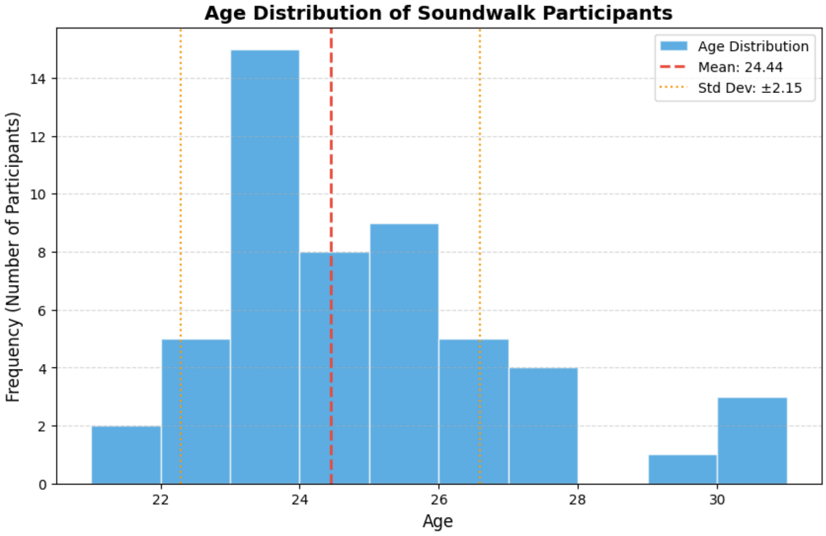


Figure 3.9: Age distribution from soundwalk.

Table 3.5: Sample size for each spot.

Spot	**Sample size**
RB01	27
RB02	16
BL	14

The investigation into sound source dominance (**Figure 3.11**) reveals that traffic noise is highly dominant at the Rebaudengo Station (RB01) and, to a significant extent, at Spazio 211 (RB02). Other sources exhibit comparable levels, with the exception of the Via Bologna Park (BL01), where traffic noise and natural sounds show similar prevalence. This shift in dominance may be attributed to the presence of birdsong and the sound of the small fountain within the soundscape. The architectural and urban environment ratings (**Figure 3.12**) were significantly higher for Spazio 211 (RB02) and Via Bologna Park (BL01). This trend may be attributed to the poorly preserved state of Rebaudengo Station (RB01), which received a very low score due to its degraded aesthetic and structural condition. Regarding the appropriateness rating of the surrounding sound environment (**Figure 3.13**), RB01 was perceived as the most appropriate, although results were generally consistent across all sites. This higher rating for Rebaudengo Station (RB01) may be attributed to its environmental consistency; the site lacks the "dual identity"

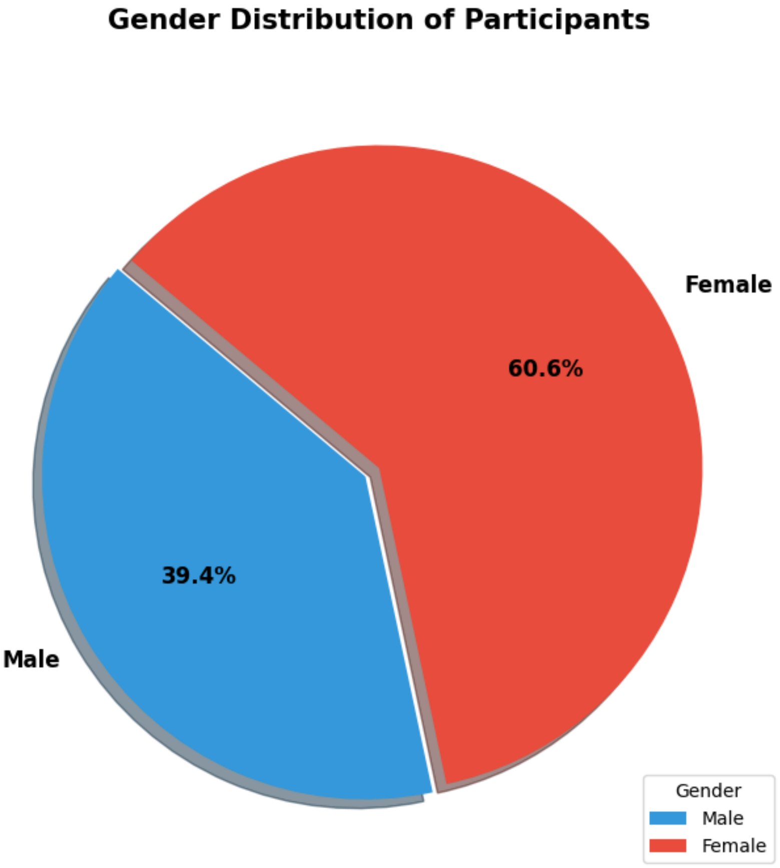


Figure 3.10: Gender distribution.

found in RB02 and BL01, where the conflicting presence of both traffic-heavy roads and green areas may reduce the perceived congruency of the soundscape. Regarding the perception of the sound environment (**Figure 3.14**), the Via Bologna Park (BL01) was rated as the most pleasant location, followed by Spazio 211 (RB02) and Rebaudengo Station (RB01). The latter, characterized by its dominant traffic noise patterns, received the lowest pleasantness scores. These findings suggest a potential correlation between sound source dominance and the perceived pleasantness of the soundscape.

The results of the PAQ analysis, derived from the questionnaire data, are illustrated through scatter (**Figure 3.15**), density (**Figure 3.16**), and joint (**Figure 3.17**) plots. These visualizations were generated using Soundscapy [39], a specialized Python library designed for the collection and analysis of soundscape data. The plots represent the relationship between the Pleasant and Eventful dimensions.

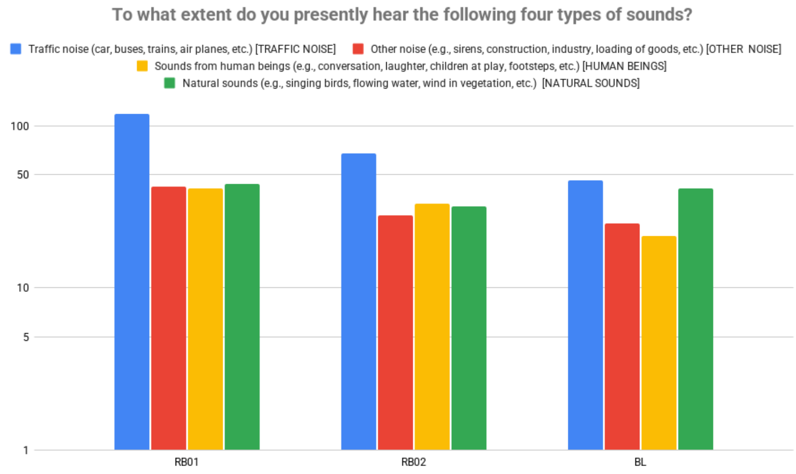


Figure 3.11: Sound source domination distribution.

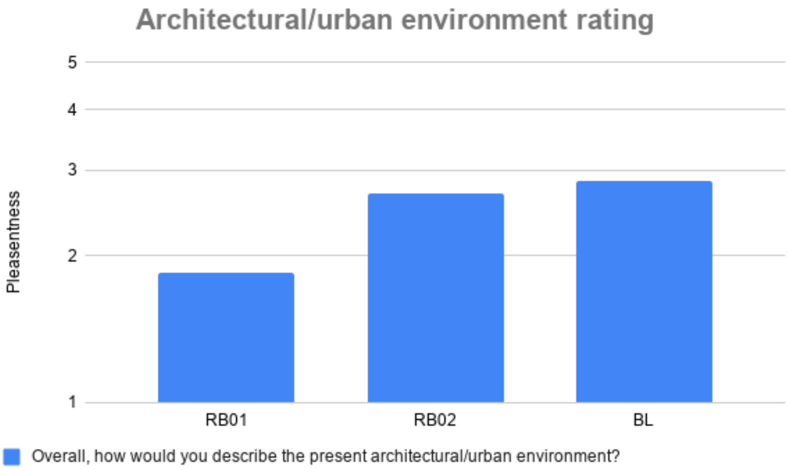


Figure 3.12: Architectural and urban design rating.

The data clearly indicate that the BL01 site is the most pleasant location, predominantly characterized as *calm* and *monotonous*.
Furthermore, the results demonstrate that RB01 exhibits the lowest Pleasant/Eventful ratio, as the most frequently reported descriptor is *monotonous*.

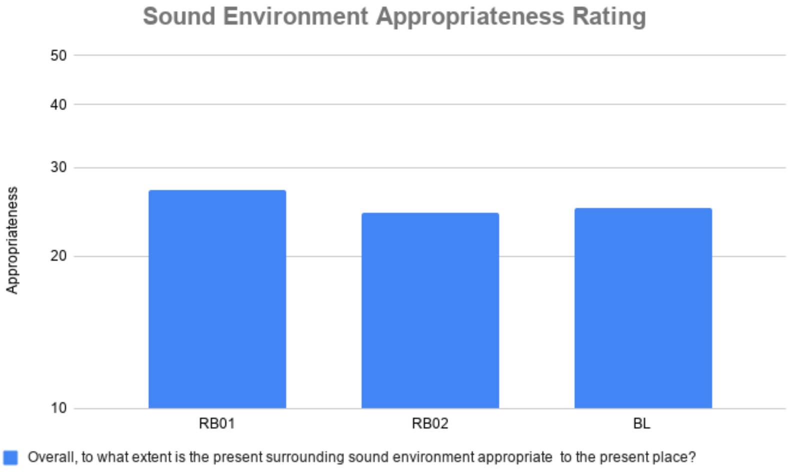


Figure 3.13: Sound environment appropriateness rating.

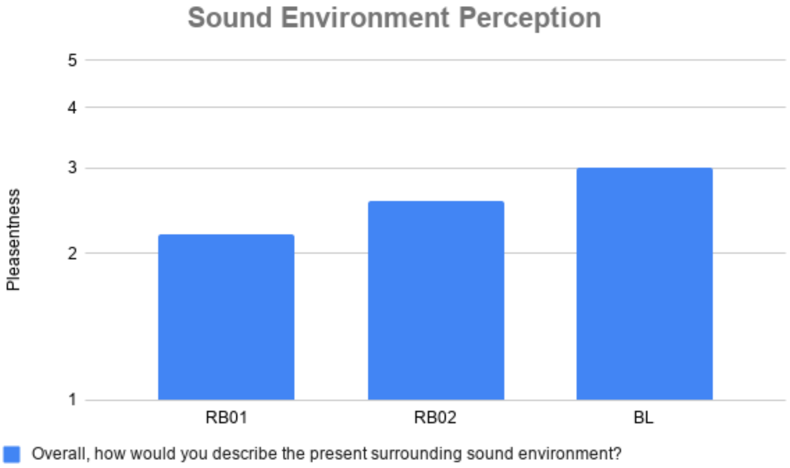


Figure 3.14: Sound environment perception.

3.2.2 Psychoacoustic parameters measurements (in field)

Following the field experiments, a laboratory analysis of psychoacoustic parameters was conducted using the binaural recordings acquired in situ.

The recordings were processed using Simcenter Testlab, calibrated with the sound pressure level (SPL) measurements obtained via the sound level meter. For each recording, the software extracted 200 samples from a 100-second segment to

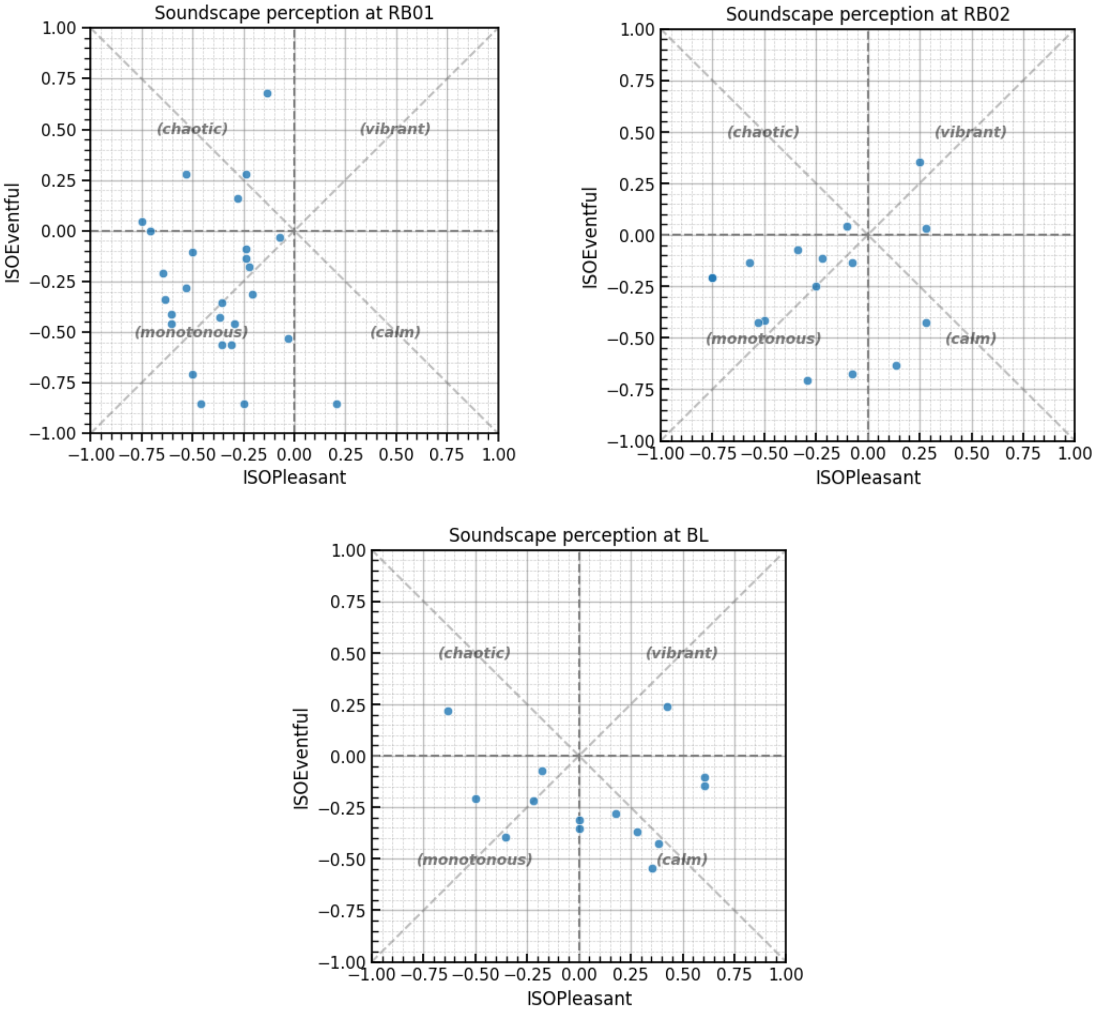


Figure 3.15: Scattering plots from soundwalk spots.

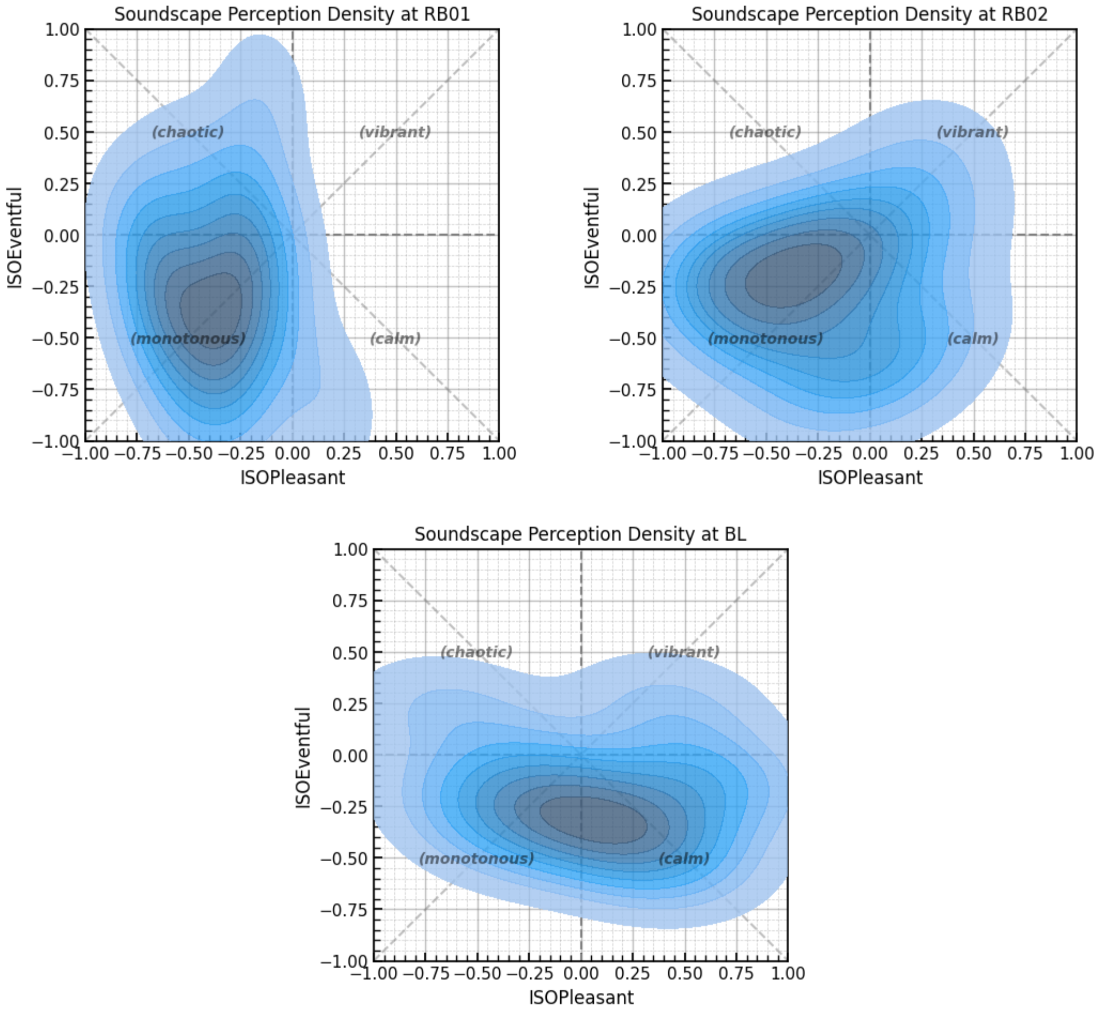


Figure 3.16: Density plots from soundwalk spots.

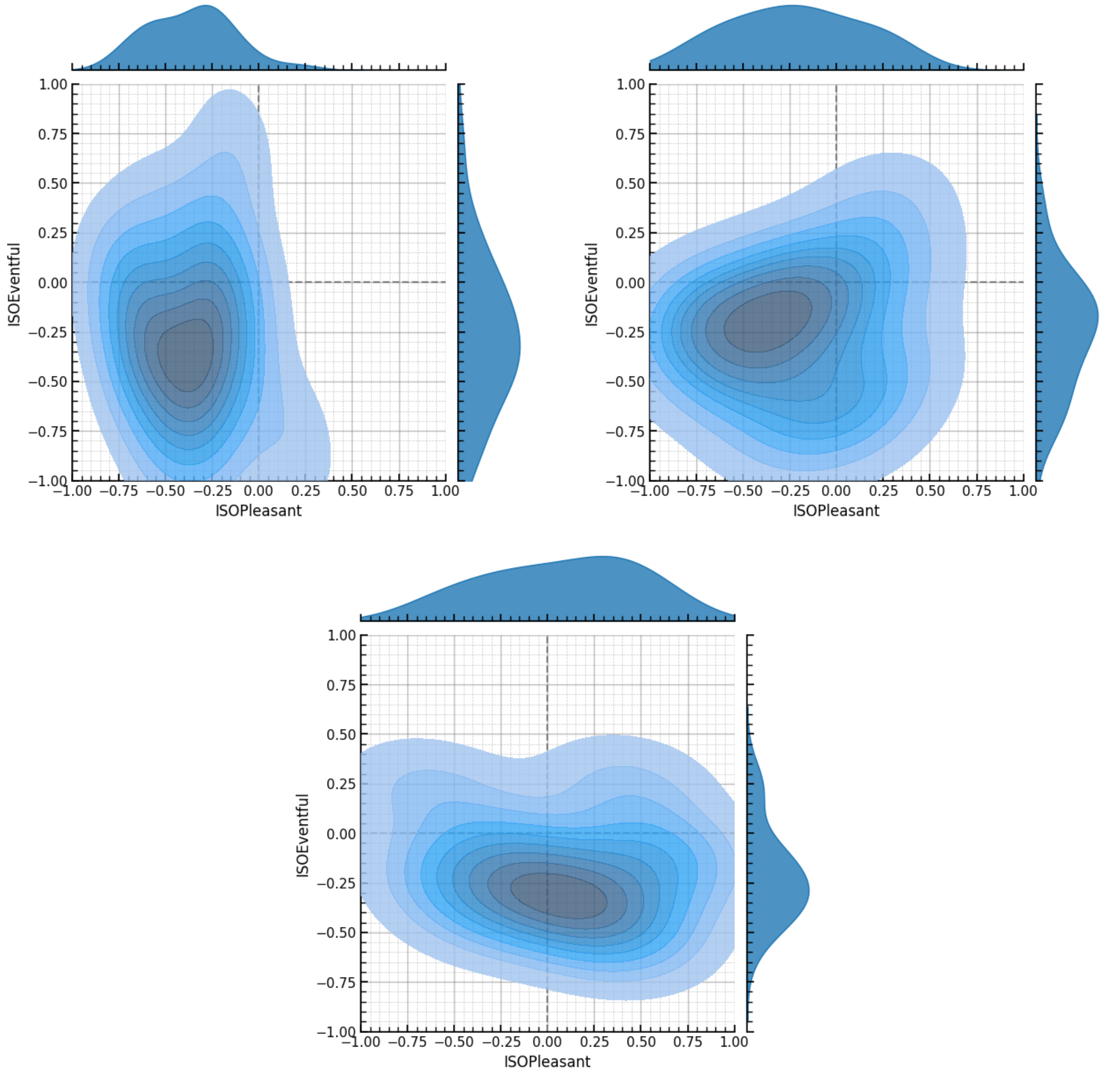


Figure 3.17: Join plots from soundwalk spots.

calculate the psychoacoustic parameters of loudness, sharpness, roughness, and fluctuation strength, following the methodology described by *Zhang et al.* [40].

Subsequently, the extracted data were processed in Python to compute the mean values, standard deviations, and statistical descriptors N_5 and N_{95} for each parameter.

Starting with the **SPL values** shown in **Figure 3.18**, the highest L_{Aeq} levels were recorded at RB02. The L_5 value reflects the impact of intense traffic noise, specifically the passage of motorcycles and heavy vehicles. Conversely, the quietest location was BL01; this evidence suggests a potential correlation between these objective acoustic data and the subjective PAQ values derived from the questionnaires.

Table 3.6: Complete SPL data for each location.

Location	$L_{A_{eq}}$	L_5	L_{95}	$L_{C_{eq}}$
RB01	59,4	65.8	50.3	71.4
RB02	65.0	71.0	58.2	71.6
BL	57.1	60.8	49.3	70.3

The **Loudness** analysis (**Table 3.7**) shows a linear correlation with the SPL data. The loudest location is RB02, confirming the previous acoustic findings. As illustrated in **Figure 3.19**, this result is strongly influenced by singular acoustic events, such as the previously mentioned passage of motorcycles or trucks. Similarly, RB01 exhibits high loudness levels due to continuous traffic noise. Conversely, the BL01 site is significantly less loud than the other locations, likely due to its greater distance from the main road axes. **Roughness** levels are most significant at the RB02 location, with an average value of $R_{avg} = 0.245$ (calculated as the mean between ch.1 and ch.2), as shown in **Table 3.8**. This result is primarily driven by a specific acoustic event during the recording that led to an R_5 level of 0.425 (mean of ch.1 and ch.2), a value considerably higher than the R_5 levels recorded at RB01 and BL01. Furthermore, the table reveals a homogeneous trend across all sites regarding the R_{95} values, indicating a consistent baseline roughness for each environment. **Sharpness** presents a different trend, with the highest levels observed at BL01, specifically reaching $S_5 = 1.43$ Acum on channel 2, as highlighted in **Table 3.9**. This peak is likely attributable to a high-frequency noise occurring during the recording—specifically a "whistle-like" sound—as clearly evidenced by the temporal evolution shown in **Figure 3.21**.

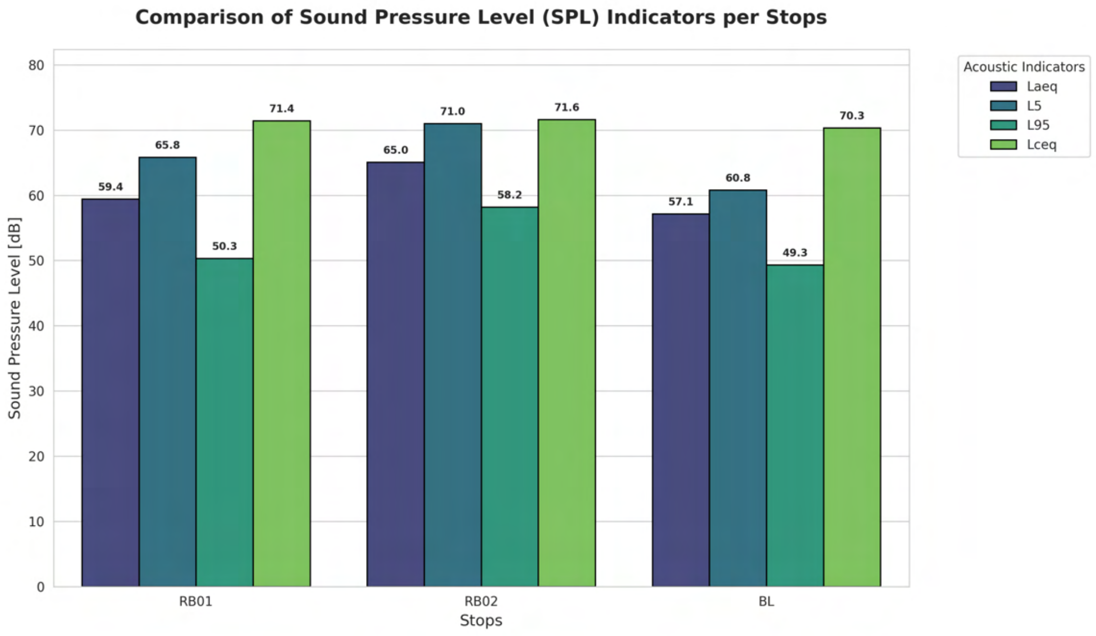


Figure 3.18: Comparison of Sound Pressure Level (SPL) Indicators per Stops.

Table 3.7: Complete loudness data for each location.

Location	N_{avg}	N_5	N_{95}
RB01 ch.1	9.9	17.1	6.1
RB01 ch.2	10.4	18.0	6.0
RB02 ch.1	13.9	21.1	9.0
RB02 ch.2	13.8	20.8	8.8
BL ch.1	9.0	13.0	5.8
BL ch.2	9.2	13.0	6.3

Fluctuation Strength exhibits higher values at the RB02 location, as shown in **Table 3.10**. This trend is likely attributable to the prominent sound of footsteps within the recorded sample. The effect is further accentuated by the presence of dry leaves on the grass, which increased the acoustic modulation during the walk. A notable level was also observed at the RB01 site; this may be due to the background music from a loudspeaker, which was audible during the playback of the recording. Regarding **tonality**, no significant data were obtained during the

Table 3.8: Complete roughness data for each location.

Location	R_{10}	$R50$
RB01 ch.1	0.26	0.18
RB01 ch.2	0.25	0.18
RB02 ch.1	0.34	0.23
RB02 ch.2	0.31	0.22
BL ch.1	0,24	0.17
BL ch.2	0.28	0.19

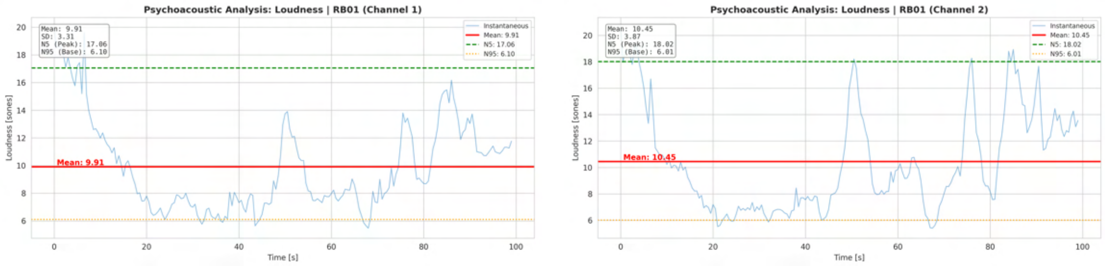
Table 3.9: Complete sharpness data for each location.

Location	S_{avg}	S_5	S_{95}
RB01 ch.1	1.19	1.39	0.97
RB01 ch.2	1.28	1.62	1.01
RB02 ch.1	1.39	1.70	1.19
RB02 ch.2	1.40	1.70	1.23
BL ch.1	1.32	1.51	1.21
BL ch.2	1.43	1.69	1.24

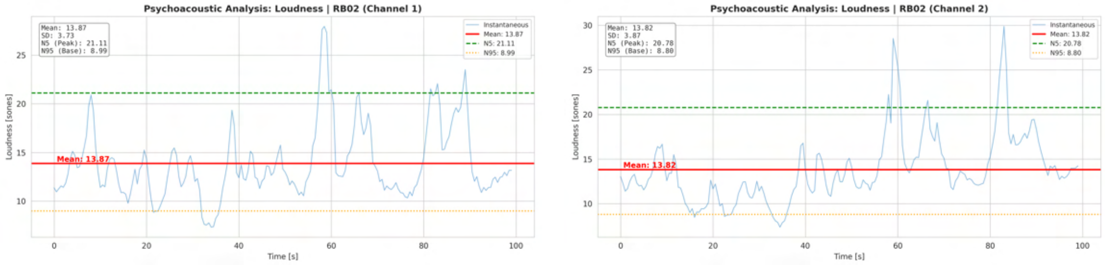
analysis, indicating a lack of prominent tonal components within the recordings. This absence suggests that the acoustic environments are primarily characterized by broadband noise sources rather than narrow-band or discrete frequency emissions.

Table 3.10: Complete fluctuation strength data for each location.

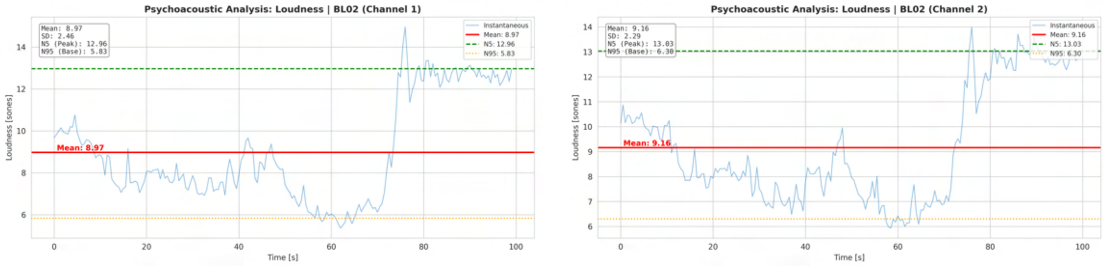
Location	FS_{10}	FS_{50}
RB01 ch.1	1.38	1.21
RB01 ch.2	1.35	1.21
RB02 ch.1	1.43	1.24
RB02 ch.2	1.35	1.19
BL ch.1	1.30	1.17
BL ch.2	1.35	1.20



(a) Loudness from RB01 channel 1. **(b)** Loudness from RB01 channel 2.

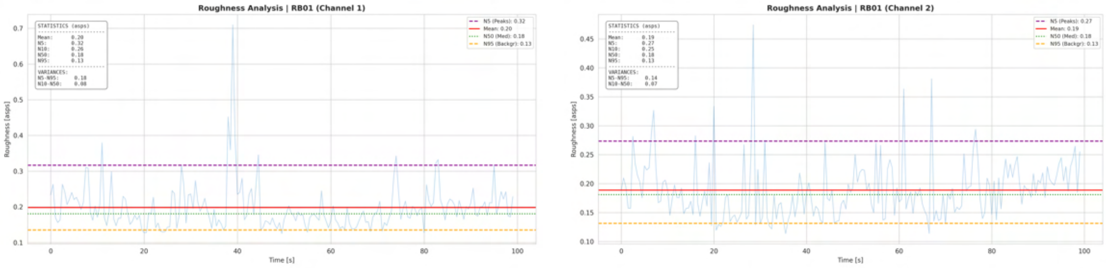


(c) Loudness from RB02 channel 1. **(d)** Loudness from RB02 channel 2.



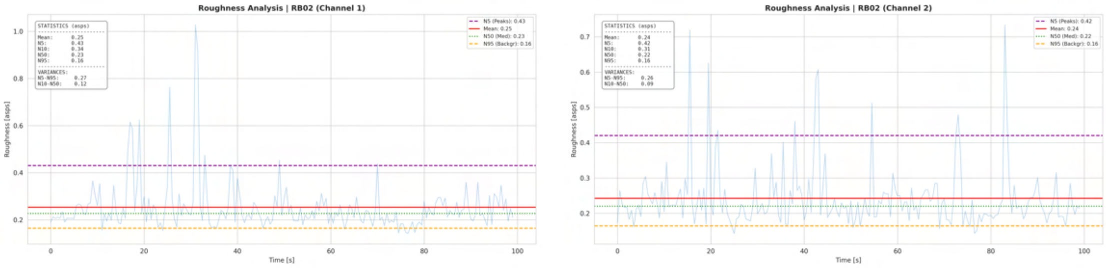
(e) Loudness from BL channel 1. **(f)** Loudness from BL channel 2.

Figure 3.19: Loudness values from every spot.



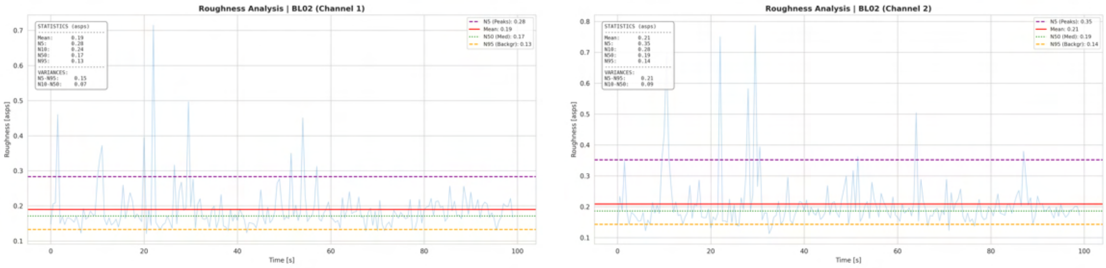
(a) Roughness from RB01 channel 1.

(b) Roughness from RB01 channel 2.



(c) Roughness from RB02 channel 1.

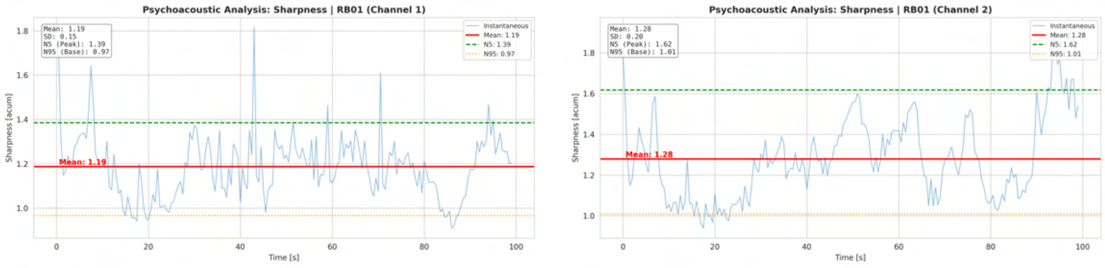
(d) Roughness from RB02 channel 2.



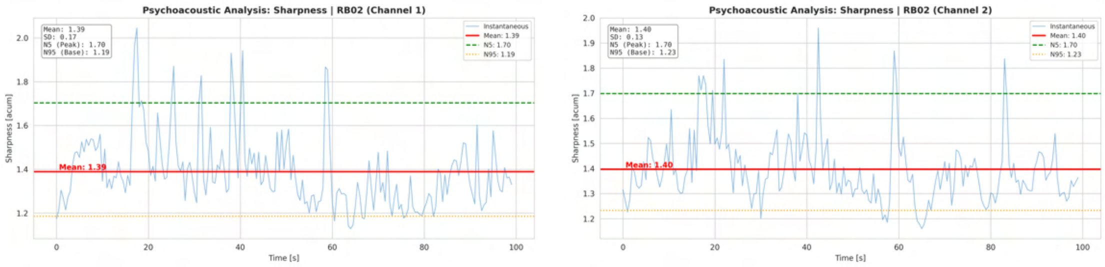
(e) Roughness from BL channel 1.

(f) Roughness from BL channel 2.

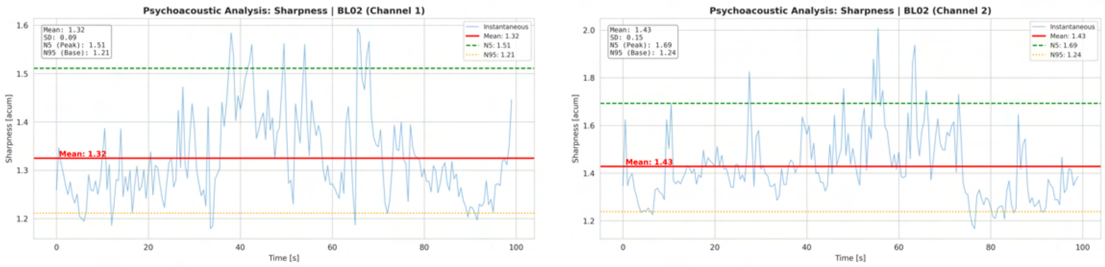
Figure 3.20: Roughness values from every spot.



(a) Sharpness from RB01 channel 1. **(b)** Sharpness from RB01 channel 2.

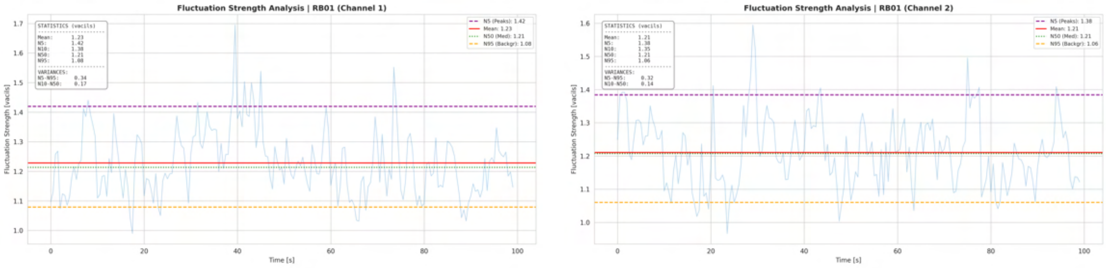


(c) Sharpness from RB02 channel 1. **(d)** Sharpness from RB02 channel 2.

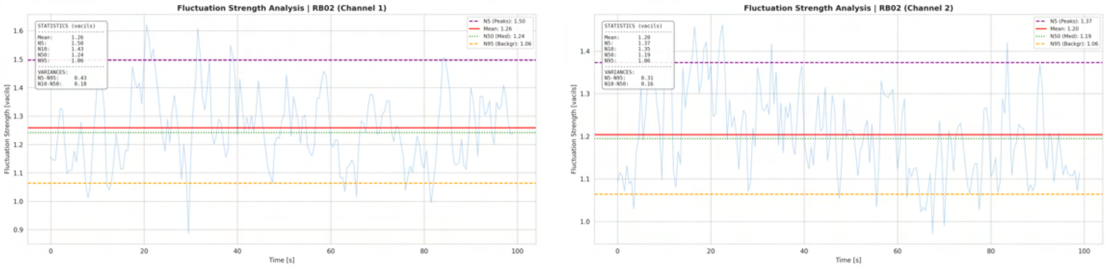


(e) Sharpness from BL channel 1. **(f)** Sharpness from BL channel 2.

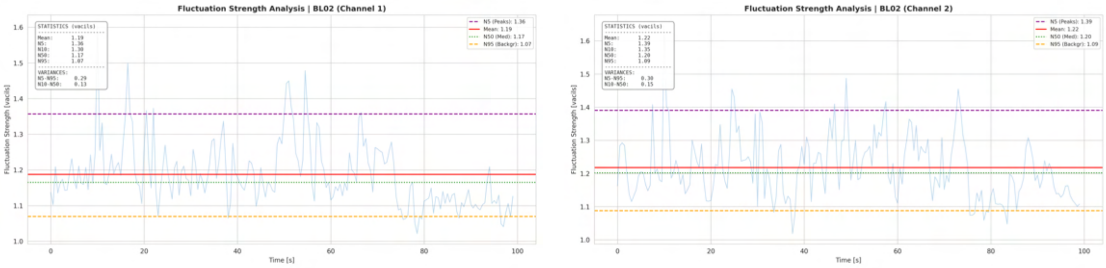
Figure 3.21: Sharpness values from every spot.



(a) Fluctuation Strength from RB01 channel 1. **(b)** Fluctuation Strength from RB01 channel 2.



(c) Fluctuation Strength from RB02 channel 1. **(d)** Fluctuation Strength from RB02 channel 2.



(e) Fluctuation Strength from BL channel 1. **(f)** Fluctuation Strength from BL channel 2.

Figure 3.22: Fluctuation Strength values from every spot.

3.2.3 Data Correlation (in field)

In order to evaluate the correlation between psychoacoustic parameters, *Pearson correlation* (**Figure 3.23**) was employed for continuous variables, while *Spearman correlation* (**Figure 3.24**) was used for non-continuous data. For this analysis, the data from the two binaural channels were averaged and treated as a single representative value.

The results indicate a strong correlation between L_{Aeq} and N_{avg}, as well as between SPL and loudness, which is consistent with standard soundscape analysis. Interestingly, FS_{50} shows a significant correlation with other parameters; this is likely attributable to the humans-made noise generated by the soundwalk participants, such as footsteps, speech, and general movement. Furthermore, as expected, roughness exhibits a high correlation with L_{Aeq} and N_{avg}, confirming the link between signal amplitude modulation and perceived intensity.

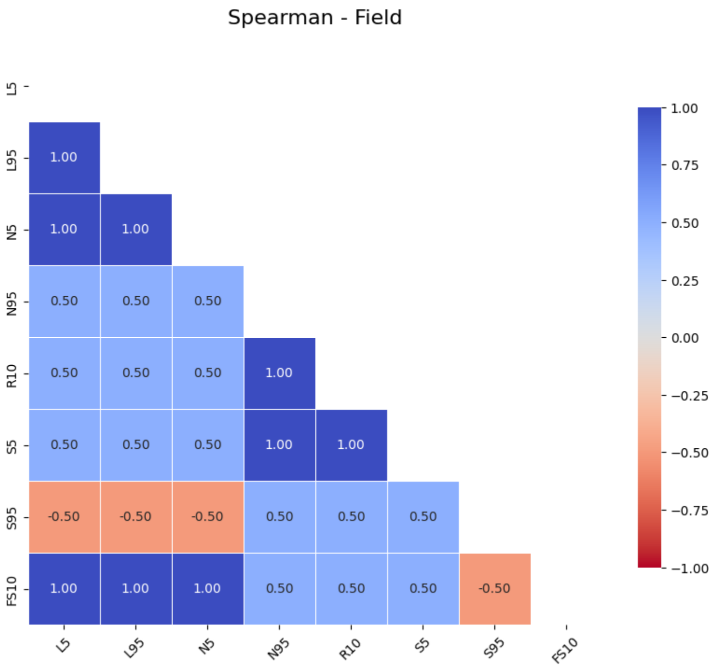


Figure 3.23: Pearson field corelogram.

Subjective data, specifically the PAQ scores, were analyzed using the *Soundscapy* Python library.
As shown in **Figure 3.25**, all three locations exhibit relatively low values for both pleasantness (P_{ISO}) and eventfulness (E_{ISO}). While the BL01 site received higher pleasantness ratings compared to the other spots, it remains characterized as "uneventful" in the soundscape circumplex.
Overall, the data suggest that the three locations share a similar profile in terms of eventfulness, showing no significant variation across the different urban contexts.

Further details regarding the questionnaire evaluations and demographic distributions have been discussed in the preceding sections.

The comprehensive data summarized in **Table 3.11** further validate the trends observed in the previous plots, confirming the internal consistency of the soundscape assessments across all surveyed locations.

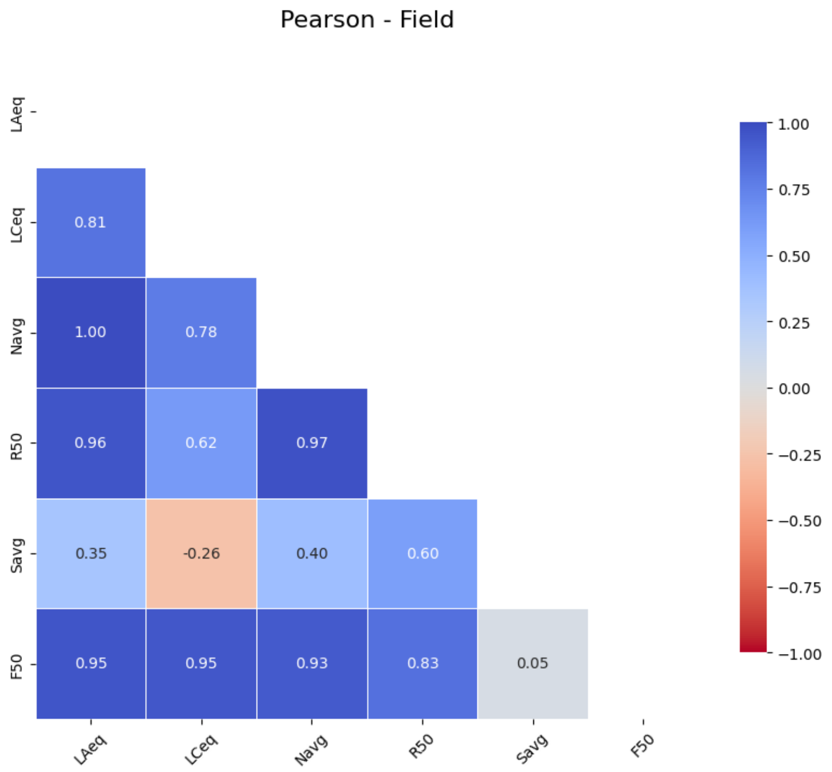


Figure 3.24: Spearman field corelogram.

Table 3.11: Descriptive statistics and ISO 12913-2 attributes for each location.

Location	Count	ISOP.	ISOE.	Pleas.	Even.	Vibr.	Chao.	Mono.	Calm
RB01	27	−0.364	−0.269	0.037	0.222	0.000	0.222	0.741	0.037
RB02	16	−0.220	−0.249	0.250	0.188	0.125	0.062	0.688	0.125
BL	14	0.067	−0.212	0.500	0.143	0.071	0.071	0.429	0.429

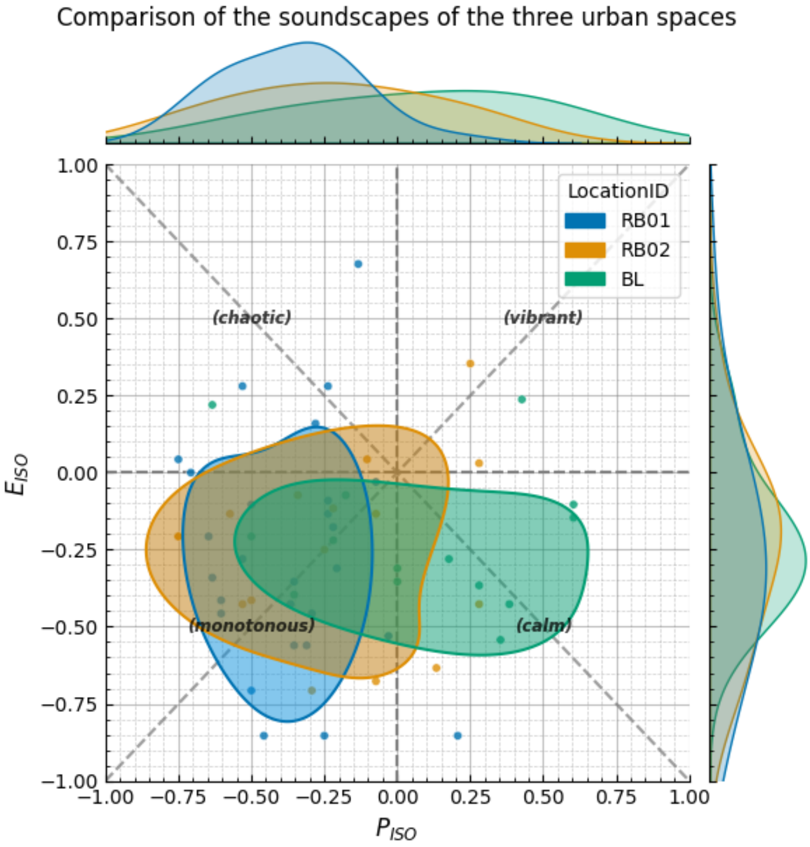


Figure 3.25: Comparison of the soundscapes of the three urban spaces.

3.3 Real and Virtual Environment Models

The laboratory experimental phase involved the development of two distinct types of environmental models:

- **Real Model**: Three scenarios were reconstructed using the audiovisual assets acquired during the soundwalk, providing a high-fidelity reproduction of the actual environments;

- **Virtual Model**: Using the data from the real scenarios as a baseline, one site was completely reconstructed using **CGI** and synthetic soundscapes to allow for controlled environmental manipulation.

3.3.1 Real model building

The 360-degree video footage was recorded in 5.7K resolution using a Log codec. This choice ensured maximum flexibility during field operations and provided a wider dynamic range for post-production. The Log codec produces a "flat" color profile, which facilitates color grading and correction. The video assets were subsequently edited and color-processed using DaVinci Resolve 20.

Audio data were captured using the Zylia ZM-1 microphone, generating 19-channel raw recordings. Using Zylia's proprietary software, these files were converted into the Ambisonics format. After being edited for length and seamlessly looped, two versions of each recording were generated: one in 3rd-order Ambisonics (HOA) and a second in 1st-order Ambisonics.

Finally, two versions of each scenario were built: a visual-only version (without audio) and a version integrated with 1st-order Ambisonics audio.

3.3.2 Virtual Model Construction

The Spazio 211 (RB02) site was selected for the virtual reconstruction due to its versatile characteristics, featuring a mix of anthropogenic noise and natural assets. Constructing virtual environments requires a focus on proposing a coherent model of the site's permanent features, rather than simply replicating a specific, transient moment from the real world. As illustrated in **Figure 3.26**, the process involved several stages to translate physical environmental data into the CGI domain. The following sections provide an in-depth analysis of this workflow.

The construction of the visual model and the associated workflow began in Blender [41]. Initially, a comprehensive base map of the real-world environment was established. For this purpose, the BlenderGIS add-on [42] was utilized to

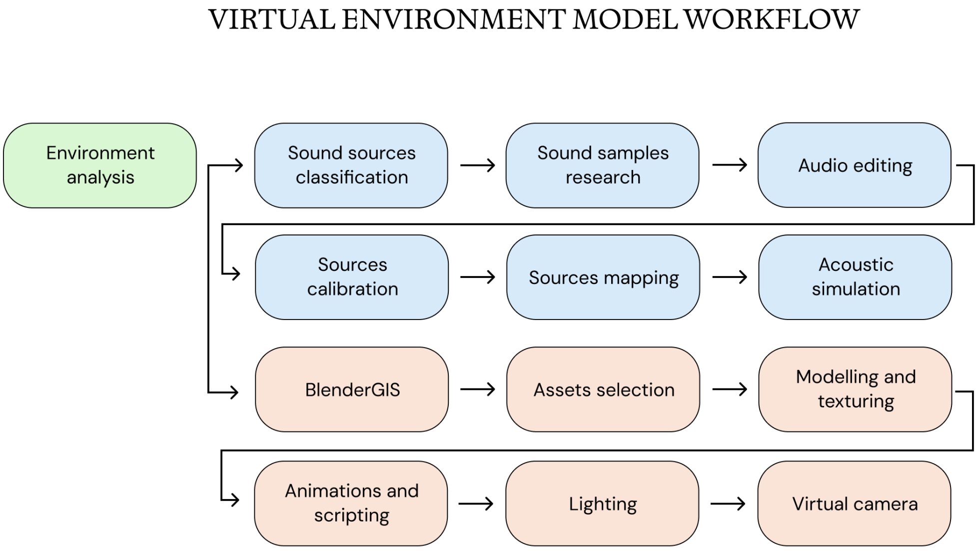


Figure 3.26: Virtual environment model workflow.

generate scaled "block meshes"; this tool leverages user-selected geographic open data to reconstruct a 3D model of the specific territory within Blender. These static meshes required manual modeling adjustments to improve accuracy. Subsequently, high-quality textures sourced from the public 3D asset library Poly Haven [43] were applied. Additional environmental elements, such as trees and fences, were integrated into the scene using the BlenderKit [44] asset library.

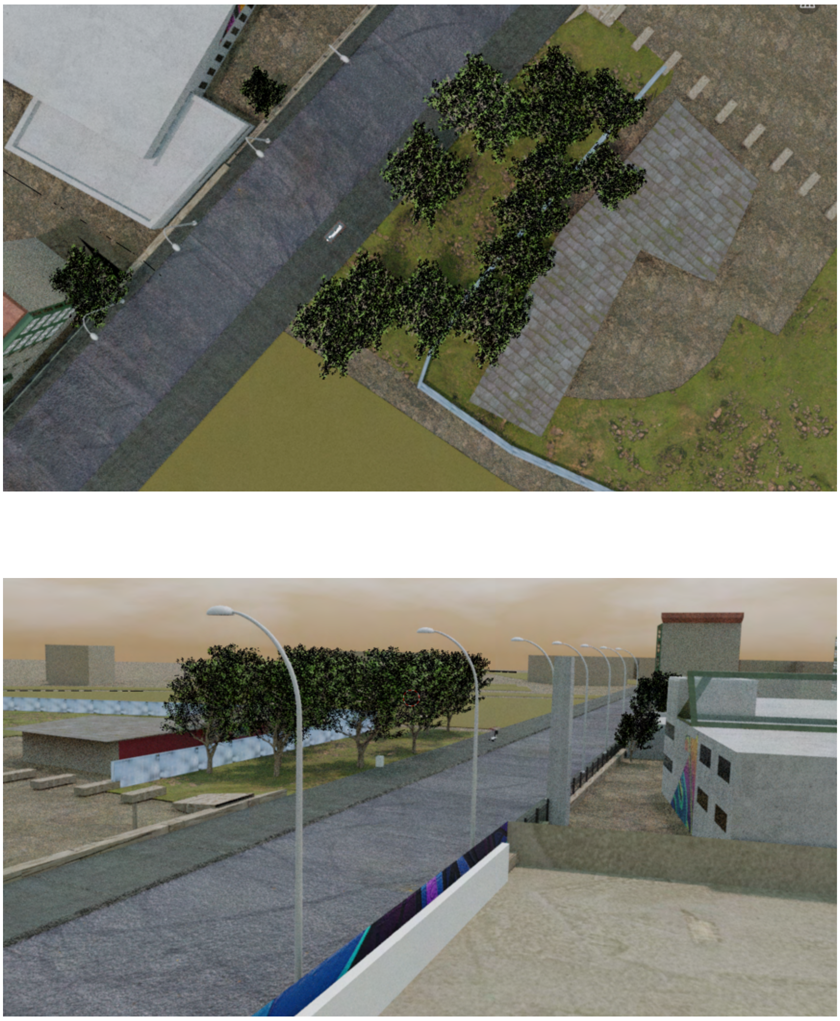


Figure 3.27: Blender's scene pictures during environment building.

Subsequently, the entire scene was exported to the Unity engine [20] for final compositing and animation. At this stage, dynamic assets, including human and vehicular models, were integrated and animated via custom scripts. The pedestrian models utilize a "follow path" script, navigating through the scene by targeting designated empty objects as waypoints. The traffic system is managed by a spawner script that regulates path direction, spawn frequency, and vehicle velocity. To ensure environmental fidelity, the vehicle speed was calibrated to 50 km/h, reflecting local traffic regulations and the site's long straightaway, which allows vehicles to reach the maximum permitted speed.

Environmental lighting was implemented using a skybox material and an HDRI texture sourced from Poly Haven [43]. The lighting configuration was meticulously

calibrated to match the visual conditions captured in the 360-degree reference video for the same location.
Finally, the virtual camera was positioned to simulate the eye-level perspective of a user seated on a bench (**Figure 3.28**). For the reconstruction of the acoustic environment, a deliberate choice was made to minimize the use of real-world recordings. The objective was to develop a methodology that remains accessible and affordable, regardless of a user's available equipment or resources. The entire audio workflow was designed to adhere to this principle of resource optimization.

```
CPU: dsp = 1.5%, studio = 0.4%
MEMORY: cur = 10MB, max = 10MB
CHANNELS: real = 5, total = 12
VOLUME: RMS = -13.96db
```

Figure 3.28: *Unity game window*

Initially, the noise map of the area (**Figure 3.29**) was used as a primary reference to identify the zones with the highest sound pressure levels. The map clearly indicates that the areas adjacent to the main road axes exhibit the highest noise concentration. This spatial analysis guided the placement of artificial sound sources within the virtual engine, specifically regarding the density and positioning of vehicular traffic.
Conversely, the lowest noise levels were identified within the green areas, particularly in the vicinity of the large disused building near the Rebaudengo railway station.

This procedure enabled a structured classification of sound sources based on

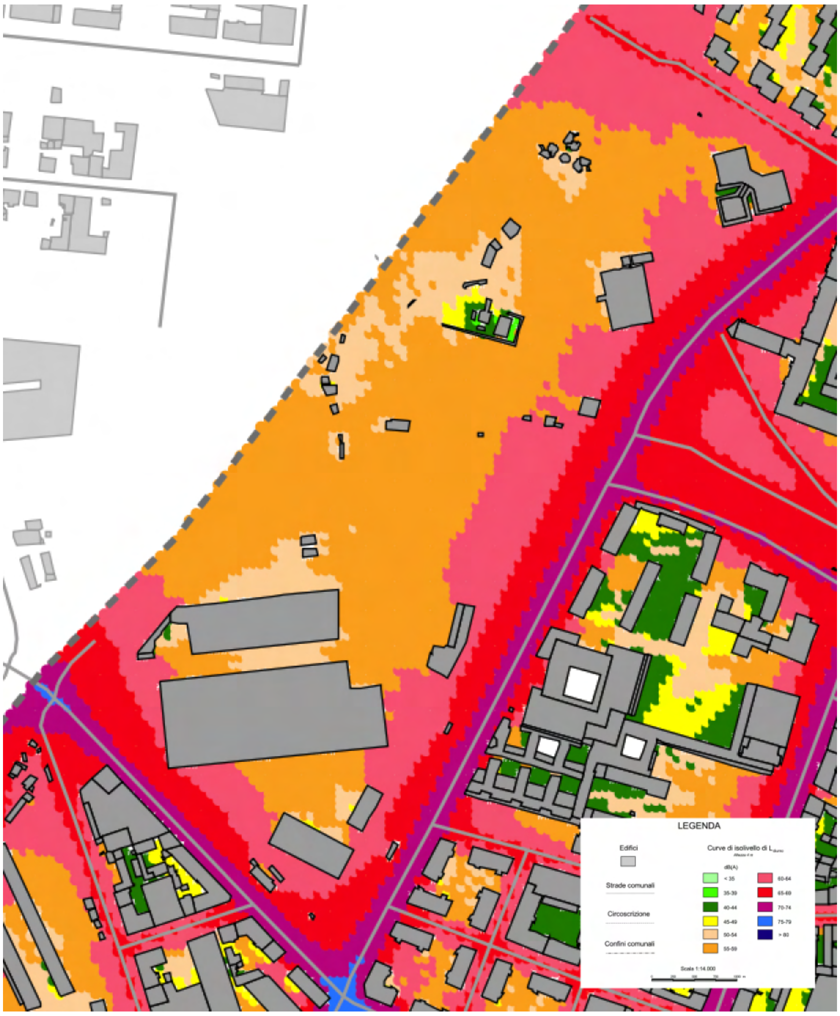


Figure 3.29: Rebaudengo noise map.

environmental keypoints and primary spatial features. Following this, a comprehensive inventory of sound sources was compiled based on assessments conducted during the initial soundwalk and subsequently distributed across the virtual map.

The organizational framework for these sources was inspired by the taxonomy proposed by *Bones et al.* [45], which categorizes soundscape components into three main domains: **people**, **nature**, and **man-made** sounds. This classification facilitated the development of the source hierarchy presented in **Table 3.12**. Nevertheless, to maintain consistency with the reference Ambisonics recordings from the real-world site, only the three most prominent sound sources were prioritized for the final implementation.

The selected sound sources for the virtual environment included: background

Table 3.12: Classification of sound sources.

People	**Nature**	**Manmade**
Voices	Animals	Meccanical
Music	Weather	Transport
		Alarm
		Music

noise, footsteps, vehicular traffic, and a fountain.
The background noise, footsteps, and fountain sounds were sourced from the *Splice* digital library [46], which provides an extensive collection of high-fidelity field recordings. In contrast, the vehicle sounds were generated through auralization techniques based on the research by *Pieren et al.* [47]. This synthesized sound is composed of two primary components: tire-road noise and engine noise, represented by the sound sources $S_1(t)$ and $S_2(t)$, respectively. These sources are spatially localized at two distinct points on the vehicle model, as defined in equations (3.2) and (3.1).
Using an approach based on additive and subtractive synthesis, combined with harmonic modulations and digital filters, a custom script was developed in the open-source environment *SuperCollider* [48]. This script allowed for the generation and recording of two distinct vehicle profiles: a "basic" engine and a "sport" engine.

$$s_2(t) = \sqrt{15} \cdot s_{\text{tire},2}(t) + s_{\text{prop}}(t) \tag{3.1}$$

$$s_1(t) = \sqrt{25} \cdot s_{\text{tire},1}(t) \tag{3.2}$$

Once the sound assets were finalized, the audio library was imported into the game audio middleware *FMOD Studio* [49] (**Figure 3.30**). This choice was made to leverage a professional tool that optimizes the workflow and provides advanced management of the audio engine. Within FMOD, all samples were processed using a low-pass and a high-pass filter, with cutoff frequencies set at 15 kHz and 20 Hz, respectively, in accordance with the psychoacoustic recommendations by *Taghipour et al.* [27].

These processed sounds were then assigned to their corresponding spatial objects within the Unity scene. The only source designated as a 2D spatializer was the background noise, providing a non-diegetic atmospheric layer, while all other sources were implemented as 3D spatialized events.

00:00.000

AcousticGain 0.00
HighEQ 1.00
LowEQ 1.00
MidEQ 1.00
OcclusionReal 0.00

Audio 1: -oo dB
Audio 2: 2.7 dB
Master: 0dB
Radius: 200
Volume: 0.00 dB
A: LP 12dB
B: Off
C: Off
Direction: 0 Deg
Extent: 360 Deg
Persistent: Off
Doppler: On
Max Instances: oo
Pitch: 0.00 st
Doppler Scale: 500%
Stealing: -
Priority: Medium
Cooldown: 0.00 ms
Min & Max Distance: 7.00 – 35.0
Platform: Desktop

Figure 3.30: FMOD Studio screen

Afterwards, the process required a meticulous calibration of individual sound sources to establish the correct volume ratios between objects. To achieve this, the methodology proposed by *Rehman et al.* [6] was adopted.

Regarding the spatial acoustics, a procedural approach developed by technical sound designer Ivan Stepanov [29] was implemented. His integration scripts for Unity and FMOD provide a real-time procedural interpretation of reverberation and occlusion, utilizing ray-casting and geometric analysis of the scene's meshes to simulate physical sound propagation. Each virtual sound source was assigned a specific SPL target, calibrated against its actual RMS (Root Mean Square) value within the digital environment. The comprehensive list of these SPL targets and their corresponding references is detailed in **Table 3.13**.

Table 3.13: SPL target for each sound source.

Sound source	**SPL target**	**Reference**
Car	85 dB	*Rehman et al.* [6]
Footsteps	55 dB	*Aletta et al.* [37]
Background noise	49 dB	*Llorca-Bofi et al.* [50]
Fountain	78 dB	*Hong et al.* [18]

As established in the literature, each sound source was individually placed within an empty Unity scene, with an FMOD Audio Listener positioned at a specific distance determined by the reference studies. Utilizing a custom-developed script, 60 RMS (Root Mean Square) samples were recorded during real-time playback for each source. From these measurements, the baseline effective pressure $\tilde{p}_{base}$ was derived. Subsequently, using the predetermined SPL_{target}, the required target pressure $\tilde{p}_{target}$ was calculated as follows:

$$\tilde{p}_{\text{base}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{3.3}$$

$$\tilde{p}_{\text{target}} = 20\,\mu\text{Pa} \cdot 10^{\frac{\text{SPL}_{\text{target}}}{20}} \tag{3.4}$$

$$G_{\text{cal}} = G_{\text{base}} \cdot \frac{\tilde{p}_{\text{target}}}{\tilde{p}_{\text{base}}} \tag{3.5}$$

In this calculation, $G_{base} = 1$ was assumed to represent the 0 dBFS reference level, as the calibration was conducted entirely within the digital domain.

Subsequently, the distance-based attenuation for each sound source was configured. Within FMOD Studio, the maximum propagation distance and the shape of the attenuation curve were defined for each event. To ensure physical accuracy, the maximum audible distance was calculated in accordance with the **ISO 9613-2** standard [51] (3.6). For these calculations, a background noise floor of $SPL_f = 49$ dB was adopted as the limit of audibility [18]. Furthermore, each source was modeled as a spherical point source, neglecting additional atmospheric or ground absorption effects to maintain model focus on geometric divergence.

$$SPL_f = SPL_i - log_{10}(r) - 11dB \tag{3.6}$$

Each sound source was assigned to its corresponding object within the scene. The dynamic environment featured a bidirectional traffic system, along with two pedestrians walking near the audio listener at different intervals. Furthermore, two "sport" vehicle events were integrated to transition through the standard traffic flow.
For the real-time acoustic simulation, the methodology developed by *Ivan Stepanov* [29] was implemented. This approach leverages ray-tracing technology to dynamically scan the environment surrounding the listener, subsequently modulating FMOD Studio parameters to adapt the audio perception based on the surrounding geometry (**Figure 3.31**).

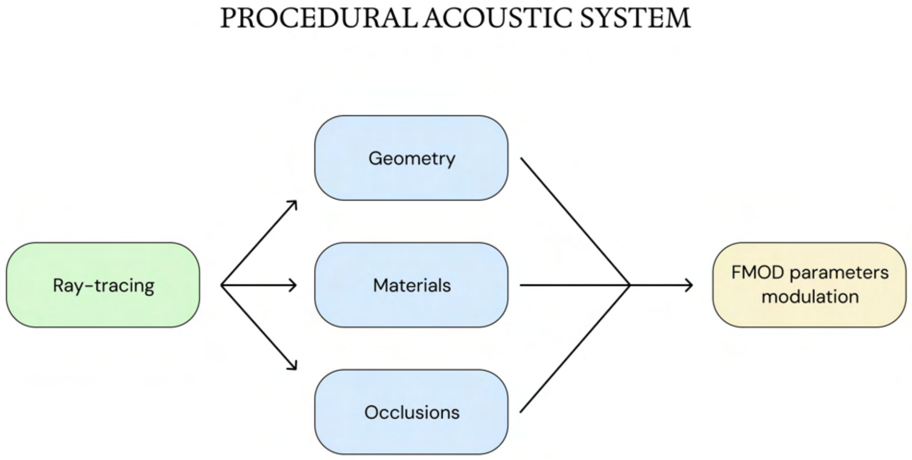


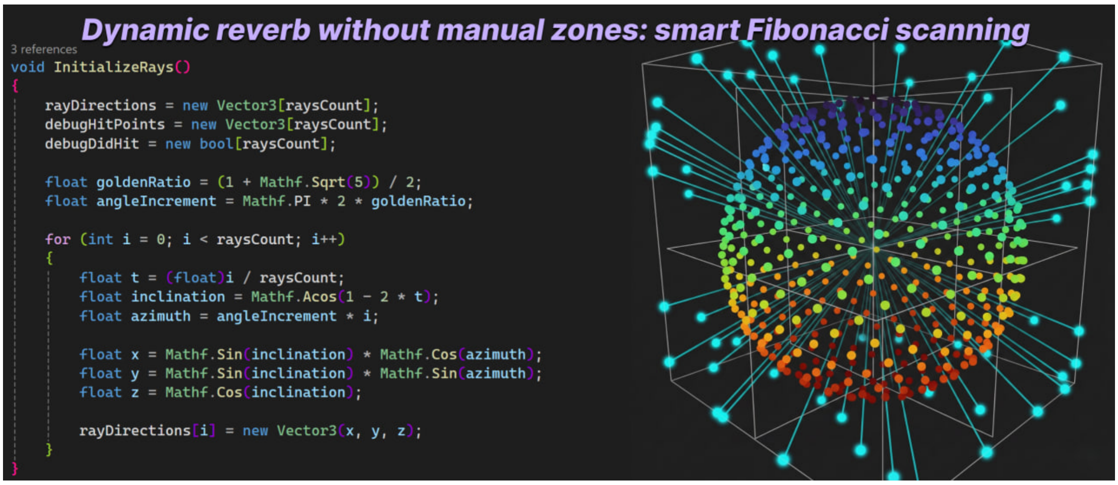
Figure 3.31: Procedural Acoustic System

The system utilizes a Fibonacci sphere geometry (**Figure 3.32**) to uniformly distribute rays into the environment starting from the audio listener's position. This configuration manages three distinct physical acoustic effects:

- **Geometry**: The system implements a script that identifies surrounding objects and their relative distances. These data are used to dynamically modulate *RoomSize* and *ReverbVolume* parameters, emulating real-world reflections. A specific threshold is applied to the ratio of "hitting rays" to total rays; below this limit, reverb parameters are deactivated to prevent artifacts or false acoustic impressions within the environment.

- **Materials**: The original script was modified to account for the acoustic properties of surfaces. Each material in the scene is associated with a template based on its specific absorption and transmission coefficients [52]. When a ray intersects an object, it retrieves the material data and adjusts the FMOD Studio *Volume* and *EQ* parameters to simulate the corresponding acoustic response.

- **Occlusion**: The system traces a direct ray connecting each sound source to the listener. If an obstacle is detected along this path, the system modulates the *Volume* and *EQ* parameters in FMOD Studio, adapting the occlusion effect proportionally to the obstacle's surface area and position.



```
void InitializeRays()
{
    rayDirections = new Vector3[raysCount];
    debugHitPoints = new Vector3[raysCount];
    debugDidHit = new bool[raysCount];

    float goldenRatio = (1 + Mathf.Sqrt(5)) / 2;
    float angleIncrement = Mathf.PI * 2 * goldenRatio;

    for (int i = 0; i < raysCount; i++)
    {
        float t = (float)i / raysCount;
        float inclination = Mathf.Acos(1 - 2 * t);
        float azimuth = angleIncrement * i;

        float x = Mathf.Sin(inclination) * Mathf.Cos(azimuth);
        float y = Mathf.Sin(inclination) * Mathf.Sin(azimuth);
        float z = Mathf.Cos(inclination);

        rayDirections[i] = new Vector3(x, y, z);
    }
}
```

Figure 3.32: Fibonacci's sphere logic [29].

Once the model was completed, it was ready to be calibrated with lab instrumentation for experiment.

3.4 Laboratory Experiment

The experiment was conducted at the **Audio Space Lab** (ASL) of the Politecnico di Torino. This facility was specifically designed to evaluate virtual environments with a primary focus on the auditory experience. The laboratory features a sophisticated spherical loudspeaker array, comprising a total of 16 monitors and two sub-woofers, as illustrated in **Figure 3.33**. The experimental design aimed to replicate the soundwalk scenarios within a controlled laboratory setting, allowing participants to re-experience the environments under varied conditions. In total, eight distinct combinations were presented:

Figure 3.33: Image from Audio Space Lab (ASL).

- **Three real models** (audiovisual recordings) with audio reproduced via the spherical loudspeaker array;
- **Three real models** with audio reproduced via high-fidelity headphones;
- **Two virtual models** (CGI reconstructions of the RB02 site) featuring distinct soundscape scenarios:
  - The first was designed to be as faithful as possible to the real-world baseline;
  - The second was inspired by urban soundscape interventions, incorporating a fountain sound source behind the "Spazio 211" area and reducing the frequency of vehicular traffic.

Each participant was tasked with assessing the scenarios through a two-minute immersive session using a Meta Quest Link 2 HMD, followed immediately by an evaluation questionnaire based on the original soundwalk design. Throughout the assessment, participants remained immersed in the soundscape, with the audio sources playing in a continuous loop to maintain environmental presence.

3.4.1 Experimental Design and Signal Flow

In the laboratory, three distinct Unity projects were developed to streamline the management of the experimental conditions. As illustrated in the signal flow diagram (**Figure 3.37**), these projects coordinated outputs across FMOD, the Unity audio framework, and Plogue Bidule (**Figure 3.34**).

Figure 3.34: Plogue Bidule screen.

For the three scenarios reproduced via the loudspeaker array in 3rd-order Ambisonics, the system utilized the **Open Sound Control** (OSC) protocol. Upon scene activation, Unity sent a trigger to Plogue Bidule, which initiated high-order Ambisonics (HOA) playback through an Antelope Orion32 audio interface.

The three real-world scenarios designated for headphone playback were managed within a project utilizing the **Meta Audio SDK**, which enabled the spatialization of 1st-order Ambisonics files directly within the Unity audio framework.

The final two virtual scenarios were managed entirely via **FMOD Studio**, allowing for real-time manipulation and editing of the procedural audio environment prior to scene loading.

To ensure that the Sound Pressure Levels (SPL) accurately reflected the parameters recorded during the field soundwalk, a rigorous calibration phase was conducted.

For the 3rd-order Ambisonics reproduction on the loudspeaker array, a monophonic signal was recorded at the listener's central position (**Figure 3.35**). This signal was processed through a custom MATLAB script to determine the required final gain. The calibration was referenced against a 94 dB SPL sinusoidal calibration tone at 1 kHz.

Figure 3.35: Picture from calibration process.

Furthermore, binaural recordings were conducted at each location using calibrated binaural microphones to ensure a direct and reliable comparison with in-field parameters during the subsequent analysis. This approach provides a reference baseline, allowing for the cross-validation of psychoacoustic metrics between the real-world acoustic environment and the simulated virtual scenarios.

For the headphone-based scenarios, a **Head Acoustics** artificial head (HMS) was employed. The signal was captured through the headphones by the binaural microphones of the artificial head (**Figure 3.36**).

Figure 3.36: Artificial head by HEADAcoustic with Meta Quest 2 visor and Senheiser HD 600 headphones.

The resulting $L_{A,eq}$ was analyzed using **ArtemiS** software, and the final gain adjustment was calculated following the methodology previously described for individual source calibration [6]. The gain values required to match the original field SPL are summarized in **Table 3.14**.

3.4.2 Questionnaire

The questionnaire administered to participants at the conclusion of each listening session was designed to maintain consistency with the field survey.
The primary modifications involved the inclusion of a field for the listening setup (speakers vs. headphones) and a dedicated **immersivity section**. This latter section is crucial for evaluating the correlation between the perceived realism of the virtual experience and real-world perception [53].

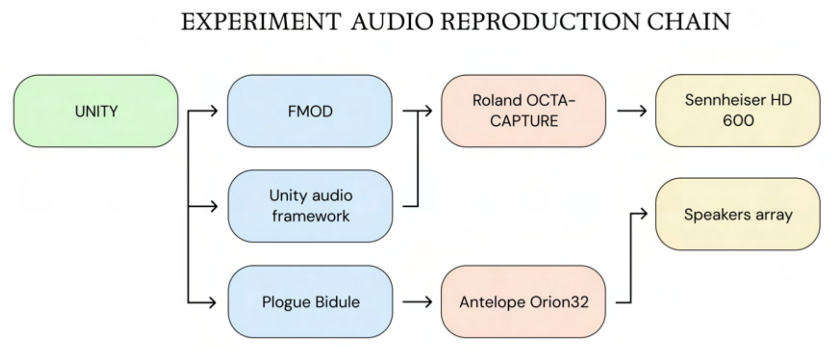


Figure 3.37: Experiment audio reproduction chain.

Table 3.14: SPL headphone calibration gain matching values.

Location	$\boldsymbol{L_{A_{eq}}}$	$\boldsymbol{L_{A_{eq}target}}$ target	$\boldsymbol{G_{match}}$
RB01	63.7	59.4	-4.3
RB02	59.4	65.0	5.6
BL	56.6	57.1	0.5
RB02 (VR)	69.7	65.0	-4.7

Since the sample group consisted of both individuals who had participated in the original soundwalk and new participants who had not experienced the physical environment, the questionnaire was provided in both Italian and English. This bilingual approach was necessary to ensure full comprehension across the diverse demographic of the sample and to avoid data distortion due to language barriers.

This distinction allowed for a further categorization of the results, distinguishing between "in-field" experienced individuals and "new" participants.

The translation of the questionnaire was meticulously executed, accounting for the linguistic biases identified in recent studies regarding the cross-cultural translation of soundscape descriptors [12].

The complete questionnaire is provided in the **Appendix A**.

Chapter 4

Results

4.1 Psychoacoustic parameters comparison

To ensure clarity and systematic comparison across the dataset, each experimental condition was assigned a unique identification code (ID). These IDs were structured based on the specific parameters of the listening setup and the scenario configuration, as detailed in **Table 4.1**. This standardized nomenclature facilitates the cross-referencing of objective acoustic data with the subjective evaluations collected during the assessment sessions.

The comparison between in-field and speaker reproduction laboratory-collected data were already validated, as the laboratory system's validity has been previously established [54].

However, the analysis focused on evaluating the variations among IF, SP360, HP360, VR, and AVR setups within each specific location.

The following tables present a comparative analysis of these configurations, reporting the relative difference (Δ) and the Just Noticeable Difference (JND) for each psychoacoustic parameter. This approach allows for a rigorous assessment of whether the observed technical variances translate into perceptible auditory differences for the user.

The comparison between the IF recordings and SP360 was conducted using data acquired via a binaural headset, supported by Siemens Simcenter Testlab software to ensure a consistent setup throughout the analysis. Additionally, further comparisons were performed using measurements obtained from an artificial head, which were subsequently processed and elaborated within the Head Acoustics ArtemiS

SUITE software.

The comparative analysis in **Table 4.1** reveals significant differences between the loudspeaker array and headphone setups. Notably, the $L_{C_{eq}}$ values show a variance of $\Delta L_{C_{eq}} = 7.8$ dB, indicating a substantial increase in the perception of low-frequency content when using headphones.

This discrepancy may be attributed to the frequency response of the headphones, as illustrated in the example in **Figure 4.1**. Furthermore, the physical design of the headphones can create a resonant system that artificially enhances low frequencies. To mitigate these effects and ensure acoustic transparency, it is appropriate to apply an **inverse filter** to the headphones. This equalization process aims to achieve a "flat" response, compensating for the transducer's coloration and aligning the playback as closely as possible with the original sound field.

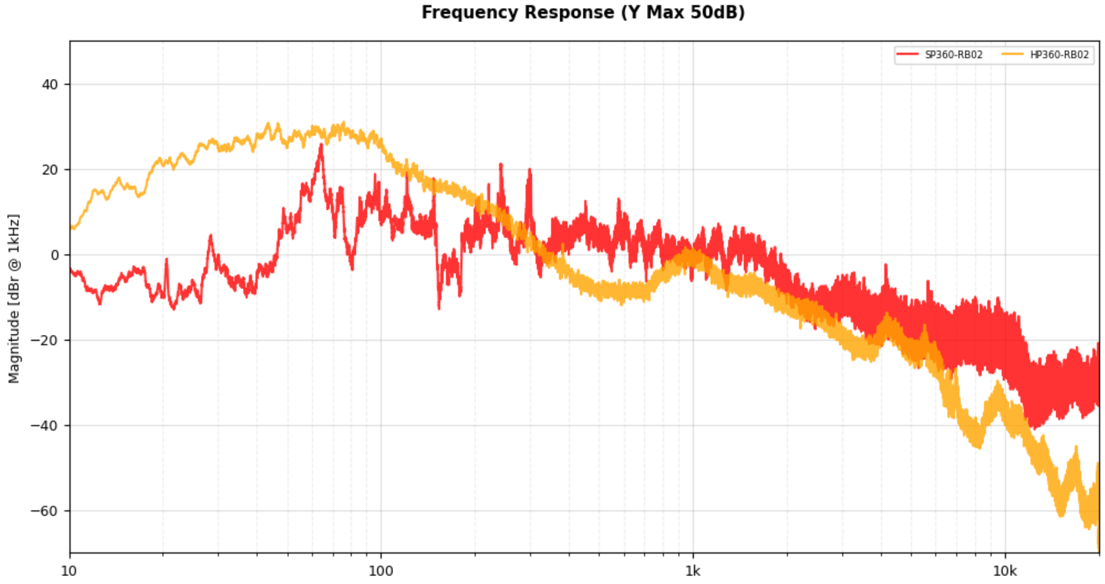


Figure 4.1: Frequency response differences between SP360 and HP360 conditions in RB02 station.

Furthermore, the data regarding average Sharpness show a variation of $\Delta S_{avg} = -0.22$ acum. This negative delta suggest a perceptible shift in the spectral balance, highlighting a different reproduction of the higher frequency components between the two listening conditions.

The trends identified in the analysis of the RB01 station remain consistent for the RB02 data. Despite a reduction in average Loudness with $\Delta N_{avg} = -5.9$ sone, the $L_{C_{eq}}$ values remain significantly elevated, showing a variance of $\Delta L_{C_{eq}} = 12.1$ dB, as detailed in **Table 4.1**.

A consistent trend is also observed in the baseline data, as presented in **Table 4.1**. In this instance, while the difference in average Loudness is negligible ($\Delta N_{avg} = -0.2$ sone), the trends for $L_{C_{eq}}$ and average Roughness (R_{avg}) remain aligned with those of the other stations.

However, the analysis yields different results for the virtual environments. As shown in **Table 4.1**, a variance of $\Delta L_{C_{eq}} = -4.34$ dB was observed, representing a notable departure from the spectral characteristics of the 360-degree video scenarios listened to via headphones.

Another significant factor is the increase in Roughness, with $\Delta R_{50} = 0.461$ asper. This is likely attributable to the specific spectral modulations of the synthesized car engine noise. A similar trend is observed when comparing the Virtual Reality (VR) baseline to the Alternative Virtual Reality (AVR) intervention.

As detailed in **Table 4.1**, the AVR scenario shows a negative ΔR_{50}, indicating a decrease in Roughness that correlates with the programmed reduction in traffic intensity. Simultaneously, N_{avg} (Loudness) and S_{avg} (Sharpness) values show evident decreases, despite negligible differences in $L_{A_{eq}}$ and $L_{C_{eq}}$. This suggests that while the overall sound pressure level remained stable, the perceived quality and "aggressiveness" of the soundscape were successfully mitigated by the design interventions.

Subsequently, a correlation analysis between the psychoacoustic parameters was conducted. The *Spearman's rank correlation coefficient* was selected due to the non-linear nature of the acoustic metrics and the potential presence of non-normal distributions in the subjective assessments, as illustrated in the correlation matrix in **Figure 4.2**. This non-parametric approach ensures a more accurate evaluation of the monotonic relationships between objective measurements (such as Loudness, Sharpness, and Roughness) and the perceived soundscape descriptors.

Since virtual environments were excluded from this analysis to maintain a balanced comparison across all three conditions, it is noteworthy that the headphone simulations consistently show a high correlation between N_{avg} and $L_{C,eq}$. This statistical evidence corroborates the previously analyzed data in the tables, reinforcing

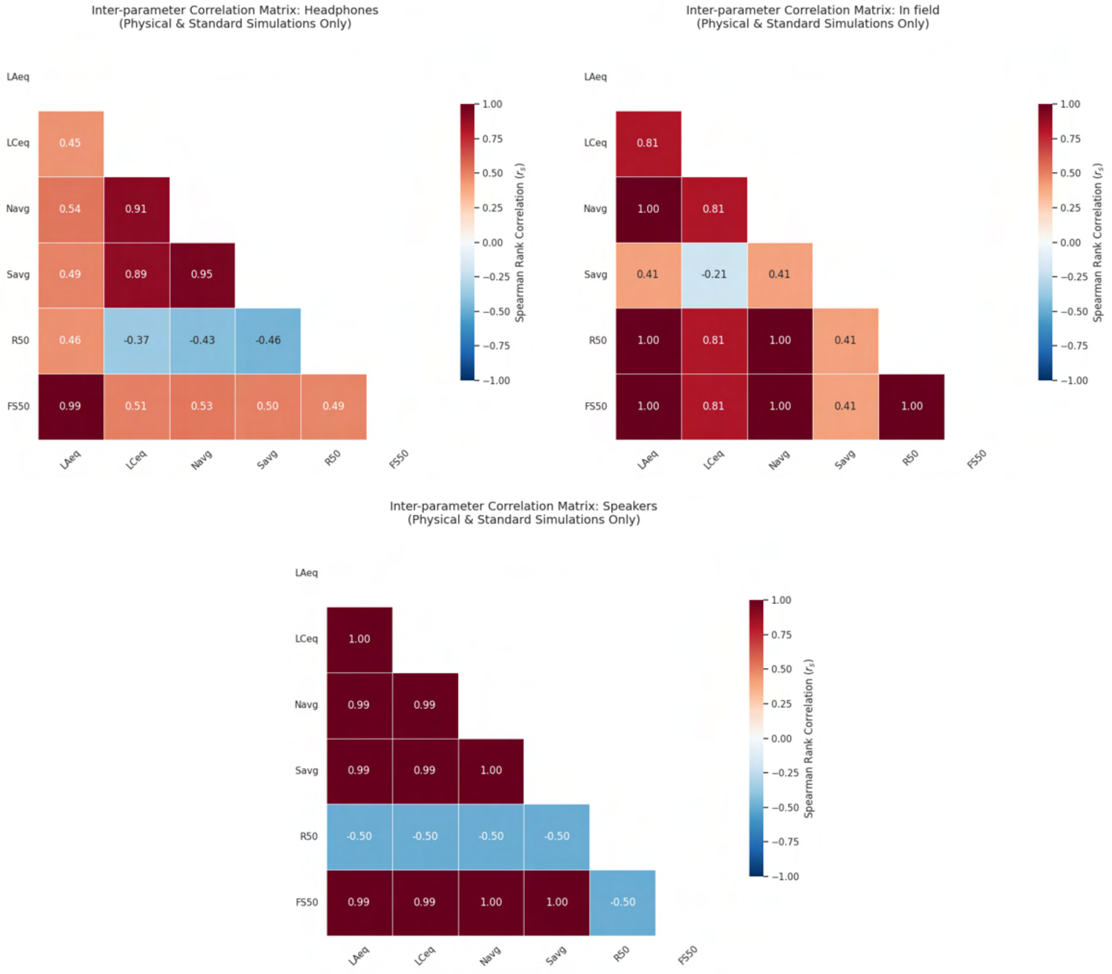


Figure 4.2: Correlation tables for psychoacoustic parameters divided in simulation methods.

the role of low-frequency energy in perceived loudness for this setup.

In the in-field assessments, the correlation between $L_{A,eq}$ and N_{avg} is significantly stronger. Conversely, S_{avg} (Sharpness) demonstrated greater variability depending on the specific location, suggesting that spectral clarity in the real world is more sensitive to environmental context than to overall pressure levels.

In the loudspeaker array setup, correlations are stronger across nearly all parameters. This phenomenon can be attributed to the acoustically treated environment of the laboratory, which minimizes external noise and uncontrolled reflections, thereby allowing for a more linear and predictable interaction between acoustic metrics.

These findings provide a fundamental basis for evaluating the correlation between objective psychoacoustic parameters and the subjective perceptual values derived from the questionnaires for each experimental condition.

4.2 Subjective parameters comparison

The comparative analysis across all stations and setups, as illustrated in **Table 4.3**, highlights the BL station as the most consistent location; it was confirmed as the quietest and most stable environment regardless of the listening configuration.

Generally, a perceptible shift exists between the original in-field conditions and the laboratory assessments. This variance may be attributed to the sample's composition, as a significant portion of the laboratory participants did not attend the initial soundwalk. This "novelty effect" likely justifies the observed shift in perception across all evaluation spots.

For the RB01 station, a clear distinction emerges: the real environment listened to via headphones is perceived as more **eventful**, whereas the loudspeaker array reproduction leads to a more **calm** and **monotonous** assessment. Overall, RB01 exhibits the highest degree of variability across the different devices and simulated environments.

The RB02 station consistently ranks as the **least pleasant** across all evaluations. The real-world recordings (both via speakers and headphones) emphasize the **eventfulness** of the site, though reproduction through the loudspeaker array was judged as significantly more pleasant than the headphone counterpart.

Within the virtual environments, the VR setup was perceived as the most **monotonous**. In contrast, the AVR intervention—which included the fountain and reduced traffic—shifted the perception closer to that of the real-world headphone experience, suggesting that the sound design successfully bridged the gap between synthetic and recorded realism.

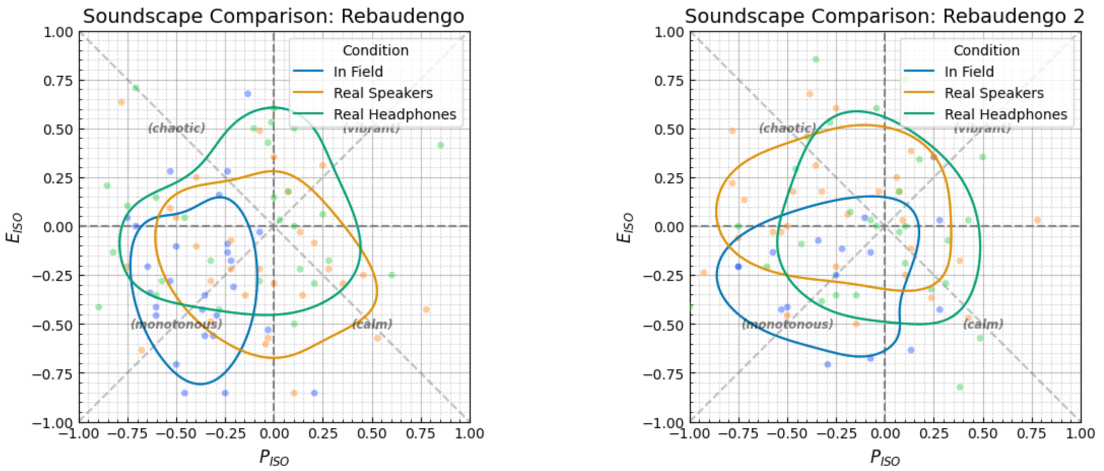
All these observations are represented even in **Tables 4.10**, **4.11** and **4.12**.

Table 4.10: Descriptive statistics and ISO 12913-2 attributes for each condition in RB01.

Condition	**Count**	**ISOP.**	**ISOE.**	**Pleas.**	**Even.**	**Vibr.**	**Chao.**	**Mono.**	**Calm**
IF	27	-0.364	-0.269	1.407	1.889	2.037	2.778	3.481	2.074
HP360	28	-0.155	0.039	2.214	3.107	2.643	3.179	3.286	2.357
SP360	29	-0.075	-0.165	2.310	2.793	2.310	2.483	3.655	2.759

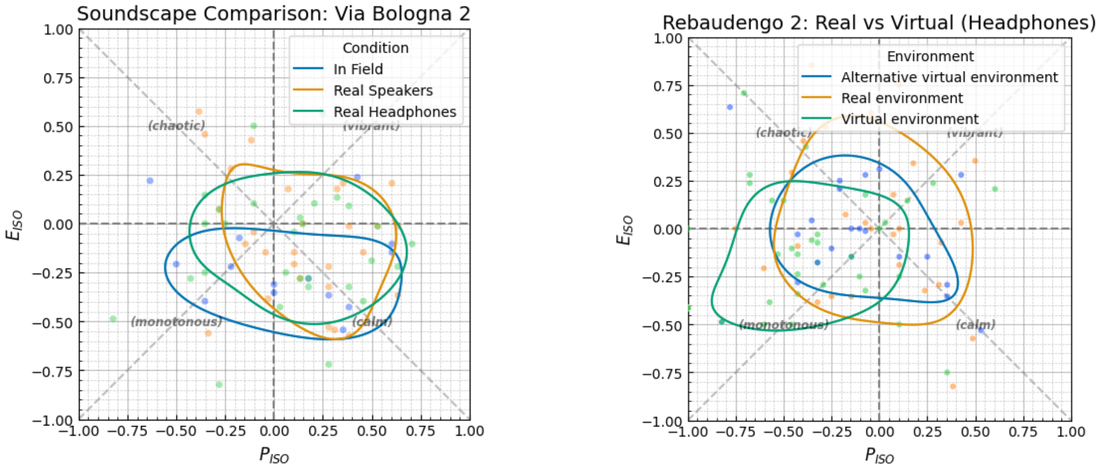
Table 4.11: Descriptive statistics and ISO 12913-2 attributes for each condition in RB02.

Condition	**Count**	**ISOP.**	**ISOE.**	**Pleas.**	**Even.**	**Vibr.**	**Chao.**	**Mono.**	**Calm**
AVR	21	−0.111	−0.033	2.524	2.905	2.619	2.905	3.333	2.714
IF	16	−0.220	−0.249	2.562	2.062	2.000	2.562	3.562	2.188
HP360	28	−0.072	0.014	2.607	2.929	2.821	3.036	3.321	2.500
SP360	30	−0.189	0.064	2.367	3.200	2.533	3.300	3.533	2.233
VR	36	−0.351	−0.116	1.806	2.750	2.167	3.333	3.722	2.222



(a) Subjective comparison plot for different setups in RB01 station.

(b) Subjective comparison plot for different setups in RB01 station.



(c) Subjective comparison plot for different setups in BL station.

(d) Subjective comparison plot for different environments in RB0 station with headphones setup.

Figure 4.3: Subjective comparisons between setups and environments plots.

Table 4.12: Descriptive statistics and ISO 12913-2 attributes for each condition in BL.

Condition	**Count**	**ISOP.**	**ISOE.**	**Pleas.**	**Even.**	**Vibr.**	**Chao.**	**Mono.**	**Calm**
IF	14	0.067	−0.212	2.857	2.357	2.286	2.143	3.071	3.143
HP360	29	0.079	−0.126	2.897	2.897	2.586	2.345	3.276	3.138
SP360	28	0.143	−0.073	3.179	3.036	2.821	2.393	3.250	3.107

To evaluate the perceptual differences across each listening condition, the *Mann-Whitney U test* was employed. This non-parametric test was selected to perform pairwise comparisons between conditions, accounting for the independence of the participant samples between the in-field and laboratory settings.

As shown in **Table 4.13**, the results demonstrate a high degree of consistency across the different descriptors. The absence of statistically significant differences in several key parameters suggests a linear alignment in the assessment of each condition, indicating that the subjective perception remained stable regardless of the specific listening setup used for those descriptors.

Table 4.13: Intra-Laboratory Comparison (HP360 vs SP360): Mann-Whitney U test results and descriptive statistics.

Attribute	HP360 (M)	SP360 (M)	*p*-value	Inference
Pleasant	2.58	2.65	0.75	Consistent
Vibrant	2.68	2.56	0.51	Consistent
Eventful	2.98	3.01	0.86	Consistent
Chaotic	2.85	2.71	0.48	Consistent
Annoying	2.48	2.34	0.45	Consistent
Monotonous	3.29	3.46	0.44	Consistent
Uneventful	2.92	2.94	0.98	Consistent
Calm	2.67	2.74	0.76	Consistent

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. M = Mean score.

A markedly different situation is presented in **Table 4.14**, where several parameters exhibit statistically significant differences. The most pronounced divergence is observed along the **Eventfulness** axis. This finding may be attributed to the inherent psychological difference between being in an open, physical space versus an isolated, simulated environment. The latter often prompts participants to focus more intensely on specific auditory events due to the lack of external distractions. Simultaneously, the **Vibrant** and **Annoying** descriptors show significant variations, further supporting the hypothesis that the transition from a real-world setting to a controlled laboratory environment alters the perceived intensity and character of the soundscape. A similar trend is observed in **Table 4.15**, with the most significant difference emerging in the **Pleasant** descriptor. This shift suggests

Table 4.14: In field-In lab Comparison (IF vs HP360): Mann-Whitney U test results and descriptive statistics.

Attribute	**IF (M)**	**HP360 (M)**	***p*-value**	**Inference**
Pleasant	2.09	2.58	0.011*	Significant Diff.
Vibrant	2.09	2.68	0.0023**	Significant Diff.
Eventful	2.05	2.98	0.0000***	Significant Diff.
Chaotic	2.56	2.85	0.14	Consistent
Annoying	3.12	2.48	0.0024**	Significant Diff.
Monotonous	3.40	3.29	0.49	Consistent
Uneventful	3.67	2.92	0.0003***	Significant Diff.
Calm	2.37	2.67	0.12	

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. M = Mean score.

that SP360 condition, while maintaining spatial coherence, may alter the perceived hedonic quality of the soundscape compared to the in-field experience. Such a discrepancy could be linked to the acoustic characteristics of the laboratory—where the absence of non-auditory environmental stressors and the controlled nature of the room acoustics might lead participants to rate the environment as more (or less) pleasant than the original location, depending on the specific sound sources involved.

Regarding the RB02 station, a focused comparison between the HP360 (real-world recording via headphones) and the VR (virtual reconstruction) conditions was conducted. As shown in **Table 4.16**, significant differences emerge across the **Pleasant**, **Eventfulness**, and **Annoying** parameters. These statistical findings are consistent with qualitative feedback provided by participants, who highlighted the intrusive nature of vehicular noise in the VR scenario. The disparity in the **Annoying** descriptor can be attributed to the constant speed and acoustic emission of the vehicles within the Unity/FMOD environment; since the cars follow a predefined path at a constant speed of 50 km/h, they lack the natural fluctuations and "spectral life" of real-world traffic, resulting in a more persistent and taxing auditory stimulus. The AVR scenario exhibited further significant departures from the standard VR condition. As detailed in **Table 4.17**, the most prominent difference concerns the **Pleasant** descriptor. This result was anticipated, as the AVR environment was specifically designed to enhance participant well-being through targeted soundscape interventions.

Table 4.15: In field-In lab Comparison (IF vs SP360): Mann-Whitney U test results and descriptive statistics.

Attribute	**IF (M)**	**SP360 (M)**	***p*-value**	**Inference**
Pleasant	2.09	2.65	0.0052**	Significant Diff.
Vibrant	2.09	2.56	0.013*	Significant Diff.
Eventful	2.05	3.01	0.0000***	Significant Diff.
Chaotic	2.56	2.71	0.38	Consistent
Annoying	3.12	2.34	0.0001***	Significant Diff.
Monotonous	3.40	3.46	0.95	Consistent
Uneventful	3.67	2.94	0.0004***	Significant Diff.
Calm	2.37	2.74	0.11	Consistent

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. M = Mean score.

Table 4.16: In lab Comparison (HP360 vs VR): Mann-Whitney U test results and descriptive statistics (RB02 stop).

Attribute	**HP360 (M)**	**VR (M)**	***p*-value**	**Inference**
Pleasant	2.61	1.81	0.0012**	Significant Diff.
Vibrant	2.82	2.17	0.030*	Significant Diff.
Eventful	2.93	2.75	0.57	Consistent
Chaotic	3.04	3.33	0.31	Consistent
Annoying	2.57	3.31	0.029*	Significant Diff.
Monotonous	3.32	3.72	0.17	Consistent
Uneventful	2.82	3.56	0.033*	Significant Diff.
Calm	2.50	2.22	0.38	Consistent

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. M = Mean score.

Furthermore, the **Uneventful** descriptor was significantly higher in the VR condition compared to the AVR one. This finding is corroborated by qualitative feedback from participants, who noted an improved ability to perceive subtle sounds, such as footsteps, in the AVR scenario. This enhanced auditory clarity is likely due to the reduction in vehicular traffic intensity, which in the VR model acted as a dominant masker, obscuring low-level informational sound sources that contribute to a sense of presence and environmental detail.

Table 4.17: In lab Comparison (VR vs AVR): Mann-Whitney U test results and descriptive statistics (RB02 stop).

Attribute	**VR (M)**	**AVR (M)**	p**-value**	**Inference**
Pleasant	1.81	2.52	0.0039**	Significant Diff.
Vibrant	2.17	2.62	0.0709	Consistent
Eventful	2.75	2.90	0.6462	Consistent
Chaotic	3.33	2.90	0.1751	Consistent
Annoying	3.31	2.95	0.2314	Consistent
Monotonous	3.72	3.33	0.1362	Consistent
Uneventful	3.56	2.86	0.0300*	Significant Diff.
Calm	2.22	2.71	0.0720	Consistent

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. M = Mean score.

4.3 Comparison between Objective and Subjective Parameters

To evaluate the correlation between psychoacoustic parameters and participant assessments, the *Spearman's rank correlation coefficient* was employed, accounting for the non-linear nature of the data.

The dataset was segmented by listening condition to isolate the perceptual shifts related to specific contexts and setup configurations. As illustrated in **Figure 4.4**, several strong correlations emerge across the different conditions (excluding the VR and AVR scenarios from this specific comparison).

In the **In-Field (IF)** condition, a robust correlation was found between **Pleasantness** and S_{avg} (Sharpness), as well as between **Annoying** scores and the other psychoacoustic metrics. The **SP360** (Speaker Array) condition demonstrated a strong correlation between nearly all parameters and the **Chaotic** descriptor, with the exception of the R_{50} (Roughness) value.

In the **HP360** (Headphone) condition, a specific and significant correlation was identified between the **Chaotic** descriptor and the $L_{A,eq}$, FS_{50} (Fluctuation Strength), and R_{50} parameters. This suggests that in headphone-based simulations, the perception of environmental disorder is more closely tied to temporal fluctuations and spectral roughness than in other setups.

Comprehensive analyses of the statistically significant correlations are provided in **Tables 4.18, 4.19**, and **4.20**. These tables highlight again how the listener's relationship with the acoustic environment changes depending on the reproduction system and the physical context.

Table 4.18: Spearman's significance correlations (HP360).

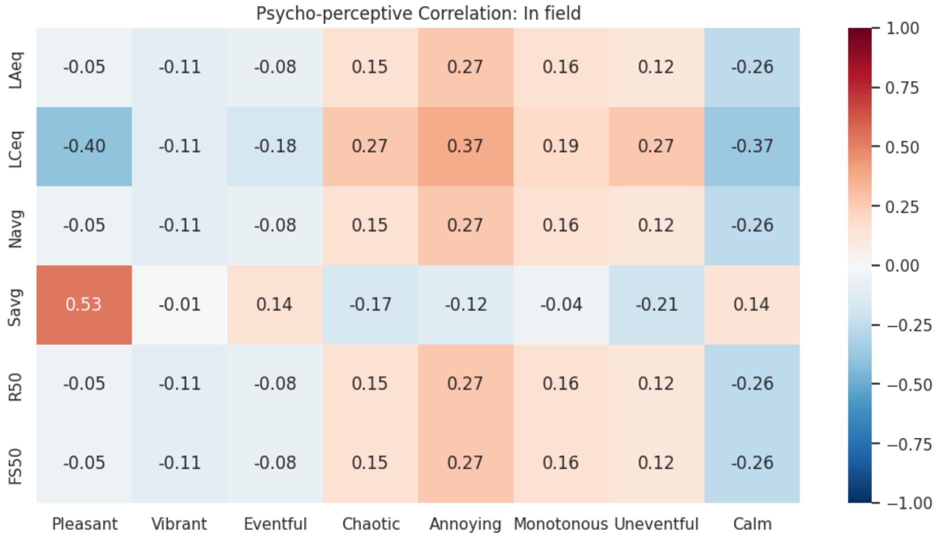
Parameter (X)	**Descriptor (Y)**	r_s	p**-value**
R_{50}	Chaotic	0.273	0.010**
R_{50}	Pleasant	-0.229	0.032*
FS_{50}	Chaotic	0.227	0.034*
$L_{A_{eq}}$	Chaotic	0.220	0.039*
R_{50}	Calm	-0.218	0.041*

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.4 Immersivity

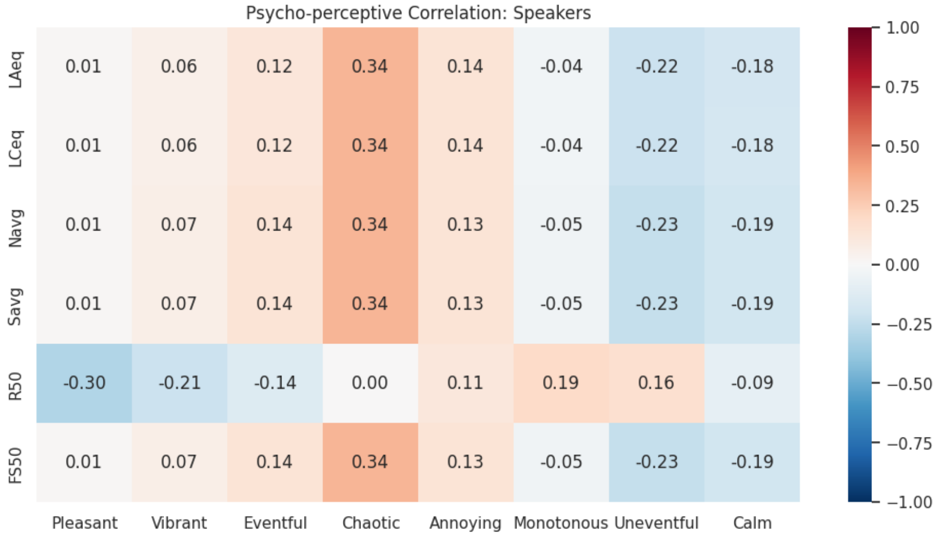
A critical factor that cannot be overlooked is the perception of immersivity by the participants. To evaluate the differences between the real and virtual environments while maintaining a balanced comparison, the RB02 station was analyzed across all headphone-based conditions. This approach ensures consistency, as the virtual environment (VR/AVR) does not include a loudspeaker reproduction setup.

As illustrated in **Figure 4.5**, notable differences emerge between the conditions. The **Immersivity** rating is higher in the real environment (HP360). This disparity is likely due to the inherent nature of Computer Graphics (CG) in VR, which may



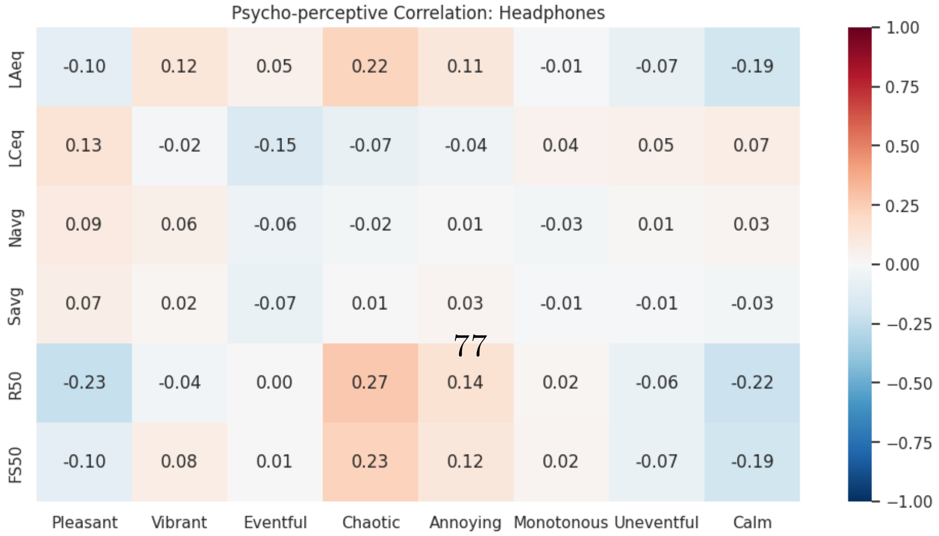
	Pleasant	Vibrant	Eventful	Chaotic	Annoying	Monotonous	Uneventful	Calm
LAeq	-0.05	-0.11	-0.08	0.15	0.27	0.16	0.12	-0.26
LCeq	-0.40	-0.11	-0.18	0.27	0.37	0.19	0.27	-0.37
Navg	-0.05	-0.11	-0.08	0.15	0.27	0.16	0.12	-0.26
Savg	0.53	-0.01	0.14	-0.17	-0.12	-0.04	-0.21	0.14
R50	-0.05	-0.11	-0.08	0.15	0.27	0.16	0.12	-0.26
FS50	-0.05	-0.11	-0.08	0.15	0.27	0.16	0.12	-0.26

(a) In-Field condition correlation.



	Pleasant	Vibrant	Eventful	Chaotic	Annoying	Monotonous	Uneventful	Calm
LAeq	0.01	0.06	0.12	0.34	0.14	-0.04	-0.22	-0.18
LCeq	0.01	0.06	0.12	0.34	0.14	-0.04	-0.22	-0.18
Navg	0.01	0.07	0.14	0.34	0.13	-0.05	-0.23	-0.19
Savg	0.01	0.07	0.14	0.34	0.13	-0.05	-0.23	-0.19
R50	-0.30	-0.21	-0.14	0.00	0.11	0.19	0.16	-0.09
FS50	0.01	0.07	0.14	0.34	0.13	-0.05	-0.23	-0.19

(b) Speaker condition correlation.



	Pleasant	Vibrant	Eventful	Chaotic	Annoying	Monotonous	Uneventful	Calm
LAeq	-0.10	0.12	0.05	0.22	0.11	-0.01	-0.07	-0.19
LCeq	0.13	-0.02	-0.15	-0.07	-0.04	0.04	0.05	0.07
Navg	0.09	0.06	-0.06	-0.02	0.01	-0.03	0.01	0.03
Savg	0.07	0.02	-0.07	0.01	0.03	-0.01	-0.01	-0.03
R50	-0.23	-0.04	0.00	0.27	0.14	0.02	-0.06	-0.22
FS50	-0.10	0.08	0.01	0.23	0.12	0.02	-0.07	-0.19

(c) Headphones condition correlation.

Figure 4.4: Correlation tables between psychoacoustic and subjective parameters for each condition.

Table 4.19: Spearman's significance correlations (IF).

Parameter (X)	**Descriptor (Y)**	r_s	p**-value**
S_{avg}	Pleasant	0.532	0.0000^{***}
$L_{C_{eq}}$	Pleasant	-0.401	0.0020^{**}
$L_{C_{eq}}$	Calm	-0.371	0.0045^{**}
$L_{C_{eq}}$	Annoying	0.368	0.0048^{**}
$L_{C_{eq}}$	Chaotic	0.272	0.041^{*}
N_{avg}	Annoying	0.271	0.042^{*}
$L_{C_{eq}}$	Uneventful	0.268	0.044^{*}
$L_{A_{eq}}$	Calm	-0.262	0.049^{*}

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.20: Spearman's significance correlations (SP360).

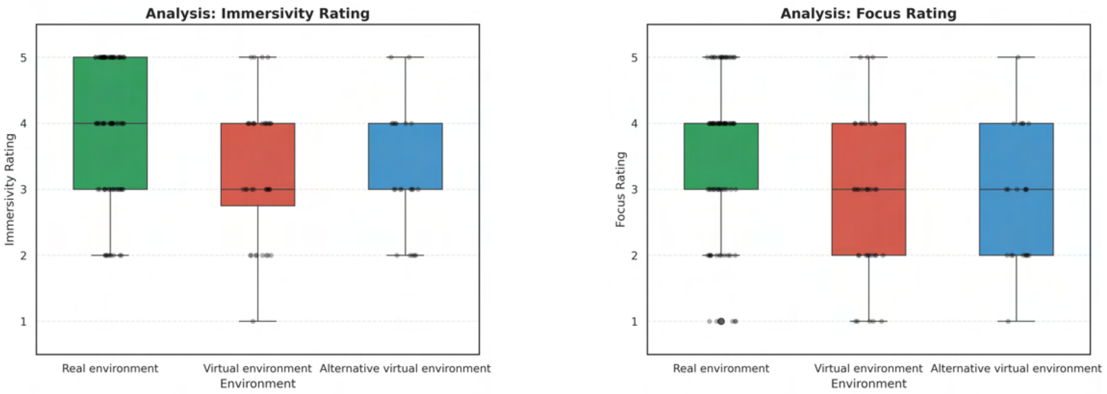
Parameter (X)	**Descriptor (Y)**	r_s	p**-value**
N_{avg}	Chaotic	0.343	0.0011^{**}
$L_{A_{eq}}$	Chaotic	0.339	0.0013^{**}
$L_{C_{eq}}$	Chaotic	0.337	0.0014^{**}
R_{50}	Pleasant	-0.303	0.0044^{**}
N_{avg}	Uneventful	-0.235	0.028^{*}
$L_{A_{eq}}$	Uneventful	-0.217	0.044^{*}
R_{50}	Vibrant	-0.212	0.048^{*}

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

not always achieve the high-fidelity visual realism of a 360-degree video, especially in the stylized representation proposed for this project.

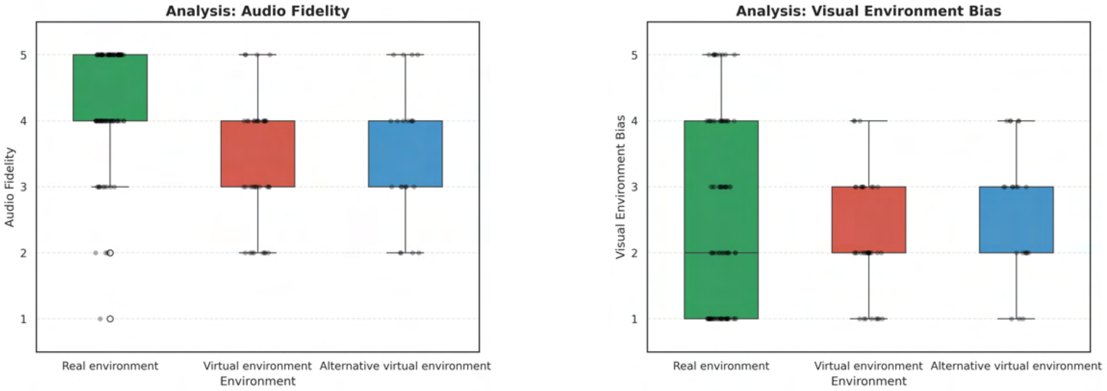
Similarly, the **Focus** rating remains higher for the real environment, whereas in the virtual conditions, the responses are more widely distributed. Regarding **Audio Fidelity**, the real environment recordings were rated more highly than the synthesized virtual environments.

Perhaps the most unexpected result is found in the **Visual Environment Bias**, where the real environment shows a significant distribution of scores. This could be correlated with participant feedback indicating difficulty in focusing on a real-world scene when bystanders or known individuals are visible or looking directly into the camera, an effect that is absent in the controlled, albeit less "realistic," virtual reconstruction.



(a) Immersivity rating comparison between virtual and real environments.

(b) Focus rating comparison between virtual and real environments.



(c) Audio fidelity comparison between virtual and real environments.

(d) Visual environment bias comparison between virtual and real environments.

Figure 4.5: Subjective comparisons between setups and environments plots.

Chapter 5

Conclusion

This research, despite certain discrepancies in experimental setups and participant demographics, provides significant insights into the ecological validity of soundscape assessments within immersive environments.

In defining a methodology for the development of Virtual Reality (VR) soundscapes, contemporary tools offer a wide range of assets capable of achieving a high degree of fidelity. However, several factors still require meticulous attention, as they represent the critical links between virtual and real-world perception.

While various software platforms can effectively manage visual fidelity [20] [23] and complex physical acoustic features—such as reflections [26], distance-based attenuation [49], and material absorption—the selection of sound sources and their synchronization with in-scene animations remain a significant challenge.

This issue was particularly evident in the present research: several participants expressed perplexity during the evaluation of the VR condition, specifically regarding the mechanical fidelity of vehicle sounds. The constant speed and lack of dynamic auditory feedback were identified as areas requiring further refinement. This finding is consistent with previous studies that highlighted similar limitations in procedural or sample-based traffic simulations [21].

Consequently, the fundamental comparison between real-world environments and Computer-Generated (CG) environments in VR remains a major challenge. It necessitates a delicate balance between absolute fidelity to the physical world and the inherent flexibility provided by virtual audiovisual representations.

The comparison between real and virtual conditions in the assessment of soundscapes yielded specific, quantifiable results. The analysis of psychoacoustic data revealed significant discrepancies, particularly during headphone-based listening sessions. This gap is likely attributable to the non-linear frequency response of the headphones, which can introduce spectral differences.

As demonstrated in this study, such effects could be mitigated through the application of an inverse filter designed to equalize the headphones and achieve a more balanced spectral content. This finding is closely aligned with established literature; for instance, the IHTA Park case study [24] emphasizes how headphone calibration is a prerequisite for achieving high perceptual agreement between in-situ and laboratory environments.

Closely related to these psychoacoustic variations is the perception of eventfulness. Laboratory-based listening elicited more eventful responses from participants, as the controlled environment appeared to facilitate better focus. However, this observation must account for the inherent differences between the in-field and laboratory participant samples; furthermore, it is important to note that these results do not align with findings from previous research in the field.

As before mentioned, in Virtual Reality (VR) and Alternative Virtual Reality (AVR) conditions, the results shifted. In the VR scenarios specifically, pleasantness levels were lower, while annoyance ratings were higher. Participant feedback from the questionnaires suggests this may be due to the sound and animation selection; for instance, the specific timbre and constant speed of the simulated vehicles created a more disturbing and monotonous environment compared to the real world. In the AVR condition, however, the implementation of soundscape interventions successfully re-stabilized these perception levels.

In general, in-field conditions remain perceptually distinct from laboratory settings across most parameters, with the exception of the [BL] station, which showed consistent results across both environments.

Furthermore, the "immersivity" section of the laboratory questionnaires provided important reflections. Although the real environment reproduction offered a superior acoustic simulation, a visual bias was persistent, as it is even reported in other researches [16]. Participants reported that the presence of people looking directly at the camera during the 360-degree recordings was distracting, a bias that was notably absent in the CG virtual reality environment.

Consequently, VR simulation emerges as a highly flexible and promising solution, though it demands meticulous environmental modeling [22]. The goal is to find a "sweet spot" that captures the fundamental elements of a real environment while leveraging the customization potential of the technology.

Nevertheless, laboratory reproductions of real environments remain a robust solution for exact representation, even without the ability to control individual scene elements. As current research confirms, the future challenge lies in creating integrated solutions that can be effectively applied across transversal fields such as urban soundscape design.

DISCLAIMER

The author confirms that Artificial Intelligence (AI) technologies were utilized during the development of this research. Specifically:

- **Textual Review:** AI tools were employed for linguistic refinement, grammar correction, and to improve the overall flow and clarity of the English manuscript.
- **Coding Support:** AI-assisted platforms provided technical support in writing and debugging scripts used for data processing, psychoacoustic analysis, and experimental simulation.

The author remains fully responsible for the originality of the research, the accuracy of the data, the scientific interpretations, and the final conclusions presented in this thesis.

Bibliography

[1] Jian Kang et al. «Ten questions on the soundscapes of the built environment». In: *Building and Environment* 108 (Nov. 2016), pp. 284–294. ISSN: 03601323. DOI: `10.1016/j.buildenv.2016.08.011`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0360132316303067` (visited on 12/30/2025) (cit. on pp. 3, 5, 6).

[2] Jian Kang and Brigitte Schulte-Fortkamp, eds. *Soundscape and the Built Environment*. CRC Press, Dec. 15, 2015. ISBN: 978-1-4822-2631-7 978-1-4822-2632-4. DOI: `10.1201/b19145`. URL: `http://www.crcnetbase.com/doi/book/10.1201/b19145` (visited on 12/30/2025) (cit. on p. 3).

[3] Arianna Latini, Simone Torresin, Tin Oberman, Elisa Di Giuseppe, Francesco Aletta, Jian Kang, and Marco D'Orazio. «Virtual reality application to explore indoor soundscape and physiological responses to audio-visual biophilic design interventions: An experimental study in an office environment». In: *Journal of Building Engineering* 87 (June 2024), p. 108947. ISSN: 23527102. DOI: `10.1016/j.jobe.2024.108947`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S2352710224005151` (visited on 12/12/2025) (cit. on pp. 3, 13).

[4] Hatice Kurukose Cal, Francesco Aletta, Jian Kang, and Peter Clarke. «Comparing Teachers' And Students' Perceptions Of School Soundscapes Through A Questionnaire-Based Study». In: *Proceedings of the 11th Convention of the European Acoustics Association Forum Acusticum / EuroNoise 2025*. 11th Convention of the European Acoustics Association Forum Acusticum / EuroNoise 2025. Málaga, Spain: European Acoustics Association, Dec. 25, 2025, pp. 4889–4895. ISBN: 978-84-87985-35-5. DOI: `10.61782/fa.2025.0230`. URL: `https://dael.euracoustics.org/confs/landing_pages/fa2025/000230.html` (visited on 12/30/2025) (cit. on p. 3).

[5] Andrew Mitchell, Tin Oberman, Francesco Aletta, Mercede Erfanian, Magdalena Kachlicka, Matteo Lionello, and Jian Kang. «The Soundscape Indices (SSID) Protocol: A Method for Urban Soundscape Surveys—Questionnaires with Acoustical and Contextual Information». In: *Applied Sciences* 10.7

(Apr. 1, 2020), p. 2397. ISSN: 2076-3417. DOI: `10.3390/app10072397`. URL: `https://www.mdpi.com/2076-3417/10/7/2397` (visited on 12/30/2025) (cit. on pp. 3, 7, 8, 22).

[6] Rouben Rehman, Christian Dreier, Jonas Heck, Josep Llorca-Bofí, and Michael Vorländer. «Comparison of virtual reality and web-based listening experiments on the perception in complex auralized environments». In: *The Journal of the Acoustical Society of America* 157.3 (Mar. 1, 2025), pp. 2001–2017. ISSN: 1520-8524. DOI: `10.1121/10.0036147`. URL: `https://pubs.aip.org/jasa/article/157/3/2001/3340465/Comparison-of-virtual-reality-and-web-based` (visited on 12/15/2025) (cit. on pp. 3, 15, 19, 55, 56, 62).

[7] ISO. «ISO 12913-1». In: (2014). URL: `https://www.iso.org/obp/ui/en/#iso:std:iso:12913:-1:ed-1:v1:en` (cit. on p. 5).

[8] ISO. «ISO 12913-2». In: (2018). URL: `https://www.iso.org/obp/ui/en/#iso:std:iso:ts:12913:-2:ed-1:v1:en` (cit. on p. 5).

[9] ISO. «ISO 12913-3». In: (2025). URL: `https://www.iso.org/obp/ui/en/#iso:std:iso:ts:12913:-3:ed-2:v1:en` (cit. on p. 6).

[10] M Adams and N Bruce. «SOUNDWALKING AS METHODOLOGY FOR UNDERSTANDING SOUNDSCAPES». In: *Spring Conference Acoustics 2008*. Spring Conference Acoustics 2008. Rugby: Institute of Acoustics, Nov. 28, 2023. DOI: `10.25144/17576`. URL: `https://www.ioa.org.uk/catalogue/paper/soundwalking-methodology-understanding-soundscapes` (visited on 03/01/2026) (cit. on p. 6).

[11] Giovanni Brambilla, Massimiliano Masullo, and Aniello Pascale. «LA PASSEG-GIATA SONORA AD ALGHERO THE SOUNDWALK AT ALGHERO». In: () (cit. on p. 6).

[12] Giuseppina Emma Puglisi, Louena Shtrepi, Marco Carlo Masoero, and Arianna Astolfi. «Evaluating soundscape in the Italian language: Validation of the translation of the English standardized perceptual attributes of the ISO/TS 12913-2:2018 and comparison with other Romance languages». In: *Applied Acoustics* 222 (June 2024), p. 110050. ISSN: 0003682X. DOI: `10.1016/j.apacoust.2024.110050`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0003682X24002019` (visited on 03/01/2026) (cit. on pp. 9, 63).

[13] Wikipedia. «Sound pressure level». In: (). URL: `https://en.wikipedia.org/wiki/Sound_pressure` (cit. on p. 10).

[14] Wikipedia. «Loudness». In: (). URL: `https://en.wikipedia.org/wiki/Loudness` (cit. on p. 11).

[15] Marlene Wessels, Sophie Kröling, and Daniel Oberfeld. «Audiovisual time-to-collision estimation for accelerating vehicles: The acoustic signature of electric vehicles impairs pedestrians' judgments». In: *Transportation Research Part F: Traffic Psychology and Behaviour* 91 (Nov. 2022), pp. 191–212. ISSN: 13698478. DOI: `10.1016/j.trf.2022.09.023`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S1369847822002200` (visited on 01/06/2026) (cit. on pp. 13, 14, 16, 19).

[16] Özlem Gök Tokgöz, Margret Sibylle Engel, Cherif Othmani, and M. Ercan Altinsoy. «Virtual Reality Application in Evaluating the Soundscape in Urban Environment: A Systematic Review». In: *Acoustics* 7.4 (Oct. 17, 2025), p. 68. ISSN: 2624-599X. DOI: `10.3390/acoustics7040068`. URL: `https://www.mdpi.com/2624-599X/7/4/68` (visited on 12/27/2025) (cit. on pp. 14, 82).

[17] Massimiliano Masullo, Luigi Maffei, Aniello Pascale, Vincenzo Paolo Senese, Simona De Stefano, and Chi Kwan Chau. «Effects of Evocative Audio-Visual Installations on the Restorativeness in Urban Parks». In: *Sustainability* 13.15 (July 26, 2021), p. 8328. ISSN: 2071-1050. DOI: `10.3390/su13158328`. URL: `https://www.mdpi.com/2071-1050/13/15/8328` (visited on 01/06/2026) (cit. on pp. 14, 19).

[18] Joo Young Hong, Bhan Lam, Zhen-Ting Ong, Kenneth Ooi, Woon-Seng Gan, Jian Kang, Samuel Yeong, Irene Lee, and Sze-Tiong Tan. «The effects of spatial separations between water sound and traffic noise sources on soundscape assessment». In: *Building and Environment* 167 (Jan. 2020), p. 106423. ISSN: 03601323. DOI: `10.1016/j.buildenv.2019.106423`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S036013231930633X` (visited on 01/06/2026) (cit. on pp. 14, 56).

[19] Xiaojie Long, Nazli Che Din, Yuliang Lei, and Norhayati Mahyuddin. «The restorative effects of outdoor soundscapes in nursing homes for elderly individuals». In: *Building and Environment* 242 (Aug. 2023), p. 110520. ISSN: 03601323. DOI: `10.1016/j.buildenv.2023.110520`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0360132323005474` (visited on 01/06/2026) (cit. on pp. 15, 20).

[20] Ubuntu Canonical. «Unity». In: (). URL: `https://it.wikipedia.org/wiki/Unity_(software)` (cit. on pp. 15, 51, 80).

[21] Like Jiang, Massimiliano Masullo, Luigi Maffei, Fanyu Meng, and Michael Vorländer. «A demonstrator tool of web-based virtual reality for participatory evaluation of urban sound environment». In: *Landscape and Urban Planning* 170 (Feb. 2018), pp. 276–282. ISSN: 01692046. DOI: `10.1016/j.landurbplan.`

2017.09.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169204617302086 (visited on 01/06/2026) (cit. on pp. 15, 16, 19, 80).

[22] Richard Yanaky, Darcy Tyler, and Catherine Guastavino. «City Ditty: An Immersive Soundscape Sketchpad for Professionals of the Built Environment». In: *Applied Sciences* 13.3 (Jan. 27, 2023), p. 1611. ISSN: 2076-3417. DOI: 10.3390/app13031611. URL: https://www.mdpi.com/2076-3417/13/3/1611 (visited on 01/06/2026) (cit. on pp. 15, 19, 21, 82).

[23] Epic Games. «Unreal Engine». In: (). URL: https://it.wikipedia.org/wiki/Unreal_Engine (cit. on pp. 15, 80).

[24] Josep Llorca-Bofí, Christian Dreier, Jonas Heck, and Michael Vorländer. «Urban Sound Auralization and Visualization Framework—Case Study at IHTApark». In: *Sustainability* 14.4 (Feb. 10, 2022), p. 2026. ISSN: 2071-1050. DOI: 10.3390/su14042026. URL: https://www.mdpi.com/2071-1050/14/4/2026 (visited on 01/06/2026) (cit. on pp. 16, 19, 81).

[25] Giso Grimm. «TASCAR (Toolbox for Acoustic Scene Creation and Rendering)». In: (). URL: http://www.tascar.org (cit. on p. 16).

[26] Institute for Hearing Technology and RWTH Aachen University Acoustics (IHTA). «Virtual Acoustic». In: (). URL: https://www.virtualacoustics.org (cit. on pp. 16, 80).

[27] Armin Taghipour, Tessa Sievers, and Kurt Eggenschwiler. «Acoustic Comfort in Virtual Inner Yards with Various Building Facades». In: *International Journal of Environmental Research and Public Health* 16.2 (Jan. 16, 2019), p. 249. ISSN: 1660-4601. DOI: 10.3390/ijerph16020249. URL: https://www.mdpi.com/1660-4601/16/2/249 (visited on 01/06/2026) (cit. on pp. 16, 18, 19, 54).

[28] ODEON A/S. «ODEON». In: (). URL: https://odeon.dk (cit. on p. 16).

[29] Ivan Stepanov. «Procedural Acoustic and Volumetric Occlusion». URL: https://github.com/ibandurist (cit. on pp. 19, 55, 57, 58).

[30] Markus Berger and Ralf Bill. «Combining VR Visualization and Sonification for Immersive Exploration of Urban Noise Standards». In: *Multimodal Technologies and Interaction* 3.2 (May 13, 2019), p. 34. ISSN: 2414-4088. DOI: 10.3390/mti3020034. URL: https://www.mdpi.com/2414-4088/3/2/34 (visited on 01/06/2026) (cit. on p. 19).

[31] Yichun Lu, Yoshimi Hasegawa, Johann Kay Ann Tan, and Siu-Kit Lau. «Effect of audio-visual interaction on soundscape in the urban residential context: A virtual reality experiment». In: *Applied Acoustics* 192 (Apr. 2022), p. 108717. ISSN: 0003682X. DOI: `10.1016/j.apacoust.2022.108717`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0003682X22000913` (visited on 01/06/2026) (cit. on p. 19).

[32] Like Jiang, Massimiliano Masullo, Luigi Maffei, Fanyu Meng, and Michael Vorländer. «How do shared-street design and traffic restriction improve urban soundscape and human experience? —An online survey with virtual reality». In: *Building and Environment* 143 (Oct. 2018), pp. 318–328. ISSN: 03601323. DOI: `10.1016/j.buildenv.2018.07.005`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S036013231830413X` (visited on 01/06/2026) (cit. on p. 19).

[33] Chunyang Xu, Tin Oberman, Francesco Aletta, Huan Tong, and Jian Kang. «Ecological Validity of Immersive Virtual Reality (IVR) Techniques for the Perception of Urban Sound Environments». In: *Acoustics* 3.1 (Dec. 25, 2020), pp. 11–24. ISSN: 2624-599X. DOI: `10.3390/acoustics3010003`. URL: `https://www.mdpi.com/2624-599X/3/1/3` (visited on 12/15/2025) (cit. on p. 20).

[34] Özlem Gök Tokgöz. «The Effects of User Experience on the Questionnaire to Detection of the Soundscape of Historical Areas - Example of Eskisehir Factories Region». In: *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* 265.6 (Feb. 1, 2023), pp. 1708–1716. ISSN: 0736-2935. DOI: `10.3397/IN_2022_0243`. URL: `https://www.ingentaconnect.com/content/10.3397/IN_2022_0243` (visited on 03/02/2026) (cit. on p. 20).

[35] Raymond Lavoie, Kelley Main, Corey King, and Danielle King. «Virtual experience, real consequences: the potential negative emotional consequences of virtual reality gameplay». In: *Virtual Reality* 25.1 (Mar. 2021), pp. 69–81. ISSN: 1359-4338, 1434-9957. DOI: `10.1007/s10055-020-00440-y`. URL: `http://link.springer.com/10.1007/s10055-020-00440-y` (visited on 03/02/2026) (cit. on p. 20).

[36] Wen-Jie Tseng. «Understanding Physical Breakdowns in Virtual Reality». In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23: CHI Conference on Human Factors in Computing Systems. Hamburg Germany: ACM, Apr. 19, 2023, pp. 1–5. ISBN: 978-1-4503-9422-2. DOI: `10.1145/3544549.3577064`. URL: `https://dl.acm.org/doi/10.1145/3544549.3577064` (visited on 03/02/2026) (cit. on p. 20).

[37] Francesco Aletta, Jian Kang, Samuele Fuda, and Arianna Astolfi. «THE EFFECT OF WALKING SOUNDS FROM DIFFERENT WALKED-ON MATERIALS ON THE SOUNDSCAPE OF URBAN PARKS». In: *JOURNAL*

OF ENVIRONMENTAL ENGINEERING AND LANDSCAPE MANAGEMENT 24.3 (Aug. 19, 2016), pp. 165–175. ISSN: 1648-6897, 1822-4199. DOI: `10.3846/16486897.2016.1172076`. URL: `https://journals.vilniustech.lt/index.php/JEELM/article/view/1746` (visited on 02/28/2026) (cit. on pp. 21, 56).

[38] Massimiliano Masullo and Luigi Maffei. «The Multidisciplinary Integration of Knowledge, Approaches and Tools: Toward the Sensory Human Experience Centres». In: *Vibrations in Physical Systems* 33 (2022), p. 20221081. DOI: `10.21008/J.0860-6897.2022.1.08`. URL: `https://sin.put.poznan.pl/publications/details/i50815` (visited on 03/02/2026) (cit. on p. 21).

[39] Andrew Mitchell, Francesco Aletta, and Jian Kang. «How to analyse and represent quantitative soundscape data». In: *JASA Express Letters* 2.3 (Mar. 1, 2022), p. 037201. ISSN: 2691-1191. DOI: `10.1121/10.0009794`. URL: `https://pubs.aip.org/jel/article/2/3/037201/2845750/How-to-analyse-and-represent-quantitative` (visited on 03/05/2026) (cit. on p. 33).

[40] Dongxu Zhang, Mei Zhang, Daping Liu, and Jian Kang. «Sounds and sound preferences in Han Buddhist temples». In: *Building and Environment* 142 (Sept. 2018), pp. 58–69. ISSN: 03601323. DOI: `10.1016/j.buildenv.2018.06.012`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S036013231830355X` (visited on 12/19/2025) (cit. on p. 39).

[41] Blender Foundation. «Blender». In: (). URL: `https://en.wikipedia.org/wiki/Blender_(software)` (cit. on p. 49).

[42] domlysz. «BlenderGIS». In: (). URL: `https://github.com/domlysz/BlenderGIS` (cit. on p. 49).

[43] Poly Haven. «Poly Haven». In: (). URL: `https://polyhaven.com` (cit. on pp. 50, 51).

[44] BlenderKit. «BlenderKit». In: (). URL: `https://www.blenderkit.com` (cit. on p. 50).

[45] Trevor J. Cox Oliver Bones and William J. Davies. «Sound Categories: Category Formation and Evidence-Based Taxonomies». In: *Frontiers in Psychology* 9.1277 (2018), pp. 1–17 (cit. on p. 53).

[46] Splice. «Splice». In: (). URL: `https://splice.com` (cit. on p. 54).

[47] Reto Pieren, Thomas Bütler, and Kurt Heutschi. «Auralization of Accelerating Passenger Cars Using Spectral Modeling Synthesis». In: *Applied Sciences* 6.1 (Dec. 24, 2015), p. 5. ISSN: 2076-3417. DOI: `10.3390/app6010005`. URL: `https://www.mdpi.com/2076-3417/6/1/5` (visited on 02/28/2026) (cit. on p. 54).

[48] SuperCollider. «SuperCollider». In: (). URL: `https://supercollider.github.io` (cit. on p. 54).

[49] Firelight Technologies. «FMOD Studio». In: (). URL: `https://www.fmod.com` (cit. on pp. 54, 80).

[50] Josep Llorca-Bofi; Jonas Heck; Christian Dreier; Michael Vorlaender. «Urban background sound recordings for virtual acoustics under various weather conditions». In: *J. Acoust. Soc. Am. 155, A178* (2024). URL: `https://pubs.aip.org/asa/jasa/article/155/3_Supplement/A178/3301526/Urban-background-sound-recordings-for-virtual` (cit. on p. 56).

[51] Elden F Ray. «Applications of attenuations and reflections in ISO 9613-2, Acoustics - Attenuation of sound during propagation outdoors - Part 2: General method of calculation». In: (2004) (cit. on p. 56).

[52] JCW. «Absorption coefficients of common building materials and finishes». In: (). URL: `https://www.acoustic-supplies.com/absorption-coefficient-chart/` (cit. on p. 58).

[53] Fabrizio Lamberti, Andrea Bottino, Gianluca Paravati, and Oddera Fabrizio. «Un confronto tra diverse soluzioni per la visualizzazione tridimensionale in contesti industriali». In: () (cit. on p. 62).

[54] Lorenzo Dell'Anna. «Confronto tra valutazioni soggettive nella percezione del soundscape tra ambienti reali e loro riproduzioni in laboratorio». In: (2025) (cit. on p. 64).

[55] Marshall Long. «Human Perception and Reaction to Sound». In: *Architectural Acoustics*. Elsevier, 2014, pp. 81–127. ISBN: 978-0-12-398258-2. DOI: `10.1016/B978-0-12-398258-2.00003-9`. URL: `https://linkinghub.elsevier.com/retrieve/pii/B9780123982582000039` (visited on 03/14/2026).

[56] Jin You and Jin Yong Jeon. «Just noticeable difference of psychoacoustical and autocorrelation function parameters of refrigerator noise». In: ().

Appendix A

Soundscape Assessment Questionnaire

Question	**Scale**
Informed consent	I consent
Sound dominance	Traffic, Other, Humans, Nature *Likert 1-5*
Perceived affective quality	Chaotic, Annoying, Monotonous, Uneventful, Calm, Pleasant, Vibrant, Eventful *Likert 1-5*
Overall, how would you describe the present surrounding sound environment?	*Likert 1-5*
Overall, to what extent is the present surrounding sound environment appropriate to the present place?	*Likert 1-5*
Overall, how would you describe the present architectural/urban environment?	*Likert 1-5*
Overall, how would you describe the present surrounding sound environment?	*Likert 1-5*
Overall, to what extent is the present surrounding sound environment appropriate to the present place?	*Likert 1-5*

Overall, how would you describe the present architectural/urban environment?	*Likert 1-5*
What is the location you are assessing?	Rebaudengo, Via Bologna, Other
What is your sex?	M, F, Other / I'd rather not say
Age	*Paragraph*
During the assessment, did you notice any significant differences between the various listening positions (if there were any)?	*Paragraph*

Table A.1: In field questionnaire.

Question	**Scale**
Informed consent	I consent
Which listening setup are you using?	Headphones, Speakers
Which type of environment are you assessing?	Real environment, Virtual environment, Alternative virtual environment
Sound dominance	Traffic, Other, Humans, Nature *Likert 1-5*
Perceived affective quality	Chaotic, Annoying, Monotonous, Uneventful, Calm, Pleasant, Vibrant, Eventful *Likert 1-5*
Overall, how would you describe the present surrounding sound environment?	*Likert 1-5*
Overall, to what extent is the present surrounding sound environment appropriate to the present place?	*Likert 1-5*
Overall, how would you describe the present architectural/urban environment?	*Likert 1-5*
Overall, how would you describe the present surrounding sound environment?	*Likert 1-5*
Overall, to what extent is the present surrounding sound environment appropriate to the present place?	*Likert 1-5*

Overall, how would you describe the present architectural/urban environment?	*Likert 1-5*
I felt completely immersed in the virtual environment, as if I were physically present.	*Likert 1-5*
The experience was so involving that I completely forgot about the outside world.	*Likert 1-5*
The sounds and audio effects closely matched how I would perceive them in real life.	*Likert 1-5*
The visual environment significantly biased soundscape assessment.	*Likert 1-5*
What is the location you are assessing?	Rebaudengo, Via Bologna, Other
What is your sex?	M, F, Other / I'd rather not say
Age	*Paragraph*
During the assessment, did you notice any significant differences between the various listening positions (if there were any)?	*Paragraph*

Table A.2: In lab questionnaire.

Acknowledgements

Oggi si conclude uno di quei percorsi che ti guardi indietro e sembra essere passato solo qualche secondo.

Ho temporeggiato a scrivere queste ultime pagine perché avevo paura di guardarmi allo specchio, ma ho capito che certe volte bisogna lasciarsi andare.

Oggi ci sono molte persone che hanno fatto, fanno e ancora faranno parte della mia vita. Persone che mi hanno dato qualcosa, anche se poco, indispensabile per costruirmi come persona ed essere umano. Perché la conclusione di questo percorso non è solo una titolo, ma un atto di consapevolezza verso me stesso e il mondo che mi circonda.

Ringrazio innanzitutto i miei genitori, che mi hanno insegnato ad essere giusto e a non mollare mai. Mi avete sempre fatto seguire la mia strada senza mai forzarmi, con premura e amore. Siete l'esempio di cosa voglia dire dedizione e sacrificio. Questa laurea è vostra.

Voglio ringraziare mia sorella, mia complice in molte situazioni. Siamo le classiche persone che si rimproverano per errori che loro stesse fanno, e siamo pronti a tutto per difenderci a vicenda. Ti voglio bene, spero di essere un esempio per il tuo percorso che hai appena iniziato.

Voglio ringraziare nonno Nicola e nonna Antonietta, per quel fuoco dentro che mi porto da quando sono piccolo. Per le manifestazioni del Primo Maggio, per la resposabilità di prendere parte nelle cose senza farsi piegare ma rimanendo giusti. Questa è la laurea che non avete mai avuto ma che vi meritate più di chiunque altro.

Ringrazio nonno Gianpaolo e nonna Carla per l'amore e l'equilibrio che servono nelle cose. Grazie nonna per la fiducia, so che sei fiera di me. Ed io fiero di essere tuo nipote per sempre.

Ringrazio Giulia, per essere il Sole quando mi trasformo in Luna. Illumini le mie giornate grigie, e in questi anni ce ne sono stati parecchi. Ti guardo e sorrido,

perché vedo le immagini di un futuro insieme.

Ringrazio i PDL, la mia seconda famiglia. Questa tesi è per le serate da Carmen, le estati a Torino e i bei momenti insieme. Siamo la prova reale che una persona possa scegliere il suo destino, indipendentemente da dove provenga.

Ringrazio Matte per tutti i bei momenti passati insieme dalle elementari ad oggi. Ci siamo sempre fatti compagnia nel bene e nel male. So che ci saremo sempre per una chiacchierata. Grazie.

Ringrazio Ale per la cultura e la tua curiosità. Questa è per tutta la musica che ci siamo ascoltati e consigliati.

Ringrazio Ricky perché è rimasto dal primo giorno. È bello oggi poter chiudere insieme questo ciclo, so che ci saremo sempre l'uno per l'altro.

Ringrazio il Fanta, sono fortunato ad avere un gruppo come il nostro. Per quanto diversi ci rivediamo sempre. Grazie per tutti gli spunti, le schitarrate e i pomeriggi al Campus. La vostra amicizia fa la differenza sempre.

Ringrazio Giulio e Lorenzo, alle strade che ci dividono e ai nostri pensieri che ci tengono insieme. Quando siamo insieme è come se il tempo si fosse fermato a dieci anni fa. Non si cambia mai.

Ringrazio Ali e Simo. Ripenso alle mattina al freddo sul 29 verso futuri incerti. Ora ci vedo.

Alla Power Smile, chi fuori, chi dentro il tatami. Siete l'esempio di resilienza e di cosa voglia dire non mollare mai. Alle nostre trasferte e alle nostre risate. Insieme si fatica di meno.

A Fabio e Brigitta, alle chiacchierate in VSD2, ai drammi d'amore. Mi avete regalato i due anni più belli della mia vita. Grazie.

Ai Cinemini, cuore pulsante di un posto a tratti inospitale come il Politecnico. Ai giovedì alle panche e alle pause con il sole fuori dalle aule I. Grazie per aver portato il caldo a Torino.

A Lorenzo per l'innestimabile contributo in questa tesi. Senza di te sarebbe stato decisamente più triste.

Un ringraziamento va anche alla professoressa Shtrepi. Ci sono persone che hanno una vocazione e lei è una di quelle.

Ringrazio anche il professor Strada per il supporto.

Ringrazio Luca per l'ispirazione. La tua continua curiosità è di insegnamento.

Ringrazio Gigi con la sua energia trascinante. Sei la prova vivente di dedizione e volontà.

Ringrazio tutte le persone che hanno preso parte alla realizzazione di questa tesi, dai dottorandi a chi ha provato l'esperimento. Grazie a voi ora è realtà.

Ringrazio tutto il resto della mia famiglia e chiunque abbia contribuito un minimo a questi anni davvero intensi.

Infine ringrazio quel ragazzino che si alzava la mattina per farsi un'ora di bus per andare a scuola. Quei pomeriggi in camera con la chitarra, i sottili momenti di bilico tra l'arrendersi e farcela.

Questa laurea è per i ragazzi di zona come esempio di riscatto. Non permettete a nessuno di definirvi.

Non disunitevi, mai.
Il futuro è in mano ai deboli che si sono fatti coraggio.