VR-based Soundscape Evaluation

Auralising the Sound from Audio Rendering, Reflection Modelling to Source Synthesis in the Acoustic Environment

Chunyang Xu

A thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Institute for Environmental Design and Engineering, The Bartlett
University College London

April 22, 2022

I, Chunyang Xu, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work. I also declare that some parts of this thesis have been presented at the conferences and journals, listed as follows:

[1] C. Xu and J. Kang. Soundscape evaluation: Binaural or monaural? *The Journal of the Acoustical Society of America*, 145(5):3208–3217, 2019.

[2] C. Xu and J. Kang. Simplification of reflection orders in virtual soundscapes through a subjective evaluation. In *Proceedings of the 23rd International Congress on Acoustics, Aachen, Germany*, pages 6432–6437. Deutsche Gesellschaft für Akustik (DEGA), 2019.

[3] C. Xu, T. Oberman, F. Aletta, H. Tong, and J. Kang. Ecological validity of immersive virtual reality (IVR) techniques for the perception of urban sound environments. *Acoustics*, 3(1):11–24, 2021.

[4] C. Xu, H. Tong, and J. Kang. Perceived width evaluation on interpolated line sources in a virtual urban square. In *Proceedings of the International Conference on Immersive and 3D Audio, Bologna, Italy*, pages 1–4. IEEE, 2021.

Abstract

Soundscape has been growing as a research field associated with acoustics, urban planning, environmental psychology and other disciplines since it was first introduced in the 1960s. To assess soundscapes, subjective validation is frequently integrated with soundscape reproduction. However, the existing soundscape standards do not give clear reproduction specifications to recreate a virtual sound environment. Selecting appropriate audio rendering methods, simulating sound propagation, and synthesising non-point sound sources remain major challenges for researchers.

This thesis therefore attempts to give alternative or simplified strategies to reproduce a virtual sound environment by suggesting binaural or monaural audio renderings, reflection modelling during sound propagation, and less synthesis points of non-point sources. To solve these unclear issues, a systematic review of original studies first examines the ecological validity of immersive virtual reality in soundscape evaluation. Through recording and reproducing audio-visual stimuli of sound environments, participants give their subjective responses according to the structured questionnaires. Thus, different audio rendering, reflection modelling, and source synthesis methods are validated by subjective evaluation.

The results of this thesis reveal that a rational setup of VR techniques and evaluation methods will be a solid foundation for soundscape evaluation with reliable ecological validity. For soundscape audio rendering, the binaural rendering still dominates the soundscape evaluation compared with the monaural. For sound propagation with consideration of different reflection conditions, fewer orders can be employed during sound reflection to assess different kinds of sounds in outdoor sound environments through VR experiences. The VR experience combining both HMDs and Ambisonics will significantly strengthen our immersion at low orders. For non-point source synthesis, especially line sources, when adequate synthesis points reach the threshold of the minimum audible angle, human ears cannot distinguish the location of the synthesised sound sources in the horizontal plane, thus increasing immersion significantly. These minimum specifications and simplifications refine the understanding of soundscape reproduction, and the findings will be beneficial for researchers and engineers in determining appropriate audio rendering, sound propagation modelling, and non-point source synthesis strategies.

Acknowledgements

On this journey, I would like to particularly thank my principal supervisor, Prof. Jian Kang, who has been guiding me so positively and has always made me feel confident in my abilities to complete this PhD thesis. My sincere gratitude also goes to Dr Sung-Min Hong, my subsidiary supervisor, for his encouragement and insightful comments.

I would be most grateful for generous supports from Dr Francesco Aletta, Dr Tin Oberman, Dr Huan Tong, Dr Lee Young, Dr Tingting Yang, Mithuna Maran, Dr Gioia Fusaro, and Mengting Liu. I would like to extend sincere appreciation to all participates and team members who took the time to complete my tests and who contributed so thoroughly through their in-depth comments.

Finally, I would like to thank my parents and family for supporting me during the compilation of this thesis. This thesis is for my aunt and grandpa, and all I want to say is enveloped in the first and second verses of the 13th track in the album of 'E=MC^2'.

Impact Statement

Research into the effect of soundscapes on the built environment has contributed to a fundamental shift in the global understanding of noise control in urban spaces. The research on soundscape ecology has demonstrated a positive effect on human well-being over the past few decades.

In 2008, the International Organization for Standardization (ISO) began to prepare to develop a series of standards in soundscape. Of these standards, ISO 12913-2 highlights major methodologies for soundscape evaluation. For the section on soundscape reproduction, the standard gives generic methods and strategies, but these strategies are not well practised in soundscape research and management.

This thesis provides evidence on the selection of binaural or monaural audio rendering strategies, optimisation of reflection modelling, and synthesis of non-point like sources, which are not precisely specified in ISO 12913-2. For soundscape evaluation, engineers can select monaural or binaural audio rendering methods based on contextual and spatial information. When attempting to reproduce similar spatial layouts and acoustic conditions in this thesis, engineers can refer to the simplified conditions of reflection modelling and source synthesis under VR. Fewer reflections considering urban square areas and fewer synthesis points for non-point sources can be performed during auralisation.

The appropriate selection of these methods and parameters will ensure that soundscape reproduction is as consistent as possible with subjective perception. Therefore, the exploration and simplification from audio rendering, reflection modelling to source synthesis have positive implications not only for soundscape evaluation, but also for urban management and game development.

Contents

List of Figures

List of Tables

List of Acronyms

ANOVA Analysis of Variance

ASA Acoustical Society of America

BEM Boundary Element Method

CAD Computer-Aided Design

CAVE Cave Automatic Virtual Environment

CPU Central Processing Unit

DPA Data Protection Act

DWM Digital Waveguide Mesh

EDT Early Decay Time

EPSRC Engineering and Physical Science Research Council

ERC European Research Council

FDTD Finite-Difference Time-Domain

FOA First-Order Ambisonics

FPS First-Person Shooters

GDPR General Data Protection Regulation

GPU Graphics Processing Unit

GUI Graphical User Interface

HATS Head and Torso Simulator

HMD Head-Mounted Display

HOA Hihger-Order Ambisonics

HR Heart Rate

HRIR Head-Related Impulse Response

HRTF Head-Related Transfer Function

HRV Heart Rate Variability

ILD Interaural Level Difference

ISD Interaural Spectral Difference

ISO International Organization for Standardization

ITD Interaural Time Difference

IVR Immersive Virtual Reality

LCD Liquid Crystal Display

MAA Minimum Audible Angle

MAMA Minimum Audible Movement Angle

MRT Modified Rhyme Test

PI Place Illusion

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Psi Plausibility Illusion

RT Reverberation Time

S/N Signal-to-Noise

SAM Self-Assessment Manikin

SPL Sound Pressure Level

STI Speech Transmission Index

TS Technical Standard

TSST Trier Social Stress Test

TWA T-Wave Amplitude

UCL University College London

VBAP Vector Base Amplitude Panning

VR Virtual Reality

VRISE VR-Induced Symptoms and Effects

VRML Virtual Reality Modelling Language

WFAE World Forum for Acoustic Ecology

WFS Wave Field Synthesis

WHO World Health Organization

3DoF Three Degrees of Freedom

6DoF Six Degrees of Freedom

Chapter 1

Introduction

1.1 Background

We are surrounded and enveloped by sound all the time, and we hear all kinds of sounds, consciously or unconsciously. We hear music in an arena, vehicles on a roadway, voices in a meeting, birdsong in a park, and sometimes baby cries on a plane. Our two ears hear these sounds, and our brains image the whole world in which we receive the information, just as our eyes do. Sound is one of the most important mediums between humans and the world, and it is also of great significance to human health, emotional communication and cultural development.

In the ancient Greek period, Pythagoras, Aristotle and other sages had already started to discuss and study audible sounds. In recent centuries, as classical physics evolved into modern physics, the physical phenomena and laws of audible sounds have long ceased to be mysterious. With the advent of industrialisation and the post-industrial era, people are aware that noise and sound environments in cities significantly affect human well-being [1, 2]. Researchers are starting to pay more attention to how daily sounds affect us [3, 4], as physics does not fully explain the sensations that people feel in response to these sounds. Humans consist of extremely complex systems with countless degrees of freedom and variables. It is almost impossible to describe how we perceive and react to these sounds with simple physical formulas. Thus, sound in the environment has gradually become one of the subjects of research for a wide range of researchers. Various research areas have also emerged, for instance, room acoustics, which mainly studies indoor sound, and environmental acoustics, which mainly focuses on outdoor sound, from the perspective of human activity space.

As a type of human settlement, a city clusters various human activity spaces. In this regard, urban public spaces with different environmental contexts tend to integrate all people in a city. The connectivity and functionality of such spaces lead to complex communication and perception of sound by people. Compared to indoor spaces, outdoor spaces have more uncontrollable and time-variant environmental conditions and variables. For a given scale of outdoor space, spatial changes in the temperature distribution, for example, may affect the velocity of sound propagation, but this is very rare for an indoor space where the mean temperature gradients are small. In principle, we have been able to use various numerical models to objectively simulate or predict the acoustic performance of sound fields in real spaces. The exploration of urban sound environments is thereby uniquely challenging and fascinating.

Psychoacoustics is a branch of acoustics that primarily focuses on how humans perceive sounds, in contrast to research on the mechanical phenomena of wave propagation. Acoustic concepts, such as the frequency thresholds for hearing from 20 Hz to 20 kHz, A-weighting sound level, and equal-loudness contour, which are now commonly accepted, were explored by early acousticians and psychologists who found that sound pressure alone could not accurately describe the human perception of sound. Researchers in psychoacoustics frequently used pure tones or combinations of pure tones to test how humans respond to such sounds. With the progressive deepening of psychoacoustics, more environmental variables were introduced, moving beyond the limitation of the sole use of pure tones, and the concept of soundscapes emerged in the 1960s [3] as the times required. The soundscape is a concept related to how humans perceive an acoustic environment. The conceptualisation of a soundscape is also a paradigm shift from noise control to a resource of the acoustic environment [5, 6]. Soundscape research draws upon a strand of thinking in environmental psychology advocating interdisciplinary integration of the interaction between humans and nature [7]. Different from only reducing noise, the soundscape concept concentrates on improving the overall comfort associated with acoustic environments. Soundscape evaluation is a vital component in sound environment research and involves physical, psychological, social, cultural and architectural aspects. A well developed framework for soundscape indices based on psychological and physiological evaluation is still an ongoing goal for soundscape researchers.

With rapid progress in another area, virtual reality (VR), researchers are beginning to apply this technology to the field of urban planning and design in accordance with its particular characteristics, i.e., users can perceive and interact with the virtual environment. By creating a realistic-looking environment, developers can fully immerse users in this environment. Significantly since the advent of portable immersive virtual reality devices with higher computational speed, VR has boosted the technical revolution of sound environments. Lightweight wearable VR devices have gradually shifted from their early applications in space and the military to daily life. As the earliest studies of VR were closely associated with visual display solutions, most VR studies and applications focus on visual rendering with the leading technologies. Computer-aided design (CAD) tools for VR visualisation have gone through many technical developments. Audio is usually added as supplementary information in a VR experience without plausible or authentic reference to the virtual environment [8]. Auralisation is a procedure to reproduce the experience of acoustic environments for humans, including sampling, signal transform, calibration, and a series of audio techniques [8, 9]. The technologies of auralisation in VR are not as well developed as those of visualisation. Following further work on vision and auditory studies, researchers found that when visual and auditory stimuli are matched, visual perception is greatly enhanced [10].

In addition, sound field modelling in urban spaces is exceptionally complicated. A variety of acoustic phenomena, such as reflection, diffusion, and diffraction, need to be considered within a wide frequency band with reasonable system latency. Creating a reliable audio system is even more challenging than creating a visualisation system, given the technology currently available. This augmented VR experience has significant implications for architectural and environmental acoustics. One of the significant applications of VR-based soundscape reproduction is that people can perceive virtual urban environments in advance, and the results of VR-based studies can be considered a prediction to evaluate real sound environments. Researchers can conduct virtual walks to assign environmental contexts for a space that they aim to design or investigate.

VR-based soundscape evaluation provides a promising tool with enhanced immersion for citizens to improve the future creation, planning and regeneration of urban spaces. For the future direction of soundscape research, Kang et al. in 2016 [11] pointed out the following:

> More studies about the optimisation of soundscape data collection in an ecologically valid way that does not disturb the usual context of perceiving the acoustic environment are desirable, as there are relatively few studies dealing with these methodological aspects of soundscape research.

The minimum reporting requirements on data collection, sound recording methods, and human-computer interaction are still unclear under the current soundscape standards for VR experiences in which the researchers can explicitly control the experimental conditions for each user. The development of a comprehensive and rational standard of soundscape evaluation for off-site reproduction and evaluation requires a systematic and in-depth investigation.

1.2 Aims and objectives

Soundscape evaluation studies tackle more than just the minimisation of annoyance via noise control. Pleasant sounds in urban acoustic environments or sounds that we hear unconsciously, accompanied by human activities or natural events in our city, significantly contribute to how we perceive and judge the acoustic environment we live in. The holistic understanding achieved by recreating a virtual environment from the existing standards is still limited for soundscape researchers. Therefore, the aim of this thesis is to improve the understanding and gaps of auralisation for soundscape reproduction based on the existing standards. These research gaps are frequently associated with details in audio rendering strategies, reflection modelling, and sound source synthesis. This thesis will explore alternative or simplified strategies to recreate or reproduce a virtual sound environment for the future liveable environment by comparing subjective responses with objective parameters.

To assess these potential strategies, research objectives will follow:

1. Determine the ecological validity of existing technologies in soundscape evaluation and identify practical problems in the existing standards for audio rendering strategies, reflection modelling, and sound source synthesis.

2. Compare the subjective responses of different audio rendering methods, especially monaural and binaural static rendering, in soundscape evaluation.

3. Investigate whether fewer reflections can be applied to sounds in urban open spaces, especially urban squares.

4. Explore the effect of sound synthesis points, particularly for the perception of non-point like sources in urban squares on subjective perception of soundscape evaluation.

1.3 Research overview

To answer these questions, different research methods are employed to establish a deeper understanding of soundscape reproduction and evaluation methods in future soundscape ecology. Soundscape studies frequently employ hybrid approaches of numerical simulation, measurements, modelling and subjective tests, which are broadly in line with my research flow. When facing the issues of the evaluation of soundscapes in virtual reality, we use both objective parameters and subjective response data. The methods of soundscape reproduction and subjective evaluation are modified according to different environmental contexts, research questions and testing procedures of existing studies to ensure the validity of the research design.

To assess the ecological validity of IVR in soundscape evaluation, a systematic review of original studies examines the approaches and technologies utilised in sound environment research. Compared with a traditional review, a systematic review with specific inclusion and exclusion criteria can avoid subjective judgments during the search of the literature. To reproduce the listening experience, I first choose appropriate audio rendering methods. Through a field trip with audio-visual recordings, in situ measurements, and a subjective test, I compare the performance of binaural and monaural methods in soundscape evaluation. Then, I investigate the optimisation of auralisation in VR during sound propagation, especially how reflection orders affect our perception. I undertake a field trip to collect essential spatial and contextual information, reproduce the environments, and conduct a subjective test in VR to obtain subjective responses to sounds with different reflection orders. Finally, I examine the simulation of non-point sources in VR. I virtually design a public square through modelling software, auralise sounds with different events, synthesise audio-visual stimuli in VR and conduct a subjective test.

A series of subjective evaluation experiments verify the influence of acoustic or spatial parameters during auralisation. During this examination, deductive reasoning based on subjective evaluation is carried out in association with audio-visual reproduction, acoustic parameters, geospatial information and other factors. In this way, the methodological process combining subjective and objective results improves technical specifications of auralisation for VR-based soundscape evaluation.

1.4 Structure of the thesis

The diagram of the thesis structure is shown in Figure 1.1. The relationship among the chapters reflects a progressive and integrated flow. Chapters 2, 3 and 4 present a literature review, the methodologies, and an examination of the ecological validity of IVR in soundscape evaluation. Chapters 5, 6 and 7 perform deductive reasoning through subjective evaluation integrated into a route for auralisation mapping from audio rendering strategies, reflection modelling in environmental sound propagation simulation to the synthesis of non-point like sources, as mentioned above. Chapter 8 summarises the auralisation strategy in this thesis for quick-decision making in urban planning, game design, and future soundscape research.

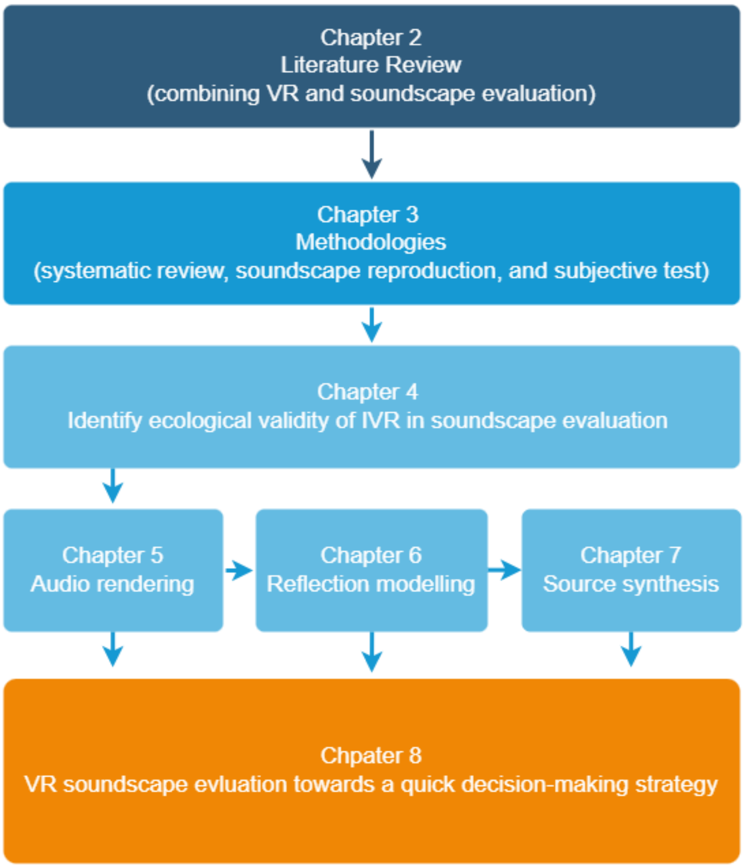


Figure 1.1: Diagram showing the overall research flow of this thesis.

Chapter 2 discusses the soundscape research development in general, including the definition, environmental contexts in soundscapes, and the broader implications of soundscapes; how we perceive, evaluate and improve a soundscape; VR technologies, including its origins, visual-audio rendering methods, and human-computer interaction in VR; and a discussion of combining VR and soundscape evaluation. An emphasis is placed on the auralisation methods, reflection modelling in environmental sound propagation simulation to the non-point like source synthesis.

Chapter 3 introduces the overall research methodological framework which is split into a systematic review, reproduction and subjective evaluation. This systematic review aims to collect secondary data to analyse the issues of the ecological validity of IVR. Soundscape reproduction primarily focuses on fieldwork, in situ measurements, area functional and contextual identification, modelling methods of urban environments, and sound auralisation. The subjective evaluation elaborates on the procedure for the identification of perceptual indicators, survey/questionnaire design, ethics review, participant recruitment and formal testing.

Chapter 4 identifies the ecological validity of IVR through a systematic review. Through a search protocol, commonly used approaches such as subjective response surveys, cognitive performance tests and physiological responses, can all contribute to the assessment of urban sound environment studies using IVR.

Chapter 5 explores the two soundscape audio rendering methods, i.e., binaural and monaural, and examines the performance of these two rendering methods in soundscape evaluation. This chapter compares the acoustic parameters in situ and the subjective ratings in twelve public spaces. Both binaural static and monaural rendering methods present good agreement on usual soundscape evaluation indicators, such as overall impression, acoustic comfort, pleasantness, annoyance, eventfulness, and loudness. Binaural static rendering would bring a significantly higher level in realism, reverberance and directionality than the monaural.

Chapter 6 investigates the reflection modelling in environmental sound propagation simulation and how the subjective response is influenced by reflection orders in urban squares. Four public squares are selected in London and they are virtually reproduced. VR significantly enhances our sense for realism and immersion at a very low reflection order. It is feasible to employ fewer reflection orders or even use attenuated direct sound during the auralisation when the urban squares is large enough to render sounds under VR experience with similar realism and immersion ratings.

Chapter 7 discusses a synthesis method on non-point like sound sources in an urban environment. This chapter investigates the implication of the synthesis of a line source in a virtual open urban space on the perceived width, immersion and distance. Due to the impact of COVID-19 from 2020, the vast majority of offline tests did not meet the requirements of local authorities for social distancing. The listening tests in this chapter could not be conducted in the audio laboratory. A remote listening test was carried through an online platform. The results reveal that width and immersion ratings are significantly affected by the synthesis points of the line source, and a small audible angle ($< 1^{\circ}$) will significantly improve perceived width and immersion ratings. The perceived distance is not relevant to the variation of synthesis points of the sound source under VR.

Chapter 8 summarises the research findings and answers the research questions in the thesis. This chapter also presents the implications for urban planning and design, game development, and soundscape evaluation. The limitations and future work of this thesis are discussed in the last section.

Chapter 2

Literature Review

2.1 Overview

This chapter reviews the development of soundscape research and VR technology, both as a multidisciplinary field and how these subjects are beginning to interact and fuse. A wide range of issues in studies on soundscapes and sound environments will be discussed. These issues are commonly associated with acoustic computational simulation, subjective perception, urban planning and design, software and hardware, and evaluation methods.

This chapter starts with a series of discussions of soundscape development and outreach. The systematic introduction of perception, evaluation, modelling, design and planning is of particular benefit in understanding how soundscapes can protect and improve the existing urban environment. Then, the literature review introduces VR technology, including its definition, limitations, audio and visual rendering, and interaction between humans and computers, which are considered fundamental aspects of VR implementation. Finally, a discussion of environmental psychology applications based on VR technology is given, and insights into combining both VR and soundscape evaluation are provided.

2.2 Sound environment and soundscape

2.2.1 Sound and soundscape

Being transient and invisible, sound is ubiquitous in our lives, and indeed, it is a significant component in urban environments and is closely related to human activities, urban ecology and public health. Urban noise is generally from human activities, and millions of people suffer from noise, which has long been known to be harmful and has negative effects on human health [1]. Researchers have found that noise can cause sleep disturbance [12], increased blood pressure, heart rate (HR) and other physiological indicators, as well as annoyance in both adults and children [13]. Due to these adverse effects, laws and regulations relating to noise control have been in place in various countries for the last century. The World Health Organization (WHO) Quality of Life Group in 1998 proposed six domains of quality of life, including the influence of noise [14]. Subsequently, the WHO in 1999 [15] published the guidelines on community noise, including discussions of measurements, negative effects, guideline values, and management of noise. In England, the Department for Environmental Food and Rural Affairs also conducted strategic noise mapping across the nation [16]. In a local region, the City of London presented a ten-year noise plan from 2016 to 2026 to minimise noise nuisances [17]. Therefore, at international, national and local levels, countless researchers, engineers, and policymakers have made great efforts to reduce noise in urban areas over the years. It is clear that there are substantial administrative and financial costs behind these policies.

Many studies [18, 19, 20] have found that reducing the sound level does not necessarily result in better acoustic comfort in urban public spaces. Soundscape researchers regard the sound environment, including noise, as a whole. A sound environment is a place including all sounds that can be heard by people in that place, and people commonly share this environment [6]. Sounds are commonly considered to be a perpetual and dynamic characteristic of all landscapes[21]. The soundscape term is derived from the landscape. By analogy to landscapes, many similar interpretations can be applied to soundscapes. The concept of soundscapes is also evoked by the sound environment, and the latter generally describes the physical sonic environment [5].

The soundscape concept was introduced in the 1960s by Michael Southworth [3]. In the 1970s, Raymond Murray Schafer, a Canadian musician and composer, conducted ground-breaking research on the World Soundscape Project [22]. The World Forum for Acoustic Ecology (WFAE) was founded in 1993, and aims to promote interdisciplinary research on acoustic ecology [5]. In 2006, Kang [5] provided a comprehensive review of soundscapes with different types of urban open spaces and evaluation methods. Since 2017, new index-based exploration methods have been used to search for a new framework to describe the soundscape by taking psychological, (psycho)acoustic, physiological, and contextual factors into account [23]. Researchers are seeking a more effective way for easy assessments of the soundscape and the impacts of noise in cities by developing soundscape indices.

The theoretical and practical development of the soundscape concept has been achieved by exploiting different research methodologies and fields of study, but these processes cannot be separated from the discussion of the centrality of soundscapes. The central role of a soundscape is the listener in a sound environment, so the personal and social meanings of environmental sounds [24] and the historical or aesthetic appreciation [25] should be taken into consideration. Activities in spaces, the expectation of a person with regard to the place, cultural background, and the person's prior knowledge [11] all affect how sounds are perceived and noticed by people in urban environments. With the evolution of urbanism, the increasing loss of natural sounds has weakened the connection between humans and nature [21]. These environmental contexts and factors are not normally taken into account in noise control. The management concept of a sound environment is gradually transferring from a 'waste' to a 'resource' [11]. As with other environmental resources, soundscapes of natural harmony have many social values, including but not limited to cultural, recreational, therapeutic, educational, research, artistic, and aesthetic values [21].

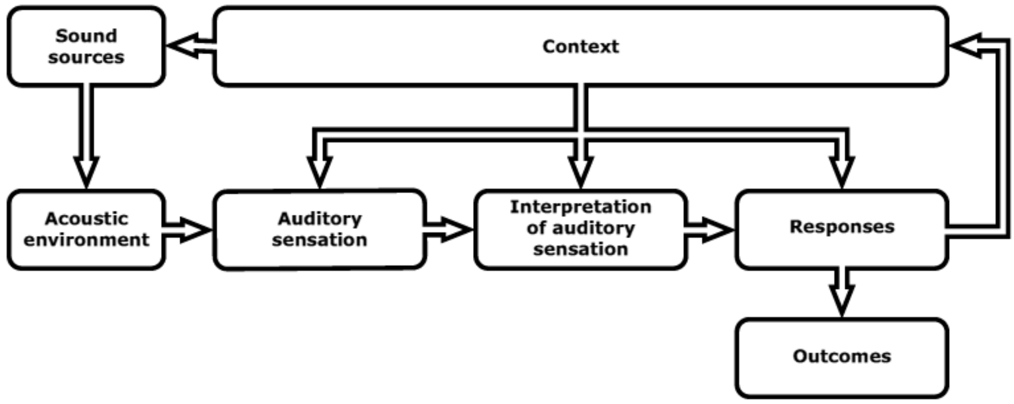


Figure 2.1: Schematic diagram of the perceptual construct of a soundscape. [1]

In relation to the perceptual construct of the soundscape, a schematic diagram [26] is shown in Figure 2.1. Both environmental contexts and human responses occupy an important position. They can influence each other, and they represent the interaction between people and the whole environment. Compared with reducing sound levels, perceptual results originate from people's subjective responses rather than a sound level metre. The contexts in the soundscape may involve spatial-temporal dynamics, environmental covariates [21], and socioeconomic factors [27]. For the built environment discipline, spatial and temporal dimensions must be very important yardsticks to assess the objects of the research. Everyone perceives the environment from a different space at every moment, so perception differs between the morning and the evening and between London and New York. Additionally, environmental covariates with spatial-temporal dynamics can also vary as a result of spatial-temporal dynamics on Earth, e.g., the illumination and temperature in the morning and the evening. Various socioeconomic factors have also developed as variates and covariates as history has progressed in different cities, countries, and continents. People perceive sound differently in these differing natural and social contexts.

The ecology of the acoustic environment and soundscape exhibits a wide variety of forms worldwide. Thus, in response to this widespread variability, re-

[1] Image republished in a thesis/dissertation with permission of the rights holder, British Standards Institution (Copyright licence in Appendix A).

searchers have been attempting to summarise the possible regularity and test their conjectures through deductive methods, and the soundscape is fascinating in this way.

2.2.2 Soundscape outreach

As an interdisciplinary field, the notion of a soundscape has been expanded to many areas of social development, as mentioned above. This integration, blending with the artistry, policies and research in these fields, has produced new avenues of thought. Furthermore, the soundscape concept has been recognised by more nations and regional organisations, and numerous research projects have been carried out to establish more comfortable, liveable and harmonious urban environments.

Many types of soundscape projects in Europe have greatly facilitated the development of urban sound environments, and they have also expanded the boundaries of the overall soundscape ecology, such as the *SONORUS* Project funded by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7 2007-2013, the *Soundscape support to health* project funded by the Swedish Foundation for Strategic Environmental Research [5, 11], and a series of projects funded by UK EPSRC (Engineering and Physical Science Research Council) and ERC (European Research Council). In the field of nature conservation, much of the soundscape research in the US has focused on national park protection and management [25, 28, 29]. Meanwhile, researchers from China, Singapore, Japan, Korea and other parts of the world have paid attention to common issues and local contexts of soundscape research.

The outreach of soundscape research has also focused on public health and quality of life. The positive health-related effects of soundscapes were systematically reviewed recently [2]. Rather than considering the negative effects of noise only, the positive aspects of soundscapes are also relevant to their positive health-related effects (e.g., enhanced restoration [6, 30, 31] and reduced stress-inducing mechanisms [32, 33]).

In the field of artistic creation, musicians, writers, historians and others have rethought and reshaped the soundscape in cities. Sound artists seek to investigate the potential ways in which sonic art can redraw boundaries in Belfast historically [34]. In 2014, Bennett and Rogers [35] endeavoured to blend the performance of street musicians and contemporary urban soundscapes through digital music tools. Researchers have likewise been able to determine the aural experience and culture through the thriving development of audio technology from the 1900s to the 1930s [36], and this process is considered to be the modernisation of the soundscape.

The soundscape concept has developed into an emerging science worth exploring and practising. The attractive, healthy, social, and aesthetic prospects will accelerate the development of soundscapes to perfect a more sustainable and ecological environment.

2.3 Soundscape perception and evaluation

Through the review of the conceptualisation and outreach of soundscapes, we realise that soundscapes are not only a gathering of urban sounds, but also closely associated with various aspects of our lives. There are three essential issues in sound environment research: how people listen to a sound environment, how people perceive and evaluate it, and how we can improve and protect soundscapes. At the moment, as soundscape designers or planners, we should be concerned with what we can improve through quantitative or qualitative analyses to establish a more harmonious soundscape for a public space, e.g., a square or street.

2.3.1 Psychoacoustics and spatial hearing

Hearing is a prominent social sense of humans, and it also delivers explicit information to human beings, similar to vision. When the movement of a person in a 3D space varies, the sound signals that people receive also change, resulting in different perception. The binaural auditory system has different acuities in terms of locating a sound source in the horizontal plane, vertical plane, and distance [37]. It is most sensitive to the horizontal plane and poorest in terms of distance perception [37]. Perceiving these differences is inseparable from an in-depth understanding of the auditory system and psychoacoustics. The peripheral auditory system of humans mainly includes the outer, middle, and inner ear. They all have different functions that have been widely discussed by psychoacousticians, e.g., Brian Moore [37] and Eberhard Zwicker [38]. Our eyes rest every day at bedtime, but our ears do not stop working even when we are asleep. Hearing, therefore, plays a vital role in providing warnings of danger for humans. In urban sound environments, the most common example of a warning is the beep of a car.

Establishing a mental image of urban environments is also considered an essential function of soundscapes [39]. Auditory attention can be distracted to another sound with enhanced saliency, and it is one of the basic principles for the utility of soundscapes to improve sonic environments. Irrelevant sound can draw our attention away from the current task towards the irregular event, and auditory attention distraction describes the degree to which the auditory stimuli violate phenomenological, behavioural and electrophysiological expectations [40]. The importance of auditory attention is emphasised in the perception of complex acoustic environments [41]. Auditory scene analysis is the great ability of humans to decompose complex sounds from different sources into individual streams and components [42]. Saliency-based models were initially designed for speech processing and have high resolution in the time and frequency domains [43, 44] based on the structure of visual saliency models relying on the spatial domain only. Simplified models with higher computational speed have been proposed for soundscape perception [41, 45]. The temporal and spectral contrasts are processed with long timescales (a signal of 2 minutes used in [41]), providing enough auditory experience for an assumed scene.

Masking is also an essential concept in soundscape research to improve sound environments. Pollack [46] gave a definition of masking in 1975, stating that 'masking is typically defined in terms of the threshold change in signal level resulting from the presence of a neighboring noise'. Informational masking and energetic masking are two typical methods in sound masking. Informational masking does not generate sounds with higher sound pressure levels, so it has a more significant meaning in soundscape research. The definition of informational masking was given by Pollack [46] in 1975:

> Informational masking is defined in terms of the threshold change in statistical structure resulting from the presence of a neighboring signal of the same amplitude.

Effective masking is related to personal sound preferences in sound environments. To select appropriate masks in road traffic noise and construction noise, Jeon et al. [47] chose a variety of masking sounds in an auditory experiment in 2010. The preference for bird sounds in a forest and a port shows a significant gap, and different forms of water sounds reveal variable subjective preferences. Moreover, the significance of urban morphology on birdsong loudness was reported by Hao et al. [48] in 2015, and they stated that the visibility of green areas shows a positive correlation with birdsong loudness. Magnitude estimation is considered an essential approach in measuring the audibility of sounds, and this method estimates the psychological magnitudes of a series of stimuli by assigning different numbers from a subject [49, 50]. This method has been applied to the estimation of perceived loudness [49], audibility [51] and other psychoacoustic indicators. For the estimation of audibility, Nilsson et al. [51] reported that the subject provides a direct numerical estimate of a series of sound samples, and if more than 50% of the magnitude estimates for an individual sound are zero, the corresponding sound is defined as inaudible.

Spatial hearing is an important capacity to identify the direction of sound and localise the source. Over the past hundred years, there have also been important studies and conclusions to interpret how the auditory system perceives sound in 3D spaces. Lord Rayleigh in 1907 proposed the Duplex theory of sound localisation. The Duplex theory interprets the the ability of humans to localise the sound by the interaural time difference (ITD) and interaural level difference (ILD) [37]. The ITD is the time difference for a sound to reach both ears. The ILD is the intensity difference between the two ears for the same sound [52]. The minimum value of ITD that can be perceived by humans is 10 μs and the minimum value of ILD is 0.5 dB [53]. In azimuth in the horizontal plane, the angle at which the human ears can localise two sound sources is called the minimum audible angle (MAA). The MAA for static sound sources is at approximately 1°, and when the sound sources is moving, this angle is about 2-5°, also called the minimum audible movement angle (MAMA) [54].

One of the limitations of the Duplex theory is the incapability of localising a sound source directly in front of and behind the listener. The sound source outside the horizontal plane around the head can not be fully interpreted by the Duplex theory, and the pinna filtering effect emerged. When the elevation of the sound source changes, the shape of human pinnae has a significant impact on localisation [55]. The interaural spectral differences (ISD) thus describes the difference in frequency distribution between the two ears due to the shape of the torso, head and pinnae [38]. The sound is filtered through these parts of the body affecting the sound arriving in each ear differently, and this filtering is referred to as a head-related transfer function (HRTF) [52]. A deeper understanding of these spatial hearing insights will contribute to a better comprehension of how people perceive sound in urban environments and the implications for applications and research questions in possible VR scenarios.

2.3.2 Soundscape evaluation development

The urban sound environment is closely associated with the well-being of local residents and bears the historical, cultural and social imprints of a city. City managers, policymakers, and local residents all seek to determine the possible implications of this association. The process of quantitatively or qualitatively exploring these implications is called soundscape evaluation or soundscape assessment.

Soundscape evaluation is still a complicated process with various hypotheses and debates. These hypotheses and debates derive from the complexity of real-world environments and human perception. This has also given rise to a diversity of evaluation methods, and this process is a reinterpretation of the dichotomy between inductive and deductive reasoning [56]. The main role of the inductive reasoning in soundscape research is to develop soundscape theory through extensive observations. The methods used in such observations generally demonstrate the researchers' reflections on the association between sound itself and human subjective and objective responses. Early soundscape researchers used generalisation, prediction, argument from analogy, and other approaches to draw some essential inferences, e.g., the conceptualisation of soundscapes in the 1970s [4, 57]. These inferences and predictions were derived from established psychoacoustic findings and urban planning theories. Since inductive reasoning cannot ensure the precision of the summaries, the role of deductive reasoning is to collect data to test the hypothesis based on existing theories. Thus, the findings in the soundscape studies through deductive reasoning can be falsified. More deductive methods are gradually merging with inductive methods in soundscape research. Researchers seek to assess soundscapes by integrating environmental psychology, ecology and acoustics, while using a combination of subjective and objective methods [7].

Researchers are attempting to follow a path of standardisation in soundscape evaluation. Local authorities, regional unions and international organisations are also becoming involved in soundscape evaluation and providing appropriate guidance. Methodologies for soundscape preference assessment have been discussed over the years, and various methods have been proposed. In recent years, the conceptualisation, development and refinement of soundscape standards have been progressively undertaken worldwide. In 2008, a working group was established by the International Organization for Standardization (ISO) in preparation for the development of a series of soundscape standards [56]. Researchers and engineers from around the world worked together, and this series is what is now known as ISO 12913-1:2014 [26], ISO/TS 12913-2:2018 [58], and ISO/TS 12913-3:2019 [59]. These three sequential steps represent the conceptualisation and definition of soundscape, evaluation and measurement methods, and analyses of results. This standard-setting process, on the one hand, provided a valuable opportunity for a broad discussion across the soundscape field and, on the other hand, offered a quick guide for subsequent researchers, consultants and policymakers in soundscape evaluation.

ISO/TS 15666: 2003 [60] specified a method of social and socioacoustic surveys in assessments of residential noise annoyance. The noise-induced annoyance normally refers to someone's individual adverse reaction to noise, and the reaction can be presented in various ways including dissatisfaction, bother, annoyance and disturbance [60]. This specification has been adopted by many engineering projects and numerous studies in noise annoyance and exposure assessments. However, soundscape researchers noticed that significant situational differences exist among the noise annoyance and soundscape preference measurements [61]. They pointed out that annoyance measurements typically consider a single annoyance outcome for residents who live at a certain place suffering from outdoor noise, but soundscape measurements take place under situations in which multiple activities, types of sound sources, and a wide range of levels of sound exist without specified assumptions regarding aggregation of perception. In other words, soundscape measurements require further consideration of perceived indicators, location selection, sound source types, scene activities and spatial functions.

2.3.3 Typical evaluation methods

Currently, there are three major approaches broadly employed in soundscape measurements, i.e., soundwalks, in situ surveys, and laboratory-based auditory experiments. In terms of where participants conduct evaluations, soundwalks may occur in situ, online or in the laboratory, and surveys usually occur in situ or online. All three methods have their own characteristics, and they are often combined with each other as a hybrid approach in assessments.

A soundwalk is a method to obtain a person's sensations through a participatory group walk led by a moderator following a pre-defined walking route and using a structured protocol [62]. The route of a soundwalk should conventionally include various places with different public functions and representative acoustic scenarios. During the walk, participants carefully listen to the environment, and they are asked to answer certain questions at each selected location by filling out structured questionnaires with quantitative or qualitative questions. Pioneering research on soundwalks started in the 1970s [4, 57], and a practical method stressing listening experience in urban sound environments was developed in these studies. Some research projects have employed in situ soundwalks as a tool to interpret and evaluate soundscapes [62, 63], and some researchers have also conducted virtual soundwalks in the laboratory [64].

A survey collects specific data from a group of people through a well-structured list of questions or a questionnaire. Setting a specific research scope and a series of questions, the researchers can collect large amounts of personal responses to these questions without having to go through a complicated process of interpretation. In the soundscape domain, in situ surveys coexist with online surveys, and both of them are regarded as the most widely used investigative tools. In 2013, Liu et al. [65] collected 580 pieces of data through in situ surveys in five public city parks in Xiamen, China. Through their analyses of the survey data, they found that the landscape effects on overall soundscape preference are more affected by artificial sounds than natural sounds. In relation to online methods, Jiang et al. [66] conducted an online survey in 2018 to explore how street design and traffic restrictions improve urban soundscapes. The results of their study indicate that shared-street design makes urban soundscapes calmer and that traffic restrictions make soundscapes more pleasant. The ability to survey large cross-country soundscape samples [67, 68, 69] is also a key advantage of the survey method. A large number of samples with high statistical power in surveys also provide a benchmark for other research methods in soundscape evaluation.

Laboratory-based auditory experiments (also called off-site experiments) are widely used in soundscape evaluation and typically refer to the situation in which a questionnaire-like evaluation of some visual-auditory features of soundscapes is carried out in a laboratory environment with stable conditions of light, sound, temperature, humidity and other parameters [5, 58]. The participants can receive similar visual, auditory and other sense stimuli during the laboratory-based experiment and give their subjective evaluation based on a well-structured questionnaire or an interview. Compared with in situ studies, studies in the laboratory can control more explicit variables during the evaluation [70, 71, 72, 73, 74], e.g., same sounds of equal time duration. A cross-national study in France, Korea, and Sweden was conducted through laboratory experiments in 2018 [75]. The researchers in that study concluded that for cross-national studies, it is almost impossible to use identical equipment and identical laboratory conditions simultaneously, and without strict compliance under the same test conditions, there is a potential risk that the results from the different linguistic versions will not be comparable. The results of cross-national studies are closely related to participants' textual understanding of the questionnaires. These cross-national studies will not be limited to in situ questionnaires and soundwalks, but will also include the listening test or virtual soundwalks in the laboratory. This represents a challenge that cannot be ignored during international data collection of soundscape experiments. The translation process is generally referred to different nation's soundscape protocols, noise evaluation standards, and the discussion through focus groups and panels of experts in soundscape studies to ensure semantic consistency of different language versions. [75, 76].

2.3.4 Perceptual indicators of soundscape evaluation

How to construct reasonable and effective questions with indicators to describe people's feelings regarding soundscapes is an issue in the design of a structured survey or questionnaire. Both holistic hearing and descriptive hearing differently affect how people process sound while listening in urban environments [77]. The former without semantic processing considers the soundscape as a whole, and the latter aims to identify sound events or sources. According to the cognitive perspective, it is meaningful to identify how people perceive sound events or sources and to refine researchers' understanding of the full picture of soundscape perception. The semantic differential method is a useful tool to characterise perceptual attributes (e.g., refs [5, 18, 78, 79]), and its basic concept is to scale the connotative meanings of sounds with different perceived attributes including comfort, pleasantness, annoyance, eventfulness, etc. Hence, semantic differential analysis is commonly used in soundscape research to link people's sensations and specific sound events or sources at both linguistic and psychological levels [79, 80, 81]. In 2010, Kang and Zhang [79] investigated the performance of 18 pairs of indicators, such as 'comfort—discomfort', 'pleasant—unpleasant', 'directional—everywhere', and 'natural—artificial', and according to the results, it was possible to identify some significant factors, including relaxation, communication, spatiality and dynamics in urban soundscapes through semantic differential analysis.

On the basis of the discussion of the semantic differential method, the results in these studies also provide a valuable benchmark for the proposal of standardised soundscape evaluation indicators. To use these perceptual indicators to portray a visual image of the soundscape data, a study to identify the underlying dimensions of soundscapes was done by Axelsson et al. [82] in 2010, and a model widely used in later studies (e.g., refs [83, 84, 85, 86]) was established, i.e., a system where pleasantness and eventfulness are orthogonal.

There has been much discussion [85, 87, 88, 89] of soundscape dimensions or classification in terms of emotional or affective representations of soundscapes and dimensions. Emotional responses can be measured by physiological and psychological assessments. The physiological measuring, e.g., heart rate variability (HRV) [90, 91, 92], describes the causation of specific behaviours evoked by the environmental sounds [93]. For psychological measuring, some studies using graphical representation of reduced dimensional variables have explored the dimensions of soundscape classification. The Self-Assessment Manikin (SAM) is a picture-oriented questionnaire assessment that directly measures an emotional response to a wide variety of stimuli [94]. The results obtained from such an assessment will provide a understanding of how soundscapes affect people's emotions in different psychological dimensions, i.e., valence, arousal and dominance [95]. The psychological approach converts underlying mechanisms into explicit behavioural manifestations, arising from the perception of sound environments. [93].

The quantification of these perceptual indicators into one or several variables that involve simple expressions of the whole soundscape with psychological, physiological, socioeconomic, social relations, demographic, objective sound level and other indicators together is a vital goal pursued by soundscape researchers.

2.3.5 Acoustic modelling in sound environment

To accurately evaluate acoustic environments or soundscapes, it is important to understand and measure the sound environment in the physical space. For any urban space with sounds, physical oscillations and waves are still the foundation of our discussion. Measurement, quantification and modelling methods for these oscillations and waves have been developed by numerous scientists and engineers through the language of mathematics over the centuries. The physical measurements of sounds are based on wave properties to calculate or simulate objective parameters, e.g., the sound pressure level (SPL). In the A-weighted SPL, the modification of the equal-loudness contour is taken into consideration. Several parameters of psychoacoustics are often discussed in sound perception, such as loudness, roughness, fluctuation strength, pitch strength, and other hearing related parameters [37, 38]. For simple boundary conditions and time-invariant spaces in which the medium is homogeneous, it is possible to obtain an analytical solution of the sound field based on the wave function. In most urban spaces with complex boundary conditions, analytical solutions cannot be solved, and through computer simulation, numerical solutions can be calculated to obtain a single value at a point or render 2D or 3D sound field images.

Acoustic environment modelling is an essential tool to predict, improve and create urban sound environments. The computer-based modelling of room acoustics began in the 1960s [96, 97]. With the rapid development of computers, various modelling approaches, e.g., geometrical and wave-based methods, accompanied by corresponding hypotheses and limitations have become widely applied in the simulation of different urban spaces.

Geometrical acoustic methods have been developed over the years, and they have been extensively used in noise mapping and room acoustic prediction for engineering purposes. The basic concepts of geometrical acoustics are similar to geometrical optics, and two conventional methods of geometrical acoustics are the image source method and ray tracing [5]. Image source methods consider boundaries to be mirrors and uses image source sounds to replace reflected sounds. This means that the receiver sound is the sum of the real source and all image sources. Image source methods require validity tests and realistic auralisation [98], because the physical properties of waves are ignored. Validity tests are designed to compare the numerical simulation results with the measured data and to calibrate the auralised sound [99]. The energy-based model is an approach of considering the dissipation of sound energy in three-dimensional space, for example, the dissipation of the energy of a point source in a free field along a sphere. For low-frequency sound, the wavelengths and spatial scales can be comparable. The wave properties of sound, e.g., diffraction, are not accounted for in the energy-based methods. Moreover, boundary conditions with complex diffuse structures are rarely reflected in geometrical models. An energy-based image source method was applied to simulate the sound fields of urban squares in 2005 [100]. The brief mathematical expression of energy-based image source methods provides an available and fast approach to predict the sound pressure level reduction, reverberation time (RT) and early decay time (EDT) in public squares, and the size of squares, ground absorption and boundary conditions can be incorporated into image source methods [101]. The direct sound and early reflections can be precisely predicted. Ray tracing generates a spread of rays reflected around a space [97] and has been applied to model sound environments, such as calculating the sound distribution in interconnected streets [5]. For narrow spaces and low frequencies, geometrical methods are still limited and further acoustic analyses are needed.

Wave-based methods generated from classical wave theory solve physical equations to obtain sound pressure and particle velocity. Principally, sound wave properties are included in wave-based methods, and this method is suitable for the analysis of narrow spaces and low frequencies. Wave-based equivalent source methods expand sound sources on symmetrical boundaries as image sources, and consider the wave properties when dealing with the sound propagation. An equivalent source approach was investigated to study sound propagation in a simple 2D model of city canyons [102]. This simplified model is still computationally heavy, although it is less computationally intensive than the standard finite-difference time-domain (FDTD) method and boundary element method (BEM).

Numerical modelling methods were initially extensively applied in room acoustics. Focusing on the properties of low frequency, Botteldooren [103] in 1995 used a numerical time-domain simulation based on FDTD approximation to solve room acoustic problems. Moreover, graphics processing units (GPUs) with parallel computation capabilities have enhanced the computational speed of real-time 3D FDTD simulation of room acoustics [104]. A central processing unit (CPU) runs the process of signal input-output, filtering and underlying systems, and the FDTD iteration is processed by a GPU. Thus, auralisation with arbitrary 3D geometries can be achieved in real-time through different algorithms and hardware. In addition, a diffusion field approach generated from the 3D diffusion equation implemented in a finite-difference scheme was developed in 2012 to simulate urban streets, and the results showed good agreement with those of the FDTD wave-based method and geometrical acoustic method [105].

The digital waveguide mesh (DWM) is also an active research area in numerical sound simulation and auralisation [106]. It can discretise space into different dimensions, and according to the d'Alembert solution of the acoustic wave equation for calculating the sound pressure of discrete points, sound propagation can be simulated on a mesh of different dimensions, e.g., 1D, 2D and 3D [107]. Based on the DWM, a simulation tool called RoomWeaver was developed in 2004 that can generate virtual room impulse responses and achieve auralisation based on this numerical simulation [108]. This method is widely used in room acoustics to study boundary conditions [109], impulse response generation [110, 111], and real-time auralisation [112].

These geometric, numerical and hybrid simulation methods offer researchers tools that are not influenced by outdoor ambient noise with a low signal-to-noise (S/N) ratio or outdoor activities not controlled by the researchers. Researchers should choose appropriate methods with consideration of their time cost, computational cost, accuracy and applicability for the implementation of spatial sound field modelling based on practical concerns, which is also an important process in the understanding of the physical aspects of the soundscape.

2.3.6 Soundscape design and planning

Through soundscape evaluation and acoustic modelling, city managers should think about how to improve and protect existing sound environments according to the results. At this point, policymakers, designers, engineers, and researchers can work collaboratively to explore an effective strategy with consideration of aesthetic design.

Both soundscape design and noise control are essential components in urban sound environment management, as evidenced by the fact that the Welsh government issued the noise and soundscape action plan 2018 to 2023 [113] juxtaposing these two terms. These terms sometimes run in parallel and sometimes intersect with each other. People have no patience for highly severe noise, and noise can also have a negative physiological and psychological effect on people in the short or long term. Decision makers certainly need to improve the noise immediately under these conditions. On the other hand, the improvement and enhancement of the overall sound environment cannot be achieved without an in-depth understanding and knowledge of the local soundscapes. They are both interlinked and promote each other's development.

For specific urban design, water installations frequently serve as part of landscapes as well as soundscapes. Two cases in Sheffield are often reported as successful and representative examples of active exploration of soundscape design, the Peace Gardens [5, 79, 114, 115] and Sheaf Square [116, 117]. The Peace Gardens is located in the city centre of Sheffield. The area around the central fountain is paved, and the grassy areas with additional benches are distributed between the centre and the outer circle. A fountain is located at the centre of the square, and cascade fountains are located on the outer edge. Its unique water features actively improve acoustic comfort in the commercial and official centre in Sheffield and attract many visitors and residents [5]. Sheaf Square is another good example, and it is situated outside the Sheffield train station, serving as a stunning city entrance and transport hub space. It won the coveted Project of the Year Award in the 2006

National Rail Awards through the use of a dramatic cascade of water to give the plaza an exciting atmosphere with consideration of sound and light [118]. The water feature of the long cascade acts as an acoustic signpost to the train station by imitating the rhythms of a running train [116].

Aside from waterscapes, other architectural and landscape installations have been introduced by researchers. The gabion wall, benches with a loudspeaker system, and ring-shaped chairs with a loudspeaker system have been placed in the public space Nauener Platz in Berlin to improve soundscape perception [6, 119, 120]. These installations and re-arrangement of public space were well accepted by local residents one year after the completion of the redesigned space. In 2018, researchers utilised 3D printed materials to design urban furniture and investigated the acoustic suitability of the printing materials and the soundscape impact of urban furniture on human perception [121]. They stated that the proposed urban furniture positively influences psychoacoustic perception in urban sound environments, which also confirmed the feasibility of this new form of soundscape installation in soundscape design and improvement.

In addition, Fowler [122] in 2013 introduced a soundscape design strategy for landscape architectural praxis through the analysis and examination of examples from three landscape architecture design studios taught at RMIT University, Melbourne Australia. He stated one of the difficulties:

> Perhaps a difficulty in using the theory of soundscape as the basis for generating landscape architecture is the reliance on particular visual modes of communication and dissemination within the field of design.

He also mentioned that a great number of verbal presentations are required to communicate the soundscape in architecture design because the delicate acoustic qualities of the design are obscured by the focus on the visual presentations. One of his affirmative comments is to use onomatopoeia in the visual language to interpret the functions of the design in which it can better communicate the auditory qualities of the design intents. Therefore, one question posed here is whether there is a multimedia way of efficiently presenting the visual and aural expressions needed to design a soundscape blended with a landscape architecture.

2.4 VR technology

VR serves as a multimedia means for conveying design concepts, and the further introduction of the possibilities of applying VR for soundscape assessments and design are discussed in terms of VR definition and development, visualisation and displays, auralisation and spatial audio, and interaction.

2.4.1 What is VR?

VR is a medium based on various software and hardware to create a realistic-looking world [123]. This virtual world is dynamic, and users can interact with it. The users can respond to the system based on the corresponding VR output. With the increasing development of theory and technology, VR technologies have been extensively applied in various industries, including entertainment, education, games, healthcare, and the military. As a multidisciplinary technology field, VR has gradually been utilised in those fields and beyond to achieve significant successes.

Pioneering work on VR was done by Morton Heilig in the 1950s, and he developed a multisensory machine called Sensorama that integrated a stereoscopic display, stereo speakers, fans, and a vibrating chair to provide an individually immersive theatre experience [124, 125]. Ivan Sutherland [126] in 1965 described 'the ultimate display' that required an interaction system between the users and a computer to give the feeling of a realistic world. Myron Krueger [127] proposed 'artificial reality' in the 1980s. This is a term that was discussed and used in the early stages of VR. Currently, the term 'virtual reality' has generally been more accepted by academia and industry since the late 1980s.

To simulate objects and their states in the real world in a computer, scientists need a relationship map to transition from the environment to digital spaces. The mapping process is also the process of modelling the virtual environment. Virtual Reality Modelling Language (VRML) 1.0 was first conceptualised at the World

Wide Web conference in 1994 [128]. VRML is a programming language for 3D interactive vector graphics. The evolution of this kind of standard (e.g., VRML 2.0 and X3D) has enabled the development of a common and standardised language for VR modelling. The sources of simulation data are typically from actual measurements (e.g., a panoramic camera) or artificial constructions (e.g., AutoCAD).

Burdea and Coiffet [123] summarised three essential features in VR techniques, i.e., immersion, interaction and imagination. These three features combine advances in interdisciplinary research in fields as diverse as control engineering, computer science, psychology, and electronics. Immersion is a term widely used in the entertainment industry. Its definition usually varies with the context. Immersion is normally accompanied by sensory engagement and engrossment to describe the reproduction degree of participatory experience [129]. Based on the different degrees of freedom of VR systems, 'immersive' VR or IVR has a more narrow interpretation in this thesis, i.e., three degrees of freedom or more of audio-visual representations. Other 2D or static representations are considered 'non-immersive'. In addition, VR techniques are centred on human perception and aim to provide a reliable practical tool for research in these disciplines. VR applications initially focused on extreme or complex industrial, military or space scenarios [130, 131]. Other domains, such as healthcare, education, planning and design, entertainment, and training, have also built their own VR applications to solve a wide range of practical issues.

There are some health and safety implications of VR related to its side effects. In 1993, Nickols et al. [132] divided the source of the health and safety implications of VR into four factors, i.e., VR system (e.g., temporal/spatial resolution and visual quality), virtual environment (e.g., permitted movement and interactivity), task characteristics (e.g., duration of exposure and user training), and user characteristics (e.g., age, sex, personality and health). Eye strain, blurred vision, nausea, oculomotor function and disorientation are frequently reported symptoms in virtual environments [133, 134, 135, 136], and a new term was also coined, i.e., VR-induced symptoms and effects (VRISE). For most people, these symptoms are mild and short-lived, and better prior guidance and termination mechanisms should be identified via VR experiments [133].

2.4.2 Visualisation and visual display for VR

VR is tightly associated with the development of virtual visual displays. The visual aspect of VR is an important channel through which users receive information from the virtual environment. Much of the early VR research focused on visual renderings [124, 126], although they also emphasised multisensory perception. Based on the current 3D computer graphics languages, e.g., X3D, generic modelling data allow for good compatibility and transfer among different software programs, such as SketchUp, AutoCAD, and 3dsMax. These models are graphically rendered by a computer and finally presented to our eyes through a display device. In recent decades, a diverse range of visual display methods for VR have been explored, and used in high-tech applications, such as medicine, training and education. The state-of-the-art display types that VR is still employing today include head-mounted displays (HMDs), cave automatic virtual environments (CAVEs), desktop displays, and other customised patterns.

An HMD is a device with two video displays corresponding to both eyes, integrated into a helmet or a pair of glasses worn on the head. Both displays are composed of modulated light sources with drive electronics to create the illusion of depth for both eyes [8]. To ensure users' gaze point based on their head position, the visual system must be coupled [137]. In other words, the visual presentations of the two lenses have certain phase and spatial differences depending on the position and motion of both eyes. Consequently, a motion tracker to collect these positions and motions is always used in HMDs. The initial HMD prototype was pioneered by Ivan Sutherland in the 1960s [138]. Fisher et al. [130] introduced a head-mounted, wide-angle, stereoscopic, lightweight display system in 1986 to simulate complex operational tasks by operators in space stations. The HMD headset they used was not significantly different in appearance from those still in use today, as shown in Fig. 2.2.

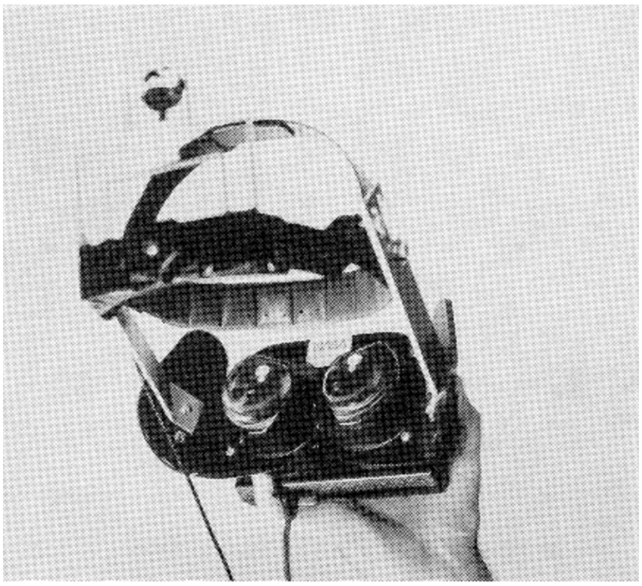


Figure 2.2: An HMD unit with wide-angle optics and LCD displays in the 1980s.[2]

As state-of-the-art display systems, HMDs are experiencing numerous theoretical innovations and technical challenges. One of the significant challenges of HMDs was stated by Rolland and Hua [137] in 2005:

> An ergonomically designed headband that properly secures the display on the user's head is perhaps one of the biggest challenges for designers of HMDs, a challenge that is intrinsically coupled to that associated with the optical design.

The wide-angle optical presentation at a close vision distance and the comfort of the wearable device are constant challenges for HMD developers.

The CAVE was invented in 1992 [139, 140]. It is a surrounding video display system in which the walls and the floor are projection screens. Users wearing 3D glasses stand in the space surrounded by these projection screens. An electromagnetic or infrared tracing system is used in the CAVE for the adjustment of binocular and binaural signals [8]. The CAVE was invented to provide a visualisation experience for one to many people utilising large projection screens [141]. These

[2] Image republished in a thesis/dissertation with permission of the rights holder, Association for Computing Machinery (Copyright licence in Appendix A).

projection screens are constrained in terms of the acoustic space, loudspeaker arrangement, brightness, sharpness, contrast and polarisation, so developers have to make many compromises in audio and visual presentations [8].

The large footprint, the cost of high-resolution projectors and the human-computer interaction are also reported to be limitations of a CAVE system [142]. Compared with a CAVE system, an HMD has its drawbacks, especially when one user is trying to interact with other users, and it does not offer interaction with real objects aside from VR control devices [142]. However, both rendering methods have a higher level of immersive experience than desktop displays.

2.4.3 Auralisation and spatial audio

For room or environmental acoustic modelling, objective acoustic parameters can be solved mathematically, presenting data-based results. On the other hand, we expect to experience an interactive acoustic environment rendered by these virtual acoustic setups. Sounds may be recorded in actual places in reality or virtually created with a computer, and these sounds are finally rendered to participants by a series of audio techniques. For room acoustics or environmental acoustics, the scale of objects is normally comparable to the wavelength of the sound, generating complex wave problems and numerical solutions. Solving these physics solutions for sounds in real time requires extremely high computing power. In the case of sound in VR, perceptual accuracy is more meaningful than physical accuracy [8]. As a consequence, a great number of approximations have been adopted for auralisation.

Spatial audio is a technique of creating sound in a 3D space; then, a listener can hear the sound from any direction in a sphere [143]. Because of this feature, it is often combined with VR to render auditory stimuli. In virtual environment reproduction, ensuring that the sound is correctly received by participants is a laborious task for soundscape auralisation.

Binaural rendering through a headphone is a conventional method in soundscape reproduction, and headphone presentation can deliver a reproduced sound field at the entrance of the ear canal that is similar to the subjective impression of the sound field directly without the headphone. Due to its simplicity, this method has been widely employed in soundscape auralisation research [75, 144]. It also requires specialised recording equipment to pair the headphone rendering, i.e., an artificial head or binaural microphones. For head tracking, three degrees of freedom (3DoF) means that the rotation of the object in three axes can be tracked. The 3DoF enables to track the angle change of head rotation for a 360 view of the surrounding environment. Six degrees of freedom (6DoF) include additional motion tracking in three axes, x, y and z in addition to rotation tracking.

Ambisonics is a sound reproduction technique used for recording and playing-back spatial audio, and it is based on the spherical harmonic decomposition of the sound field [145]. Ambisonics enables a listener to experience a spatially accurate perception of the sound field [146], and this reproduction technique was originally introduced by Gerzon in the 1970s [147, 148]. The recordings made by Ambisonics are known as B-format, and in the first-order Ambisonics (FOA), the information recorded by four channels are A-format. The spatial sound field is represented as spherical harmonics, and the information recorded in the A-format is synthesised into the B-format. The B-format can be decoded into various rendering formats such as 2-channel stereo, 5.1 surround sound and 7.1 surround sound matching the needs of dynamic auralisation under IVR. Higher-order Ambisonics (HOA) with a higher spatial resolution uses more microphones based on higher-order spherical harmonics to record the spatial audio. Sounds can be recorded at a certain location with an Ambisonic microphone (e.g., SoundField SPS200), and participants can experience the ambient sound environment with an appropriate playback system. A playback system is usually a set of devices where the audio signal is transmitted and processed through audio hardware to the listener. The speakers in this system, e.g., loudspeaker arrays or headphones, will eventually generate sound for human ears.

With the development of encoding and decoding methods, Ambisonic techniques have received more attention in broadcasting, room acoustics, and 3D audio [148, 149, 150, 151, 152]. Owing to the complex composition in a soundscape, Ambisonics, with the capability of recording a specific sound field, has been adopted in various soundscape studies. In 2005, Guastavino et al. [73] validated that the semantic features in soundscape reproduction by Ambisonics are similar to those of the original soundscapes. Soundscape reproduction and synthesis through eight-loudspeaker Ambisonic loudspeaker playback were investigated by Davies et al. [74] in 2014, and they stated that Ambisonic reproduction is ecologically valid in semantic aspects of soundscape perception. In 2016, Sudarsono et al. [153] conducted a subjective experiment comparing an Ambisonic reproduction system with a real soundwalk. They stated that the reproduced sound level should be adjusted to −9.5 dB below the actual level. For the stereo UHJ format referring to the C-format in Ambisonics, Stevens et al. [154], in 2017, examined whether the use of this format is valid for soundscape reproduction. Head-tracking binaural rendering allows 3DoF for head rotation, resulting in a higher immersive experience, and the applications of Ambisonics show a broad range of soundscape evaluation scenarios during reproduction.

Other spatial audio techniques have been developed, e.g., vector base amplitude panning (VBAP) [155] and wave field synthesis (WFS) [156, 157], and they are also gradually showing potential applications in VR-based soundscape research. Auralisation of an urban soundscape can be controlled to investigate sound perception in a reproduced environment. Spatial audio also provides the practical foundation of acoustics for real-time audio-visual interaction and synchronisation in VR.

2.4.4 Audio-visual and human-computer interaction in VR

Audio-visual interaction can significantly influence how people perceive an urban environment during a virtual experience. When people hear a sound, the inclusion of visual cues will distract attention from the sound perception, and the results may also be subtly different. Psychological variation and distraction are the same for visual perception with consideration of auditory stimuli [3].

Both visual and audio stimuli can deliver specific instruction information for human beings, and these two types of stimuli contribute most to our psychological perception of urban environments. Some studies have focused on the coupled audio-visual interaction on noise or soundscape perception with the aid of a visual display. Carles et al. [158], in 1999, pointed out that the coherent combinations of the sounds and images of a landscape are rated higher than the component stimuli by subjects. The noise-masked stimuli were easier to identify and more intelligible under audio-visual interactive conditions than under audio-only conditions with respect to extra visual information delivering acoustic cues [159]. Under coupled audio-visual rendering, the pleasantness of water sounds was found to be more preferable than bird sounds for Singaporeans in a study conducted by Hong et al. [160] in 2017. The reproduction with appropriate audio-visual interaction will ensure the accuracy of the evaluation results.

In addition to hearing and vision, other senses may be added into VR if possible, such as olfaction, haptics and tactiles. Because a VR laboratory cannot reproduce all environmental conditions, the researchers in various studies (e.g., refs [161, 162, 163, 164]) still regard audio-visual coupling as the dominant factor that significantly influences how people perceive the sound environment at this stage. As soundscape practices have been commonly implemented in urban construction and regeneration projects, audio-visual interactive experiences will play a more important role in assessments of soundscapes and environmental quality.

The purpose of VR is to achieve an immersive, interactive, and dynamic environment in real time for users. To accomplish this, we need more than just visual and auditory outputs from the VR system. Real-time input is an essential aspect of VR interaction as well. Being able to interact effectively with the environment in a visually immersive environment is an important area of research in the field of human-computer interaction. Input via hands or gestures is probably still the most common solution used by VR researchers. Hand selection patterns through tracked hand controllers have been broadly adopted by major consumer grade products, such as the Oculus Rift Touch Controller and Sony PlayStation Move Motion Controller [165]. Additionally, some other tools, e.g., dynamic gesture recognition and infrared imaging, are also grounded in the flexibility of the hands. Although voice input is now commonplace, most of us still use keyboards, touch mobile phones or write by hand to produce efficient output. During human-computer interaction in a VR experience, headtracker latency between head movements and output signals to the ears in virtual audio displays is inevitable. The critical amount of headtracker latency was reported to be less than 30 ms to ensure undetectable delays for users [166, 167].

A well-developed audio-visual and human-computer interaction will deliver a high level of immersion to users, and this high immersion is one of the key arguments for maintaining the ecological validity of the VR experience for soundscape evaluation. These interactions can be utilised effectively in research, and environmental psychologists have provided some examples of work studying the relationship between people's perception and virtual environments.

2.5 Discussion on combining VR and soundscape evaluation

2.5.1 Current VR research on urban environments

Due to the methodological overlap between VR and environmental research, there have been countless studies using VR technology to assess environmental issues. One of the significant applications of VR-based environmental reproduction is that people can experience and perceive virtual environments in advance, and the results of VR-based studies can be considered predictions allowing the identification of potential risks, solutions, choice-making and strategies through in situ practice. VR-based evaluation provides an alternative tool with strong immersion for citizens, benefiting the future creation, planning and regeneration of urban spaces.

Multisensory information in the real world is replaced by synthetic stimuli in VR [168, 169]. Visual imagery, stereophonic sound, tactile feedback and other sensory stimuli can be integrated into VR-related research to investigate the relation between human perception and virtual environments [170, 171, 172]. The comparison of experiences between the real world and the virtual world has been discussed in many studies (e.g., refs [173, 174, 175, 176, 177]). These studies compare the realism of virtual versus real environments and the practicability of reproduced systems from an environmental psychology perspective. For instance, Bishop and Rohrmann [174], in 2003, investigated whether the evoked cognitive and affective responses to simulations are similar to those when exposed to reality during a day or night walk. They proposed a significant limitation: the affective responses generated by the simulated world do not always match the response pattern induced in the real world, but the difference between the simulated and real worlds is considered to be acceptable and reliable in such an urban environment with rational audio-visual rendering. Some of these studies mentioned above incorporate audio-visual experiences but do not discuss how people perceive and evaluate the sounds in these virtual environments systematically.

For the perception of urban sound, the researchers also used a paradigm shift analogy to environmental psychology studies. Researchers have a wide range of research interests, such as road/railway noise, overall sound environmental quality, the performance and design of noise barriers, audio reproduction techniques, audio/visual preferences and interaction, and soundscape planning associated with affective or cognitive performances. For instance, acoustic and visual congruence in VR-based urban environmental research also benefits community planning [178], revealing that the huge potential of VR-based soundscape research represents an innovative tool for predicting the impact of sound environments on human beings. an aesthetic preference for people resulting in preconceptions of perceiving less noise. A case study of an urban public space to assess the sound environment in VR was conducted by Sanchez et al. [179] in 2017. Participants were positioned as pedestrians with different visual designs of noise barriers. Thus, owing to the different acoustic performances of noise barriers, the process of sound propagation passing through these noise barriers can be calculated by a simplified 2D FDTD model, and the sounds received by participants matched the visual designs of the noise barriers. They also stated that the visual design has a strong influence on the evaluation of the overall appreciation: a shorter noise barrier with a better visual design can effectively enhance the pleasantness.

A growing number of soundscape evaluation studies, e.g., refs [161, 162, 180, 181], have embraced VR technology. The discussion of complex environmental contexts and reproduction techniques still needs intensive work, and practising soundscape reproduction based on the existing standards still faces challenges.

2.5.2 Challenges facing VR-based soundscape research

Currently, researchers in the field of soundscapes may be less concerned with how to measure common soundscape indicators for an unknown environment or space. They attempt to summarise the relationship between the acoustic environment and human perception with consideration of social, political, economic, cultural and reproduced factors. In the process of this exploration, auralising and reproducing a sound environment efficiently and accurately in VR is of particular importance. A great deal of simplification and compromise is inevitable to achieve real-time dynamic interactions between humans and computers. The study of the minimum details necessary with limited computational power to implement VR auralisation is valuable practice for urban planning, noise and soundscape policies, soundscape assessments, game design, and urban interactive art.

In urban planning or design, designers are striving to communicate design concepts based on acoustic comfort, and they are also communicating audio-visual presentations through multisensory interactive technologies such as VR. In this communication process, they show uncertainty and concern about the accuracy of the acoustic representation for the reproduced scenes, and this uncertainty is likely to lead to poorer perceptions of the final presentation scheme.

For game design, 3D game developers have long sought to immerse the player in a virtual world, or a narrative thread by using various visual and auditory aspects to create a soundscape and environment that is defined by the context of the game. They need to give deep thought to what kind of sound is appropriate for the players in the right ambience to immerse themselves in the scenario or to provide the emotional resonance intended by the developers, such as happy, sad, excited, or tranquil. This procedure may sometimes be judged on the basis of artistry and may very often also entail the intervention of a systematically subjective approach to evaluation.

When soundscape researchers or consultants assess soundscape evaluation in VR, the technical specifications of soundscape reproduction are not elaborated in the series of ISO 12913. Different audio rendering [88, 182], reflection modelling [183, 184] and source synthesis [185] methods conducted by various researchers have not been systematically compared. The relationship between the reproduced performance and subjective responses is not well studied and interpreted in existing protocols and standards. Therefore, the exploration of reproduction strategies in urban sound environments has important informative and practical implications.

2.6 Conclusions

This chapter reviewed the cutting-edge development of soundscape research and VR technology. Through the discussions of soundscape definition, acoustic modelling, and urban design, there are compelling needs for participatory soundscape assessment and management. VR thus serves as a multisensory approach for conveying design concepts of soundscapes, providing an effective interaction between humans and environments. Subjective tests combining both environmental psychology and VR have been widely adopted by researchers looking for evidence of people's perception and responses to environments.

Although the advent of soundscape standards in the last decade has provided some technical guidance for acousticians, our knowledge of soundscape reproduction from the perspective of human perception in VR is still limited. There is still a lot of uncertainty for soundscape researchers and engineers when it comes to reproducing soundscapes based on existing standards. This uncertainty derives from the inadequate comparison between different renderings and simulations. These critical issues for improvement frequently focus on the ecological validity between reproduction methods, audio rendering strategies, reflection modelling during sound propagation simulation, and sound source synthesis especially for non-point like sources.

Based on the discussion of the reproduction strategies, this thesis aims to investigate the following:

1. How is the ecological validity of IVR identified in soundscape evaluation?
2. How to choose appropriate monaural and binaural rendering methods in soundscape evaluation?
3. What kind of simplifications can be made to optimise the reflection modelling in environmental sound propagation simulation?
4. What kind of compromises or simulations can be made to synthesise non-point like sources to improve the perception of width, immersion and distance under virtual environments?

Chapter 3

Methodologies

This chapter introduces the conventional research methods performed across the next several chapters and the justifications for choosing these methods. To investigate the reproduction strategies in VR-based soundscape evaluation, methodological integration including a systematic review, soundscape reproduction and subjective tests is employed.

3.1 Overall workflow

The overall process of VR-based soundscape evaluation is illustrated in Fig. 3.1. For the identification of the ecological validity of VR in soundscape assessments, a systematic review will be conducted. This process collects secondary data from prior studies and analyses them through a repeatable analytical method. The procedure of the systematic review follows a deductive approach based on a portico of evidence synthesis. A deductive approach based on subjective and objective data verifies the issues raised in the auralisation of soundscape during VR experiences. The off-site experiments occupy a central position in VR-based soundscape evaluation.

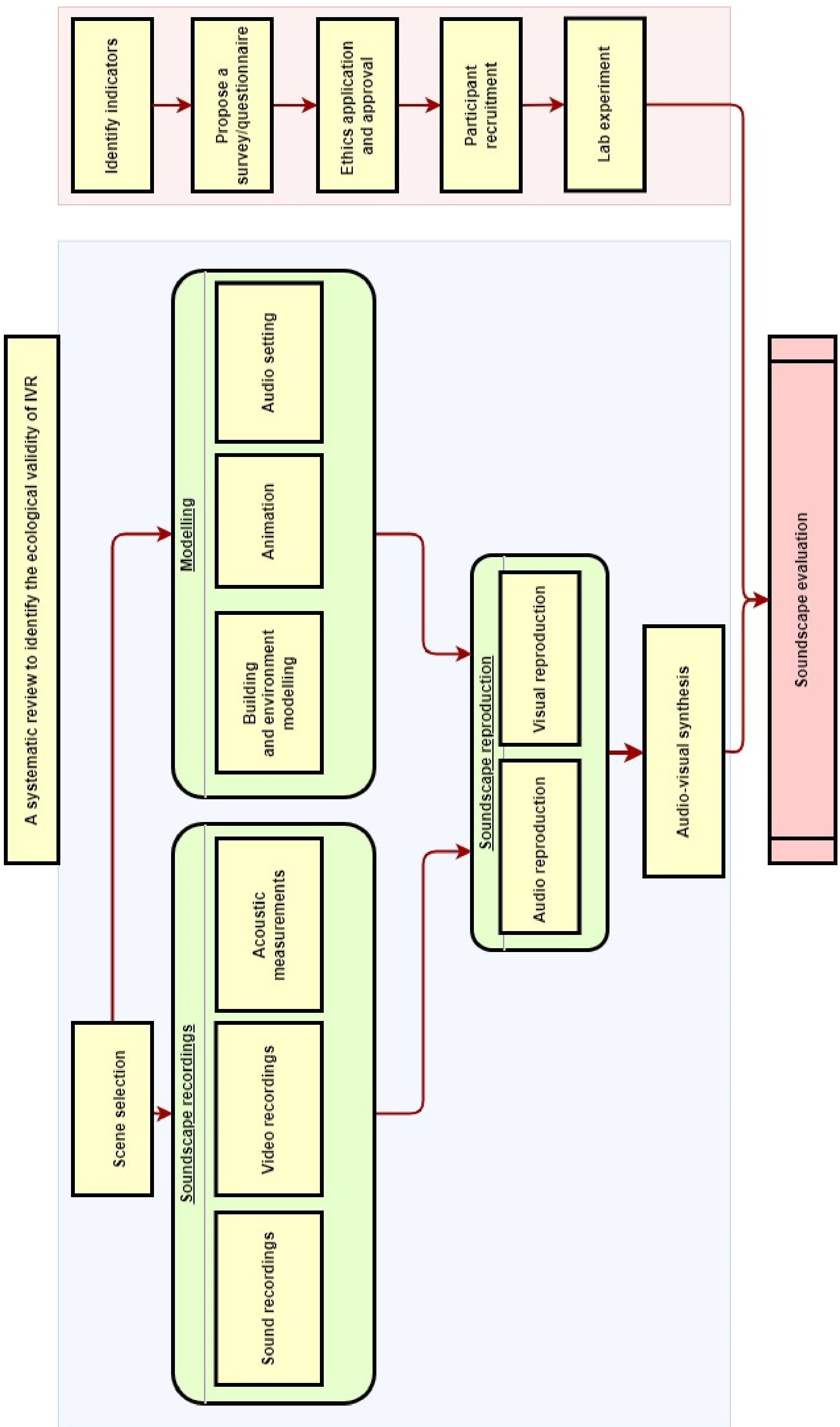


Figure 3.1: Methodology workflow of the thesis.

To acquire subjective responses of sounds, the experiments throughout Chapters 5-7 were divided into two main sections as shown in Fig. 3.1. One section of the experiments illustrates soundscape reproduction and acoustic measurements and the other represents questionnaire design in the laboratory. For soundscape reproduction, the main task is to achieve audio-visual stimuli by recording and modelling contextual and acoustic information. For soundscape recording, it may include sound recordings, video recordings and acoustic measurements. For soundscape modelling, it will implement environmental modelling, object animation, audio setups (e.g., different reflections and synthesis points). The experiments will broadly follow these two sections for the investigation of the comparison between monaural and binaural static rendering in Chapter 5, the effect of reflection order in Chapter 6, and the synthesis of line sources in Chapter 7. For questionnaire design in the laboratory, the workflow is from indicator identification, questionnaire draft, ethics application, participant recruitment to the listening test. The justification of this workflow is illustrated in Section 3.4, and the specific scenarios of using different indicators and testing conditions are explained separately in Chapters 5-7. Through soundscape reproduction and subjective evaluation, the perceptual responses can be collected for further analysis. The undesired variables (e.g., background noise and weather) can be controlled effectively during the experiment.

3.2 Review methods

A traditional review is predominantly subjective, and it mainly relies on the author's knowledge and experience [186]. Without commonly agreed guidelines and steps to follow in a review, different researchers may produce widely varying results at risk of bias or systematic error [186, 187]. It is not comprehensive for the identification and incorporation of specific research findings. Systematic reviews provide a systematic, repeatable and objective means of searching for a particular research question. The process of a systematic review is well defined and accepted by global researchers. The Preferred Reporting Items for Systematic Reviews and

Meta-Analyses (PRISMA) provides a checklist for researchers to conduct a systematic review [188]. Its guidance can effectively identify, appraise and synthesise results of primary studies under the process of data screening and extraction to minimise the risk of error and bias.

In the fields of noise control and soundscape evaluation, subjective data from original studies are usually obtained from deductive reasoning to search for convincing evidence on a particular research topic. The systematic review has been used in some soundscape studies to draw inductive conclusions, such as positive health-related effects [2], psychophysiological implications [93], prediction models [189], and audio-visual bimodal and interactive effects [190]. For the investigation into the ecological validity of VR in soundscape evaluation, I need to seek primary sources rather than secondary sources and data. Thus, a systematic review is an effective, repeatable and updatable review method to avoid subjective judgments for specific research methods and technical tools of soundscapes.

3.3 Soundscape reproduction

Soundscape reproduction aims to utilise audio-visual techniques to create perceptually equivalent simulations for people. For the vast majority of urban regenerated or built areas, the urban spatial morphologies are ever-changing, and people unconsciously move through and consciously engage in a number of activities in these spaces. An urban public space may carry out many functions, such as office, education, business, and relaxation. These different functions are determined by human behaviours and activities in our cities. People, together with their natural and built environments, jointly shape what we can hear in our cities. Urban soundscapes naturally arise under these conditions.

The purpose of soundscape reproduction is to strive to replicate this experience by audio-visual rendering. The identification of typical spaces in cities is of great importance in soundscape reproduction, as it can represent many of the same types of spaces that are involved in virtual environments through human-computer interaction. Soundscape reproduction is not only about replicating a built environment presented by audio-visual stimuli, but also about evoking a sense of identity with the space that the city bears and the historical and cultural contexts behind it.

For in situ recordings, a suitable observation or recording location needs to be found in an urban space. Multiple sound sources in urban spaces determine the complexity and diversity of the acoustic environment. The recording location should not be too close to a specific sound source, and there should not be too many pedestrians around the recording location. On the one hand, many pedestrians create a lot of unnecessary dynamic sounds, and on the other hand, the physical movements or facial expressions of these pedestrians significantly attract participants' attention at short distances.

For visual modelling, spatial and architectural proportions should be carefully visualised. Incongruous spatial proportions can influence spatial perception and thus audio-visual interaction. Popular modelling software (e.g., SketchUp) demonstrates superior compatibility and visual representation for a wide range of architectural structures and texture renderings. These tools with various plug-ins will be used to accomplish the modelling of architectural spaces. The synthesis of animation or particle effects can be accomplished by 3D animation or game engines (e.g., 3ds Max and Unity).

When we only record videos for visual rendering, the videos will contain a large number of environmental components, including various pedestrians, vehicles, events and other factors. It becomes difficult for us to modulate some of the additional acoustic and spatial variables in these videos. For example, when we attempt artificially to add a fountain or a music band to a realistic urban space in a recorded video, it is difficult to show human interaction with such a fountain or band. Thus, I adopt both recorded videos and environmental visual modelling for visual rendering. In auralisation, the sound will be selected for appropriate visual renderings, depending on different methods of sound recordings and audio processing.

3.4 Subjective evaluation procedures

The concept of soundscapes is centred on human perception in sound environments. Objective numerical simulation and measurements are generally conducted by modelling physical parameters such as sound pressure (e.g., refs [100, 101, 191, 192]). Compared with numerical simulation, subjective tests are regarded as an essential tool for researchers to understand how people perceive sound in urban spaces and to assess soundscape quality. For audible sound, people have different subjective preferences and physiological responses to various sounds in urban or natural environments. The sound pressure cannot accurately represent people's perception of sound, especially in complex acoustic spaces. Researchers of psychoacoustics have pioneered the use of subjective evaluation methods to obtain systematic evidence for the cognition and perception of sound based on our auditory nervous system. Many fundamental studies, such as equal loudness contours (e.g., ref [193]) and hearing loss tests (e.g., ref [194]), have widely adopted pure tone evaluation. For soundscape research, pure tones do not exist in natural environments, and it is meaningless to use pure tones to validate subjective perception in virtual sound environments. The subjective evaluation of soundscapes should thus take into consideration both the sound rationality and the validity of the evaluation process.

It is significant to identify what perceptual indicators are to be discussed in the subjective evaluation and what these indicators can interpret according to different research questions. The frequently used perceived affective attributes including 'pleasant', 'chaotic', 'vibrant', 'uneventful' 'calm', 'annoying', 'eventful' and 'monotonous' can be measured according to appropriate psychometric scales [58, 82]. The psychoacoustic indicators, e.g., loudness [37], will be taken into account. Related to soundscape reproduction, the reproduced indicators, such as immersion and realism, are essential for testing the reproduced environments [195, 196]. In terms of spatial characteristics of urban environments, the indicators including perceived distance, width, and reverberance [37] will be examined in different situations.

It is essential to tailor these indicators to our studies in readable sentences to ensure that the survey or test is easy for the participants to read and understand. A well-structured subjective test is one of the key factors in ensuring high measurement validity. This step is equally important for the participant information sheet. The psychometric scales and question design will follow the guidance from international standards and other soundscape or overall environmental quality research.

As participants evaluate the soundscape we are in essence collecting personal data, which can be real or anonymous depending on the situation. Ethical compliance and approval are also necessary for the relevant committee. The entire testing process and information, data protection, risk assessment, and potential ethical issues during the subjective evaluation are all reported to the committee. Recruitment of participants is usually through emailing to university student lists, and the recruitment information is approved and agreed in advance by the committee as well.

Data protection is an important dimension of subjective evaluation involving data collection, processing and storage. It is related to the fair and proper use of information about people, and ensures their right to control their own identity [197, 198]. The collection, processing and storage of personal data are also in compliance with the University College London (UCL) data protection guidance on Data Protection Act (DPA 2018) and General Data Protection Regulation (GDPR).

3.5 Conclusions

This chapter introduces the methodological workflow in this thesis from review methods, soundscape reproduction approaches to subjective evaluation procedures. I present the justifications for choosing the assessment methods running through the thesis, including the comparison between a traditional review and a systematic review, audio-visual rendering methods, and differences among traditional psychoacoustic evaluation and soundscape evaluation. Soundscape studies frequently employ a hybrid approach from these methods to reproduction and evaluation, which are broadly in line with our research workflow.

Faced with the issues of the evaluation of soundscapes in virtual reality, I will not only use objective indicators or simulation but also incorporate subjective evaluation. The methodological process will combine subjective and objective indicators to improve our understanding of auralisation in VR experiences. It is essential to tailor the methods of soundscape reproduction and subjective evaluation in our studies according to different environmental contexts, research questions, and test procedures of existing studies to ensure the validity of research design. The elaboration and selection of these methodologies will set the foundation for the rest of our research, and in the next chapters, I will present the corresponding method details for identification, reproduction and evaluation.

Chapter 4

A Systematic Review to Identify the Ecological Validity of IVR

This chapter[1] aims to review the approaches that are utilised to assess the ecological validity of IVR for the perception of urban sound environments and the necessary technologies during audio-visual reproduction to establish a dynamic VR experience that ensures ecological validity. Section 4.1 introduces the background of ecological validity and how it is connected with urban sound environment research. Section 4.2 conducts the systematic reviewing flow to identify the possible research involved within the field of ecological validity of urban sound environments. The urban sound environment in this chapter refers primarily to sound sources originating outdoors or in urban public spaces. Section 4.3 compares the subjective responses among different studies and their reproduction systems. Section 4.4 discusses the assessment methods, other visual rendering methods, verisimilitude and veridicalidity, and some limitations in the systematic review. Section 4.5 outlines the approaches of laboratory tests and the audio playback methods from the review and discussions.

4.1 Background

Ecological validity was introduced in the 1980s to evaluate the outcomes of a laboratory experiment focused on visual perception [200]. Ecological validity describes

[1]This chapter was partially published in Acoustics [199].

the degree to which results obtained in a controlled laboratory experiment are related to those obtained in the real world [201]. The discussion of the ecological approach regarding its internal validity and experimental control began in the 1980s with cognitive and behavioral psychology research [200, 202], and these two factors are still significant factors in the design and undertaking of an ecological approach study. Under laboratory conditions, researchers should give participants corresponding environmental cues and instructions to enable the reactivation of the cognitive processes of participants that were determined in actual situations [73]. For high ecological validity, the findings in the laboratory can be generalised into real-life settings [201]. As a simulated technology, IVR places the user inside an experience, which allows the impact on participants of a new environment with complex social interactions and contexts to be assessed [170, 178, 203]. In 2001, Bishop et al. [204] reported their non-IVR assessments of path choices on a country walk, and they agreed that faster computers and better display systems make the virtual environment experience more credible. Thus, low ecological validity resulting from non-sufficient immersiveness could be a limiting factor for the generalizability of data collected from laboratory experiments. The need for more research that explores applications of perceptual simulations in general and related questions of validity and reliability has been stressed ever since the emergence of environmental simulation as a research paradigm.

Ecological validity has been conceptualised into two approaches: verisimilitude and veridicality. Verisimilitude refers to the extent of similarity of a virtual experience to relevant environmental behaviors [205]; it reflects the similarity of the task demands between the test in the laboratory and the real world [206]. This approach attempts to create new evaluation assessments with ecological goals [207]. Veridicality refers to the degree of accuracy in predicting some environmental behaviors [208, 209]; the establishment of veridicality is required to assess the results from the laboratory test and the measures in the real-world. There are some limitations for both approaches. One limitation of the veridicality approach is that, for those conditions which are not likely to be reproduced in the real world or that have a high cost, the outcomes from real-world measures cannot correlate with experimental results. When using the verisimilitude approach alone, no empirical data are needed to claim that the evaluation is similar to real life settings [207]. For soundscape reproduction, researchers expect the high consistency between the data obtained in the laboratory and the data obtained in the real world. For the listening environment in the laboratory, researchers also attempt to approximate the real-world experience in terms of recreating a more realistic virtual experience. Thus, the results of the experiment and design of evaluation (e.g., a virtual soundwalk and a real soundwalk) jointly determine the ecological validity of soundscape research.

VR has revealed a functional rapprochement that fuses the boundary between the laboratory and real life as discussed in Chapter 2. Through multisensory stimuli with experimental control, participants tend to respond realistically to virtual situations as if they were in a real environment [164, 172, 210, 211]. The responses to a virtual environment are generated when place illusion (PI) and plausibility illusion (Psi) occur at the same time [164, 172, 210]. The ecological approach studies based on VR provide controlled dynamic presentations of background narratives to enhance the affective experience and social interactions [202, 212]. From a methodological viewpoint, environmental conditions and test results can be ecologically validated through VR technologies according to a subjective evaluation framework. Numerous researchers have examined ecological validity in different topics and fields with the comparison of a virtual environment and real life [173, 174, 175, 177, 213].

The multisensory evaluation method shows enormous significance in helping participants to perceive environments holistically [214, 215, 216, 217, 218]. The reproduction system of listening tests needs to be adapted to the purpose of the study to allow the subjects to treat the test samples as potentially familiar experiences through cognitive processes elaborated in actual situations. With the aid of IVR, the installation of laboratory conditions was performed with the aim of reproducing urban sound environments and presenting a multisensory experience to participants. A subjective test of IVR reproduction in urban sound environment assessments would show high veridicality if it correlated well with measures of perceptual responses in the real world.

The concept of ecological validity has been extended from psychological experiments to the domain of complex sound environment perception. It is not only related to the evaluation methods during laboratory tests, but also closely associated with the developing IVR technologies. Attempting to establish a standardised soundscape evaluation protocol with high veridicality under an immersive virtual environment has a broader impact on the practice of soundscape planning and design. The research on soundscape standardisation has discussed the definitions, variety of contexts, evaluation methods and reporting requirements [61, 219].

The ISO Technical Standard (ISO/TS) 12913-2:2018 [58] introduced two common recording techniques in soundscape research: binaural and Ambisonics. The standard states that if some environmental factors are not present or differ during playback, the outcomes could possibly result in different impressions to those received in the original context. In terms of the statement of ISO/TS 12913-3:2019 [59], the validity of these auralisation techniques combined with other environmental factors still presents some uncertainty. The ISO/TS 12913-3:2019 [59] stated that the key factors to consider when conducting ecologically valid laboratory studies are the effect of memory, the duration of exposure to each of the stimuli and the auditory immersiveness. IVR could deliver more degrees of freedom for users than non-immersive rendering methods as discussed in Chapter 2. A comparison of the ecological validity using IVR for urban sound environments with different reproduction techniques and research topics is therefore made. This review aims to investigate (1) which kinds of approaches can be utilised and integrated to assess the ecological validity of IVR when humans perceive urban sound environments and (2) which technologies are necessary during audio-visual reproduction to establish a dynamic IVR system to assess the perception of urban sound environments. Through the discussion of the ecological validity of different techniques and evaluation methods, I will explore the insufficiently discussed gaps for audio rendering and modelling strategies in the existing soundscape standards.

4.2 Methods

4.2.1 Search strategy and eligibility criteria

There were no pre-defined protocol registrations for this review. The basic process and data extraction forms were agreed upon at the beginning of the review work. The study was performed under the guidance of the PRISMA [188].

Given the exploratory nature of this study, as many studies do not directly mention 'ecological validity', and they may not include the terms 'ecologically valid', 'ecologically validate' and similar expressions, the studies were selected manually according to the following inclusion criteria: (1) original participatory studies using VR techniques conducted in a laboratory, and (2) studies collecting subjective responses under virtual environments. The subject areas included 'ecological validity', 'acoustics' and 'virtual reality'. Some studies did not directly mention 'ecological validity', but the workflow was under the framework of the virtual sound environment evaluation described above. These studies were selected in the review with full-text scanning.

Studies were identified by searching the electronic database, scanning reference lists of articles and in consultation with experts in the field. A literature search was conducted using the Web of Science. Only peer-reviewed journal articles published in English were considered. This search was applied to the Web of Science (1980-2020). The last search was run on 01 July 2020. I used the following search terms to search the databases: 'sound', 'perception', 'participant' and 'virtual reality'.

4.2.2 Data extraction

Information was extracted from each included document regarding (1) the research focus of the studies, (2) participant numbers, (3) in situ responses vs. laboratory experimental data and (4) the main parameters selected in the studies. Considering ecological validity across the selected studies with various topics and different outcomes, a qualitative approach was adopted to answer the review questions.

4.3 Results

The initial results showed 65 documents. Fifty-three items were excluded because the topic of the papers failed to meet the eligibility criteria, which included (1) the studies not using a VR head-mounted device, and (2) the studies not involving sound-related perception. The full texts of the remaining 12 papers were accessed, and these 12 papers were included in the review.

4.3.1 Ecological validity with subjective responses

Table 4.1 shows the research focus, the participant numbers, in situ responses vs. experimental data, the main parameters and variables in these studies. These studies with IVR had different emphases on their subjective evaluation and research focuses, and they assessed ecological validity with subjective responses varying from environmental preferences/quality, audio/visual indicators, coupled interactions and reproduction quality.

Generally, these studies were not only limited to one topic, and several topics were often integrated together. The audio-visual interaction was also one of the sub-topics of these studies. Most of these works addressed the importance of audio-visual interaction in IVR-based soundscape or noise assessments. The audio-visual interaction in these studies was discussed in an attempt to interpret how participants perceived the virtual environments, and the ecological validity was also tested with their research questions. Global environmental evaluation, visual and acoustic coherence and familiarity and visual and acoustic congruence were compared, respectively, through the field survey and the laboratory experiment to jointly validate the acoustic and visual congruence between the simulated and real world [178]. Both groups in the in situ session and the laboratory session with 16 participants, respectively, were recruited for the in situ and laboratory sessions. Both comparison groups showed robust similarity in visuo-acoustic coherence and familiarity and the visuo-acoustic salience of urban, human and natural activities.

Related to audio-visual interactions, in 2020, Jeon and Jo [181] examined the contribution of audio and visual stimuli in the evaluation of urban environment satisfaction under an immersive virtual environment. Three conditions were considered: (1) audio only, (2) vision only and (3) audio-visual interaction. The contributions of audio and visual information on overall satisfaction were 24% and 76%, respectively. The study by Ruotolo et al. [164] in 2013 asked the participants to answer questions about auditory and visual annoyance, respectively. The results presented in their study showed both auditory information and visual information in a close interaction, supporting participants perceiving the virtual environment holistically. Aletta et al. in 2016 [220] carried out a study to investigate the chiller noise involved with the distance to a source and the visibility of a source. They found that the visibility of a source is not a significant influencing factor for noise perception for the kinds of chillers examined in the study.

Table 4.1: Virtual environment evaluation studies with subjective responses.

References	Research Focus	Participant Number	In Situ Responses vs. Experimental Data	Auditory Perception	Visual Perception	Coupled or Other Variables
(Iachini et al., 2012) [163]	Acoustic comfort aboard metros	51	x	Annoyance	–	–
(Ruotolo et al., 2013) [164]	Noise assessment for a motorway	20	x	Audio annoyance	Visual annoyance	–
(Maffei, Iachini, et al., 2013) [161]	The effects of visual and acoustical aspects of the impact of a wind farm in a quiet area	46	x	Perceived attributes	Visual pleasantness	General environment evaluation
(Maffei, Masullo, et al., 2013) [162]	The influence of visual characteristics of barriers on railway noise perception	41	x	Annoyance, perceived loudness	Visual pleasantness	–
(Aletta et al., 2016) [220]	The effect of vision on the perception of the chiller noise	26	x	Perceived loudness, noise annoyance	Visual unpleasantness	–
(Maffei et al., 2016) [178]	Global sound environmental quality	16 in situ, 16 in the laboratory	✓	Acoustic coherence, and familiarity	Visual coherence, and familiarity	Global qualitative evaluations
(Sanchez et al., 2017) [179]	The role of noise in the audio-visual design of an urban public space	71	x	–	–	The audio-visual interaction for preference and reality evaluation
(Yu et al., 2017) [221]	Noise and visual intrusion from wind parks on affective and cognitive performances	20	x	Audio annoyance	Visual annoyance	–
(Hong et al., 2019) [196]	The FOA reproduction comparison with in situ soundscape evaluation	5, 12, 13 in three days (in situ), the same participants in the laboratory	✓	Overall soundscape quality	Perceived spatial quality	Distinctiveness, immersiveness, realism, reproduction fidelity
(Jeon and Jo, 2019) [195]	Road traffic noise perception in urban high-rise residential buildings	40	x	Perceived loudness, annoyance, sound acceptance	–	Perceived distance, perceived directionality, perceived width, immersion, realism, and perceived externalization
(Sun et al., 2019) [180]	Classifying soundscapes of urban public open spaces	20 for Group 1, 20 for Group 2	x	Classification components, psycho-acoustical indicators, and saliency	Visual factors	–
(Jeon and Jo, 2020) [181]	The relationship between overall satisfaction of the urban environment and audio-visual interactions	30	x	Sound preference, soundscape attributes	Visual preference, visual attributes	Environment satisfaction

In 2019, Jeon and Jo [195] carried out a study to assess the noise in urban high-rise residential buildings. They reported that the directional and visual information generated by HRTFs and HMDs could affect sound perception and virtual environmental immersion ratings. Two parameters, HRTFs and HMDs, were coupled into four cases: no HRTF-no HMD, no HRTF-HMD, HRTF-no HMD and HRTF-HMD. The results showed that the contribution of the HRTF to subjective responses was 77% higher than the contribution of HMD, at 23%. This study showed the applicability and necessity of the HRTF and HMD to assess noise in terms of the audio-visual interactions under immersive virtual environments. In 2012, Iachini et al. [163] assessed acoustic comfort aboard metros through subjective annoyance and cognitive performance measures. In their findings, visual contexts could be considered a modulating method affecting noise annoyance for people. Noise barrier designs are generally associated with noise assessment, and different noise barrier designs were assessed under an immersive virtual environment [162, 179]; different project solutions concerning noise mitigation in order to obtain more reliable results on local residents were also examined.

The potential environmental risks and negative effects of wind parks, as emerging landscape projects, were also evaluated under virtual environments. Maffei, Iachini, et al. [161] in 2013 stated that the noise perception of wind turbines under IVR requires extended experiments to ensure its ecological validity, especially the results from in situ sessions. In 2017, Yu et al. [221] conducted a subjective test revealing that wind parks can increase both the aural and visual annoyance associated with personal attitudes toward wind parks. The research of VR technologies in the sound environments of wind parks ecologically validated these potential negative influences.

Soundscape evaluations show a trend of using multi-dimensional attributes to test participants' perception in a virtual environment. In 2019, Hong et al. [196] carried out a study exploring the ecological validity of reproduced acoustic environments based on three spatial audio reproduction methods. The main indicators in their study included sound preferences, visual preferences, soundscape attributes, visual attributes and environmental satisfaction, as shown in Table 4.1. In 2019, Sun et al. [180] proposed a hierarchical soundscape classification method using VR playback with a participatory experiment inside a soundproof booth. The method, based on different classification components, could be potentially validated by verification on an independent dataset.

In IVR laboratory experiments, the numbers of participants differ in different subjective studies. The minimum number of participants in the subjective test was 16, in the work by Maffei et al. [178] in 2016, and the maximum number reached 71, in the work by Sanchez et al. [179] in 2017. The number of participants for most subjective tests in the laboratory ranges from 20 to 60.

4.3.2 Reproduction systems

The reproduction systems in these studies mainly include two aspects of auralisation and visualisation, as shown in Table 4.2. To simulate an immersive auditory environment, Ambisonics is a prevailing method to record and auralise sounds, which allows various decoding patterns with the flexibility to lay out loudspeaker positions or headphones. In the headphone-based reproduction method, the recorded stimuli captured in Ambisonic formats are most usually presented as either head-tracked or binaural static renderings [181, 196]. In the loudspeaker array-based reproduction method, there is no need for software to compensate for head movement in real time [196]. VR HMDs have an important role to play in the evaluation of reproduced quality, e.g., realism, and the realism rating in soundscape evaluation was found to be significantly improved with HMDs compared with the condition without HMDs [196]. Simulated visual environments can also be built using software including but not limited to 3ds Max [162, 179], Google SketchUp [163, 164, 178, 195], Unity [179, 221], Kubity [195] and WorldViz [161, 163, 220]. Unity (developed by Unity Technologies) is a game engine widely applied in 2D, 3D and VR games. Unity is supported on numerous platforms, such as Oculus Rift, Gear VR, PlayStation VR, Steam VR, Google VR and other possible developer platforms, and it has been widely adopted by many studies in VR-based planning research [222, 223, 224, 225, 226]. In 2019, Hong et al. [196] reported no

Table 4.2: Auralisation and visualisation during the participatory experiments.

Auralisation	
Recordings	Playback
Binaural audio signal recordings [161, 162, 163, 164, 220, 221]	Headphones [163, 164, 179, 180, 181, 195, 221]
Ambisonic recordings [178, 179, 180, 181, 196]	A number of loudspeakers, and a sub-woofer [161, 162]
	Headphones with a sub-woofer [220]
	5.1-format loudspeaker configuration [178]
Visualisation	
Visual construction methods	Visual rendering
3ds Max [162, 179]	HMD [161, 162, 163, 164, 178, 179, 180, 181, 195, 196, 220, 221, 227]
Google SketchUp [163, 164, 178, 195]	
WorldViz [161, 163, 220]	
Unity [179, 221]	
Kubity [195]	
Panoramic views [180, 181, 195, 221]	

significant differences in perceived dominant sound sources and affective soundscape quality between reproduction and in situ results. These findings are in agreement with previous studies showing that IVR HMDs with Ambisonics could be a reliable tool for soundscape assessment as an alternative to in situ surveys.

Some devices have been introduced to record information and render stimuli. A panoramic camera is usually used to record omnidirectional videos as visual stimuli in the laboratory test [181, 196]. A hybrid and simultaneous audio and video recording setup was used in the study by Sun et al. [180] in 2019. This setup consists of binaural audio (an artificial head with windshield and binaural recording device), an FOA microphone and a 360° video camera. A mobile device (a Google Cardboard headset) was also used in the evaluation of the audio-visual perception of wind parks. This portable HMD also showed the potential to provide an immersive experience in response to participants' head movements.

Notably, owing to the fact that the entire IVR industry is driven by both hardware and software upgrades, older ecological validity studies on virtual environments face limitations in terms of their utility or efficacy. It would be expected that the advancement in the computation of IVR simulations would ultimately increase the ecological validity of participatory studies conducted in laboratories. A comparison of the technical parameters of all IVR systems in these studies shows the limitations of initial research and how these limitations are gradually improved by subsequent studies. However, due to the lack of control measures across the analysed studies, it was not possible to conduct such a comparison. I cannot systematically assess the differences between the studies.

4.4 Discussions

4.4.1 Subjective response, cognitive performance, and physiological response

Many studies have suggested that urban noise can negatively affect people's cognitive functions and influence their daily life [5, 13, 228]. Subjective responses may not show annoyance regarding urban noise, but the cognitive performance may be affected. Thus, during the laboratory test, some studies also used cognitive tasks to evaluate the cognitive performance caused by the virtual environment [163, 164, 220, 221]. Related to stress recovery, researchers have used measures based on the physiological responses of participants. Annerstedt et al. [229] in 2013 conducted a study to investigate the sounds of nature inducing physiological stress recovery, and the trier social stress test (TSST), as a highly standardised protocol for inducing stress, was applied in their study. Cortisol, HR, HRV and T-wave amplitude (TWA) were tested to analyse the physiological stress recovery induced by the sounds of nature. Hedblom et al. [230] in 2019 adopted mild electrical shocks and skin conductance measurements to evaluate the stress recovery under virtual environments with a birdsong-traffic noise interaction. Compared with subjective responses, physiological responses do not directly reflect the relationship between subjective sound preferences and characteristics of acoustic environments. Thus, these three methods can jointly assess the ecological validity of complex sound environment perception.

4.4.2 Other visual rendering methods

For visual rendering, many studies used non-HMD options. Some of them adopted non-immersive methods, such as a monitor screen [66, 174, 204, 225, 231, 232], visual screen [233] and 2D projection [234, 235]. Some of the studies utilised the immersive CAVE system [229, 236].

Studies without visual stimuli were also conducted [73, 99, 237]. A visual component presents rationality when examining the ecological validity of auditory perception. The coupled audio-visual interaction is associated with the spatial attributes of sound perception, e.g., distance, width and directionality [195], and it also provides an animated visual anchor, improving the sense of presence and immersiveness during the subjective evaluation [99, 238].

4.4.3 Verisimilitude and veridicality

Verisimilitude and veridicality in IVR-based sound environment research have different emphases according to their definitions. Establishing verisimilitude and veridicality in a subjective evaluation experiment allows a virtual sound environment to be perceived with reliable ecological validity. The IVR research involved with verisimilitude in soundscape or noise assessments assumes that the stimuli in the test and the cognitive processing are sufficiently similar to the psychological construct of corresponding scenarios in the real world. The verisimilitude approach is likely to focus on specific tasks in the laboratory test similar to the task demands in the real world. The evaluation indicators and questionnaire design can be formatted in a quite similar way to a participatory experiment. Sanchez et al. [179] in 2017 pointed out that their study did not strictly prove that audio-visual designs in a virtual environment would lead to the predicted pleasantness of real environments. Establishing verisimilitude in soundscape evaluation is more intuitive compared with establishing a new cognitive task or a clinic neuropsychological assessment. However, when researchers discuss the relationship between subjective responses, cognitive performance and physiological responses, they need to carefully examine the verisimilitude approach with which some aspects of testing conditions limit the applicability of a method without empirical data to the real world.

A few studies validated veridicality in IVR-based soundscape or noise assessments. The pioneering studies examined several fundamental playback systems. In 2005, Guastavino et al. [73] explored the linguistic analysis of verbal data in soundscape reproduction through a field survey and two listening tests. Both listening tests compared exposure to the stimuli reproduced via stereophonic and Ambisonic approaches. They pointed out that both neutral visual elements and a good sense of spatial immersion should be provided to ensure ecological validity when testing the effects of urban background noise. Both reproduction methods have been demonstrated to be ecologically valid tools in terms of source identification. However, IVR was not applied in their study. Many perceptual attributes and indicators have been selected to describe the similarity between the real world and the laboratory conditions. In 2016, Maffei et al. [178] compared the congruence between audio and visual elements, and there was no significant difference in the perceived global quality of the environments in both the simulated and real world in their results. The global quality of the environments was shown to have high veridicality under the framework of subjective evaluation. The findings are consistent with the results of audio-visual interaction evaluation studies conducted in urban sound environments. In 2019, Hong et al. [196] validated three Ambisonic reproduction methods and tested their veridicality under a virtual sound environment related to the performance of reproduction methods. IVR has been shown to be a valid tool to simulate multisensory environments not only by acousticians but also in clinical neuroscience, cognitive psychology and other research fields. When researchers adopt the verisimilitude approach, they believe that the reproduction system and the subjective test have veridicality. In addition, there are also some difficulties to validate veridicality resulting from the complex contexts and unpredictability of outdoor sound environments. For outdoor sound environments, it is sometimes impossible to measure the real-world; e.g., a projected area without construction. Some contextual conditions cannot be changed independently in the real world as well.

It is notable that two studies addressed realism in their subjective experiments. The study by Jeon and Jo [195] in 2019 validated that the usage of HMD significantly increased the impact on the recognition of realism. In 2019, Hong et al. [196] conducted both in situ and laboratory experiments to assess the performance of different Ambisonic reproduction systems in perception. They both successfully assessed realism in their studies. The former de-emphasised the verisimilitude to the real world, and they underlined the realism difference brought by HMD compared with the non-HMD condition. The latter conducted a veridicality study with in situ responses, and they described the degree to which different reproduction approaches were similar to reality. When both verisimilitude and veridicality are examined, the most ecologically valid studies [178, 196] revealed the congruence between immersive virtual experience and real experience along with multisensory stimuli.

4.4.4 Limitations

An IVR system in soundscape or noise assessment should be adapted to the relationship between human cognition and subjective perception during the laboratory experiment. The diversity of IVR rendering techniques also brings an unnormalised experience to participants. An online survey has been introduced as a non-IVR tool to evaluate soundscape and noise perception [225]. Web-based VR was constructed in computationally cheap ways, and it could be improved with higher auralisation and visualisation quality. The one-to-one nature of tests also showed that the laboratory test cannot reach the sample size of traditional surveys. More economical and vivid reproducing systems following the development in hardware and software show higher veridicality.

HRTFs significantly contribute to localisation performance [185, 239], e.g., sound recognition of the direction and width of the source [195]. Compared with the non-HRTF rendering, i.e., the sound signal not convolved with the head-related impulse response (HRIR), the ratings of immersion, realism and externalisation are higher in the HRTF rendering [195]. Individualised and non-individualised HRTFs were utilised to assess various perceptual attributes by Simon et al. [240] in 2016. It is necessary to select a suitable HRTF that is well matched to the listener's own HRTF [240] to ensure ecological validity in terms of sound source localisation, and it can be an individualised HRTF or from an HRTF database. For different sound environments, such as a lively urban square with multiple water features, a quiet park or a park adjacent to a motorway, whether sound source localisation is considered a key feature or not [70, 241], the choice of an HRTF could differ in terms of ecological validity, and further studies are still needed.

At the moment, a head-tracking display system synchronising FOA-tracking binaural playback shows reliable validity under immersive virtual experiences for complex sound environment perception. Compared with FOA, HOA significantly improves the quality of this experience [242]. Different systems of HOA have already been implemented as hearing aids research for subjects with hearing loss [243, 244]. HOA is becoming popular in industrial applications such as Youtube360 and Facebook360 [245], and it shows great potential for the ecological validity of IVR in further urban sound environment studies.

4.5 Conclusions

This chapter aims to review the approaches to assess the ecological validity of IVR for the perception of urban sound environments and the necessary technologies during audio-visual reproduction ensuring ecological validity. The review qualitatively shows that IVR techniques have the potential to contribute greatly as an ecologically valid tool in soundscape or noise assessments. The ecological validity of IVR to assess urban sound environments is multimodal, dynamic and contextual. The main conclusions of this work reveal the following:

1. Through the approaches of laboratory tests including subjective response surveys, cognitive performance tests and physiological responses, the ecological validity of complex sound environment perception can be assessed for IVR. With participatory experiments in situ and in a laboratory, the veridicality of IVR can be verified through subjective responses including environmental preferences/quality, audio-visual indicators (e.g., pleasantness and annoyance), coupled interactions and reproduction quality (e.g., realism and immersiveness).

2. A head-tracking unit with a display and synchronised spatial audio (e.g., HMD with FOA-tracking binaural playback) is advantageous to assess ecological validity in immersive virtual environments. When the urban sound environment research involves interaction among multiple users, a CAVE system should be considered. With higher spatial resolutions, HOA also shows increasing potential for the ecological validity of IVR in urban sound environment research.

These studies on ecological validity with the utilised evaluation methods also go beyond the outcomes gained towards a normalised framework in soundscape and noise assessment protocols. For standardised soundscape evaluation, the ISO 12913 series should give more detailed guidelines and specifications on the establishment of reproduction. Binaural recording is mentioned in the ISO 12913-2 **??** for soundscape reproduction. For the purpose of soundscape reproduction, this type of binaural recording is is described as a method using an artificial head to record binaural static signals in the standard. However, monaural recording and binaural tracking rendering still have a wide range of applications during soundscape reproduction. The performance between binaural static and binaural tracking rendering has been explored by Guastavino et al. [73]. The performance gap between monaural and binaural static rendering is still unclear. This gap will drive us to reflect on these rendering strategies in soundscape evaluation that will be discussed in detail in the next chapter.

Chapter 5

Soundscape Rendering: Binaural or Monaural?

After identifying the ecological validity of virtual reality in urban sound environment studies, I will investigate the comparison of audio rendering methods. Choosing a suitable audio rendering method is an essential aspect to reproduce sound, as this is closely linked to our auditory system. This chapter[1] is to explore the performance of binaural and monaural audio rendering in soundscape evaluation. As discussed in ISO/TS 12913-2, binaural recording or rendering will refer specifically to the binaural static recording with headphone representations in this chapter. The monaural rendering will refer to the single microphone recording with headphone representations. Section 5.1 introduces the background of binaural and monaural methods in soundscape research. Section 5.2 illustrates the specific experimental methods utilised to verify the performance differences between the two rendering methods. Section 5.3 compares the results of the subjective evaluation with consideration of different indicators. Section 5.4 discusses the relationship among different results and other factors that may influence the performance of both methods. Section 5.5 summarises the difference between the binaural and monaural methods in soundscape evaluation and gives advice on how to select the appropriate method for assessments.

[1]This chapter was partially published in the Journal of the Acoustical Society of America [70].

5.1 Background

The initial discussion about monaural and binaural audio rendering is related to threshold sensitivity and hearing loss. In 1948, Pollack [246] reported that the binaural threshold was significantly lower than the monaural threshold under the equated intensity for two ears, and the difference between monaural and binaural threshold sensitivity for a pure tone is greater than noise. In subsequent psychoacoustic research, monaural and binaural perception modes were compared to find the discrimination under different signal types, signal-to-noise (S/N) ratios and reverberation conditions. S/N ratio refers to the ratio of the meaningful input power to the power of background noise. A higher S/N ratio normally means that the signal contains more useful input information, and vice-versa a lower S/N ratio contains more unwanted information which is also called noise. Hirsh [247] in 1971 concluded that binaural perception is not a great advantage than monaural in quiet or at high S/N ratios. More spatial information of binaural recordings is rendered for subjects, and useful signals could be extracted from original sounds mixed up with noise. When in quiet or at high S/N ratios, the discrimination of two rendering modes cannot be effectively distinguished by normal hearing subjects.

Speech perception is considered a significant component in hearing aids closely relevant to monaural and binaural perception. The effect of noise and RT for these two perception modes was discussed between normal and hearing-impaired subjects by Nabelek and Pickett [248] in 1974. RT is the time required for the sound pressure level to decay to a certain level as the sound is continuously reflected in a space. The time required for the sound to decay by 60 dB is usually noted as RT or T_{60}. They proposed advantages of binaural hearing in quiet or at high S/N ratios only for hearing-impaired subjects with the consideration of reflected speech energy in a closed space. RT is also an essential parameter reported by Nabelek and Pickett widely utilised in acoustic design especially for hearing-impaired people [248, 249]. Word intelligibility shows a slight decrease with a longer RT (0.3 and 0.6 s used in their study). The reverberation impairing people's speech perception in room acoustics has been a focus relevant to monaural and binaural audio rendering. Allen et al. [250] designed a multi-microphone signal processing system in 1977 to test the effect of binaural speech hearing by removing reverberation in a room. Therefore, this fundamental work provided an available method to compare two rendering modes under different reverberation. In 1982, Nabelek and Robinson [251] conducted a study based on monaural and binaural perception under different reverberation and age groups. The modified rhyme test (MRT) was processed in their research, and 5% binaural advantage was showed in MRT scores. The results suggested that more words in the MRT could be correctly identified by the subjects through binaural rendering and thus speech perception in a reverberation field could be improved. Moreover, research based on loudness between monaural and binaural rendering is reported by Marks [252] in 1987. A basic concept of binaural summation is that, for a pure 1,000 Hz tone, binaural rendering is twice louder than monaural rendering. From the perspective of soundscapes and psychoacoustics, the sound events in environments can be classified separately by binaural rendering, and directional rendering and selectivity play a significant role in auditory scene analysis [253].

Previous research on monaural and binaural rendering focuses on pure tone and speech, and in soundscape research, it is still significant to further investigate the effects of contextual components on monaural and binaural rendering, including spatial information, acoustic parameters, non-auditory stimuli, and other factors in a complex urban environment. The two rendering modes are primarily determined by the recordings whether they are binaural or monaural. Both binaural and monaural recordings are widely used in the assessment of urban soundscapes. The laboratory-based auditory test through rendering recorded sounds is conventional to model human perception on sound environments [47, 73, 234].

For noise control in urban spaces, monaural microphones are applied in sound level monitoring, and a set of objective single-value parameters could be obtained. Berglund and Nilsson [254], in 2004, assessed adverse perceived indicators. They utilized binaural and monaural recordings for a listening test, but there was no specific conclusion between two recordings in soundscape evaluation drawn from their research. Berglund and Nilsson [255], in 2006, used monaural recordings to calculate the acoustic parameters of soundscapes, and these results were correlated with perceived soundscape quality evaluation from a structured soundwalk. The sounds of birds and fountains were applied to improve the overall soundscape quality for freeways, minor, and major roads by Coensel et al. [256] in 2011. They designed and recorded a binaural birdsong sound by playing a monaural birdsong in a reverberation chamber, and mixed this artificial birdsong with soundscapes. Hao et al. [182], in 2016, conducted an assessment of the masking effect of birdsong for traffic noise, and monaural recordings were made in their research to compare occurrence frequencies of birdsong and the distance to a road. It was stated that owing to the high frequency of birdsong, binaural recordings, including more spatial characteristics, would cause more uncontrolled variables. For indoor sound environments, monaural sound sources and directivity-applied sound sources have been examined through recent research [195, 257].

Some research focused on soundscape categorisation conducted by binaural renderings. A principal components model was conducted to identify dimensions of soundscape perception by an auditory test through binaural recordings [82]. Rychtáriková and Gerrit, in 2013, [258] utilised binaural recordings to categorise soundscapes based on an automatic clustering algorithm. Binaural recordings were made by Jeon et al. [75] to assess cross-national urban soundscapes under different cultural backgrounds with the use of principal component analysis and cluster analysis. Other studies emphasised the perceptual qualities accompanied with typical urban sound environments. A head and torso simulator (HATS) was used to investigate the effect of natural sounds on traffic noise by binaural recordings [256]. Genuit and Fiebig [253], in 2006, recorded soundscapes with an artificial head to explore the use of psychoacoustics in the evaluation of soundscape quality, and they stated that the binaural recordings could reproduce aurally accurate acoustic scenarios. Soundscapes were binaurally recorded by Cain et al. [88] in 2013 to study the emotional dimensions, e.g., calmness and vibrancy. The soundscapes of three urban parks in Rome were assessed through binaural recordings to investigate a place with the higher sound level, and this place still led to a 'good' environment [259]. The binaural headphones worn by the operator connected with an audio recorder was utilised. Jambrošić et al. [260], in 2013, assessed urban soundscapes combining in situ surveys and these two recording methods, and binaural recordings were only used to calculate the sound level differences between two ears. Recently, an international soundscape standard ISO/TS 12913-2:2018 [58] proposed an approach for soundscape measurements based on binaural recordings. Spatial information could be recorded through the means of calibrated binaural measurement systems, and subjects in guided interviews would be presented with the same acoustic stimuli.

There is still a debate between the use of binaural and monaural recordings in soundscape evaluation, but the research discussed above showed no systematic conclusions explaining the performance of binaural and monaural recordings for various perceived indicators especially in outdoor sound environments. The subjective auditory test was conducted in this chapter to compare binaural and monaural audio renderings in soundscape evaluation. Monaural audio rendering is specified in this chapter as recording through a single microphone and delivering the single channel sound to the headphones. Binaural audio rendering is specified as recording through binaural microphones or an artificial head and delivering the binaural audio to the headphones, or as it is referred to, binaural static rendering. The binaural head-tracking rendering will be noted separately in this thesis. Soundscape indicators, psychoacoustics, and acoustic parameters were assessed in binaural and monaural audio rendering to determine the proper recording approach according to the auditory test.

The aim of this chapter is to explore (1) the overall comparison between two rendering methods on soundscape evaluation, (2) the relationship among different perceived indicators, (3) the relationship among various sites, (4) the effect of acoustic parameters, and (5) the effect of contextual parameters on soundscape evaluation given by the two rendering methods.

5.2 Methods

5.2.1 Site selection

The criteria for the selection of locations include: (1) sites with more diverse urban functions, e.g., education and relaxation, (2) sites with different sound levels and sound compositions, and (3) not all sites centralised in one area of the city. Twelve public sites of representative functions were chosen in Sheffield, United Kingdom, and the views, typical sound events, and functions in these sites are shown in Fig. 5.1. These sites crossed a wide geographical range in Sheffield (1,000 m × 2,500 m) from the railway station, university campus, city centre, city hall, cathedral to local parks. There is a distinct variation of acoustic performances, space functions, soundscape composition and building installations across these 12 sites. Water features exist in site 1 (Crookes Valley Park), site 2 (Weston Park), site 4 (City Hall), sites 5 and 10 (railway station), and sites 7 and 12 (Peace Gardens). Multiple water features, including the pools, fountains, and water curtains, are abundant in the centre of Sheffield, which are mixed with urban spaces for relaxation, recreation, culture, offices, etc. Site 11 (Winter Garden) is an indoor space open all the year round to the public, and it is also considered a public space visited by citizens with the functions for relaxation and culture. From the perspective of room acoustics, it should have the highest RT owing to its closed space and glass façades. Thus, it was chosen as a particular sample among other environments.

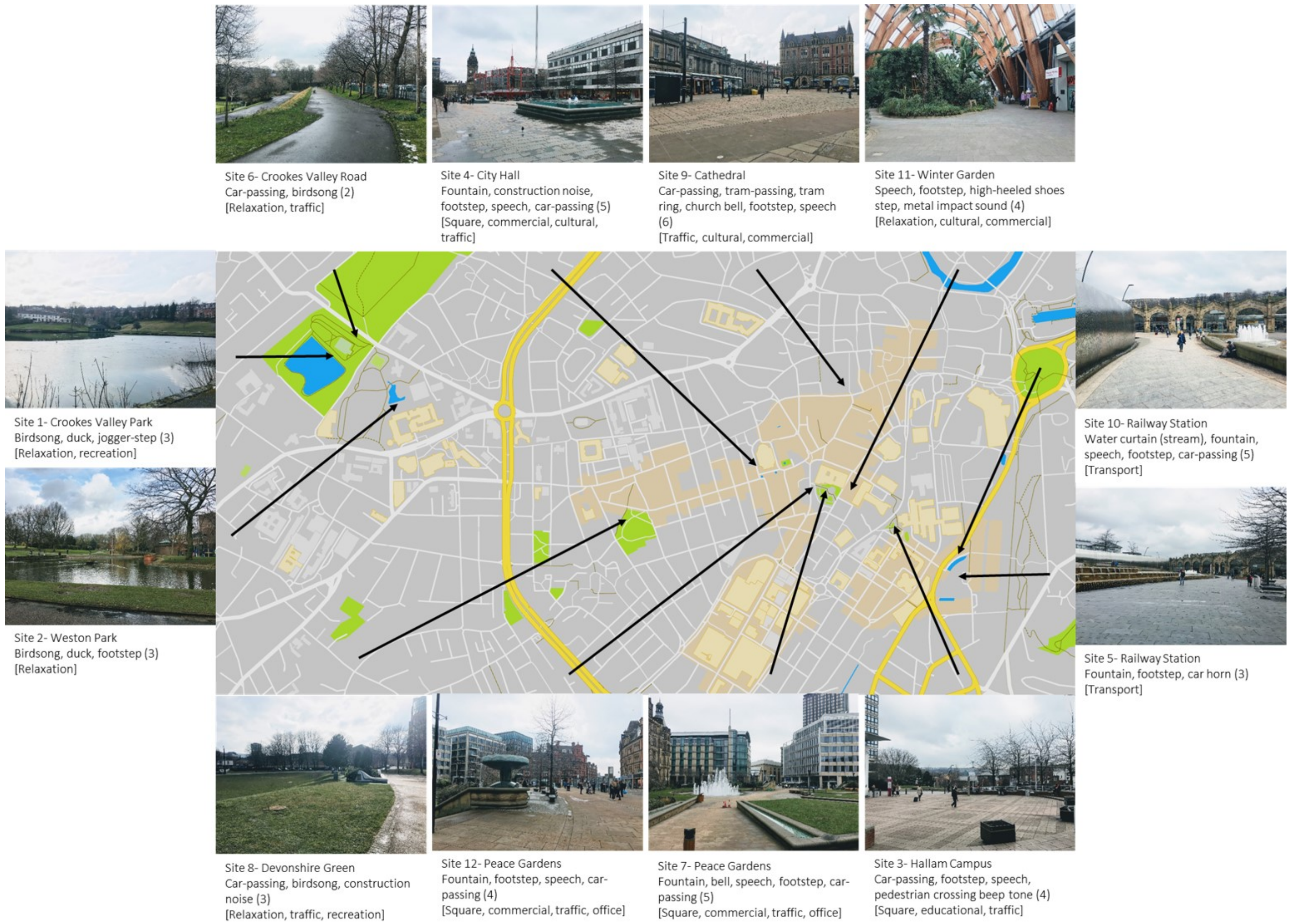


Figure 5.1: Twelve sites chosen in Sheffield with views and the description of sound events and representative functions. (The number of sound events in these spaces is counted in parentheses, and the functions of these public spaces are illustrated in brackets.)

5.2.2 Audio and visual recordings

At each location, the acoustic environment was recorded by a four-channel digital recorder (Roland R-44) connected with in-ear microphones (DPA 4060) for binaural signals and an omnidirectional microphone (BSWA MP201) for monaural signals simultaneously. Recorded sound events were the same for both recordings, and the duration of recordings was 3 min, containing typical sound events in these public spaces. The sampling rate was 48 kHz, and the depth of the recordings was 24 bits. A tripod fixed with the omnidirectional microphone was used to ensure the same height as the operator. Meanwhile, videos at these 12 locations were recorded by a camera (Sony Handycam DCR-DVD115E) as visual stimuli with the same head orientation as the operator. Residents in Sheffield should be familiar with the recorded views when they saw these videos.

5.2.3 Acoustic parameter measurements

Acoustic parameters were also recorded in these sites by a sound level metre (01dB Solo). The results and definitions of acoustic parameters are illustrated in Table 5.1. L_{Aeq} is the most commonly used sound level in environmental acoustics. The numerical order of sites illustrated in this study is determined by L_{Aeq}. The overall range of A-weighted equivalent sound pressure levels for these sites is from 49.9 to 70.8 dB. L_{A10}, L_{A50} and L_{A90} were used to evaluate the characteristics of the temporal statistical distribution of sound levels during the measurement period. $L_{\mathrm{A10}}-L_{\mathrm{A90}}$ reveals the temporal variability of sound in the selected sites. When this value is smaller, it indicates less variation in sound pressure in the time domain. Compared with L_{Aeq}, L_{Ceq} has a greater weighting for low frequency components. Thus, $L_{\mathrm{Ceq}}-L_{\mathrm{Aeq}}$ was used to describe low frequency contents of sound in the selected sites. $L_{\mathrm{A10}}-L_{\mathrm{A90}}$ and $L_{\mathrm{Ceq}}-L_{\mathrm{Aeq}}$ show a large variation presenting a wide range of sound environments chosen. The lowest L_{Aeq} is in site 1 (Crookes Valley Park) near the pool, and the highest is in site 12 (Peace Gardens) near Pinstone Street. The time duration of measurements was kept the same with soundscape recordings. All recordings and measurements were made during weekdays from 10:00 a.m. to 2:00 p.m. to ensure that these public spaces were recorded

Table 5.1: Acoustic parameters (dB), binaural sound level difference Δ (dB) and loudness (sone) in twelve sites.

Site	L_{Aeq}[1]	L_{A10}[2]	L_{A50}[3]	L_{A90}[4]	$L_{A10} - L_{A90}$	$L_{Ceq} - L_{Aeq}$[5]	Δ[6]	Loudness[7]
1	49.9	51.0	47.6	46.1	4.9	20.8	2.1	14.0
2	51.6	52.8	51.4	49.9	2.9	15.6	2.6	12.3
3	58.1	60.7	56.9	53.9	6.8	6.3	2.9	19.6
4	59.1	60.7	57.9	55.4	5.3	19.2	3.6	19.9
5	60.3	62.1	59.8	57.3	4.8	14.1	3.0	34.4
6	60.7	65.1	58.2	50.5	14.6	11.9	5.6	24.9
7	62.9	64.6	61.7	60.0	4.6	19.1	3.1	41.7
8	63.2	65.4	61.8	58.0	7.4	15.3	3.1	13.2
9	67.1	68.7	66.6	65.0	3.7	15.3	3.3	19.3
10	67.8	68.0	67.4	67.0	1.0	12.0	1.4	33.3
11	68.5	71.2	67.2	64.8	6.4	12.1	3.2	14.4
12	70.8	71.6	71.0	68.5	3.1	2.7	1.8	31.3

[1] Equivalent continuous, A-weighted sound pressure level.
[2] A-weighted, sound pressure level exceeded for 10% of the measurement period.
[3] A-weighted, sound pressure level exceeded for 50% of the measurement period.
[4] A-weighted, sound pressure level exceeded for 90% of the measurement period.
[5] L_{Ceq} refers to the equivalent continuous, C-weighted sound pressure level.
[6] The absolute value of the difference between L_{Aeq} of the two channels of binaural recordings.
[7] Perceived loudness with a unit (sone).

under their commercial, residential, cultural, relaxation, or transport functions. Moreover, the sound level difference of binaural recordings and loudness were calculated from recorded sounds by ArtemiS SUITE (HEAD acoustics), as shown in Table 5.1.

5.2.4 Reverberation time of sound environments

To investigate the perceived reverberance, the RT of these spaces was calculated by the empirical formula proposed by Kang [5, 261]:

$$\mathrm{RT} = \frac{0.16V}{-S_0 \ln(1-\bar{\alpha}) + 4MV}\left(88.6 + 49\alpha_b + 2.7\frac{\sqrt{LW}}{H}\right) \tag{5.1}$$

where V is the volume of the space, S_0 is the total surface area (m^2), $\bar{\alpha}$ is the mean absorption coefficient, M is the sound attenuation constant in air, α_b is the average absorption coefficient of boundaries, and L, W and H are the length, width and height, respectively, of urban areas. Absorbers mainly include the sky, trees, vegetation, bricks, etc. The RT for 12 sites was estimated as shown in Table 5.2 with the description of major absorbers.

Table 5.2: Major absorbers and RT for twelve sites.

Site	S_0 (m^2)	Major absorbers	M(dB/km)[1]	$\sqrt{LW}/H$	RT (s)
1	11310	sky, trees and vegetation	4.2	18.0	1.98
2	10850	sky, trees, vegetation and bricks	4.2	8.2	1.64
3	10460	sky, trees, ground and bricks	4.2	3.6	2.70
4	6500	sky, ground and bricks	4.2	4.2	2.74
5	8200[2]	sky, ground and bricks	4.2	18.5	3.26
6	11310	sky, trees, vegetation and ground	4.2	22.0	1.92
7	9630	sky, ground and bricks	4.2	3.6	3.37
8	11310	sky, trees, vegetation and ground	4.2	12.5	1.92
9	11870	sky, ground and bricks	4.2	4.2	3.38
10	8200	sky, ground and bricks	4.2	18.5	3.26
11	7200	trees, vegetation and ground	4.2	2.8	4.81
12	8900	sky, ground and bricks	4.2	3.4	3.43

[1] The value refers to ISO 9613-2:1996 [262] at 15 °C and 50% relative humidity.
[2] Same as site 10 due to the same semi-open space.

The applicability of the formula needs to be additionally pointed out that the areas should be diffusely reflecting boundaries with an absorption coefficient of 0.1–0.9, and the area should be from 400 to 40,000 m^2 [5, 261]. For busy areas in the city centre, the background noise is always kept at a high level. In this case, T_{60} is usually not measurable. The time required for the sound to decay by 20 dB noted as T_{20}, and RT is estimated by multiplying T_{20} by three [263]. During the in situ repeated measurements by puncturing balloons, T_{20} fluctuates too much for the large area sites, e.g., Crookes Valley Park and Devonshire Green. By calibrating locations including Winter Garden and Peace Gardens, the parameters of α_b and $\bar{\alpha}$ were extrapolated from these locations.

5.2.5 Evaluation indicators

Perceptual auditory indicators are key components describing how people perceive, experience, or understand a soundscape [83, 264, 265]. Previous studies [79, 266, 267] selected overall impression and acoustic comfort as major indicators to evaluate soundscapes. Therefore, the two indicators chosen in this study will examine binaural and monaural recording performances in overall soundscape evaluation. Pleasantness, annoyance, eventfulness, loudness, reverberance, and directionality were also addressed by numerous research to assess soundscapes [58, 61, 75, 79, 82, 267, 268]. Owing to the fact that laboratory-based studies cannot render all stimuli compared with in situ studies, realism affected by acoustic and non-acoustic factors is also addressed to explore the realism perception difference between binaural and monaural recordings in soundscape evaluation. These non-acoustic factors in sonic environments are also reported by plentiful research [158, 218, 267, 269].

The description of these nine indicators is shown in Table 5.3. These indicators represent the performance of the overall acoustic environment, perceptual attributes, reproduction, and technical specifications. Therefore, the selection of these indicators for comparing monaural and binaural rendering in soundscape evaluation will give a more comprehensive picture of the possible differences between the two rendering methods. For convenience, nine soundscape indicators are categorised into four groups: O, overall evaluation indicators (overall impression and acoustic comfort); G, generally perceived indicators (pleasantness, annoyance, eventfulness and loudness); R, reproduced perceived indicator (realism);

and T, technically perceived indicators (reverberance and directionality).

Table 5.3: Nine indicators used in the evaluation.

Indicator	Group	Description
Overall impression	O	The extent of the overall impression of the sound environment.
Acoustic comfort	O	The extent to which people perceive the comfort of sound in the external environment.
Pleasantness	G	The sound is perceived as pleasant or not.
Annoyance	G	The sound is perceived as annoying or not.
Eventfulness	G	The sound is perceived as eventful or not.
Loudness	G	The sound is perceived as loud or not.
Realism	R	The sound is perceived as realistic or not.
Reverberance	T	The sound is perceived as reverberant or not.
Directionality	T	The extent to which people perceive which direction sound is coming from.

5.2.6 Auditory experiment and procedure

The active noise-cancelling headphone (Bose QuietComfort 35) was used with a headphone amplifier (Lake People PHONE-AMP G109) connected to a laptop via an audio interface (Roland UA-101). The background noise in the acoustic laboratory was 34.0 dB(A), and the sound level was below 20.0 dB(A) for subjects. The MIT HRTF data [270] were employed for monaural rendering, and the azimuth and elevation angles ware set to 0°. The recorded videos were shared on a monitor by streaming with the laptop. Thus, the participants could evaluate sound environments according to visual and audio stimuli. VR videos were not recorded. When participants interact with the VR videos, monaural and binaural static rendering cannot allow for audio-visual synchronisation, i.e., 3DoF or 6DoF tracking. Although the 2D videos may reduce the realistic experience, such visual renderings avoided the possible confusion aroused from the unsynchronised audio-visual stimuli when participants gave their subjective responses. All sound recordings were calibrated through an artificial head (Neumann KU100) before the auditory experiment.

Five-point unipolar continuous category scales were used in evaluation questionnaires suggested by ISO/TS 12913-2:2018 [58], and the verbal labelling was provided below each scale as shown in Appendix B. Twenty-five subjects aged from 18 to 30 yr, living in Sheffield, gave their subjective evaluation to these sound environments, and they were familiar with these places. The total number of participants in the auditory test was in a considerable and conventional range according to previous soundscape research [84, 182, 267]. The hearing of subjects was tested before they gave their evaluation, and all subjects had normal hearing with the normal threshold for 125, 1,000, and 4,000 Hz. In addition, they all received simple acoustic training before the formal evaluation, and they had a basic understanding of acoustic indicators used in the evaluation. Such a selection would reflect the rendering performance from two methods of the local adult residents, and the results and analyses did not apply to children and people with hearing loss. The inter-rater reliability among the subjects is 0.896 (Cronbach's α). Meanwhile, the Spearman-Brown coefficient is 0.881 for the split half method. The reliability analyses reflect the high consistency of the subjects' evaluation results. The participants heard 12 pairs of sounds in total. They could directly compare two sounds by different rendering methods, one after the other, but the playback sequence of two sounds in each pair was randomised. The consent forms and appraisal forms were obtained from participants. The participant information sheet is attached in Appendix C.

5.2.7 Statistical analyses

In order to assess the correlation between subjective evaluation and acoustic parameters in this study, SPSS Statistics 24 and OriginPro 2017 were utilised to analyse Spearman's rho correlation coefficients, the independent t-test, and linear regression.

5.3 Results

The main results of soundscape subjective evaluation are presented in five parts: (1) the mean subjective evaluation comparison between binaural and monaural recordings, (2) the comparison among evaluation indicators, (3) the comparison among different sites, (4) the effect of acoustic parameters, and (5) the effect of contextual parameters, including sound events, binaural sound level difference, and RT, on soundscape evaluation by the two rendering methods.

5.3.1 Overall comparison between two rendering methods

Fig. 5.2 shows the correlation coefficients of mean ratings between binaural and monaural recordings and the mean subjective ratings of 9 indicators over 12 sites in Sheffield. The semantic rating scale is normalised from 0 to 1 (not at all — extremely). Thus, semantic responses can be quantified on the Y-axis from 0 to 1 as shown in Fig. 5.2. Twelve sites in spatial scales are considered unordered categorical variables, so Spearman's rho is utilised to analyse this rank correlation. For the evaluation of overall impression, acoustic comfort, pleasantness, annoyance, eventfulness, and loudness, binaural and monaural recordings are statistically significant (p<0.01) with the correlation coefficients over 0.5. There are also significant correlations for realism, reverberance, and directionality between these two rendering methods with lower coefficient values of 0.362 (p<0.01), 0.496 (p<0.01), and 0.243 (p<0.01), respectively.

According to the results of the independent t-test, the mean binaural subjective ratings of overall impression, acoustic comfort, pleasantness, annoyance, and eventfulness are approximate to the monaural. The t-test results also reveal that the mean rating differences are statistically significant (p<0.01) for loudness, realism, reverberance, and directionality. For these four indicators, the mean binaural ratings are 9%, 19%, 22%, and 39% higher than the monaural. Unsurprisingly, the overall realism and directionality subjective ratings of the binaural recordings are higher than the monaural. The audio information delivered by the monaural recordings is only through one channel less than the binaural.

The standard deviation of binaural and monaural recordings for overall im-

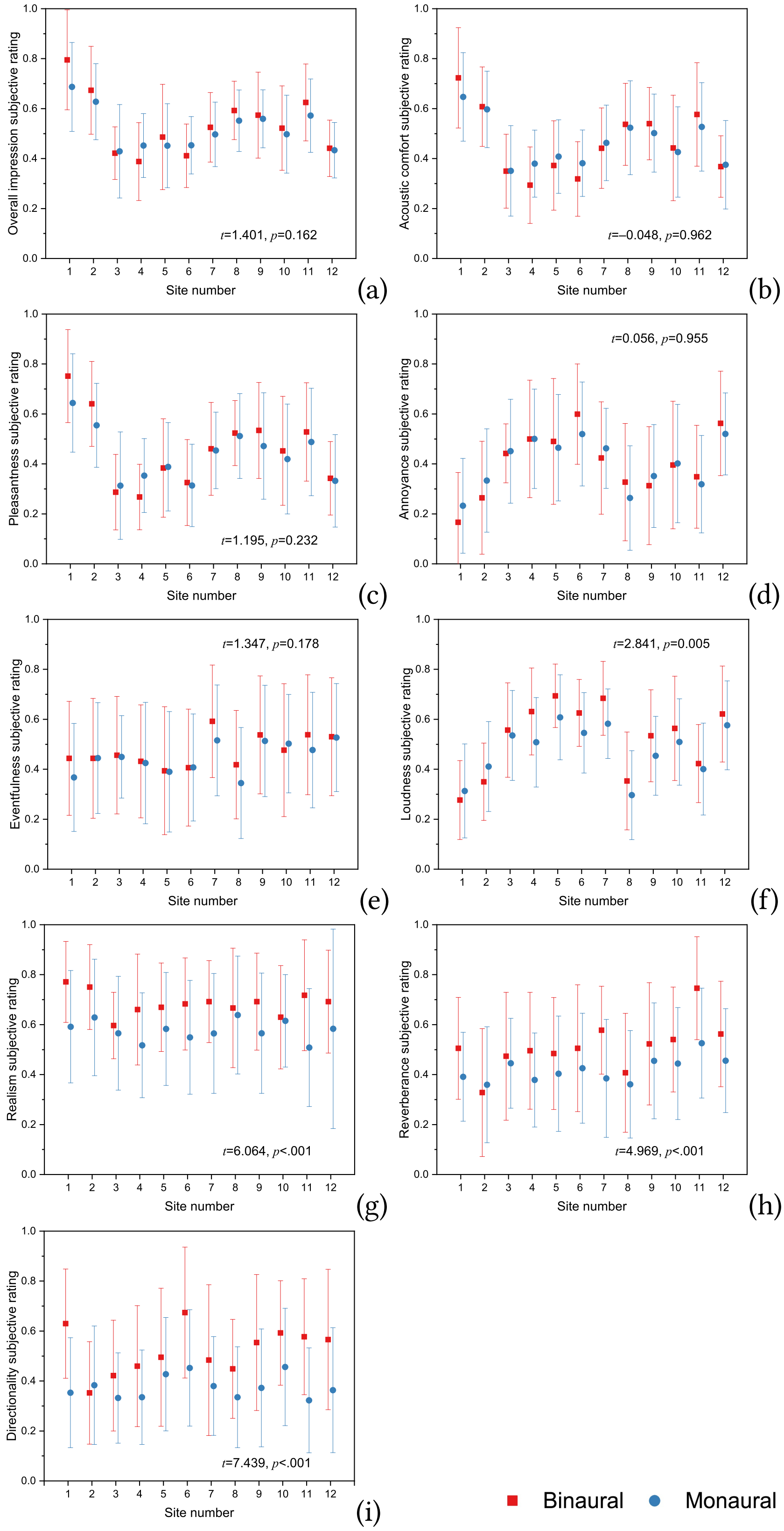


Figure 5.2: Comparison between binaural and monaural recordings in 12 sites with 9 perceived indicators.t and p values of the independent t-test and Spearman's rho coefficients are listed (**p<0.01). (a) Overall impression, r_s =0.614**; (b) acoustic comfort, r_s =0.569**; (c) pleasantness, r_s =0.670**; (d) annoyance, r_s =0.606**; (e) eventfulness, r_s =0.739**; (f) loudness, r_s =0.712**; (g) realism, r_s =0.362**; (h) reverberance, r_s =0.496**; (i) directionality, r_s =0.243**.

pression, acoustic comfort, pleasantness, annoyance, eventfulness, and loudness is approximately shown as error bars. It reveals that the overall variations of two rendering methods to evaluate these perceived indicators are similar, although these subjective fluctuations in some sites present a slight difference.

5.3.2 Comparison among evaluation indicators

A comparison between different evaluation indicators was performed by binaural and monaural ratings shown in Table 5.4. Correlations between overall impression, acoustic comfort, pleasantness, annoyance, and loudness are statistically significant (p<0.01) for the two rendering methods, and these positive or negative correlation coefficients are approximate with the diagonal contrast. There are also slight differences between two rendering methods in the interactions between these five indicators. Notably, the comparison between binaural and monaural recordings for acoustic comfort-annoyance ($r_s = -0.528$ and -0.439) and acoustic comfort-loudness ($r_s = -0.425$ and -0.323) implies that binaural recordings show a more negative tendency than the monaural for these two indicators' interactions. In addition, the correlation coefficients for eventfulness-loudness and reverberance-directionality under the monaural recordings are higher than in the binaural. The significant correlations between directionality and the other indicators in the binaural recordings are less than the monaural.

Table 5.4: Comparison of binaural and monaural ratings between indicators. (The lower left of the matrix is the correlation coefficients between different indicators in binaural recordings, and the upper right is the correlation coefficients between different indicators in monaural recordings.) $^{*}p<0.05$, $^{**}p<0.01$ (two-tailed test of statistical significance).)

Monaural	*Overall impression*	*Acoustic comfort*	*Pleasantness*	*Annoyance*	*Eventfulness*	*Loudness*	*Realism*	*Reverberance*	*Directionality*
Binaural									
Overall impression		0.724**	0.688**	-0.538**	0.042	-0.323**	0.203**	0.203**	0.119*
Acoustic comfort	0.775**		0.791**	-0.439**	0.063	-0.323**	0.236**	0.194**	0.181**
Pleasantness	0.763**	0.790**		-0.432**	0.112	-0.254**	0.284**	0.248**	0.205**
Annoyance	-0.593**	-0.528**	-0.565**		0.069	0.520**	-0.006	-0.077	0.002
Eventfulness	0.004	0.006	0.040	0.124*		0.486**	-0.240**	0.527**	0.086
Loudness	-0.422**	-0.425**	-0.358**	0.544**	0.148*		0.276**	0.147*	0.260**
Realism	0.170**	0.234**	0.304**	-0.121*	0.247**	0.142*		0.415**	0.414**
Reverberance	0.118*	0.164**	0.128*	0.008	0.310**	0.166**	0.194**		0.505**
Directionality	0.039	0.105	0.120*	0.004	0.141*	0.138*	0.392**	0.321**	

$^{*}p<0.05$, $^{**}p<0.01$ (two-tailed test of statistical significance).

5.3.3 Comparison among different sites

Ratings of each subject were taken into consideration, and Spearman's rho correlation coefficients and the independent t-test for each site are shown in Table 5.5. Not all sites show correlations with statistical significance between two rendering methods for the evaluation of overall impression, acoustic comfort, pleasantness, annoyance, and loudness in spite of their correlation coefficients in the overall comparison over 0.5 (p<0.01). All sites show correlations (p<0.01) for eventfulness, while the correlation coefficients ranged from 0.616 to 0.888. There are six sites whose mean realism differences are significant (p<0.05). These sites with typical acoustic scenarios can be perceived as entirely different through the rendering of two rendering methods resulting in mean rating differences for realism.

Eight sites show positive correlations (p<0.01 or p<0.05) between two rendering methods on reverberance. The mean binaural reverberance subjective ratings are 22% higher than the monaural, and the highest subjective rating of reverberance occurs in site 11 (Winter Garden) shown in Fig. 5.2 (h). Site 11 was installed with closed glass façades resulting in the longest RT, and indeed, this setup in 12 sites made subjects perceive the reverberance difference from binaural and monaural recordings. The mean reverberance difference on site 11 is also significant according to the result of the independent t-test (p<0.001). There is only one site with statistical significance in the directionality correlation, and seven sites present mean directionality differences with statistical significance (p<0.05). The mean rating difference of directionality is noteworthy, and binaural recordings still dominate directionality in the soundscape evaluation.

Table 5.5: Spearman's rho correlation coefficients (r_s) and independent t-test (t and p) between two rendering methods for each site. $^*p < 0.05$, $^{**}p < 0.01$ (two-tailed test of statistical significance).

Site number		1	2	3	4	5	6	7	8	9	10	11	12
Overall impression	r_s	0.585**	0.650**	0.544**	0.288	0.701**	0.516**	0.359	0.545**	0.525**	0.580**	0.571**	0.348
	t	0.227	1.330	-0.923	-1.621	0.355	-0.228	0.857	0.871	1.425	0.542	0.537	2.581
	p	0.825	0.199	0.370	0.119	0.729	0.824	0.405	0.398	0.169	0.598	0.601	0.015
Acoustic comfort	r_s	0.751**	0.494*	0.597**	0.282	0.400*	0.358	0.284	0.316	0.503*	0.398*	0.377	0.384
	t	1.632	0.202	-0.024	-2.117	-0.837	-1.422	-0.458	0.315	0.698	0.288	1.134	-0.180
	p	0.109	0.841	0.981	0.039	0.407	0.161	0.649	0.754	0.489	0.775	0.262	0.859
Pleasantness	r_s	0.675**	0.598**	0.558**	0.092	0.581**	0.509**	0.289	0.429*	0.697**	0.755**	0.435*	0.723**
	t	2.047	1.580	-0.608	-2.167	-0.084	0.251	0.121	0.254	1.070	0.530	0.829	0.210
	p	0.047	0.121	0.546	0.035	0.934	0.803	0.904	0.801	0.290	0.599	0.411	0.835
Annoyance	r_s	0.774**	0.489*	0.413*	0.447*	0.642**	0.597**	0.575**	0.653**	0.610**	0.514**	0.543**	0.344
	t	-1.383	-1.257	-0.140	-0.007	0.402	1.301	-0.589	1.199	-0.676	-0.086	0.536	0.733
	p	0.175	0.215	0.889	0.995	0.689	0.200	0.559	0.236	0.502	0.932	0.595	0.467
Eventfulness	r_s	0.720**	0.684**	0.856**	0.721**	0.836**	0.643**	0.777**	0.888**	0.663**	0.663**	0.711**	0.616**
	t	1.332	-0.019	0.109	0.102	0.062	-0.012	1.087	1.148	0.366	-0.393	0.973	0.057
	p	0.189	0.985	0.913	0.919	0.951	0.990	0.282	0.257	0.716	0.696	0.335	0.955
Loudness	r_s	0.472*	0.868**	0.606**	0.609**	0.293	0.692**	0.615**	0.879**	0.585**	0.637**	0.443*	0.540**
	t	-0.698	-1.237	0.408	2.469	1.774	1.686	2.390	1.084	1.657	1.005	0.543	1.079
	p	0.488	0.222	0.685	0.017	0.082	0.098	0.021	0.284	0.104	0.320	0.590	0.286
Realism	r_s	0.468*	0.518**	0.341	0.367	0.623**	0.325	0.265	0.139	0.649**	0.335	-0.068	0.459*
	t	3.404	2.195	0.459	2.341	1.401	2.312	2.213	0.316	1.954	0.259	3.607	1.862
	p	0.002	0.033	0.648	0.023	0.168	0.025	0.032	0.753	0.057	0.796	0.001	0.069
Reverberance	r_s	0.368	0.649**	0.410*	0.106	0.775**	0.641**	0.451*	0.438*	0.386	0.548**	0.451*	0.349
	t	1.822	-0.577	0.437	1.956	1.197	1.159	2.955	0.775	1.128	1.565	3.733	1.589
	p	0.075	0.567	0.664	0.056	0.237	0.252	0.004	0.442	0.265	0.124	0.001	0.119
Directionality	r_s	0.244	0.196	0.334	0.151	0.226	-0.083	0.388	0.569**	0.239	0.340	0.167	0.242
	t	4.815	-0.496	1.583	2.026	0.944	3.451	1.450	1.516	2.903	2.172	3.529	2.893
	p	<.001	0.622	0.120	0.048	0.350	0.001	0.154	0.136	0.006	0.035	0.001	0.006

5.3.4 Effect of acoustic parameters on subjective evaluation

Table 5.6 shows the correlation coefficients of binaural and monaural subjective ratings to acoustic parameters, including nine indicators, sound levels measured by the sound level meter, and loudness calculated by ArtemiS SUITE according to DIN 45631/A1 (HEAD acoustics). The sound pressure level correlated with both rendering methods is based on the measurement by a sound level meter.

L_{Aeq} is an essential acoustic parameter to describe a sound environment, and relatively low correlations were found on the relationship between L_{Aeq} and perceived indicators for both rendering methods. The agreement tendencies of overall impression, annoyance, loudness, reverberance, and directionality are higher for binaural recordings. The increasing temporal variability of $L_{\mathrm{A10}}-L_{\mathrm{A90}}$ causes the decline of overall impression and pleasantness with $r_s = -0.114$ ($p<0.05$) and -0.169 ($p<0.01$) for binaural recordings. $L_{\mathrm{A10}}-L_{\mathrm{A90}}$ shows no agreement tendency with reproduced indicator and technically perceived indicators. $L_{\mathrm{Ceq}}-L_{\mathrm{Aeq}}$ renders a series of significant correlations for two rendering methods. As the low frequency content becomes more dominant, perceived annoyance and loudness are likely to be higher. It is interesting to note that the highest correlation coefficients are 0.568 ($p<0.01$) for the binaural subjective ratings and 0.442 ($p<0.01$) for the monaural regarding the correlation between perceived loudness and calculated loudness. Moreover, the calculated loudness has higher correlation coefficients for the two rendering methods compared with other conventional sound levels. This reveals the importance of psychoacoustic parameters, especially loudness in the soundscape evaluation.

As a result, statistically significant correlations were obtained between acoustic parameters and perceived indicators. Binaural recordings are more sensitively correlated with $L_{\mathrm{Ceq}}-L_{\mathrm{Aeq}}$ than the monaural on overall impression, acoustic comfort, pleasantness, annoyance, and loudness. The time variability of $L_{\mathrm{A10}}-L_{\mathrm{A90}}$ does not significantly affect indicator ratings. Only a few acoustic parameters show significant correlations with realism and directionality, and these objective acoustic parameters would not directly affect realism and directionality ratings.

Table 5.6: Spearman's rho correlation coefficients between acoustic parameters and subjective ratings of two rendering methods. *p<0.05, **p<0.01 (two-tailed test of statistical significance).

Indicator		L_{Aeq}	$L_{A10}-L_{A90}$	$L_{Ceq}-L_{Aeq}$	L_{A10}	L_{A50}	L_{A90}	Loudness[1]
Overall impression	Binaural	-0.127*	-0.114*	0.325**	-0.134*	-0.131*	-0.151**	-0.365**
	Monaural	-0.104	-0.080	0.276**	-0.105	-0.115*	-0.129*	-0.319**
Acoustic Comfort	Binaural	-0.069	-0.107	0.293**	-0.068	-0.080	-0.098	-0.418**
	Monaural	-0.028	-0.001	-0.031	-0.035	-0.030	-0.028	0.041
Pleasantness	Binaural	-0.105	-0.169**	0.337**	-0.110	-0.108	-0.126*	-0.378**
	Monaural	-0.090	-0.098	0.307**	-0.103	-0.089	-0.096	-0.282**
Annoyance	Binaural	0.166**	0.101	-0.297**	0.171**	0.160**	0.143*	0.337**
	Monaural	0.073	-0.009	-0.198**	0.070	0.076	0.094	0.335**
Eventfulness	Binaural	0.142*	-0.101	-0.004	0.130*	0.135*	0.159**	0.061
	Monaural	0.158**	-0.170**	-0.081	0.147*	0.160**	0.192**	0.119*
Loudness	Binaural	0.146*	-0.045	-0.210**	0.131*	0.173**	0.216**	0.568**
	Monaural	0.076	-0.117*	-0.207**	0.063	0.103	0.145*	0.442**
Realism	Binaural	-0.058	-0.027	0.134*	-0.054	-0.071	-0.081	-0.123*
	Monaural	-0.026	-0.091	-0.012	-0.028	-0.010	-0.008	-0.009
Reverberance	Binaural	0.236**	0.032	-0.089	0.226**	0.213**	0.202**	0.142*
	Monaural	0.133*	-0.006	-0.125*	0.142*	0.125*	0.119*	0.036
Directionality	Binaural	0.127*	0.045	-0.072	0.140*	0.112	0.067	0.091
	Monaural	0.001	-0.070	-0.036	-0.001	0.013	0.001	0.100

[1] Loudness of recorded signals calculated by ArtemiS according to DIN 45631/A1.

5.3.5 Effect of contextual parameters on subjective evaluation

Eventfulness, realism, and reverberance are correlated with contextual parameters, respectively, including sound events, sound level difference, and RT.

Eventfulness is a particularly perceived indicator not closely correlated with overall impression, acoustic comfort, pleasantness, and annoyance. Thus, the relationship between the number of sound events and eventfulness subjective rating is analysed in Fig. 5.3 (a). The results of linear regression present that there is a positive correlation for both binaural and monaural rendering methods with $R = 0.444$ and 0.497, respectively. The overall subjective ratings for binaural and monaural rendering methods are approximate, and the proximate correlation coefficients in linear regression between the number of sound events and perceived eventfulness are reasonable. The meaning of 'eventful' used in the auditory test will cause vagueness for subjects. Although there is a positive correlation between the number of sound events and perceived eventfulness, the types of sound events and the sound level in environments are also involved with the evaluation of eventfulness. Although there is a positive correlation between the number of sound events and perceived eventfulness, the types of sound events and the sound level in environments are also involved with the evaluation of eventfulness. Fig. 5.3 (b) presents the linear regression between the binaural sound level difference as shown in Table 5.1 and the realism subjective rating difference. Indeed, the binaural sound level difference is an impact factor affecting perceived realism, but realism is a reproduced indicator involved with multiple stimuli. Sound environments are time dependent, and the single-value binaural sound level difference cannot fulfil the realism gap between binaural and monaural rendering methods. Eventfulness and realism are more complicated in the auditory test with the involvement of the sound composition, sound levels, personal understandings, and other non-acoustic factors.

Reverberance and directionality are categorised as technical perceived indicators in Section 5.2.5. The relationship between RT in these public spaces and

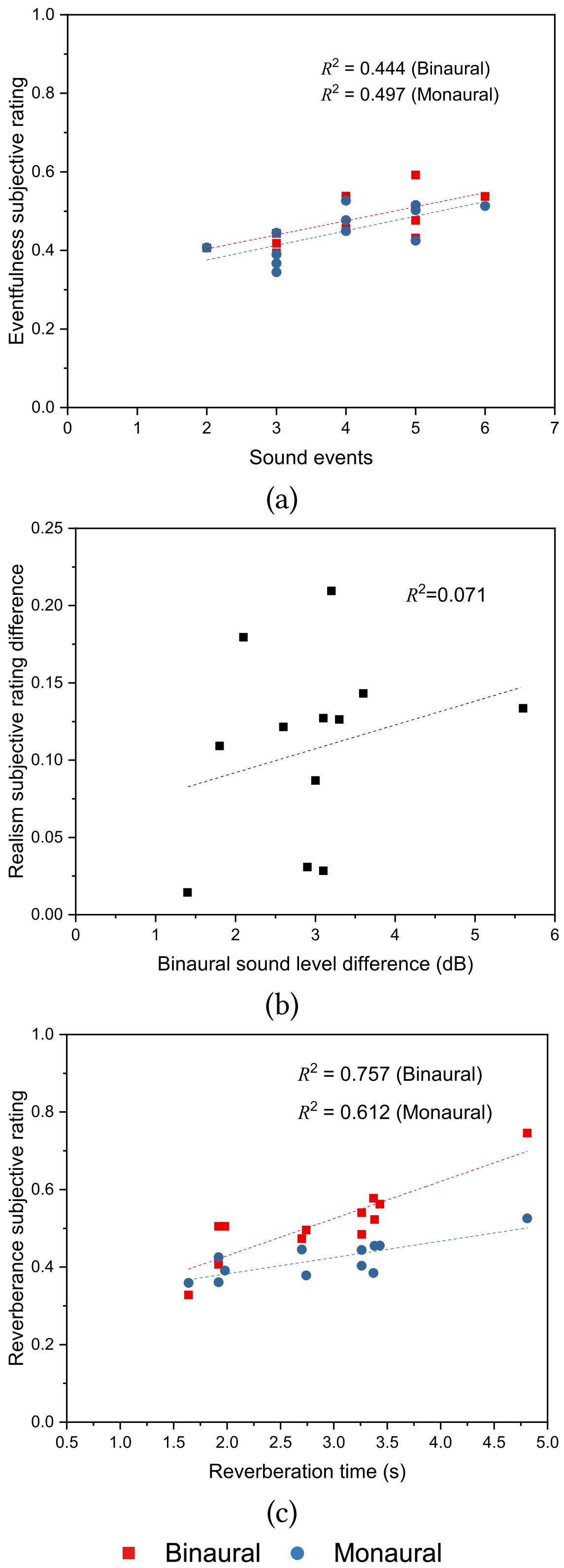


Figure 5.3: Relationships between (a) the number of sound events and eventfulness subjective rating, (b) the binaural sound level difference and realism subjective rating difference, and (c) RT and reverberance subjective rating.

perceived reverberance is shown in Fig. 5.3 (c). The coefficients of determination are 0.757 and 0.612 for binaural and monaural rendering methods, respectively. It needs to point out that the RT has some errors determined by parametric extrapolation, and the estimations are inevitable. However, for a significantly reverberant space, e.g., Winter Garden with the RT longer than 4-5 s, the variation of the two rendering methods occurs obviously. Reverberance in such public spaces is dominantly perceived through binaural recordings.

5.4 Discussions

5.4.1 Overall comparison among different analyses

A summary of binaural and monaural rendering methods of different analyses is illustrated in Table 5.7. The comprehensive performance of the two rendering methods on overall impression, acoustic comfort, pleasantness, annoyance, eventfulness, and loudness is similar. Eventfulness is an indicator depending on how subjects understand the meaning of eventful, and it presents a different tendency on the comparison among other usual indicators. It also revealed that eventfulness was previously classified into independent scales compared with pleasantness and annoyance [75].

Table 5.7: Summary of binaural and monaural recordings over different analyses. $^{*}p<0.05$, $^{**}p<0.01$ (two-tailed test of statistical significance).

Group[1]	Indicator	r_s^2	Rating difference[3]		Site (n/12)[4]	Indicator comparison (h/8)[5]		Acoustic parameters (i/7)[6]	
			t	p		Binaural	Monaural	Binaural	Monaural
O	*Overall impression*	0.614**	1.401	0.162	9/12**	5/8** 1/8*	6/8** 1/8*	3/7** 4/7*	2/7** 2/7*
	Acoustic comfort	0.569**	-0.048	0.962	2/12** 4/12*	6/8**	7/8**	2/7**	0/7
G	*Pleasantness*	0.670**	1.195	0.232	8/12** 2/12*	5/8** 2/8*	7/8**	3/7** 1/7*	2/7**
	Annoyance	0.606**	0.056	0.955	8/12** 3/12*	4/8** 1/8*	4/8**	6/7** 1/7*	2/7**
	Eventfulness	0.739**	1.347	0.178	12/12**	2/8** 3/8*	3/8**	1/7** 3/7*	4/7** 2/7*
	Loudness	0.712**	2.841	0.005	9/12** 2/12*	6/8** 2/8*	7/8** 1/8*	5/7** 2/7*	2/7** 2/7*
R	*Realism*	0.362**	6.064	<.001	3/12** 2/12*	5/8** 3/8*	7/8**	2/7*	0/7
T	*Reverberance*	0.496**	4.969	<.001	4/12** 4/12*	5/8** 2/8*	6/8** 1/8*	4/7** 1/7*	5/7*
	Directionality	0.243**	7.439	<.001	1/12**	2/8** 3/8*	5/8** 1/8*	2/7*	0/7

[1] Indicator classification in Section 5.2.5.

[2] Spearman's rho between two rendering methods for 12 sites.

[3] t and p values of the independent t-test between two rendering methods.

[4] n is the number of correlation coefficients with statistical significance between two rendering methods for 12 sites.

[5] h is the number of correlation coefficients with statistical significance correlated with other eight indicators in the comparison between indicators.

[6] i is the number of correlation coefficients with statistical significance between subjective ratings and seven acoustic parameters.

5.4.2 Directionality of binaural and monaural rendering

Directionality in urban sound environments is still a complicated indicator associated with multiple objective and subjective factors. Most sound sources in public spaces are not stationary. The sound levels and positions of these sound sources will vary with time, and these movable sources, such as pedestrians or cars-passing within a close distance could also generate higher perceived directionality. The variations of sound source strength, numbers, frequency contents, direction, distance, and other conditions will all contribute to the evaluation of directionality. The subjects in this study could give their perception for a sound environment depending on their hearing localisation abilities and familiarity to these spaces.

As expected, the subjects cannot perceive any directivity by listening to monaural recordings, but a certain number of participants did imagine it as shown in Fig. 5.2 (i). Several factors jointly influence this phenomenon: (1) the subjects are familiar with these public spaces, and they imagine the directionality generated from sound sources; (2) visual stimuli rendered by videos imply the orientation of sound sources in these environments; and (3) the subjects perceived the strong directionality from certain sites during the experiment indeed, and they cannot distinguish the environments with low directivity. For instance, there was no strong directional sound source near the operator in site 2. The subjects would not distinguish these two sounds recorded by two methods under this scenario. The mean rating of directionality for site 2 is approximate, and the mean difference of subjective ratings is also not significant (p=0.622) as shown in Table 5.5. In addition, the adjectives of 'directional' and 'reverberant' are also jargon for subjects to some extent, and this may also result in the bias in evaluation results.

Furthermore, moving sound sources like birds in site 1 and cars in site 6 have a significant impact on directionality given by binaural recordings. For site 1, the ducks were close to the pool bank and these sound events occurred near the operator. Moving cars were running parallel to the direction of the operator in site 6. These moving sound sources dominate these two sound environments, and the sound localisation of the subjects could detect these sound compositions through binaural audio rendering. Sound localisation from moving sources in urban public spaces is also an essential factor on directionality, and the large subjective difference was found among these acoustic scenarios.

Interaural cross correlation (IACC) is one of the binaural acoustic parameters used to analyse spatial impression and characteristics of sounds, having been applied in various spaces, including concert halls [271], high-speed train noise [272], and urban soundscapes [86]. The correlation between the IACC of the early sound field within 80 ms and directionality was examined and no significant correlation was found (binaural: $r_s = -0.048$, p=0.407; monaural: $r_s = -0.060$, p=0.302). The IACC of the late sound field after 80 ms was also not significantly correlated with directionality (binaural: r_s =0.017, p=0.768; monaural: $r_s = -0.030$, p=0.599).

5.4.3 Multi-factorial interaction on realism

There is a notable difference between two rendering methods for realism generated from the multi-factorial interaction. Realism is influenced by the two rendering methods in typical sites, and it is involved with different sound contexts, building installations, and public functions. Site 1 (Crookes Valley Park) and site 11 (Winter Garden) have the largest mean subjective differences in realism between the two rendering methods, and the t-test in Table 5.5 also shows that the mean differences for these two sites are significant (p<0.01). Site 1 has the lowest sound level and the highest overall impression, acoustic comfort, and pleasantness ratings. Natural sound events, e.g., ducks and other water birds near the pool bank, will attract subjects' attention and increase the overall positive ratings. Meanwhile, owing to these ducks and water birds being within a close distance, subjects will be easily able to distinguish the difference of realism between the two rendering methods. They could perceive the environment as more real under the dominant natural sound events with the low background noise in spite of its strong directivity of the sound source by binaural recordings. For site 11 having the highest RT among all other sites, binaural recordings will increase the sense of localisation and spaciousness resulting from reflected sounds recorded by binaural microphones. The multi-factorial interaction among different indicators reflects the internal connection between objective environments and the subjective evaluation given by the rendering methods.

In addition, other physical conditions, including lighting, vibration, temperature, and other factors, which have not been tested in this chapter, will also add to 'realism'. These conditions do potentially have multiple and significant impacts on realism in soundscapes. Therefore, although binaural recordings performed better than monaural recordings shown in Fig. 5.2 (g), the subjective ratings for realism did not reach maximum scores.

5.5 Conclusions

This chapter examined the performance of two rendering methods in soundscape evaluation. The subjective evaluation and comparative analyses of indicators and sites along with the effects of acoustic and contextual parameters revealed the following:

1. Binaural and monaural rendering methods showed good agreement in mean ratings of overall impression, acoustic comfort, pleasantness, annoyance, eventfulness, and loudness. In contrast with the monaural results, overall binaural subjective ratings were significantly higher in realism, reverberance, and directionality evaluations.

2. The two rendering methods were correlated with different perceived indicators in a similar way. The correlations between overall impression, acoustic comfort, pleasantness, annoyance, and loudness are statistically significant for the two rendering methods.

3. Most sites showed no correlations in directionality between the two rendering methods. It revealed that these two methods performed differently for the evaluation of directionality in most urban spaces.

4. The A-weighted sound pressure level had a weak impact on soundscape evaluation for both rendering methods.

5. The correlation between eventfulness and the number of sound events was similar with the two rendering methods. The difference in realism generated from two rendering methods did not significantly depend on the binaural sound level difference. Reverberance was perceived as more consistent with RT through binaural recordings in soundscape evaluation.

Overall, this chapter suggests that monaural recordings are sufficient to evaluate most soundscape indicators including overall impression, acoustic comfort, pleasantness, annoyance, eventfulness, and loudness. When some special acoustic scenarios (e.g., moving birdsongs or cars passing near the subjects) and built environment (e.g., RT longer than 4-5 s) occur in soundscapes, the corresponding perception, i.e., directionality and reverberance, would be much better evaluated by binaural recordings.

Chapter 6

Sound Propagation: Reducing Reflection Orders for Auralisation

This chapter[1] is to investigate how much enhancement VR will bring at a relatively low reflection order compared with binaural static audio-only rendering, and then to assess the effect of perceived indicators, sound types, and urban squares with differentiated areas and layouts under various reflection orders. Section 6.1 gives the background to sound reflection in room acoustics and urban environments. Section 6.2 introduces the visualisation and auralisation methods involved with different reflection orders to conduct the subjective test in the laboratory. Section 6.3 first compares the test results between two audio-visual conditions under different reflection orders, and then assess the perceived indicators, sound types, and layouts of urban squares under different orders. Section 6.4 discusses the significance of area and layouts of squares, the limitation in this chapter and the implementation for different roles relevant to sound design. Section 6.5 summarises the influence of reflection orders on human perception in relation to perceived immersion, realism and reverberance under a virtual urban environment.

6.1 Background

In the 1850s, the pioneering work to understand how people perceive sound reflections that arrive within a short time after the direct sound was explored [273]. For

[1]This chapter was partially presented at the International Congress on Acoustics 2019 [71].

audible sound, the reflection of sound is one of the important processes in sound propagation, regardless of the environment in which we live. In room and auditorium acoustics, sound reflection significantly affects the acoustic performance of the space designed for lectures or concerts. Clarity, reverberance [274] and speech transmission quality [275] are all involved with sound reflection in acoustic design. Both positive and negative effects of sound reflections were found on speech intelligibility in room acoustics [276]. In an urban environment, various public spaces connect people together from different occupations, families, and educational backgrounds. These spaces with multiple functions provide the necessary places for human activities. Among these public spaces, squares often exist as landmarks in a city and attract large crowds. As early as the 1960s, researchers started to study the issue of sound propagation in urban open public spaces [277]. The numerical simulation and simplifications were examined in urban squares [100] and urban streets [5, 101, 191, 278, 279], and these studies parametrically explored sound field and sound propagation in different kinds of urban spaces.

For outdoor sound environments, layouts of urban spaces, sound contexts, and propagation conditions are more complex. The presence of these issues can become problems in the perception of soundscape and the reproduction of a sound environment. Meanwhile, the human-environment interaction based on VR experiences requires high synchronisation for public participation. As discussed in Section 2.4.4 for head tracking latency, ensuring undetectable delays for users is a critical task for VR engineers. In the case of 3DoF or 6DoF, the processing of spatial audio is required to be faster to satisfy audio-visual synchronisation. Compromises and simplification during auralisation and reproduction are inevitable. Human perception in such spaces is highly significant in judging the results of auralisation and reproduced environments. It is necessary to auralise sound environments efficiently and accurately with potential simplified solutions. These solutions aim to achieve an immersive experience with fewer computational iterations. Previous research explored reflection simplifications with the audio-only condition [280]. The interactive audio-visual stimuli should be paid more attention on sound environment evaluation, as 3DoF or 6DoF rendering will deliver a more immersive experience. While there are many studies [281] focusing on numerical simulation of urban public spaces, few studies have been conducted on sounds with different reflection conditions in urban public spaces. These studies focused on the fast simulation based on different reflection layouts [183, 184], and the interpolation technique in audio-visual animation of urban space. In 2010, Richmond et al. [282] reported a linear interpolation method to auralise sound samples at discrete locations. As the subject moves across the square in a straight line, the discrete samples of sound on this trajectory are linearly interpolated between the two nearest points. Thus, the subjects experienced continuous sound stimuli in pre-calculated audio-visual renderings.

This chapter will use a subjective appraisal method, combined with VR experiences, to evaluate reverberance and reproduced indicators of sounds in different urban squares. This chapter aims to examine, compared with binaural static audio-only rendering without head tracking, how much enhancement VR (HMD and binaural tracking) will bring at a relatively low reflection order, and then to assess perceived indicators, sound types, and urban squares with differentiated areas and layouts under various reflection orders in VR experiences.

6.2 Methods

6.2.1 Site selection

For urban outdoor environments, squares are often considered open spaces of recreation and relaxation for citizens. Squares have different spatial scales, enclosing buildings, environmental contexts, and acoustic characteristics. Compared with the crowded roads where annoyance is dominated by car noise, the composition of sounds in squares is more diverse. The research of such different squares represents a wide category of outdoor environments covering multiple sounds and architectural spaces. To investigate the audio-visual interaction under VR experience in urban spaces, four squares distributed in central London were selected, as shown in Fig. 6.1: the campus square behind the Wilkins Building in Univer-

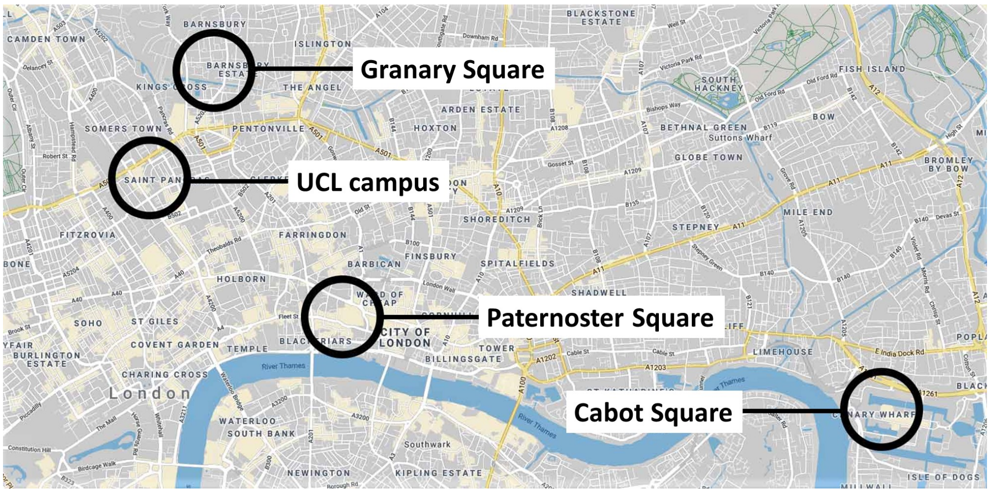


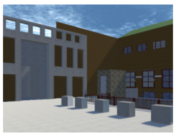
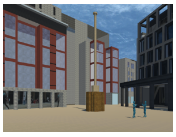
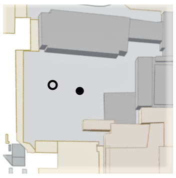
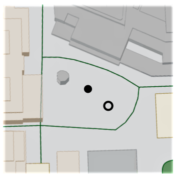
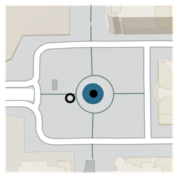
Figure 6.1: Geospatial distribution of the four squares in London.

sity College London, Paternoster Square south to St. Paul's Cathedral, Granary Square next to Caravan King's Cross, and Cabot Square in Canary Wharf. The areas of these four squares are 900 m^2, 1,700 m^2, 7,000 m^2, and 6,000 m^2. The mean heights of the buildings surrounding these four squares are 18 m, 25 m, 20 m and 65 m respectively. The square behind the Wilkins Building in University College London was chosen as the rectangular square is fully surrounded by building facades. Paternoster Square is an urban square next to St Paul's Cathedral, the prime tourist locus in London, and was chosen due to being enclosed by buildings with multiple facades and textures, which is not fully surrounded by buildings. Granary Square is a large open square in King's Cross, and it has multiple functions as a public space for education, relaxation, and retail. Cabot Square is located in Canary Wharf, which is an area famous for being a hub for significant office and commercial estates in London. Amongst these squares, UCL campus and Cabot Square have an aspect ratio of nearly 1:1. Paternoster Square and Granary Square are irregularly shaped squares.

According to the area and perimeter of the square, the 2D enclosed ratio r_{2D} is defined as the length of the surrounding buildings l divided by the perimeter of the square p:

$$r_{2D} = \frac{l}{p} \quad , \tag{6.1}$$

Table 6.1: Site selection comparison and spatial information of the squares.

Site name	UCL campus	Paternoster Square	Granary Square	Cabot Square
Photograph				
Reproduced scene				
Top view				
Figure-ground[1]				
Site No.	1	2	3	4
S (m^2)	900	1700	7000	6000
h (m)	18	25	20	65
r_{2D}	100%	70%	40%	55%
r_{3D}	71%	50%	25%	41%
Aspect ratio	1:1	irregularly shaped	irregularly shaped	1:1
Area functions	Catering Relaxation Education	Retail Sightseeing Office	Retail Relaxation Education	Retail Relaxation Office

[1] (●—source location, ○—observer location)

where l is measured in the figure-ground of the squares. Furthermore, the 3D enclosed ratio r_{3D} of a square is defined as follows:

$$r_{3D} = \frac{hl}{hp + S} \quad , \tag{6.2}$$

where h is the mean height of the surrounding buildings of a square, and S is the area of a square. According to the measurements and calculations, the spatial information of the four urban squares including their top views, figure-ground, S, h, r_{2D}, r_{3D}, aspect ratios, and area functions, is presented in Table 6.1.

6.2.2 Visualisation

Originally, both Cabot Square and Granary Square featured fountains. There were no fountains in UCL campus square and Paternoster Square. To investigate the sounds of fountains in all squares, extra fountains were virtually placed in these two squares. Meanwhile, only one fountain design was used for all squares to avoid subjective evaluation differences due to the variations in fountain appearance. The design of the fountain was based on a simplified version of the circular fountain in Cabot Square with a radius of eight metres. The original dimensions of this fountain may be too large when it was placed in two small squares, i.e., UCL campus and Paternoster Square, and the size of this fountain was uniformly reduced in all squares to a radius of three metres. The area to the south of Granary Square is a sunken plaza, and this area cannot be seen from the evaluation location at the northeast corner of the square. Thus, simplifications for this square were also made only considering the area of the square on the same level at the evaluation point. In order to eliminate the undesired scattering and absorption, limited vegetation exists in these squares, and it was not reproduced during visualisation. The visualisation of these four squares was done through the modelling software (SketchUp Pro 2018). The photographs and reproduced scenes of the four squares are shown in Table 6.1.

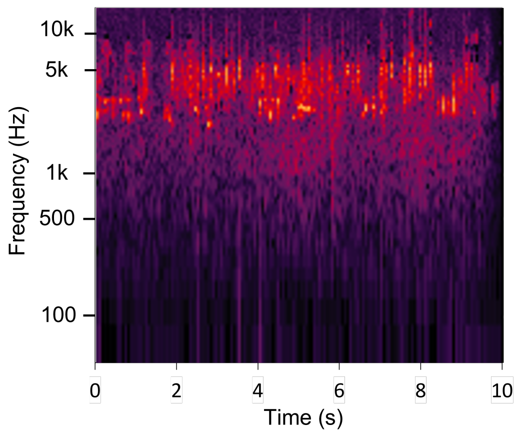
6.2.3 Auralisation

To investigate the acoustic behaviours of different sounds in these squares, several typical sounds were chosen including birdsong, clapping and the sound of a fountain (henceforth to be referred to simply as *fountain*). These three types of sounds have different acoustic characteristics. Birdsong is a natural sound, originating from animals in nature with a positive effect on urban sound environments. Birdsong has the high frequency contents of specific pitches. Clapping is a sound produced by humans, and it is discrete in the time domain analogue to an impulsive sound in this chapter. The fountain is often used as part of a landscape to beautify and decorate environments, and its sound is considered to be a natural sound as well. The water sound generated from a fountain was found to improve the acoustic environment affected by traffic noise [117, 283]. Thus, the investigation of water sound is of particular importance in soundscape research with the perspective of sound as a resource. The sound of the fountain is continuous in the time domain, and there is generally no pitch associated with it. These three sounds also represent different kinds of time-frequency characteristics. Monaural recordings were made with the birdsong and fountain recorded in Richmond park and Cabot Square respectively in the morning without cars or pedestrians. The distance from the recording position to the Cabot fountain was within half a metre to ensure that the direct sound was recorded. The clapping was recorded in a semi-anechoic room. The original audio quality for the three sounds was 24 bit / 44.1 kHz, and these sounds were all adjusted to 65 dB(A). Figure 7.2 shows the spectra of the three sounds.

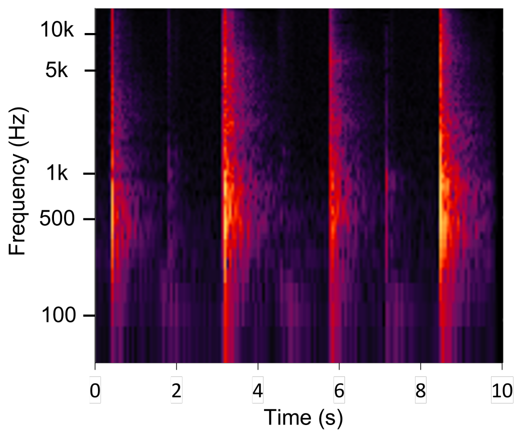
ODEON (9.2 Auditorium) was used to render the acoustic performances of these public squares. The process of auralisation was also based on the impulse responses generated from ODEON. The boundary absorption and diffusion conditions were assigned with different parameters to model the real sites. The hybrid method in ODEON could deal with the complex boundary conditions [284] and generate point-to-point B-format impulse responses.

A previous study on simplification through subjective tests validated the reflection orders of 5, 20, and 50 with aural stimuli only [280]. In this work, four reflection orders, i.e., 1, 5, 20 and 1,000, were chosen, and these four different orders were simulated in ODEON. The distance between the sound source and the receiver in the four squares was set to 8 m, as shown in Table 6.1, and it was a reasonable distance when a participant stood in these squares to observe the events. Thus, each square had a defined evaluation point and a sound source point.

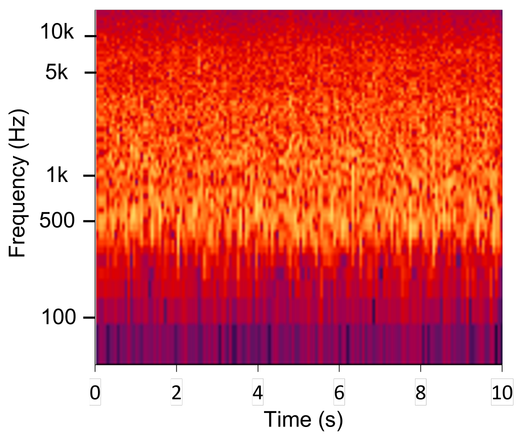
The game engine (Unity) was used to synthesise visualisation and auralisation. The lighting condition was set to a reasonable solar zenith angle and illuminance according to their geometrical locations. To synchronise the virtual visual-audio environment, a particle system of water splash was attached to the fountain, and the animation of clapping was given to the characters on the square.



(a) Birdsong



(b) Clapping



(c) Fountain

Figure 6.2: Spectra from 100 Hz to 10 kHz of the dry sounds of 10 s used in the subjective evaluation.

Birds were not rendered in the VR design. The B-format impulse responses were convolved with the three dry sounds. Thus, through decoding the B-format sound into binaural rendering convolved with the MIT HRTF data [270] through headphones, the participants could experience the spatialised audio in Unity according to the headset rotation. The VR environment was streamed through SteamVR. High performance GPU (GeForce GTX 1070) and CPU (Intel Core i7-8700K) were used to guarantee the rendering quality. The HTC VIVE as HMD was used to provide the VR experience. A headphone was connected with the VR headset through an amplifier. Through the field recording and measurements, the sound level was 65 dB(A) at the assumed listening point in Cabot Square. The artificial head was used to calibrate the playback volume for both ears at 65 dB(A) for the fountain sound of 1,000 reflections in Cabot Square. Such calibrations represent a sound volume range that could occur in these scenes, and do not indicate that the sound level at headphones during the evaluation is equal to the sound level at the actual locations. The test was conducted in a soundproofed room, and the background noise was below 25 dB(A) during the test. The participants could experience pre-rendered FOA audio and videos thus enabling head rotation with an immersive experience.

6.2.4 Indicator selection and subjective test design

Three perceived indicators were selected, including reverberance, immersion and realism. Reverberance was categorised into a technically perceived indicator in a previous study [70]. For different functions of interior spaces (e.g., a lecture room and a concert hall), acoustic performances are required to render totally different reverberance [285]. This discrepancy has been studied in room acoustics for a long time. Reverberance is more difficult to be perceived under continuous sounds, compared with impulsive sounds. Immersion and realism were considered reproduced indicators [196]. Immersion is a term to describe the virtual experience. This indicator reflects the degree of reproduction of sensory engagement. Realism also reflects this kind of sensory engagement in VR experience, but with additional comparisons to the real world for participants. Thus, these three indicators were selected to capture how variations of sound reflection in VR experience affect reverberance caused directly by sound reflection and perception of reproduction. The three adjectives 'reverberant', 'immersive' and 'realistic' were used for three indicators: reverberance, immersion and realism. The rating scale with the description is from 'not at all', 'slightly', 'moderately', 'very' to 'extremely'. The questions (in Appendix B) used in both stages are listed below:

1. How reverberant is this sound environment?
2. How immersive is this sound environment?
3. How realistic is this sound?

The sequence of these sounds was randomised in each reproduced scene. In addition, a 3D Graphical User Interface (GUI) synthesised in VR was shown in front of the participants while they were participating in the evaluation. Figure 7.3 shows the GUI in the reproduced scene for the subjective test. Whilst it was possible to wear the VR headset, the participants could give their subjective perception through a controller to pop-up or close the GUI during the evaluation.

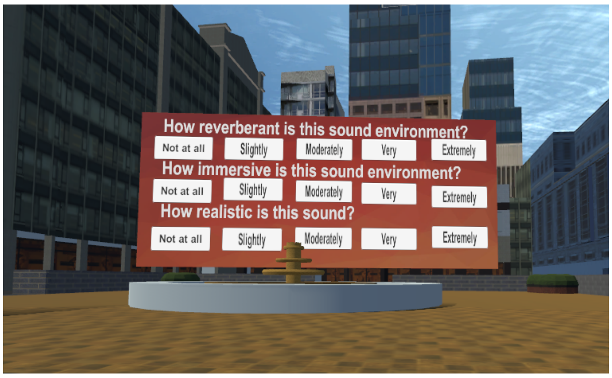


Figure 6.3: GUI in Unity 3D for the subjective test.

6.2.5 Laboratory experiment

The subjective test was divided into two stages. In stage I, the participants were informed they would hear the sounds in UCL campus square without the VR headset. For the first stage, the audio-only condition meant that the participant was not wearing the VR headset, but the VR headset was still running. The VR headset was placed aside and fixed in the direction of the sound source. In this way the participants could not be tracked on the head and could only receive sound without visual stimuli. Since all participants were familiar with UCL campus, they could have the corresponding spatial impression of this square. The participants would hear three types of sounds in turn from birdsong, clapping to fountain. In each sound heard in this scene, a reference sound was given to the participants first. The reference sound was convolved with the impulse response of the reflection order of 1,000. One thousand reflections are widely accepted as a benchmark number in relation to room acoustics modelling and prediction, e.g., speech intelligibility and diffuseness measure [286, 287, 288]. These studies present the validity of the preconditioning of 1,000 reflections during auralisation and acoustic simulation. After the reference sound, three signals of the same type of sound with different reflection orders were played. The order for these three signals was randomised in each of the same types of sound. Thus, the participants could give their subjective ratings for each signal. The headset was fixed during the audio-only test but was not worn by participants. The centre of the headset was oriented towards the source of the sound, which was also oriented towards the GUI. This means that the participants were still hearing the FOA sounds under the audio-only condition, but the binaural signals would be static and not be influenced by head rotation. In this way, the sound quality and volume for both audio-only and VR combined conditions were guaranteed at the same level.

In stage II, the participants heard these sounds with the VR headset under four reproduced scenes. For the second stage, the participants used a HMD wearing a headphone which allowed for head rotation with audio-visual stimuli of 3DoF. In each scene, the three types of sound would be played in turn as well. The evaluation procedure in each scene was the same as stage I. The time interval between each signal depended on how long participants took to complete the evaluation in the GUI at a time. After receiving the evaluation results of a single signal through a monitor in another room, the researcher would play the next signal. In general, participants completed each signal evaluation within ten seconds of the end of the sound. All sounds including the reference sound and the sounds with different reflection orders lasted 10 s. Four scenes were conducted with the same procedure in turn.

Thirty participants with normal hearing and vision took part in the subjective evaluation. All potential subjects lived in London, and they were asked if they had visited these sites with a brief description prior to the listening test. They confirmed that they would have been familiar with these scenarios based on their past visits. The consent forms were obtained from all participants. The participants were not informed by which reflection orders were applied in each sound during the evaluation. They all had a basic understanding of the perceived indicators used in the formal evaluation. The participant formation sheet is attached in Appendix E.

6.2.6 Statistical analyses

To assess the impact of different reflection orders, perceived indicators and sound types, SPSS Statistics 25 and OriginPro 2018 were utilised to analyse the repeated measures analysis of variance (ANOVA) test and linear regression.

6.3 Results

6.3.1 Comparisons between two audio-visual conditions under different reflection orders

One of the primary purposes of the chapter was to investigate the effect of the order of reflection on binaural static audio-only without head tracking and VR-combined conditions (HMD and binaural tracking). The two-stage subjective test was carried out in the reproduced UCL campus square. Thus, the comparison between the audio-only and VR-combined conditions can be made. Figure 6.4 shows the subjective ratings on the three indicators under different reflection orders and audio-visual conditions. The semantic rating scale from 'not at all' to 'extremely' was converted into 1 to 5. Thus, semantic responses can be quantified on the Y-axis from 1 to 5. Roughly, the subjective ratings of the perceived indicators of the three sounds increased with higher reflection orders under the two rendering

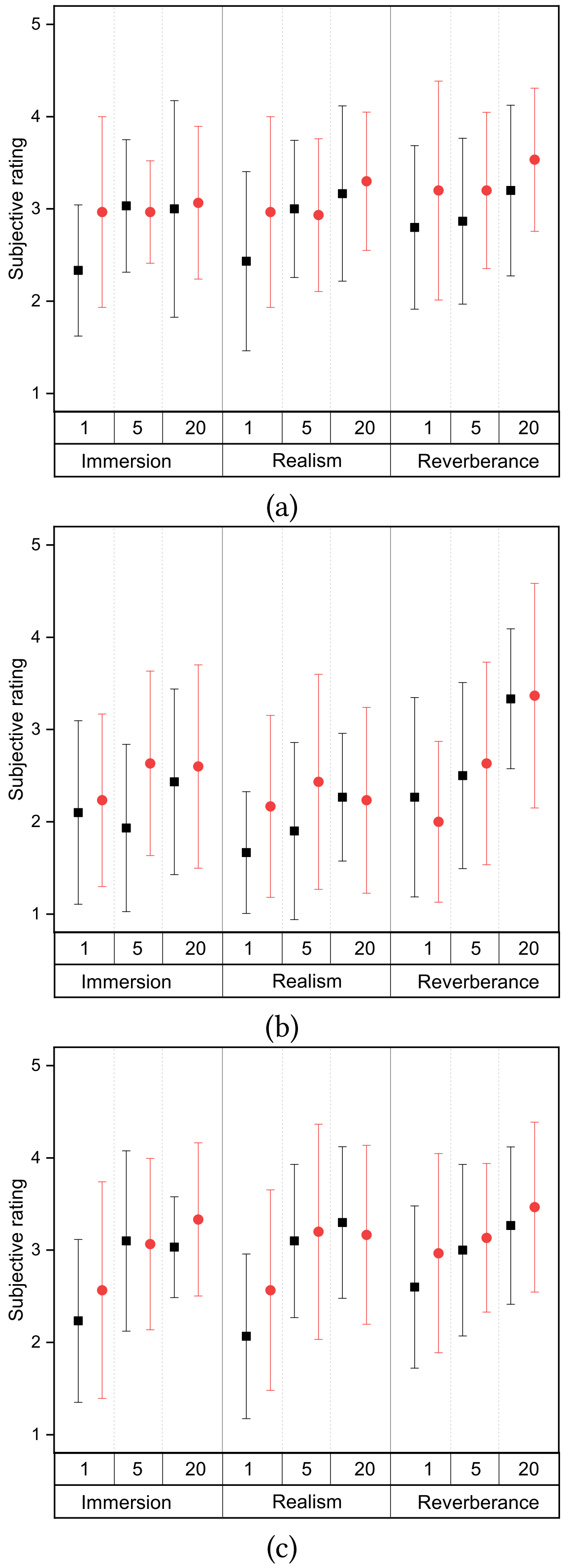


Figure 6.4: Comparison between audio-only (■) and VR-combined experience (●) conditions under different reflection orders (1, 5 and 20). (a) birdsong, (b) clapping, and (c) fountain.

conditions. There are some notable differences in the comparisons between the two audio-visual conditions. For birdsong, the rating difference between the two approaches is significant under the reflection order of 1 for all three perceived indicators. From the range of subjective ratings and overall trends, the fountain is similar to the birdsong. The rating of clapping is lower than the other two sounds in immersion and realism. Meanwhile, the variation of reverberance is the most distinct. Compared with the other two sounds, the clapping presents relatively low ratings on immersion and realism for both approaches, and the variation of reverberance is the most distinct. The clapping is discrete in the time domain and exists within a short time, and reverberation is easily perceived by the participants. For immersion and realism ratings, a number of notable differences between two audio-visual conditions occur at the reflection order of 1 or 5.

Table 6.2: Pairwise comparisons for audio-only and VR-combined conditions on different reflection orders.

Indicator	Order	MD[1]	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Reverberance	1	-0.167	0.200	0.408	-0.567	0.233
	5	-0.200	0.161	0.220	-0.523	0.123
	20	-0.189	0.175	0.285	-0.539	0.161
Immersion	1	-0.367*	0.173	0.038	-0.713	-0.021
	5	-0.200	0.125	0.115	-0.450	0.050
	20	-0.178	0.169	0.297	-0.516	0.160
Realism	1	-0.511*	0.184	0.007	-0.879	-0.143
	5	-0.189	0.156	0.229	-0.500	0.122
	20	0.011	0.168	0.947	-0.325	0.347
Rendering						
Audio-only	20∗1	0.722	0.141	<.001	0.439	1.005
VR-combined	20∗1	0.493	0.141	0.001	0.210	0.776

[1] MD—Mean Difference (I-J)

To examine the difference between these two audio-visual conditions, the repeated ANOVA analysis was conducted. As the between-subjects factor, the differences between audio-visual conditions is significant (F= 7.965, Sig.= 0.007). Based on Mauchly's Test of Sphericity, the differences between reflection orders do not meet the sphericity hypothesis (Sig.< .001). A further multivariate test is conducted, and Pillai's trace indicates that the differences among reflection orders are significantly different (F= 18.944, Sig.< .001). The effect of interaction between reflection orders and audio-visual conditions is not significant (F= 0.749, Sig.= 0.477). Thus, a further pairwise comparison was conducted to examine which reflection orders have a significant difference in the two audio-visual conditions, as shown in Table 6.2. Based on the ANOVA analysis, there are two indicators, i.e., immersion and realism, at the first order of reflection where the difference between the two audio-visual conditions is statistically significant, Sig.= 0.038 and Sig.= 0.007 respectively. The difference between these two conditions does not show statistical significance in reverberance. This analysis illustrates that compared with the audio-only condition, the introduction of VR will bring a significant enhancement on reproduced indicator perception at the first reflection order. When the reflection order was greater than or equal to five, the overall rating difference of immersion and realism between audio-only and VR-combined conditions was not significant. According to Figure 6.4, the comparison results between two audio-visual conditions reveal that both evaluation approaches present similar results on three indicators when the reflection order reaches 20.

A further pairwise comparison was made between the reflection orders of 20 and 1 shown in Table 6.2 as well. When the reflection order drops from 20 to 1, both audio-only (Sig.< .001) and VR-combined (Sig.= 0.001) renderings present a significant rating difference. The rating difference of immersion and realism under VR-combined experience is smaller than the audio-only rendering for this order drop from 20 to 1. Compared with binaural static audio-rendering rendering, the inclusion of VR will significantly improve immersion and realism ratings at low reflection orders (e.g., only the first order). According to the inclusion of visual stimuli, the participants no longer evaluated through only auditory stimuli. For visual rendering in VR, the architectural space and built environment were not been changed. With such an audio-visual presentation, participants' distracted attention and psychological cues caused such a result [289]. When it is set to a high reflection order (≥ 20), the difference between these two rendering methods is not significant.

6.3.2 Assessing perceived indicators under different reflection orders

All four squares are involved in stage II of the subjective evaluation under VR experience. The subjective ratings of the three indicators for these squares are shown in Figure 6.5. For these three indicators, the impacts of different reflection orders are shown in Table 6.3. The differences of reflection orders in $5*1$ and $20*1$ are significant for all perceived indicators. For reverberance (Sig.= 0.002) and immersion (Sig.= 0.025), there are significant differences between the orders of 20 and 5. There is no significant difference between the reflection orders of 20 and 5 in realism. Through the comparison between different indicators, the interaction between immersion and realism is significant (Sig.= 0.015). As reproduced indicators to represent the level of reproduced rendering, these two terms logically interact with each other.

Thus, the two indicators are largely convergent in rating trends during subjective evaluation. The interaction effect of reverberance with the other two indicators is not significant. It is also a good illustration of the differences in the division of functions between these two categories of indicators, i.e., the technically perceived indicator and reproduced indicators [70]. In general, the results shows that when the reflection orders drop from 20 to 1, the differences in subjective ratings are significant for these three perceived indicators. These variations caused by different reflection orders depends on the sound types and sites.

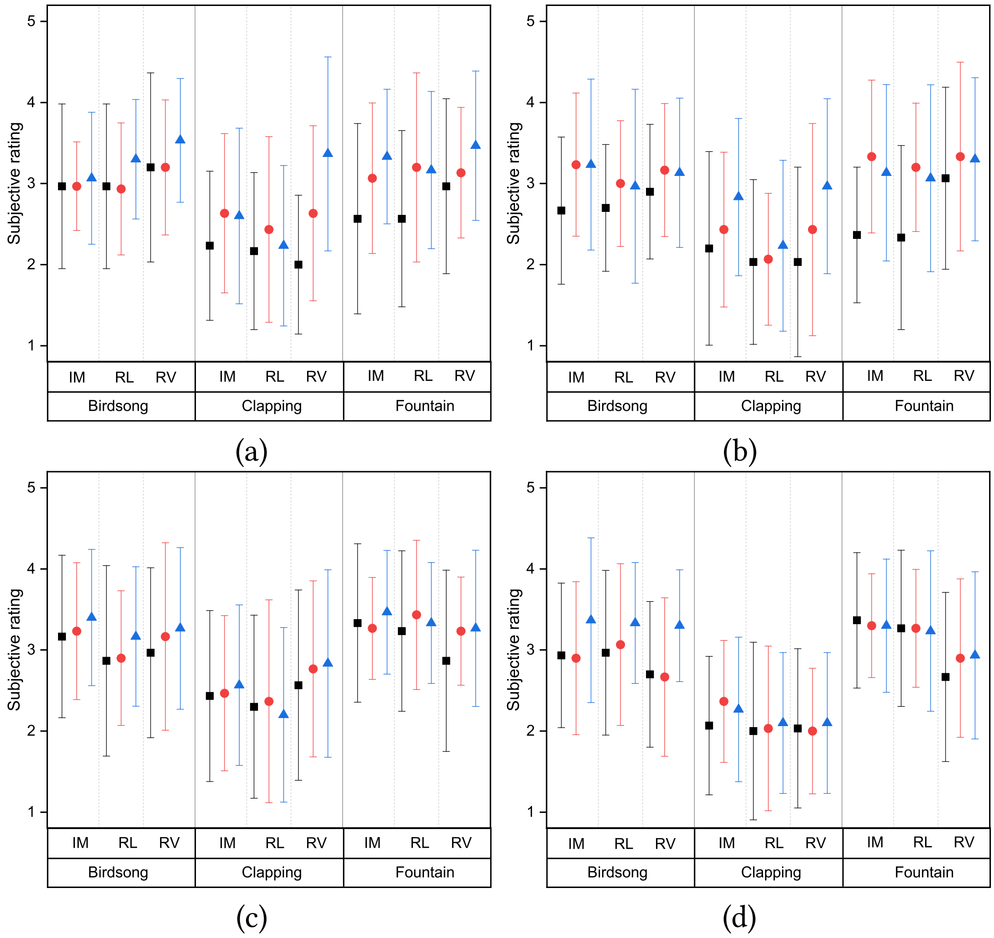


Figure 6.5: Subjective ratings under different reflection orders (■—1, ●—5 and ▲—20) among four squares. (a) UCL campus square, (b) Paternoster square, (c) Granary square, (d) Cabot square. (IM—immersion, RL—realism, RV—reverberance)

Table 6.3: Pairwise comparisons for sound types, sites and perceived indicators on different reflection orders.

		MD[1]	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Indicator	Order					
Reverberance	20∗5	0.236*	0.071	0.002	0.092	0.381
	20∗1	0.458*	0.103	<.001	0.248	0.669
	5∗1	0.222*	0.051	<.001	0.117	0.327
Immersion	20∗5	0.114*	0.048	0.025	0.015	0.212
	20∗1	0.356*	0.075	<.001	0.202	0.509
	5∗1	0.242*	0.078	0.004	0.082	0.401
Realism	20∗5	0.036	0.073	0.623	-0.112	0.185
	20∗1	0.244*	0.115	0.042	0.010	0.479
	5∗1	0.208*	0.076	0.011	0.052	0.364
Sound[2]	Order					
B∗C	—	0.708*	0.118	<.001	0.467	0.950
B∗F	—	-0.049	0.077	0.531	-0.207	0.109
C∗F	—	-0.757*	0.116	<.001	-0.995	-0.519
B	20∗5	0.219*	0.055	<.001	0.107	0.332
	20∗1	0.339*	0.085	<.001	0.165	0.513
	5∗1	0.119	0.090	0.196	-0.065	0.304
C	20∗5	0.139	0.103	0.187	-0.071	0.349
	20∗1	0.353*	0.140	0.017	0.067	0.639
	5∗1	0.214*	0.073	0.007	0.064	0.364
F	20∗5	0.028	0.079	0.726	-0.133	0.188
	20∗1	0.367*	0.092	<.001	0.178	0.555
	5∗1	0.339*	0.089	0.001	0.157	0.521
Indicator	Sound					
Reverberance	B∗C	0.622*	0.126	<.001	0.364	0.881
	C∗F	-0.617*	0.148	<.001	-0.920	-0.314
Immersion	B∗C	0.669*	0.137	<.001	0.390	0.949
	C∗F	-0.728*	0.126	<.001	-0.985	-0.471
Realism	B∗C	0.833*	0.162	<.001	0.502	1.164
	C∗F	-0.928*	0.118	<.001	-1.169	-0.687
Site	Order					
1	20∗5	0.207	0.102	0.051	-0.001	0.415
	20∗1	0.493*	0.180	0.010	0.124	0.861
	5∗1	0.285*	0.115	0.019	0.051	0.520
2	20∗5	0.074	0.092	0.426	-0.113	0.262
	20∗1	0.507*	0.124	<.001	0.254	0.761
	5∗1	0.433*	0.099	<.001	0.231	0.636
3	20∗5	0.074	0.064	0.257	-0.057	0.205
	20∗1	0.196*	0.075	0.014	0.042	0.351
	5∗1	0.122	0.072	0.101	-0.025	0.270
4	20∗5	0.159	0.082	0.063	-0.009	0.328
	20∗1	0.215*	0.081	0.013	0.049	0.380
	5∗1	0.056	0.079	0.486	-0.106	0.217

[1] Mean Difference (I-J).
[2] B—Birdsong, C—Clapping, and F—Fountain.

6.3.3 Effect of sound types under different reflection orders

As shown in Figure 6.5, the ratings of clapping present a different trend compared with the other two sounds. For birdsong and fountain, the average distribution of ratings is roughly in a comparable range, while the clapping is rated lower than these two sounds. Therefore, in order to explore the significance of the differences between sound types, the Mauchly's test of sphericity was conducted as shown in Table 6.4. According to the test results, the differences in sound (Sig.= 0.034), order (Sig.< .001) and sound*order (Sig.= 0.004) respectively are not considered to satisfy the hypothesis of sphericity. A multivariate test was conducted illustrated in Table 6.5. The results of Pillai's Trace showed that for different sounds and different reflection orders, their individual differences within the groups are statistically significant (Sig.< .001 for sound, Sig.= 0.001 for order). The effect of interaction between different sound types and reflection orders shows no overall significance (Sig.= 0.262). Thus, a pairwise comparison was made to analyse the difference between each of the two sound types illustrated in Table 6.3.

For these three types of sound, the difference between birdsong and fountain is not significant (Birdsong∗Fountain, Sig.= 0.531). The difference between clapping and the other two sounds is statistically significant (Birdsong∗Clapping, Sig.< .001; Clapping∗Fountain, Sig.< .001). According to Figure 6.5, the comparison of the subjective ratings in birdsong and fountain did not only have the smallest rating difference, but was also not significant. The lower overall rating of clapping is also significant compared with the other two sounds.

In addition, to investigate the impact of different reflection orders on sound types, the pairwise comparison was also conducted shown in Table 6.3. The differences of reflection orders from 20 to 1 for all sound types are significant (Birdsong, Sig.< .001; Clapping, Sig.= 0.017; Fountain, Sig.< .001). It is generally consistent with the overall significance of the differences of reflection orders. The differences between reflection orders of 1 and 5 are also significant for clapping (Sig.= 0.007) and fountain (Sig.= 0.001). The birdsong shows no significant difference between the reflection orders of 1 and 5, but it is significant for 5 and 20 (Sig.< .001). The types of sound have a significant influence on perception. A further pairwise comparison in Table 6.3 was made between different sound types under three perceived indicators, and the results show that the subjective ratings of clapping on realism and immersion are significantly lower than other continuous sounds. The mean difference of reverberance between clapping and the other two sounds is smaller than realism and immersion, because the low reflection order will significantly decrease the ratings of reverberance. Compared with the continuous sounds, the impulsive sound, e.g., clapping, is more easily perceived by humans according to the variation of reflection orders (e.g., from 1 to 20).

Table 6.4: Mauchly's test of sphericity for the four squares.

Within Subjects	Mauchly's W	Approx. χ^2	df	Sig.	Epsilonb[1]		
Effect					G-G	H-F	L-b
Site	0.692	10.192	5	0.070	0.795	0.871	0.333
Sound	0.785	6.764	2	0.034	0.823	0.866	0.500
Indicator	0.584	15.077	2	0.001	0.706	0.732	0.500
Site∗Order	0.159	49.163	20	<.001	0.585	0.676	0.167
Sound∗Order	0.413	24.273	9	0.004	0.660	0.732	0.250
Order∗Indicator	0.310	32.100	9	<.001	0.671	0.747	0.250

[1] G-G (Greenhouse-Geisser), H-F (Huynh-Feldt), L-b (Lower-bound).

Table 6.5: Multivariate test for the interaction between sound types and reflection orders.

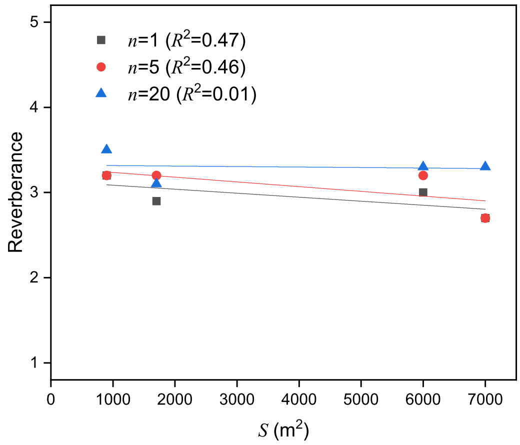
Effect	Value	F	Hypothesis df	Error df	Sig.
Sound	0.605	21.468	2.000	28.000	<.001
Order	0.382	8.660	2.000	28.000	0.001
Sound∗Order	0.177	1.398	4.000	26.000	0.262

6.3.4 Effect of urban squares with various areas and layouts under different reflection orders

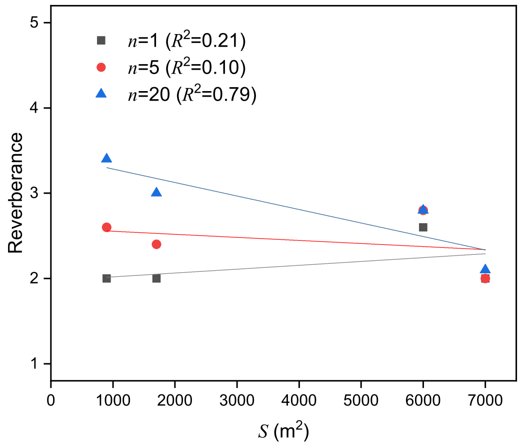
The differences of subjective ratings with sounds of different reflection orders on different sites were compared. Based on the analysis of Mauchly's test of sphericity shown in Table 6.4, the results of different sites satisfied the sphericity hypothesis (Sig.= 0.070). Thus, the within-subjects effect was considered to be statistically significant when sphericity is assumed (F= 5.132, Sig.= 0.003). Due to the selection of different squares, the subjective ratings in these different scenarios appeared to vary significantly within the group. It supports the compartmentalisation of the locations selected for this chapter. This also implies that different locations yield different perceptual ratings, yet these mean rating differences are not noticeable from a holistic view in Figure 6.5. In order to investigate the effect of each square selection on subjective evaluation, further comparisons are needed. For these four urban squares, the impacts of different reflection orders are shown in Table 6.3. When the reflection order drops from 20 to 1, the mean differences of subjective ratings are statistically significant for these four squares accordingly (Sig.= 0.010, site 1; Sig.< .001, site 2; Sig.= 0.014, site 3; Sig.= 0.013, site 4). Meanwhile, the difference between the reflection orders of 1 to 5 was found to be significant in UCL campus (Sig.= 0.019) and Paternoster square (Sig.< .001). It indicates that these two squares are more sensitive to variations in reflection orders than Granary square and Cabot square.

Using the urban square areas given in Table 6.1, an attempt to assess the relationship and trend between the area of the squares and different reflection orders on reverberance by performing linear fitting was made. Figure 6.6 shows the trend between the area of the squares and reverberance ratings of reflection orders for three types of sounds. When the area grows, there is a slight reduction in reverberance for both birdsong and fountain. However, there is a striking difference between reflection orders for clapping. At a reflection order of 20, there is a noticeable tendency for the reverberance to decrease as the area increases ($R^2 = 0.79$).

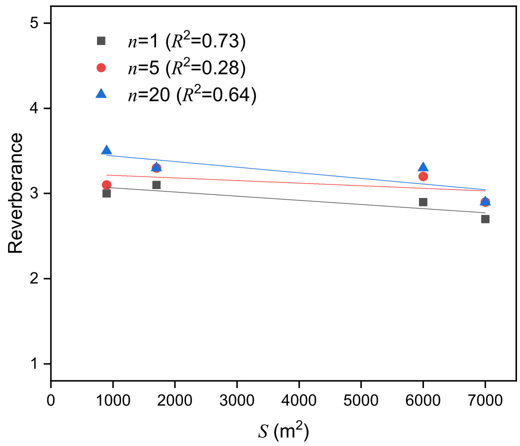
For different square layouts, the discussion is framed by the two ratios, r_{2D}



(a)

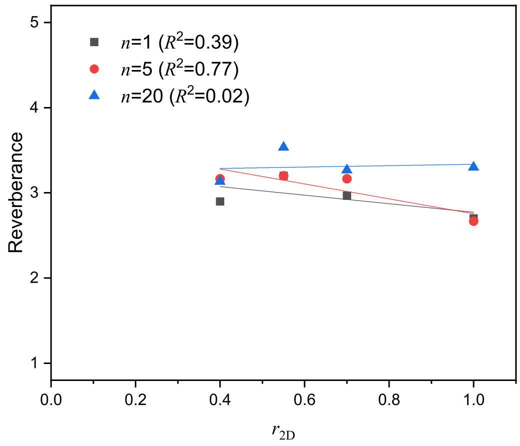


(b)

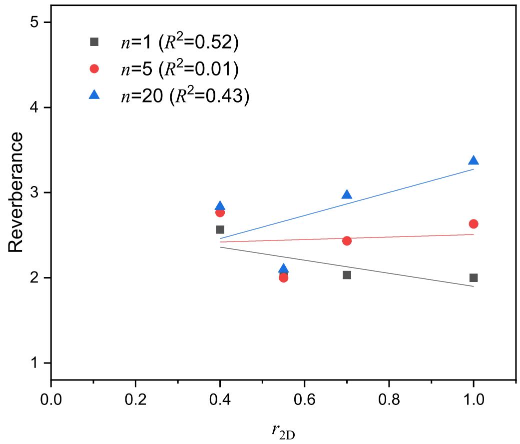


(c)

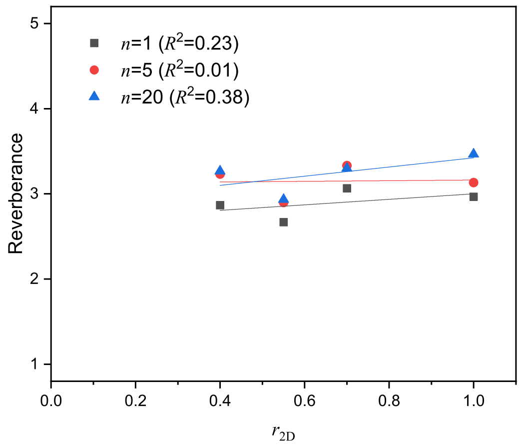
Figure 6.6: Comparison between subjective ratings and area of squares (S) under different reflection orders (1, 5 and 20). (a) birdsong, (b) clapping, (c) fountain.



(a)

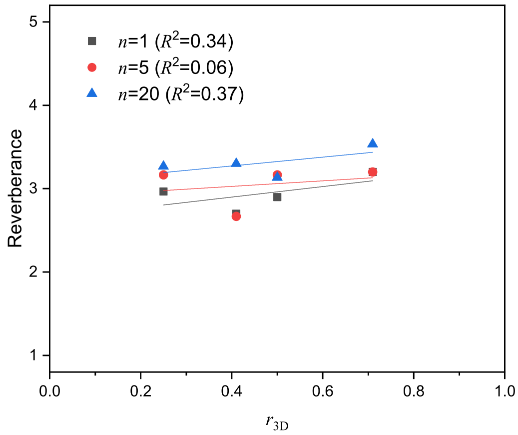


(b)

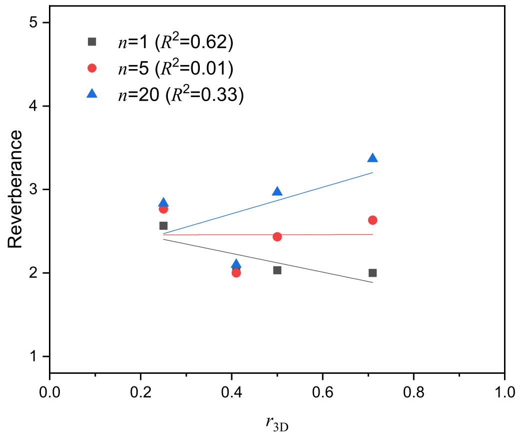


(c)

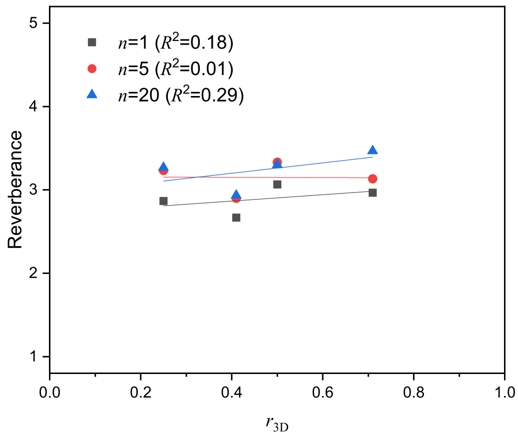
Figure 6.7: Comparison between subjective ratings and 2D enclosed ratio (r_{2D}) under different reflection orders (1, 5 and 20). (a) birdsong, (b) clapping, (c) fountain.



(a)



(b)



(c)

Figure 6.8: Comparison between subjective ratings and 3D enclosed ratio (r_{3D}) under different reflection orders (1, 5 and 20). (a) birdsong, (b) clapping, (c) fountain.

and r_{3D} shown in Table 6.1 as well. The trends between reverberance and the enclosed ratios were analysed respectively. As shown in Figure 6.7, when r_{2D} deceases, the perception of reverberation for birdsong and fountain is slightly diminished. For a high reflection order of 20, this decreasing tendency is particularly evident for clapping. When r_{3D} is taken into account, the trends of reverberance for three types of sounds under different reflection orders are largely in line with r_{2D} as shown in Figure 6.8. r_{3D} was introduced with consideration of a square surrounded by tall buildings. The ratios of r_{2D} and r_{3D} exhibit similar tendencies in the linear fitting analyses.

In UCL campus and Paternoster Square, clapping was found to have a significant variation when the reflection order is changed from 1 to 20 for reverberance. The clapping shows no notable growing or declining tendency on different reflection orders for reverberance in Granary Square and Cabot Square. Reverberation has been already difficult to perceive in these spaces.

In this chapter, the three types of sounds were found to have a significant variation on reverberance when the square area is less than or equal to the area of Paternoster Square ($S \leq$ 1,700 m^2). For the two large-sized squares, i.e., Granary Square and Cabot Square ($S \geq$ 6,000 m^2), their subjective ratings on the three indicators, especially on reverberance, were not that different. Therefore, it is feasible to use only a reflection order of one or five, or even the direct sound to render the sounds in large-sized squares ($S \geq$ 6,000 m^2).

6.4 Discussions

6.4.1 Area and layout of urban squares

For large-sized squares, like Granary Square and Cabot Square, the enclosed ratio is normally not too high. These relatively large squares are often surrounded by a variety of functional facilities which are interconnected with each other through pathways. The sound is more likely to be absorbed by the sky in these squares, and thus the subjective results are not significantly affected by reflection orders in such spaces. For the vast majority of outdoor squares, the height of a building is not the main factor in the difference in perception caused by different reflection orders. The squares covering a large area (e.g., $S \geq 6{,}000\ \text{m}^2$) are adequate enough to be able to use fewer reflections or even attenuated direct sounds, on the basis of appropriate sounds.

For small squares (e.g., $S \leq 1{,}700\ \text{m}^2$), the sound propagation is similar to indoor acoustic situations. The acoustic performance of such spaces depends more on the elaborate conditions rather than enclosed ratios. Local diffusion conditions and sound absorption coefficients of different materials will significantly influence acoustic performance. Standing waves may also exist in such small spaces affecting the results of sound evaluation. The sounds are not fully attenuated in these spaces, and early reflections should still be perceived by the participants. Many frequently used indoor acoustic indicators, such as impulsiveness, tonality, transients, and speech transmission index (STI), can be investigated in more depth in such relatively small spaces with significant early reflections.

6.4.2 Limitations

The enhancement of subjective ratings in realism and immersion comes from 3DoF. The 3DoF allows head rotation and thus synchronise the audio-visual interaction. When HMDs and Ambisonics are established, the enhancement of realism and immersion ratings is affected by other unclear factors. These factors may be related to the technical parameters during the environment reproduction, such as wearing device comfort and the number of polygons for visualisation. Quantifying these indicators clearly requires a more in-depth period of research.

For other limitations, the possible impact for some specific reflection orders is not clear, e.g., an order of 4 or 6. For relatively small open spaces, e.g., one or two hundred square metres, the results of the analyses in this chapter are not applicable. In addition, the fountain with only one design and corresponding water sound was used in this chapter. Participants' consideration of the plausibility of the same landscape installation and actual sound level rendering in different squares may also influence immersion ratings.

6.4.3 Implementation

For urban planning, designers prefer to use many perceptible or imperceptible ways to beautify the built environment, e.g., by placing a fountain or adding artificial birdsong in an urban open space. At this point, the majority of these landscape-related sounds do not exist within a short time according to their time-frequency characteristics. The designers can simplify the sound setups especially for the reflection process according to the urban open space they focus on. Virtual environments with higher immersion and realism ratings will provide the designers with a reliable tool for participatory decision-making.

For video game design, developers often attach different sounds to corresponding types of objects in 3D scenes. The game sounds utilised are sometimes different from those landscape-related sounds. The sounds, such as gunfire, striking, knocking or even a bo (a percussion instrument) in an orchestra, are very common in the scenarios of game design to create an appropriate atmosphere. Their duration time is very short and these sounds can be analogised to impulse sounds. The auralisation quality of these sounds is closely related to sound reflection orders involved. Therefore, extra care during auralisation should be taken when designing these kinds of sounds in spaces.

For soundscape studies, it is an important task for researchers to enable people to accurately perceive and evaluate the urban sound environment. In view of the fact that the whole immersive VR system has high ecological validity in sound environment research [199], these minimum details on reflection orders will contribute to a quick build of a virtual sound environment.

6.5 Conclusions

This chapter examined the influence of calculated acoustic reflection orders in a simulated environment on participants in relation to perceived immersion, realism and reverberance. The results of the subjective test revealed the following:

1. The rating differences of realism and immersion were found to be significant for both binaural static audio-only without head tracking and VR-combined

(HMDs and binaural tracking) renderings when the reflection order dropped from 20 to 1. Compared with the audio-only condition, the VR experience was found to bring higher levels of realism and immersion ratings of sounds when the reflection order was set to a relatively low range (e.g., the first order). When a high reflection order was set greater than 20, the difference between these two rendering methods was not significant.

2. When the reflection orders dropped from 20 to 1, the differences of subjective ratings were significant for reverberance, immersion and realism. The interaction effect between reverberance and realism and immersion was not significant.

3. In terms of different sound types, it should be noted that the subjective ratings of an impulsive sound (e.g., clapping) on realism and immersion was significantly lower than other continuous sounds under a reflection order ranging from 1 to 20. This type of sound significantly depended on the reflection order for reverberance, and it needed to be carefully auralised during the period of changing the reflections with consideration given to the spatial characteristics and acoustic behaviours of urban open spaces.

4. The differences in ratings for the sounds with different reflection orders were not evident in relatively large open spaces. It was possible to reduce reflection orders (e.g., ≤ 5) or even use the attenuated direct sound when the urban open space was large enough ($S \geq 6{,}000$ m^2). For relatively smaller squares, with an area ranging from a few hundred to over a thousand square metres, a higher reflection order was still required.

Overall, a VR experience combining both HMDs and Ambisonics will significantly enhance our sense for immersion ratings at low reflection orders. It is accessible to employ fewer reflections during auralisation to render sounds in VR experiences with similar realism and immersion ratings. For open urban environments, the results on perceived indicators could make a contribution to fast auralisation with fewer reflections and reasonable accuracy through VR experiences.

Chapter 7

Synthesising Line Sound Sources for Auralisation

After the studies of audio rendering and reflection modelling in urban spaces, the research will focus on the sound sources themselves. In urban sound environment simulation, many sound sources are often considered as a point source in the far field. For extensive near-field scenarios in urban open spaces, the discussion on human perception to judge the results of auralisation in such spaces is inadequate. This chapter[1] aims to investigate the relationship between subjective perception and different synthesis point setups for a line source, and whether it is possible to reduce synthesis points of a line source in a virtual open urban space. Section 7.1 introduces how we interact with different types of sound sources in urban spaces, and how I can investigate the synthesis of non-point sources, especially line sources. Section 7.2 clarifies the methods to reproduce the virtual scenes and the synthesis setups of line sources in VR. Section 7.3 analyses the results of the effect of synthesis on width, distance and immersion. Section 7.4 discusses the relationship between the audible angle and immersion, and some limitations in the studies. Section 7.5 summarises the influence of syntheses of sound sources on perceived width, immersion and distance from the subjective evaluation and analyses.

[1]This chapter was partially presented at the International Conference on Immersive and 3D Audio 2021 [290].

7.1 Background

We are surrounded and enveloped by a wide variety of sounds every day. These sounds are produced by different people or objects in diverse environments. Different types of sounds vary in frequency content, spatial position, and source volume in urban spaces with various geometric shapes. Many sound sources are considered to be point sources in the far-field [291], and their geometry is simplified to a point in Cartesian coordinates. For urban open spaces, however, there is a visible interaction between people and people or people and objects. In order to enable interaction, people do not usually maintain a far field with these sound sources. The near-field experience of these sound sources is equally important. When the near field is discussed, the scale of a sound source needs to be taken into account.

For the majority of flat urban open spaces, we are a point on a huge 2D map. The dimensions of some sound sources can be simplified to a finite or infinite length of line rather than a single point. For example, studies of road traffic noise widely adopted the simulation of line sources [292, 293, 294]. These studies using line sources to model traffic noise aim to calculate sound pressure levels and other acoustic parameters to meet the requirements of local government regulations and designers. Based on these numerical simulation, acoustic engineers worked hard to control noise. However, as the conceptualisation of soundscape became well established, researchers began putting more emphasis on how people perceive the overall acoustic environment rather than just reducing noise [5, 6].

Back to an urban public space, sound contexts are usually associated with different sound sources. When planning a square or park and placing a certain length of water curtain, for instance, designers typically consider aesthetic characteristics and public engagement first. Then, they gradually begin to consider the acoustic characteristics of such a water feature and how this sound can fit into the overall soundscape. Based on fieldwork in London, the patterns of fountains in urban spaces are not monotonic. The installation of a fountain will vary according to the demands of the spatial and aesthetic requirements. As shown in Figure 7.1, over 1,000 fountains are placed in Granary Square, and waterfall cascades surround

Cabot Square. These waterscapes are different from usual fountains. To examine such people-environment interaction and sound sources with a certain scale, some researchers have conducted VR-based auralisation studies in recent years. Richmond et al. [282] in 2010 conducted the audio-visual animation of an urban space, and they employed a linear interpolation technique to implement the pre-calculated movement for participants in an interactive audio-visual environment. An HRTF-based spatial audio technique for area and volumetric sound sources was developed in 2016, and Schissler et al. [185] utilised Monte Carlo projection to sample sound sources with the use of orthonormal basis functions to replace the analytical solutions in the source projection function and the HRTF. What we know about how the syntheses of such non-point sources affect our perception on immersion of reproduced soundscapes is still very limited.

Figure 7.1: Waterscape cases in central London. Granary Square (left), Cabot Square (right).

At this point, the issue is whether we can rationally synthesise non-point sources, especially line sources, so that sound matches vision in the near field. Previous study of different sound reflection orders has aroused interest in the compatibility of sound sources in VR [71]. Thus, different sound types were also considered, and the subjective evaluation based on VR was carried out. This chapter aims to investigate the effect of different synthesis setups of line sources on perceived width, immersion and distance under virtual environments.

7.2 Methods

7.2.1 Scene visualisation and animation

In order to assess the perceived width of line sources under a virtual experience, an urban space was pre-defined to place various sound sources. A garden square of 6,400 m^2 (80 m×80 m) was created in virtual reality, as shown in Fig. 7.2.

Figure 7.2: Figure-ground of the self-defined square.

The area of the virtual garden square is 6,400 m^2 (80 m × 80 m). According to the previous study, when the public square is large enough ($S \geq 6,000$ m^2), the reflection orders during auralisation can be highly simplified. Therefore, attenuated direct sound could be employed in this virtual square. The rectangular square was surrounded by greenery, and the central area was paved. The pavement in the square was widened to minimise the implication of the absorption or attenuation of sound by the visual representations of greenery. Kang [5] conducted the numerical simulation on a square (100 m × 100 m) in 2006 to investigate the effect of square geometry and diffusion conditions. Compared with geometrically/specularly reflecting boundaries, diffusely reflecting boundaries can significantly reduce RT from 16 s to around 2 s in a square of this scale. The virtually added trees and greenery can be considered visual diffusers. These visual diffusers will contribute to a significant reduction in RT and thus to a closer approximation of a free field.

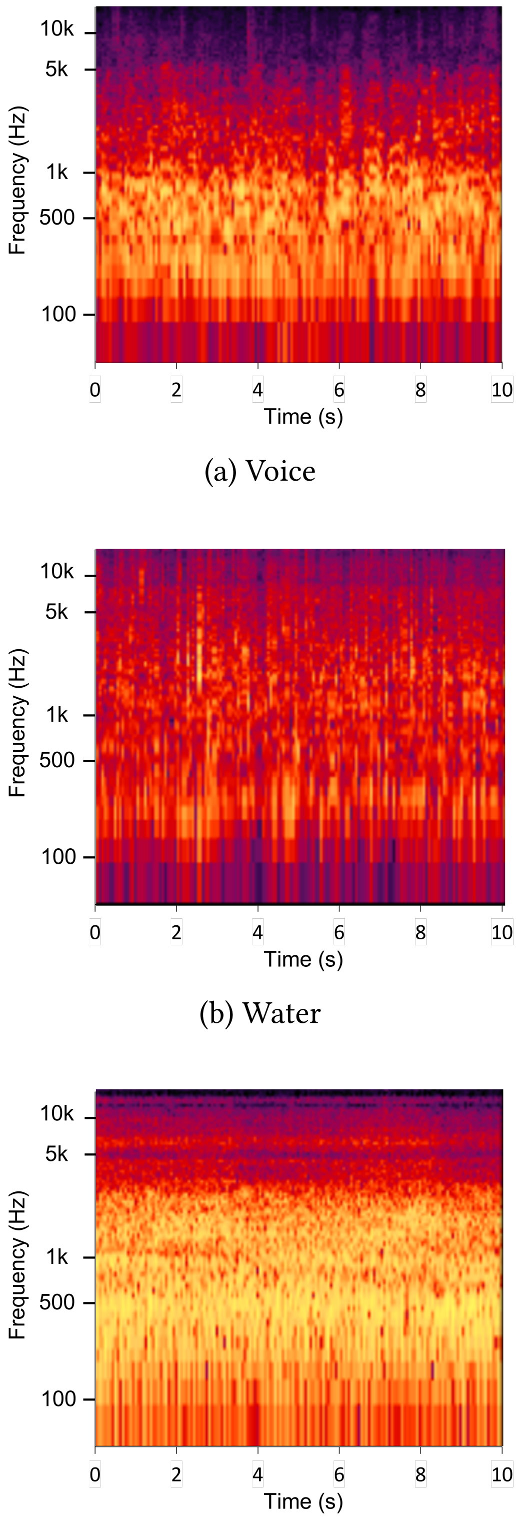
The buildings in the virtual scene were modelled in SketchUp. To provide an immersive experience, a limited number of characters were placed in the square with their individual activities attached with animation. The animation of characters was achieved by 3ds Max. These extra characters would not get close to the user location and would keep a distance of more than 30 m from the user location. All building and character models were imported into the game engine (Unity). The greenery is generated by the tree and grass creators in Unity. To further enhance immersive experience, the trees and plants were set to sway slightly with the wind. The lighting condition was set to a rational solar zenith angle and illuminance.

7.2.2 Sound event design

To investigate the acoustic behaviours of different sounds in the square, several typical sounds were chosen including the sound of a group of people talking (voice), sound from a water curtain/blade (water), and construction noise (construction). These three types of sounds have different acoustic characteristics and soundscape contexts.

Figure 7.3 shows the spectra of the three sounds. The voices utilised in this study are a sound clip that is indistinguishable for sound content. As such, the potential participants during the subjective test would not be influenced by the voice content to evaluate perceived indicators. It was considered to be a neutrally perceived sound. A water curtain, a water blade or a waterfall cascade is often used as part of the landscape to beautify and decorate environments. Sometimes it acts as an individual landscape component combined with lighting techniques to project some visual representations on it. The water sound is frequently employed to enhance the quality of the acoustic environment [62, 117, 283, 295]. Construction noise is a very common type of noise in cities, frequently reported in the studies of noise complaints [27, 296]. Because of England's lockdown regulations on non-essential field work, royalty-free sounds were chosen from Adobe Audition Sound Effects. These sounds have a high S/N ratio. The sample rate of the three sounds is 44.1 kHz, and the depth is 16 bit.

To synchronise the virtual visual-audio environment, these three sound events were also visualised and animated to enhance the audio-visual interaction experience, as shown in Figure 7.4. The characters on the square were reproduced



(a) Voice

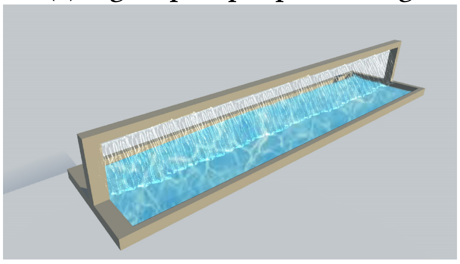
(b) Water

(c) Construction

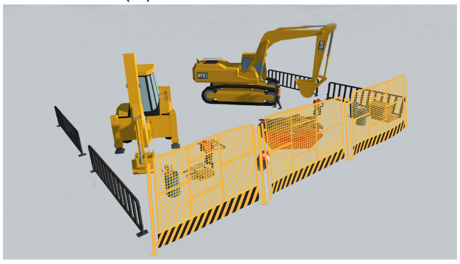
Figure 7.3: Spectra of dry sounds used in the subjective evaluation (10 s).

with animation. The water curtain was attached with a particle system to generate splash.

(a) a group of people talking



(b) water curtain



(c) construction

Figure 7.4: Reproduced people, water curtain, and construction site.

7.2.3 Synthesis setup and auralisation

To explore the effect of line source synthesis points on subjective evaluation, the line source needed to be synthesised with discrete point sources. The assumed length of the line sound source is 10 m. There are three cases in this study, i.e., 5, 21, and 101 point sources to simulate a sound source. The mean distances between the point sources for these three cases are 2.5 m, 0.5 m and 0.1 m. The distance between these synthesised sources is set to be not exactly equal to eliminate comb filtering. There are two observation points, at 5 m and 30 m away from the sound source, and both observation points move parallel to the line source at a speed of 0.5 m/s for a total time of 10 s. This movement was achieved by adding a corresponding

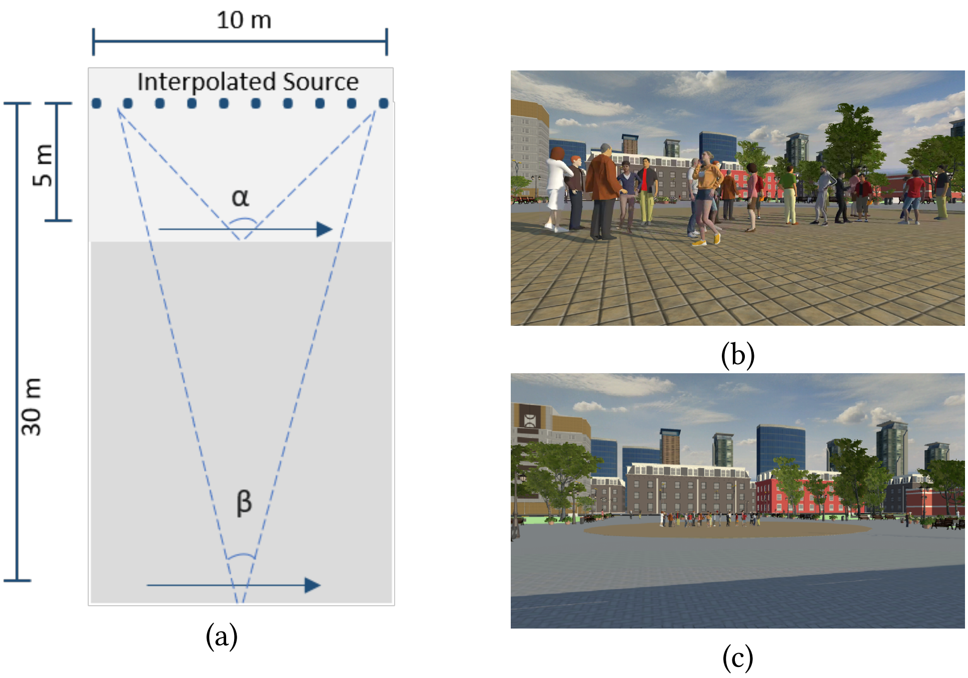


Figure 7.5: (a) Schematic diagram of the observation points and sound source location, (b) scene 5 m away from the source, and (c) scene 30 m away from the source.

displacement per frame during VR rendering to the listener's position using C# programming. The motion and audio receiver scripts were loaded simultaneously on the object of 'camera' in Unity. The object of 'camera' would match the VR headset rotation with spatial audio during VR rendering. The Doppler effect was ignored because of the slow relative velocity of the observation point with respect to the sound source. They moved from left to right, passing symmetrically through the centre of the source as illustrated in Figure 7.5.

An approximation of audible angles for these two observation points was made. The audible angle in this chapter primarily refers to the angle formed by two adjacent synthesis points to the central position of the human head. For the near case of 5 m, the total audible angle α is 90.0°, and the angle between the point source is 22.5°, 4.5° and 0.9° for 5, 21 and 101 point sources. For the far case of 30 m, the total audible angle β is around 18.9°, and the angle between the point source is around 4.7°, 0.9° and 0.2° for the three synthesis cases.

According to different number of point sound sources, the sound pressure level for these sounds at the receiving location was calibrated to the same level. The participants could experience pre-rendered FOA audio and videos thus enabling head rotation with an immersive experience. The whole VR video was recorded on a high performance desktop PC (Inter Core i7-9700k, 32 GB of RAM and NVIDIA GeForce RTX 2080) to ensure high resolutions of recording. The recorded videos are 30 frames per second.

7.2.4 Subjective test

Thirty-four participants took part in the subjective evaluation. The participants were recruited by email. The recruitment email specified that the participants should be adults and have a mobile phone capable of playing VR videos. They were asked to conduct the subjective evaluation in a poorly lit, quiet room. The participants heard these sounds using a wired earphone connected to their smart phones. Through a VR cardboard, they could watch the VR videos. The frame of the VR cardboard is made of paperboard with two plastic convex lenses. The smartphone can be placed in the VR cardboard and the eyes can perceive the pre-recorded VR videos through the lenses. Due to the different testing environments in which the subjects were exposed to, the implication of participant selection will be tested by the inter-rater reliability. The Cronbach's α among the subjects is 0.812 illustrating internal consistency of the listening test. Before the formal test, they watched a compilation video of three sounds including all synthesis situations to experience the range of perceived variation during the evaluation in advance. It was easy to hold the lightweight VR cardboard, and the participants could give their subjective ratings on their laptops or tablets after taking off the VR cardboard. Each video during the formal test with one synthesised sound lasts 10 s. They were asked to rate the perceived indicators in a structured online questionnaire for each sound.

In order to assess the impact of syntheses of sound sources, different indicators were selected. Two spatially perceived indicators were selected, including width and distance. For a non-point like source, people do not always perceive and localise such a sound as a point in space. People can perceive the location and scale of these non-point sources. As discussed in Section 2.3.1 for spatial hearing, when the variation in elevation is excluded, people perceive the variation in azimuth and distance of different sounds. For the variation in the horizontal plane and distance, perceived width and distance can illustrate the spatial perception performance of non-point sources during different synthesis setups. The reproduced indicator was chosen, i.e., immersion. Immersion was considered a reproduced indicator [196]. The questions (in Appendix B) used in the questionnaire are listed below:

1. How wide is this sound in this environment? (Narrow—Wide)
2. How do you feel about the distance to this sound? (Near—Far)
3. How immersive is this sound environment? (Not at all immersive—Extremely immersive)

Each participant watched 18 videos (three syntheses, three sounds, and two distances) in total. The sequence of different synthesis setups was randomised. The participants were not informed how many points were applied to each sound during the evaluation. They all had the basic understanding of perceived indicators used in the formal evaluation. The participants followed the instructions on their laptops or tablets, and gave their subjective ratings to experience the virtual square by holding the VR cardboard. Remote testing requirements for participants including wired earphones and file formats are in line with the initial report of the Acoustical Society of America (ASA) psychological and physiological task force on remote testing [297]. The participant formation sheet is attached in Appendix F.

Figure 7.6: A researcher demonstrating the subjective test (left), VR cardboard (right).

7.2.5 Statistical analyses

In order to assess the impact of the synthesis of line sources on urban squares, SPSS Statistics 25 and OriginPro 2018 were utilised to analyse the statistical results.

7.3 Results

7.3.1 Effect of syntheses on width

The subjective ratings of width for three types of sounds are shown in Fig. 7.7. When the distance from the sound source is at 5 m, the width ratings show a visible increase for three types of sounds with 101 points. When the distance from the sound source is at 30 m, the rating difference between 5 and 21 points is noticeable. Based on the analysis of Mauchly's test of sphericity, the results of different syntheses satisfied the sphericity hypothesis (Mauchly's W= 0.952, Sig.= 0.070). Thus, the within-subjects effect of syntheses is considered to be statistically significant (F= 19.256, Sig.< .001) when sphericity is assumed. To investigate the differences between the syntheses, the pairwise comparisons were made through the repeated measures ANOVA shown in Table 7.1.

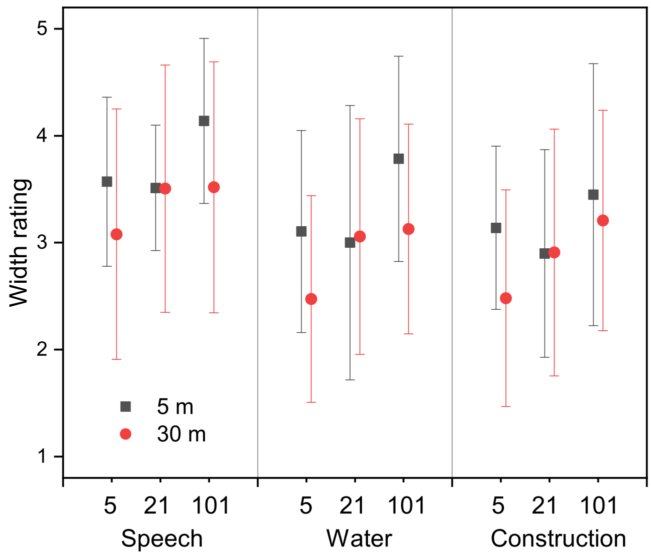


Figure 7.7: Perceived width rating under different synthesis points, distances and sound types.

When the distance is at 5 m, the rating differences between 21 and 101 points are significant for all types of sounds (speech Sig.= 0.004, water Sig.= 0.002, and construction Sig.= 0.020). The rating differences between 5 and 101 points shows statistical significance for speech (Sig.= 0.007) and water (Sig.= 0.004). When the distance is at 30 m, the rating differences between 5 and 101 points are significant

Table 7.1: Pairwise comparisons for perceived width ratings under different synthesis points, distances and sound types.

Sound	Point	MD[1]	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Distance=5 m						
Speech	5*21	0.058	0.225	0.798	-0.391	0.507
	5*101	-0.569*	0.203	0.007	-0.974	-0.163
	21*101	-0.627*	0.210	0.004	-1.046	-0.207
Water	5*21	0.105	0.257	0.685	-0.408	0.617
	5*101	-0.680*	0.225	0.004	-1.130	-0.230
	21*101	-0.785*	0.237	0.002	-1.259	-0.311
Construction	5*21	0.241	0.226	0.290	-0.210	0.691
	5*101	-0.311	0.231	0.182	-0.772	0.150
	21*101	-0.552*	0.232	0.020	-1.016	-0.088
Distance=30 m						
Speech	5*21	-0.426	0.225	0.063	-0.876	0.023
	5*101	-0.439*	0.203	0.034	-0.844	-0.033
	21*101	-0.012	0.210	0.953	-0.432	0.407
Water	5*21	-0.584*	0.257	0.026	-1.096	-0.071
	5*101	-0.655*	0.225	0.005	-1.105	-0.205
	21*101	-0.071	0.237	0.765	-0.545	0.403
Construction	5*21	-0.427	0.226	0.063	-0.878	0.023
	5*101	-0.727*	0.231	0.002	-1.188	-0.266
	21*101	-0.299	0.232	0.202	-0.763	0.164
5 m*30 m						
Speech	5	0.491*	0.246	0.050	<.001	0.983
	21	0.007	0.226	0.974	-0.444	0.458
	101	0.622*	0.245	0.013	0.133	1.111
Water	5	0.631*	0.235	0.009	0.161	1.101
	21	-0.058	0.295	0.845	-0.646	0.530
	101	0.656*	0.239	0.008	0.179	1.133
Construction	5	0.658*	0.221	0.004	0.217	1.099
	21	-0.010	0.262	0.971	-0.534	0.514
	101	0.243	0.279	0.387	-0.314	0.799

[1] Mean Difference (I-J).

for all types of sounds (speech Sig.= 0.034, water Sig.= 0.005, and construction Sig.= 0.002). In addition, the rating differences between 5 m and 30 m are also significant under 5 points for all types of sounds (speech Sig.= 0.050, water Sig.= 0.009, and construction Sig.= 0.004). The rating differences between 5 m and 30 m are also significant under 101 points for speech (Sig.= 0.013) and water (Sig.= 0.008).

7.3.2 Effect of synthesis points on perceived distance

The subjective ratings of distance for three types of sounds are shown in Fig. 7.8. The rating differences caused by the observation distance are distinct. The subjective rating at 30 m is 38% higher than at 5 m (F= 75.927, Sig.< .001). Based on the analysis of Mauchly's test of sphericity, the results of different synthesis points satisfied the sphericity hypothesis (Mauchly's W= 0.992, Sig.= 0.779). The within-subjects effect of synthesis points is considered to be not statistically significant (F= 1.414, Sig.= 0.247) when sphericity is assumed. This illustrates that when there is a significant change in spatial distance, this variation clearly dominates the perceived distance through the audio-visual experience.

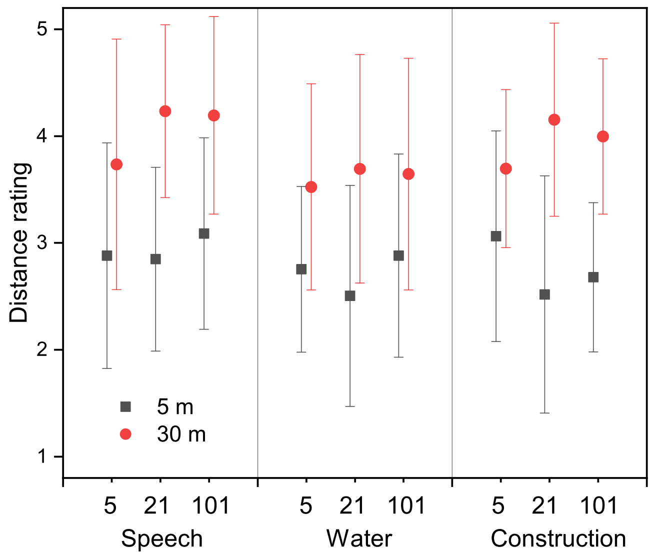


Figure 7.8: Perceived distance rating under different synthesis points, distances and sound types.

The pairwise comparisons were made through the repeated measures ANOVA shown in Table 7.2. When the distance from the sound source is at 5 m, the rating differences show no statistical significance between different synthesis points for speech and water. When the distance from the sound source is at 30 m, there are some significant rating differences between synthesis points, but these differences are not as significant as the 38% variation between two distances. Through the pairwise comparison between two distances, the perceived distance has significant rating differences under all synthesis setups. The evaluation results for perceived distance illustrate that the amount of synthesis points does not significantly affect the perception of distance, and the distance to the source combined with audio-visual rendering dominates this sense.

Table 7.2: Pairwise comparisons for perceived distance ratings under different synthesis points, distances and sound types.

Sound	Point	MD[1]	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Distance=5 m						
Speech	5∗21	0.033	0.193	0.865	-0.351	0.417
	5∗101	-0.207	0.209	0.326	-0.625	0.211
	21∗101	-0.240	0.179	0.184	-0.597	0.117
Water	5∗21	0.249	0.217	0.256	-0.185	0.682
	5∗101	-0.129	0.216	0.553	-0.559	0.302
	21∗101	-0.377	0.205	0.071	-0.787	0.033
Construction	5∗21	0.544*	0.215	0.014	0.114	0.973
	5∗101	0.384*	0.171	0.028	0.042	0.725
	21∗101	-0.160	0.199	0.425	-0.558	0.238
Distance=30 m						
Speech	5∗21	-0.499*	0.193	0.012	-0.883	-0.114
	5∗101	-0.460*	0.209	0.031	-0.878	-0.042
	21∗101	0.039	0.179	0.829	-0.318	0.395
Water	5∗21	-0.170	0.217	0.437	-0.603	0.263
	5∗101	-0.121	0.216	0.577	-0.552	0.310
	21∗101	0.049	0.205	0.813	-0.361	0.459
Construction	5∗21	-0.457*	0.215	0.037	-0.887	-0.028
	5∗101	-0.302	0.171	0.082	-0.643	0.039
	21∗101	0.156	0.199	0.438	-0.243	0.554
5 m∗30 m						
Speech	5	-0.854*	0.275	0.003	-1.402	-0.305
	21	-1.385*	0.206	<.001	-1.796	-0.975
	101	-1.106*	0.224	<.001	-1.554	-0.659
Water	5	-0.771*	0.216	0.001	-1.202	-0.341
	21	-1.190*	0.259	<.001	-1.707	-0.673
	101	-0.764*	0.251	0.003	-1.265	-0.263
Construction	5	-0.634*	0.215	0.004	-1.063	-0.206
	21	-1.635*	0.249	<.001	-2.133	-1.137
	101	-1.319*	0.176	<.001	-1.670	-0.969

[1] Mean Difference (I-J).

7.3.3 Effect of synthesis points on immersion ratings

The subjective ratings of immersion are shown in Fig. 7.9. When the distance from the sound source is 5 m, the distance ratings show a visible increase for three types of sounds with 101 points. When the distance from the sound source is 30 m, the rating difference between 5 and 21 points is noticeable. Based on the analysis of

Mauchly's test of sphericity, the results of different synthesis points satisfied the sphericity hypothesis (Mauchly's W= 0.952, df= 2, Sig.= 0.070). Thus, the within-subjects effect was considered to be statistically significant (df= 2, F= 19.256, Sig.< .001) when sphericity is assumed. To investigate the differences between the synthesis setups, the pairwise comparisons were made through the repeated measures ANOVA shown in Table 7.3.

When the distance is at 5 m, the rating differences between 5 and 101 points are significant for speech (Sig.= 0.024) and water (Sig.= 0.004). The rating differences between 5 and 101 points shows statistical significance for speech (Sig.=0.007) and water (Sig.= 0.004). When the distance is at 30 m, the rating differences between 5 and 101 points are significant for all types of sounds (speech Sig.= 0.034, water Sig.= 0.005, and construction Sig.= 0.002). In addition, the rating differences between 5 m and 30 m are also significant under 5 points for all types of sounds (speech Sig.= 0.050, water Sig.= 0.009, and construction Sig.= 0.004). The rating differences between 5 m and 30 m are also significant under 101 points for speech (Sig.= 0.013) and water (Sig.= 0.008).

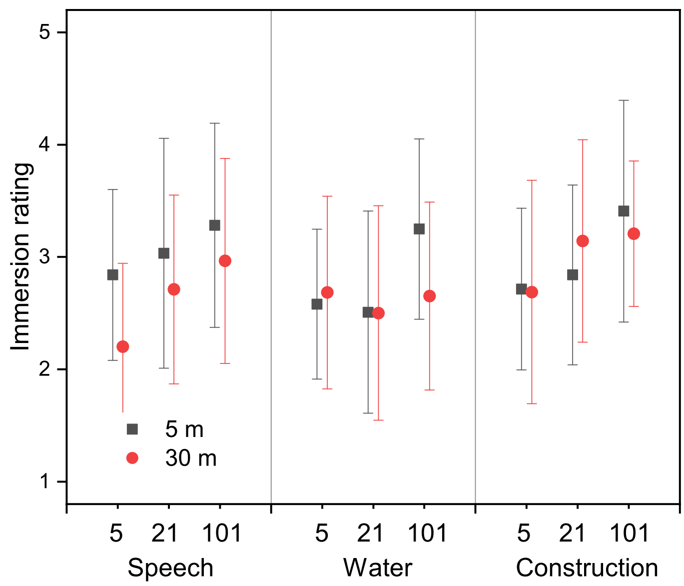


Figure 7.9: Immersion rating under different synthesis points, distances and sound types.

Table 7.3: Pairwise comparisons for immersion ratings under different synthesis points, distances and sound types.

Sound	Point	MD[1]	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Distance=5 m						
Speech	5∗21	-0.192	0.162	0.241	-0.515	0.132
	5∗101	-0.441*	0.190	0.024	-0.821	-0.061
	21∗101	-0.249	0.217	0.255	-0.683	0.184
Water	5∗21	0.072	0.166	0.667	-0.260	0.404
	5∗101	-0.669*	0.175	<.001	-1.019	-0.319
	21∗101	-0.741*	0.206	0.001	-1.151	-0.330
Construction	5∗21	-0.127	0.185	0.495	-0.496	0.242
	5∗101	-0.694*	0.207	0.001	-1.108	-0.281
	21∗101	-0.568*	0.216	0.011	-0.999	-0.136
Distance=30 m						
Speech	5∗21	-0.509*	0.162	0.002	-0.833	-0.186
	5∗101	-0.763*	0.190	<.001	-1.143	-0.383
	21∗101	-0.254	0.217	0.247	-0.688	0.180
Water	5∗21	0.182	0.166	0.278	-0.150	0.514
	5∗101	0.031	0.175	0.861	-0.319	0.381
	21∗101	-0.151	0.206	0.465	-0.562	0.259
Construction	5∗21	-0.454*	0.185	0.017	-0.824	-0.085
	5∗101	-0.520*	0.207	0.015	-0.933	-0.106
	21∗101	-0.065	0.216	0.764	-0.497	0.366
5 m∗30 m						
Speech	5	0.639*	0.185	0.001	0.269	1.009
	21	0.321	0.231	0.168	-0.139	0.782
	101	0.317	0.224	0.162	-0.131	0.765
Water	5	-0.104	0.189	0.586	-0.481	0.274
	21	0.007	0.228	0.976	-0.449	0.463
	101	0.596*	0.202	0.004	0.193	1.000
Construction	5	0.026	0.214	0.904	-0.401	0.453
	21	-0.302	0.210	0.155	-0.720	0.117
	101	0.201	0.205	0.332	-0.210	0.611

[1] Mean Difference (I-J).

7.4 Discussions

7.4.1 Auditory spatial resolution

The threshold angle for discriminating the sound positions is the MAA, and the MAA is about 1° for the sound sources in front of a subject in the horizontal plane [54]. To discuss the synthesis setups related to the audible angle, the immersion ratings under different audible angles between the synthesis points are shown in Fig. 7.10. The result shows that when the audible angle is reduced to this psychoacoustic threshold of around 1°, immersion ratings are significantly enhanced at both distance conditions. Immersion ratings do not show significant improvement when the angle is less than the threshold, i.e., when the distance is at 30 m for 0.2° and 0.9°.

When the audible angle is at 0.9° for both distance conditions, the immersion rating at 5 m is significantly higher than at 30 m. This increase is apparently more related to visual factors under the two distances. For the far case, the visual space is not given to too many objects in order to eliminate the possible presence of sound sources in the surroundings. As a result, immersion ratings are diminished by the emptier environment. For the near case, the subject can more clearly observe the sound source at a closer distance, thus significantly enhancing immersion ratings.

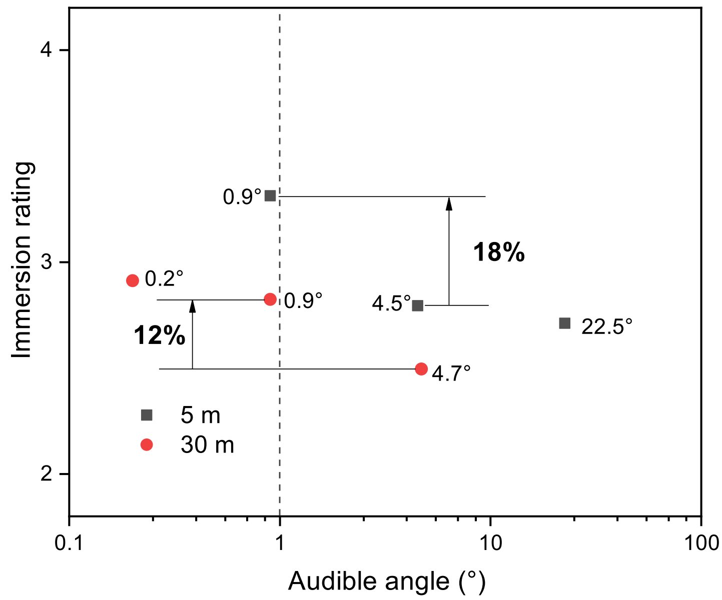


Figure 7.10: Immersion rating under different audible angles between synthesis points.

7.4.2 Limitations

For urban planning or video game design, planners or developers strive to create a more immersive acoustic environment for users. For soundscape evaluation, accurate reproduction of the acoustic environment is also an important process in soundscape standardisation [58]. A rational simulation of non-point sources under a virtual experience will better match auditory and visual information. The results of this chapter show a significant increase in immersion ratings when the auditory angle between synthesis setups reaches the MAA of 1°, although the observer in this chapter was in motion and the auditory angle was dynamic. The variation in immersion ratings gives us an insight into how to simulate and synthesise non-point sound sources in a dynamic acoustic space.

During reproduction in the square, there is still mutual motion between the sound source and the observation point. As discussed in Chapter 2 for the audible angle in spatial hearing, the MAMA in psychoacoustics is about 2-5° [54]. The immersion ratings varied by approximately 12% and 18% respectively at 5m and 30 m in the range 1-4.5°, as shown in Fig. 7.10. In this study, the immersion ratings were significantly enhanced by approximately 15% when the audible angle was reduced from 4.5° to 1°. It is still unknown how the immersion rating varies with the angles between this range with consideration of the movement and relative position between the sound source and the observation point. I simplify the dynamically varying auditory angle to a single value, which is also a limitation of my study. Therefore, when designing a sound source of a certain scale, the study advises that the auditory spatial resolution formed by the angle between the human ear and the sound source should be less than 1°.

7.5 Conclusions

Using a single point source is not adequate to simulate sound sources in urban spaces. This chapter examined subjective responses of the perceived width, distance and immersion on multiple sounds with different synthesis setups through a VR subjective test. The results of the subjective test revealed the following:

1. Both width and immersion ratings are significantly affected by the number of synthesis points for the line source. When enough synthesis points form a small audible angle (e.g., $< 1°$), immersion rating in VR will significantly improve, as human ears cannot distinguish the location of the synthesised sound sources in the horizontal plane. When the audible angle is less than the threshold of 1°, immersion rating does not vary significantly with increasing synthesis points. The immersion ratings are significantly enhanced, by approximately 15%, when the audible angle is reduced from 4.5° to 1°.

2. The perceived distance is not significantly influenced by how much the points are synthesised in the sound sources. The difference of the perceived distance remains dominated by the variation of the spatial distance between the observer and sound source.

Overall, to improve the perception of synthesised sound sources, sufficient auditory resolution for the modelling of non-point sources was needed. When the synthesis points of sound sources reach the threshold of the audible angle, the accuracy of spatial perception of the sound in VR soundscape evaluation will significantly improve, and immersion rating will be enhanced as well. The results offer positive guidance on sound design and auralisation for soundscape evaluation, game design, urban planning and other industries in VR experiences.

Chapter 8

Conclusions

8.1 Research findings

Both soundscapes and VR stress the human experience of the environment. VR-based evaluation offers an off-site approach under controlled audio-visual conditions. Soundscape reproduction clearly requires more detailed guidance regarding reproduction, and a deeper understanding of soundscape evaluation motivates us to better improve, plan and protect the urban soundscape ecology.

This thesis has performed studies on monaural-binaural rendering methods, reflection modelling during sound propagation simulation and synthesis of line sources, and derived some of the simplifications and compromises that can be made in auralisation under VR for soundscape evaluation. According to the results and analyses of the studies, it is possible to make some replacements or simplifications for the technical specifications in soundscape rendering, sound reflection modelling, and sound source synthesis. The results of subjective evaluation in this thesis on the basis of human perception will refine our knowledge and understanding of acoustic simulation and auralisation for VR-based soundscape experiences. This section draws together the key threads running through the thesis and responds to the questions proposed in Section 1.2.

1. **How is the ecological validity of IVR identified in soundscape evaluation?**

   Through laboratory test approaches, including subjective response surveys, cognitive performance tests and physiological responses, the ecological validity of complex sound environment perception can be assessed for IVR. With participatory experiments in situ and in a laboratory, the veridicality of IVR can be verified through subjective responses, including environmental preferences/quality, audio-visual indicators (e.g., pleasantness and annoyance), coupled interactions and reproductive quality (e.g., realism and immersiveness).

A head-tracking unit with a display and synchronised spatial audio (e.g., an HMD with FOA-tracking binaural playback) is advantageous for assessing ecological validity in immersive virtual environments. When urban sound environment research involves interactions among multiple users, a CAVE system should be considered. With higher spatial resolutions, HOA also shows increasing potential for the ecological validity of IVR in urban sound environment research.

2. **How to choose appropriate monaural and binaural rendering methods in soundscape evaluation?**

   Both binaural static and monaural rendering methods can be used in soundscape evaluation for perceptual indicators, including overall impression, acoustic comfort, pleasantness, annoyance, eventfulness and loudness. However, given the adaptability of binaural sound in VR, binaural rendering methods still dominate soundscape evaluation over monaural methods.

3. **What kind of simplifications can be made to optimise the reflection modelling in environmental sound propagation simulation?**

   For sound propagation with consideration of different reflection conditions, fewer orders can be employed during sound reflection to assess different kinds of sounds in outdoor sound environments in VR experiences. In urban open squares (e.g., the square larger than 900 m^2 in Table 6.1), VR combining both an HMD and Ambisonics will significantly strengthen immersion ratings at low reflection orders (e.g., the first order). It is feasible to employ a lower reflection order during auralisation to render sounds in a VR experience with similar realism and reasonable accuracy.

4. **What kind of compromises or simulations can be made to synthesise non-point like sources to improve the perception of width, immersion and distance under virtual environments?**

   For non-point sound source simulation, especially line sources, the sound sources can be synthesised depending on the relative position between the observation point and the sound source. Both width and immersion are significantly affected by the number of synthesis points for the line source. When adequate synthesis points form a small audible angle ($< 1^\circ$), immersion in VR will significantly improve, as human ears cannot distinguish the location in the synthesised sound source in the horizontal plane. When the audible angle is less than the threshold of 1°, immersion does not increase significantly with more synthesis points.

8.2 Implementation

8.2.1 Soundscape evaluation

When facing different environmental contexts and acoustic conditions, soundscape researchers and engineers do not have sufficient reproduction guidance from existing soundscape standards. Thus, this thesis provide some validated findings and specifications for context-specific soundscape reproduction. These context-specific scenarios primarily focus on urban open spaces, e.g., squares, and frequently heard sounds. e.g., water and bird sounds. These scenes and sounds often appear in research and consultancy reports as the focus of the environmental appraisal.

Soundscape researchers and engineers can choose appropriate audio rendering methods depending on the desired interactive experience, e.g., 2D desktop display or HMD. When they attempt to use non-panoramic video or images, a monaural or binaural rendering approach is feasible for soundscape evaluation. Such an evaluation should focus on a range of indicators including pleasantness, annoyance, eventfulness, loudness, etc., as discussed in Chapter 5. When soundscape evaluation requires a more immersive interactive experience, reflection modelling and non-point like source synthesis in scenes are explored in Chapters 6 and 7. Under specific conditions as mentioned above, fewer reflections and fewer source synthesis points can be applied to evaluate soundscapes in a VR experience.

It is worth pointing out that these implementations are also limited. Due to the limited exploration of enclosed or small-scale urban spaces in this thesis, audio rendering, reflection modelling and sound source synthesis methods require more elaboration and preparation in these specific spaces. As the selection was made for normal hearing participants, people with hearing impairment or hearing loss were not applicable for the implementations. In addition, as this thesis did not develop a cross-national study, designing different language translations and contextual interpretations for different country participants also needs further screening.

8.2.2 Environmental and urban planning

For urban design and planning, the regeneration of a built-up urban area is driven by multidimensional factors. The vast majority of these factors stem from the inability of the local built or natural environment to satisfy the productive quality of life of the existing inhabitants, as is the case in distressed and decaying urban areas. Urban regeneration is supported and developed by the government, the community and the business sectors. In such an opportunity, the rethinking of the urban sound environment is worthy of additional attention.

Urban planners, landscape architects or policy makers need to have a thorough understanding of the macro and micro acoustic environments. When they first think of new designs (e.g., a new fountain) or policies (e.g., car horns banned), the short-term visible impact is taken into account. It is also important to consider how these new designs or policies will manifest themselves in a specific street or square, and therefore how they will affect the overall urban acoustic environment. At this point, this thesis provides some samples that allow designers and residents to see how the improved environment will behave. These samples offer them a workflow to reproduce audio-visual environments from the perspective of subjective and objective evaluation. Possible positive impacts or potential risks can be anticipated in advance by referring to the discussion of audio rendering, reflection modelling and sound source synthesis methods in this thesis.

8.2.3 Video game design

Three-dimensional video games have been developed in barely four decades. Developers have been striving to create an immersive gaming experience for all players, thus serving the story and gameplay. Early game engines focused on supporting polygonal models, animation, particle effects, and other 3D graphic related areas. The early functional limitations in 3D video games are now viewed from the perspective of a very weak audio-visual interactive experience with visible mosaics, jaggies and low frame rates, and the graphic presentation of 3D games is now almost on par with Hollywood films.

As experiential devices such as haptics and olfaction still require additional equipment or are difficult to implement at present, auditory-visual interaction is still one of the significant aspects of video game development. For example, for first-person shooters (FPS), developers make in situ recordings of the sounds of different firearms, and, they then match these recordings to the different firearms in the game. In the film and television industry, audio engineers often process and mix these recordings. We refer to this process here collectively as pre-rendering. However, for the video game industry, developers need additional real-time rendering. This process attempts to reduce the process of processing instruction input and rendering output to a level imperceptible by game players. It is also replete with simplifications and compromises that are contingent on human perception and immersive experience.

8.3 Future research

In this thesis, the expansion of the real-time human-computer and audio-visual interaction is limited. For real-time human-computer interactions, some potentially more immersive interaction modes and devices are not applied, such as voice input, camera motion capture, and infrared imaging. The comparison of these different modes of interaction will also be a direction of soundscape evaluation research. For audio-visual interactions, real-time auralisation with visualisation is still a challenge to strive for. Additionally, the software and hardware issues often discussed in visual engineering, including colour rendering, model polygons, angle of view, anti-aliasing, and other factors, can also greatly affect the reproduction quality for soundscape evaluation.

The existing soundscape standards, i.e., ISO 12913-1:2014, ISO/TS 12913-2:2018, and ISO/TS 12913-3:2019, do not encompass the full spectrum of soundscapes. Many researchers and engineers are still facing many difficulties in reproducing soundscapes with these existing standards. These standards are not a final solution but rather a foundation for future work. The main task in the future is to improve the various details of technical specifications for soundscape standardisation during VR reproduction. he current research on issues such as scattering, absorption, and other acoustic phenomena during auralisation for soundscapes in VR is still not sufficiently subjectively verified. Investigations into these fields of auralisation will further refine our understanding of the standardisation of soundscape reproduction. Further effort will be needed to explore the implications of human-computer interaction, visual rendering quality, and wearing comfort of the devices.

Nature has given us extremely powerful and sensitive auditory perception. We use such an auditory system to perceive the world and voice our personal emotions. The research thus far is not the end but a new beginning. More possibilities will emerge with more lightweight and smart wearable devices driven by AI technology. They will be implicitly integrated into our daily lives, as smartphones have become critical to the networks of billions of people in just over a decade of evolution, and will form a vital part of research and industrial revolution to shape a sustainable and ecological future.

Bibliography

[1] J. D. Miller. Effects of noise on people. *The Journal of the Acoustical Society of America*, 56(3):729–764, 1974.

[2] F. Aletta, T. Oberman, and J. Kang. Associations between positive health-related effects and soundscapes perceptual constructs: A systematic review. *International Journal of Environmental Research and Public Health*, 15(11):2392, 2018.

[3] M. F. Southworth. *The sonic environment of cities.* PhD thesis, Massachusetts Institute of Technology, 1967.

[4] R. M. Schafer. *The tuning of the world.* Alfred A. Knopf, 1977.

[5] J. Kang. *Urban sound environment.* CRC Press, 2006.

[6] J. Kang and B. Schulte-Fortkamp. *Soundscape and the built environment.* CRC press, 2016.

[7] K. N. Irvine, P. Devine-Wright, S. R. Payne, R. A. Fuller, B. Painter, and K. J. Gaston. Green space, soundscape and urban sustainability: an interdisciplinary, empirical study. *Local Environment*, 14(2):155–172, 2009.

[8] M. Vorländer. *Auralization.* Springer, 2020.

[9] B. N. Postma and B. F. Katz. The influence of visual distance on the room-acoustic experience of auralizations. *The Journal of the Acoustical Society of America*, 142(5):3035–3046, 2017.

[10] M. S. Gazzaniga, R. B. Ivry, and G. R. Mangun. *Cognitive neuroscience: The biology of the mind (4th ed.).* Norton: New York, 2013.

[11] J. Kang, F. Aletta, T. T. Gjestland, L. A. Brown, D. Botteldooren, B. Schulte-Fortkamp, P. Lercher, I. van Kamp, K. Genuit, A. Fiebig, J. L. Bento Coelho, L. Maffei, and L. Lavia. Ten questions on the soundscapes of the built environment. *Building and Environment*, 108:284–294, 2016.

[12] E. Öhrström, R. Rylander, and M. Björkman. Effects of night time road traffic noise-an overview of laboratory and field studies on noise dose and subjective noise sensitivity. *Journal of Sound and Vibration*, 127(3):441–448, 1988.

[13] S. A. Stansfeld and M. P. Matheson. Noise pollution: non-auditory effects on health. *British Medical Bulletin*, 68(1):243–257, 2003.

[14] The WHOQOL Group. The world health organization quality of life assessment (whoqol): development and general psychometric properties. *Social Science & Medicine*, 46(12):1569–1585, 1998.

[15] B. Berglund, T. Lindvall, D. H. Schwela, W. H. O. Occupational, and E. H. Team. Guidelines for community noise. World Health Organization, 1999.

[16] Department for Environment Food Rural Affairs. Strategic noise mapping (2017). Online; accessed 01 June 2021. https://www.gov.uk/government/publications/strategic-noise-mapping-2019, 2017.

[17] The City of London Corporation. City of London Noise Strategy 2016 to 2026. Online; accessed 01 June 2021. https://www.cityoflondon.gov.uk/assets/Services-Environment/col-noise-strategy-2016-to-2026.pdf, 2016.

[18] W. Yang and J. Kang. Acoustic comfort evaluation in urban open public spaces. *Applied Acoustics*, 66(2):211–229, 2005.

[19] E. De Ruiter. Noise control in the compact city. In *Proceedings of the 7th International Congress on Sound and Vibration*, pages 2421–2426, 2000.

[20] B. Schulte-Fortkamp. The quality of acoustic environments and the meaning of soundscapes. In *Proceedings of the 17th International Conference on Acoustics, Rome, Italy*, pages 2–3, 2001.

[21] B. C. Pijanowski, A. Farina, S. H. Gage, S. L. Dumyahn, and B. L. Krause. What is soundscape ecology? an introduction and overview of an emerging new science. *Landscape Ecology*, 26(9):1213–1232, 2011.

[22] R. M. Schafer. *The soundscape: Our sonic environment and the tuning of the world.* Simon and Schuster, 1993.

[23] J. Kang, F. Aletta, T. Oberman, M. Erfanian, M. Kachlicka, M. Lionello, and A. Mitchell. Towards soundscape indices. In *Proceedings of the 23rd International Congress on Acoustics*, pages 2488–2495. Deutsche Gesellschaft für Akustik (DEGA), 2019.

[24] D. Dubois, C. Guastavino, and M. Raimbault. A cognitive approach to urban soundscapes: Using verbal data to access everyday life auditory categories. *Acta Acustica united with Acustica*, 92(6):865–874, 2006.

[25] S. L. Dumyahn and B. C. Pijanowski. Soundscape conservation. *Landscape Ecology*, 26(9):1327, 2011.

[26] ISO 12913-1:2014. *Acoustics-Soundscape-Part 1: Definition and conceptual framework.* International Organization for Standardization, Geneva, Switzerland, 2014.

[27] H. Tong and J. Kang. Relationships between noise complaints and socio-economic factors in England. *Sustainable Cities and Society*, 65:102573, 2021.

[28] J. M. Downing and E. Stusnick. Measurement of the natural soundscape in national parks. *The Journal of the Acoustical Society of America*, 108(5):2497–2497, 2000.

[29] B. Krause, S. H. Gage, and W. Joo. Measuring and interpreting the temporal variability in the soundscape at four places in sequoia national park. *Landscape Ecology*, 26(9):1247–1256, 2011.

[30] D. Shepherd, D. Welch, K. N. Dirks, and D. McBride. Do quiet areas afford greater health-related quality of life than noisy areas? *International Journal of Environmental Research and Public Health*, 10(4):1284–1303, 2013.

[31] Y. Zhang, J. Kang, and J. Kang. Effects of soundscape on the environmental restoration in urban natural environments. *Noise & Health*, 19(87):65, 2017.

[32] B. J. Park, Y. Tsunetsugu, T. Kasetani, H. Hirano, T. Kagawa, M. Sato, and Y. Miyazaki. Physiological effects of shinrin-yoku (taking in the atmosphere of the forest)-using salivary cortisol and cerebral activity as indicators-. *Journal of Physiological Anthropology*, 26(2):123–128, 2007.

[33] A. Irwin, D. A. Hall, A. Peters, and C. J. Plack. Listening to urban soundscapes: Physiological validity of perceptual dimensions. *Psychophysiology*, 48(2):258–268, 2011.

[34] G. Ouzounian. Recomposing the city: A survey of recent sound art in Belfast. *Leonardo Music Journal*, pages 47–54, 2013.

[35] A. Bennett and I. Rogers. Street music, technology and the urban soundscape. *Continuum*, 28(4):454–464, 2014.

[36] E. A. Thompson. *The soundscape of modernity: architectural acoustics and the culture of listening in America, 1900-1933.* MIT Press, 2004.

[37] B. C. Moore. *An introduction to the psychology of hearing.* Brill, 2012.

[38] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and models*, volume 22. Springer Science & Business Media, 2013.

[39] D. Botteldooren and B. De Coensel. The role of saliency, attention and source identification in soundscape research. In *INTER-NOISE and NOISE-*

CON Congress and Conference Proceedings, volume 2009, pages 2198–2205. Institute of Noise Control Engineering, 2009.

[40] E. Schröger, M.-H. Giard, and C. Wolff. Auditory distraction: event-related potential and behavioral indices. *Clinical Neurophysiology*, 111(8):1450–1460, 2000.

[41] B. De Coensel and D. Botteldooren. A model of saliency-based auditory attention to environmental sound. In *Proceedings of the 20th International Congress on Acoustics, Sydney, Australia*, pages 1–8, 2010.

[42] A. S. Bregman. *Auditory scene analysis: The perceptual organization of sound.* MIT press, 1994.

[43] C. Kayser, C. I. Petkov, M. Lippert, and N. K. Logothetis. Mechanisms for allocating auditory attention: an auditory saliency map. *Current Biology*, 15(21):1943–1947, 2005.

[44] O. Kalinli and S. S. Narayanan. A saliency-based auditory attention model with applications to unsupervised prominent syllable detection in speech. In *INTERSPEECH*, pages 1941–1944, 2007.

[45] D. Botteldooren and B. De Coensel. Informational masking and attention focussing on environmental sound. In *Proceedings of the NAG/DAGA meeting, Rotterdam, Netherlands*, pages 23–26, 2009.

[46] I. Pollack. Auditory informational masking. *The Journal of the Acoustical Society of America*, 57(S1):S5–S5, 1975.

[47] J. Y. Jeon, P. J. Lee, J. You, and J. Kang. Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds. *The Journal of the Acoustical Society of America*, 127(3):1357–1366, 2010.

[48] Y. Hao, J. Kang, and J. D. Krijnders. Integrated effects of urban morphology on birdsong loudness and visibility of green areas. *Landscape and Urban Planning*, 137:149–162, 2015.

[49] H. L. Lane, A. C. Catania, and S. S. Stevens. Voice level: Autophonic scale, perceived loudness, and effects of sidetone. *The Journal of the Acoustical Society of America*, 33(2):160–167, 1961.

[50] G. A. Gescheider. *Psychophysics: the fundamentals*. Psychology Press, 2013.

[51] M. E. Nilsson, J. Alvarsson, M. Rådsten-Ekman, and K. Bolin. Auditory masking of wanted and unwanted sounds in a city park. *Noise Control Engineering Journal*, 58(5):524–531, 2010.

[52] M. Risoud, J.-N. Hanson, F. Gauvrit, C. Renard, P.-E. Lemesre, N.-X. Bonne, and C. Vincent. Sound source localization. *European annals of otorhinolaryngology, head and neck diseases*, 135(4):259–264, 2018.

[53] A. W. Mills. On the minimum audible angle. *The Journal of the Acoustical Society of America*, 30(4):237–246, 1958.

[54] D. W. Grantham, B. W. Hornsby, and E. A. Erpenbeck. Auditory spatial resolution in horizontal, vertical, and diagonal planes. *The Journal of the Acoustical Society of America*, 114(2):1009–1022, 2003.

[55] D. W. Batteau. The role of the pinna in human localization. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 168(1011):158–180, 1967.

[56] Ö. Axelsson, C. Guastavino, and S. R. Payne. Soundscape assessment. *Frontiers in Psychology*, 10:2514, 2019.

[57] H. Westerkamp. Soundwalking. *Sound Heritage*, 3(4):18–27, 1974.

[58] ISO/TS 12913-2:2018. *Acoustics-Soundscape-Part 2: Data collection and reporting requirements*. International Organization for Standardization, Geneva, Switzerland, 2018.

[59] ISO/TS 12913-3:2019. *Acoustics-Soundscape-Part 3: Data Analysis*. International Organization for Standardization, Geneva, Switzerland, 2019.

[60] ISO/TS 15666:2003. *Acoustics-Assessment of Noise Annoyance by Means of Social and Socio-Acoustic Surveys.* International Organization for Standardization, Geneva, Switzerland, 2003.

[61] A. Brown, J. Kang, and T. Gjestland. Towards standardization in soundscape preference assessment. *Applied Acoustics*, 72(6):387–392, 2011.

[62] J. Liu, J. Kang, H. Behm, and T. Luo. Effects of landscape on soundscape perception: Soundwalks in city parks. *Landscape and Urban Planning*, 123:30–40, 2014.

[63] C. Semidor. Listening to a city with the soundwalk method. *Acta Acustica united with Acustica*, 92(6):959–964, 2006.

[64] T. Oberman, K. Jambrošić, M. Horvat, and B. B. O. Šćitaroci. Using virtual soundwalk approach for assessing sound art soundscape interventions in public spaces. *Applied Sciences*, 10(6):2102, 2020.

[65] J. Liu, J. Kang, T. Luo, and H. Behm. Landscape effects on soundscape experience in city parks. *Science of the Total Environment*, 454:474–481, 2013.

[66] L. Jiang, M. Masullo, L. Maffei, F. Meng, and M. Vorländer. How do shared-street design and traffic restriction improve urban soundscape and human experience?-an online survey with virtual reality. *Building and Environment*, 143:318–328, 2018.

[67] C. J. Yu and J. Kang. Soundscape in the sustainable living environment: A cross-cultural comparison between the UK and Taiwan. *Science of the Total Environment*, 482:501–509, 2014.

[68] A. Mitchell, T. Oberman, F. Aletta, M. Erfanian, M. Kachlicka, M. Lionello, and J. Kang. The soundscape indices (ssid) protocol: A method for urban soundscape surveys-questionnaires with acoustical and contextual information. *Applied Sciences*, 10(7):2397, 2020.

[69] L. Deng, J. Kang, W. Zhao, and K. Jambrošić. Cross-national comparison of soundscape in urban public open spaces between China and Croatia. *Applied Sciences*, 10(3):960, 2020.

[70] C. Xu and J. Kang. Soundscape evaluation: Binaural or monaural? *The Journal of the Acoustical Society of America*, 145(5):3208–3217, 2019.

[71] C. Xu and J. Kang. Simplification of reflection orders in virtual soundscapes through a subjective evaluation. In *Proceedings of the 23rd International Congress on Acoustics*, pages 6432–6437. Deutsche Gesellschaft für Akustik (DEGA), 2019.

[72] L. F. H. Cadena, A. C. L. Soares, I. Pavón, and L. B. Coelho. Assessing soundscape: Comparison between in situ and laboratory methodologies. *Noise Mapping*, 4(1):57–66, 2017.

[73] C. Guastavino, B. F. Katz, J. D. Polack, D. J. Levitin, and D. Dubois. Ecological validity of soundscape reproduction. *Acta Acustica united with Acustica*, 91(2):333–341, 2005.

[74] W. J. Davies, N. S. Bruce, and J. E. Murphy. Soundscape reproduction and synthesis. *Acta Acustica united with Acustica*, 100(2):285–292, 2014.

[75] J. Y. Jeon, J. Y. Hong, C. Lavandier, J. Lafon, Ö. Axelsson, and M. Hurtig. A cross-national comparison in assessment of urban park soundscapes in France, Korea, and Sweden through laboratory experiments. *Applied Acoustics*, 133:107–117, 2018.

[76] F. Aletta, T. Oberman, Ö. Axelsson, H. Xie, Y. Zhang, S.-K. Lau, S.-K. Tang, K. Jambrošić, B. De Coensel, K. Van den Bosch, et al. Soundscape assessment: Towards a validated translation of perceptual attributes in different languages. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 261, pages 3137–3146. Institute of Noise Control Engineering, 2020.

[77] M. Raimbault, C. Lavandier, and M. Bérengier. Ambient sound assessment of urban environments: field studies in two French cities. *Applied Acoustics*, 64(12):1241–1256, 2003.

[78] D. Welch, D. Shepherd, K. Dirks, M. Y. Tan, and G. Coad. Use of creative writing to develop a semantic differential tool for assessing soundscapes. *Frontiers in Psychology*, 9:2698, 2019.

[79] J. Kang and M. Zhang. Semantic differential analysis of the soundscape in urban open public spaces. *Building and Environment*, 45(1):150–157, 2010.

[80] J. Kang and M. Zhang. Semantic differential analysis on the soundscape of urban open public spaces. *The Journal of the Acoustical Society of America*, 112(5):2435–2435, 2002.

[81] A. J. Torija, D. P. Ruiz, and A. Ramos-Ridao. Application of a methodology for categorizing and differentiating urban soundscapes using acoustical descriptors and semantic-differential attributes. *The Journal of the Acoustical Society of America*, 134(1):791–802, 2013.

[82] Ö. Axelsson, M. E. Nilsson, and B. Berglund. A principal components model of soundscape perception. *The Journal of the Acoustical Society of America*, 128(5):2836–2846, 2010.

[83] F. Aletta, J. Kang, and Ö. Axelsson. Soundscape descriptors and a conceptual framework for developing predictive soundscape models. *Landscape and Urban Planning*, 149:65–74, 2016.

[84] J. Y. Hong and J. Y. Jeon. Designing sound and visual components for enhancement of urban soundscapes. *The Journal of the Acoustical Society of America*, 134(3):2026–2036, 2013.

[85] J. Y. Jeon and J. Y. Hong. Classification of urban park soundscapes through perceptions of the acoustical environments. *Landscape and Urban Planning*, 141:100–111, 2015.

[86] L. Hermida and I. Pavón. Spatial aspects in urban soundscapes: Binaural parameters application in the study of soundscapes from Bogotá-Colombia and Brasília-Brazil. *Applied Acoustics*, 145:420–430, 2019.

[87] A. J. Torija, D. P. Ruiz, and Á. F. Ramos-Ridao. A tool for urban soundscape evaluation applying support vector machines for developing a soundscape classification model. *Science of the Total Environment*, 482:440–451, 2014.

[88] R. Cain, P. Jennings, and J. Poxon. The development and application of the emotional dimensions of a soundscape. *Applied Acoustics*, 74(2):232–239, 2013.

[89] X. Zhang, M. Ba, J. Kang, and Q. Meng. Effect of soundscape dimensions on acoustic comfort in urban open public spaces. *Applied Acoustics*, 133:73–81, 2018.

[90] D. A. Williams, C. Brook, D. T. Murphy, A. Thomas, B. J. Cox, and C. Clark. Towards biophysiological and acoustic markers for perceived annoyance in response to reproduced environmental soundscapes. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 259, pages 6685–6694. Institute of Noise Control Engineering, 2019.

[91] F. Stevens, D. Murphy, and S. L. Smith. The self-assessment manikin and heart rate: responses to auralised soundscapes. In *Interactive Audio System Symposium, New York*, 2016.

[92] Z. Li and J. Kang. Sensitivity analysis of changes in human physiological indicators observed in soundscapes. *Landscape and Urban Planning*, 190:103593, 2019.

[93] M. Erfanian, A. J. Mitchell, J. Kang, and F. Aletta. The psychophysiological implications of soundscape: A systematic review of empirical literature and a research agenda. *International journal of environmental research and public health*, 16(19):3533, 2019.

[94] M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[95] F. Stevens, D. Murphy, and S. Smith. Soundscape categorisation and the self-assessment manikin. In *Proceedings of the 20th International Conference on Digital Audio Effects*, 2017.

[96] M. Schroeder, B. Atal, and C. Bird. Digital computers in room acoustics. In *Proceedings of the 4th International Congresses on Acoustics, Copenhagen, Denmark*, pages M21.1–M21.4, 1962.

[97] A. Krokstad, S. Strom, and S. Sørsdal. Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 8(1):118–125, 1968.

[98] M. Hornikx. Ten questions concerning computational urban acoustics. *Building and Environment*, 106:409–421, 2016.

[99] B. N. Postma and B. F. Katz. Perceptive and objective evaluation of calibrated room acoustic simulation auralizations. *The Journal of the Acoustical Society of America*, 140(6):4326–4337, 2016.

[100] J. Kang. Numerical modeling of the sound fields in urban squares. *The Journal of the Acoustical Society of America*, 117(6):3695–3706, 2005.

[101] J. Kang. Numerical modelling of the sound fields in urban streets with diffusely reflecting boundaries. *Journal of Sound and Vibration*, 258(5):793–813, 2002.

[102] M. Oegren and W. Kropp. Road traffic noise propagation between two dimensional city canyons using an equivalent sources approach. *Acta Acustica united with Acustica*, 90(2):293–300, 2004.

[103] D. Botteldooren. Finite-difference time-domain simulation of low-frequency room acoustic problems. *The Journal of the Acoustical Society of America*, 98(6):3302–3308, 1995.

[104] L. Savioja. Real-time 3d finite-difference time-domain simulation of low- and mid-frequency room acoustics. In *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx), Graz, Austria*, volume 1, page 75, 2010.

[105] S. M. Pasareanu, M. C. Remillieux, and R. A. Burdisso. Energy-based method for near-real time modeling of sound field in complex urban environments. *The Journal of the Acoustical Society of America*, 132(6):3647–3658, 2012.

[106] D. Murphy and D. Howard. 2-d digital waveguide mesh topologies in room acoustics modelling. In *Proceedings of the 3rd International Conference on Digital Audio Effects (DAFx), Verona, Italy*, pages 211–216, 2000.

[107] D. Murphy, A. Kelloniemi, J. Mullen, and S. Shelley. Acoustic modeling using the digital waveguide mesh. *IEEE Signal Processing Magazine*, 24(2):55–66, 2007.

[108] M. Beeson and D. Murphy. Roomweaver: A digital waveguide mesh based room acoustics research tool. In *Proceedings of the 7th International Conference on Digital Audio Effects (DAFx), Naples, Italy*, pages 268–273, 2004.

[109] D. Murphy and J. Mullen. Digital waveguide mesh modelling of room acoustics: Improved anechoic boundaries. In *Proceedings of the 5th International Conference on Digital Audio Effects (DAFx), Hamburg, Germany*, volume 2, pages 163–168, 2002.

[110] D. Murphy and M. Beeson. The kw-boundary hybrid digital waveguide mesh for room acoustics applications. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(2):552–564, 2007.

[111] D. Murphy, M. Beeson, S. Shelley, A. Moore, and A. Southern. Hybrid room impulse response synthesis in digital waveguide mesh based room acoustics simulation. In *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08), Espoo, Finland.*, pages 129–136, 2008.

[112] A. Southern, J. Wells, and D. Murphy. Rendering walk-through auralisations using wave-based acoustical models. In *Proceedings of the 17th European Signal Processing Conference*, pages 715–719. IEEE, 2009.

[113] Welsh Government. Noise and soundscape action plan 2018 to 2023. Online; accessed 01 June 2021. https://gov.wales/noise-and-soundscape-action-plan-2018-2023-0/, 2018.

[114] W. Yang and J. Kang. Soundscape and sound preferences in urban squares: a case study in Sheffield. *Journal of Urban Design*, 10(1):61–80, 2005.

[115] J. Cao and J. Kang. The influence of companion factors on soundscape evaluations in urban public spaces. *Sustainable Cities and Society*, page 102860, 2021.

[116] W. J. Davies, M. D. Adams, N. S. Bruce, R. Cain, A. Carlyle, P. Cusack, D. A. Hall, K. I. Hume, A. Irwin, P. Jennings, M. Marselle, C. J. Plack, and J. Poxon. Perception of soundscapes: An interdisciplinary approach. *Applied Acoustics*, 74(2):224–231, 2013.

[117] J. You, P. J. Lee, and J. Y. Jeon. Evaluating water sounds to improve the soundscape of urban areas affected by traffic noise. *Noise Control Engineering Journal*, 58(5):477–483, 2010.

[118] Sheffield City Council. Sheaf square. Online; accessed 01 June 2021. https://www.sheffield.gov.uk/home/parks-sport-recreation/public-spaces/sheaf-square/, 2019.

[119] B. Schulte-Fortkamp, R. Volz, and A. Jakob. Using the soundscape approach

to develop a public space in berlin-perception and evaluation. *The Journal of the Acoustical Society of America*, 123(5):3808–3808, 2008.

[120] B. Schulte-Fortkamp and P. Jordan. When soundscape meets architecture. *Noise Mapping*, 3(1), 2016.

[121] G. Fusaro, F. D'Alessandro, G. Baldinelli, and J. Kang. Design of urban furniture to enhance the soundscape: A case study. *Building Acoustics*, 25(1):61–75, 2018.

[122] M. D. Fowler. Soundscape as a design strategy for landscape architectural praxis. *Design Studies*, 34(1):111–128, 2013.

[123] G. C. Burdea and P. Coiffet. *Virtual reality technology*. John Wiley & Sons, 2003.

[124] M. L. Heilig. Sensorama simulator, Aug. 1962. US Patent 3,050,870.

[125] H. Rheingold. *The virtual community: Homesteading on the electronic frontier*. MIT press, 2000.

[126] I. E. Sutherland. The ultimate display. *Multimedia: From Wagner to virtual reality*, pages 506–508, 1965.

[127] M. W. Krueger and M. W. Krueger. *Artificial reality*, volume 126. Addison-Wesley Reading, MA, 1983.

[128] T. Honjo and E. M. Lim. Visualization of landscape by VRML system. *Landscape and Urban Planning*, 55(3):175–183, 2001.

[129] E. Brown and P. Cairns. A grounded investigation of game immersion. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 1297–1300, 2004.

[130] S. S. Fisher, M. McGreevy, J. Humphries, and W. Robinett. Virtual environment display system. In *Proceedings of the 1986 workshop on Interactive 3D graphics*, pages 77–87, 1987.

[131] F. P. Brooks. What's real about virtual reality? *IEEE Computer graphics and applications*, 19(6):16–27, 1999.

[132] S. Nichols, S. Cobb, and J. R. Wilson. Health and safety implications of virtual environments: Measurement issues. *Presence: Teleoperators & Virtual Environments*, 6(6):667–675, 1997.

[133] S. Cobb, S. Nichols, A. Ramsey, and J. R. Wilson. Virtual reality-induced symptoms and effects (VRISE). *Presence: Teleoperators & Virtual Environments*, 8(2):169–186, 1999.

[134] S. Sharples, S. Cobb, A. Moody, and J. R. Wilson. Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. *Displays*, 29(2):58–69, 2008.

[135] S. Nichols and H. Patel. Health and safety implications of virtual reality: a review of empirical evidence. *Applied Ergonomics*, 33(3):251–271, 2002.

[136] K. Hill and P. A. Howarth. Habituation to the side effects of immersion in a virtual environment. *Displays*, 21(1):25–30, 2000.

[137] J. P. Rolland and H. Hua. Head-mounted display systems. *Encyclopedia of Optical Engineering*, 2, 2005.

[138] I. E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I, San Francisco, CA, USA*, pages 757–764. ACM, 1968.

[139] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. The cave: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–73, 1992.

[140] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the cave. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, Anaheim, CA, USA*, pages 135–142. Citeseer, 1993.

[141] M. A. Muhanna. Virtual reality and the cave: Taxonomy, interaction challenges and research directions. *Journal of King Saud University-Computer and Information Sciences*, 27(3):344–361, 2015.

[142] P. Havig, J. McIntire, and E. Geiselman. Virtual reality in a cave: limitations and the need for HMDs? In *Head-and Helmet-Mounted Displays XVI: Design and Applications*, volume 8041, page 804107. International Society for Optics and Photonics, 2011.

[143] F. Rumsey. *Spatial audio*. Taylor & Francis, 2001.

[144] J. Y. Hong, J. He, B. Lam, R. Gupta, and W. S. Gan. Spatial audio for soundscape design: Recording and reproduction. *Applied Sciences*, 7(6):627, 2017.

[145] J. Daniel, S. Moreau, and R. Nicol. Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging. In *Audio Engineering Society Convention 114, Amsterdam, Netherlands*, page 5788. Audio Engineering Society, 2003.

[146] J. G. Tylka and E. Y. Choueiri. Models for evaluating navigational techniques for higher-order ambisonics. In *Proceedings of Meetings on Acoustics 173EAA, Boston, USA*, volume 30, page 050009. Acoustical Society of America, 2017.

[147] M. A. Gerzon. Periphony: With-height sound reproduction. *Journal of the Audio Engineering Society*, 21(1):2–10, 1973.

[148] M. A. Gerzon. Ambisonics in multichannel broadcasting and video. *Journal of the Audio Engineering Society*, 33(11):859–871, 1985.

[149] W. G. Gardner. *3-D audio using loudspeakers*, volume 444. Springer Science & Business Media, 1998.

[150] M. A. Poletti. Three-dimensional surround sound systems based on spherical harmonics. *Journal of the Audio Engineering Society*, 53(11):1004–1025, 2005.

[151] J. Daniel. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format. In *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction, Copenhagen, Denmark*, pages 1–15. Audio Engineering Society, 2003.

[152] T. Betlehem and T. D. Abhayapala. Theory and design of sound field reproduction in reverberant rooms. *The Journal of the Acoustical Society of America*, 117(4):2100–2111, 2005.

[153] A. S. Sudarsono, Y. W. Lam, and W. J. Davies. The effect of sound level on perception of reproduced soundscapes. *Applied Acoustics*, 110:53–60, 2016.

[154] F. Stevens, D. Murphy, and S. Smith. Ecological validity of stereo UHJ soundscape reproduction. In *Audio Engineering Society Convention 142, Berlin, Germany*, page 9765. Audio Engineering Society, 2017.

[155] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–466, 1997.

[156] A. J. Berkhout. A holographic approach to acoustic control. *Journal of the Audio Engineering Society*, 36(12):977–995, 1988.

[157] A. J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *The Journal of the Acoustical Society of America*, 93(5):2764–2778, 1993.

[158] J. L. Carles, I. L. Barrio, and J. V. de Lucio. Sound influence on landscape values. *Landscape and Urban Planning*, 43(4):191–200, 1999.

[159] J. L. Schwartz, F. Berthommier, and C. Savariaux. Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition*, 93(2):B69–B78, 2004.

[160] J. Y. Hong, B. Lam, Z. T. Ong, R. Gupta, W. S. Gan, S. H. Chong, and J. Feng. Suitability of natural sounds to enhance soundscape quality in urban residential areas. In *Proceedings of the 24th International Congress on Sound and Vibration, London, UK*, pages 1–6, 2017.

[161] L. Maffei, T. Iachini, M. Masullo, F. Aletta, F. Sorrentino, V. P. Senese, and F. Ruotolo. The effects of vision-related aspects on noise perception of wind turbines in quiet areas. *International Journal of Environmental Research and Public Health*, 10(5):1681–1697, 2013.

[162] L. Maffei, M. Masullo, F. Aletta, and M. Di Gabriele. The influence of visual characteristics of barriers on railway noise perception. *Science of the Total Environment*, 445:41–47, 2013.

[163] T. Iachini, L. Maffei, F. Ruotolo, V. P. Senese, G. Ruggiero, M. Masullo, and N. Alekseeva. Multisensory assessment of acoustic comfort aboard metros: a virtual reality study. *Applied Cognitive Psychology*, 26(5):757–767, 2012.

[164] F. Ruotolo, L. Maffei, M. Di Gabriele, T. Iachini, M. Masullo, G. Ruggiero, and V. P. Senese. Immersive virtual reality and environmental noise assessment: An innovative audio–visual approach. *Environmental Impact Assessment Review*, 41:10–20, 2013.

[165] M. J. Habgood, D. Wilson, D. Moore, and S. Alapont. Hci lessons from playstation vr. In *Extended abstracts publication of the annual symposium on computer-human interaction in play, Amsterdam, Netherlands*, pages 125–135, 2017.

[166] D. S. Brungart, B. D. Simpson, R. L. McKinley, A. J. Kordik, R. C. Dallman, and D. A. Ovenshire. The interaction between head-tracker latency, source duration, and response time in the localization of virtual sound sources. In *Proceedings of the 10th International Conference on Auditory Display, Sydney, Australia*, pages 1–7, 2004.

[167] D. S. Brungart, B. D. Simpson, and A. J. Kordik. The detectability of head-tracker latency in virtual audio displays. In *Proceedings of the 11th International Conference on Auditory Display, Limerick, Ireland*, pages 37–42, 2005.

[168] D. A. Bowman and R. P. McMahan. Virtual reality: how much immersion is enough? *Computer*, 40(7), 2007.

[169] M. Luigi, M. Massimiliano, P. Aniello, R. Gennaro, and P. R. Virginia. On the validity of immersive virtual reality as tool for multisensory evaluation of urban spaces. *Energy Procedia*, 78:471–476, 2015.

[170] T. Iachini, Y. Coello, F. Frassinetti, and G. Ruggiero. Body space in social interactions: a comparison of reaching and comfort distance in immersive virtual reality. *PloS one*, 9(11):e111511, 2014.

[171] T. Iachini, Y. Coello, F. Frassinetti, V. P. Senese, F. Galante, and G. Ruggiero. Peripersonal and interpersonal space in virtual and real environments: Effects of gender and age. *Journal of Environmental Psychology*, 45:154–164, 2016.

[172] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.

[173] T. Rossetti and R. Hurtubia. An assessment of the ecological validity of immersive videos in stated preference surveys. *Journal of Choice Modelling*, 34:100198, 2020.

[174] I. D. Bishop and B. Rohrmann. Subjective responses to simulated and real environments: a comparison. *Landscape and Urban Planning*, 65(4):261–277, 2003.

[175] Y. A. W. de Kort, W. A. Ijsselsteijn, J. Kooijman, and Y. Schuurmans. Virtual laboratories: Comparability of real and virtual environments for en-

vironmental psychology. *Presence: Teleoperators & Virtual Environments*, 12(4):360–373, 2003.

[176] S. L. Huang. The validity of visual surrogates for representing waterscapes. *Landscape Research*, 34(3):323–335, 2009.

[177] K. Chamilothori, J. Wienold, and M. Andersen. Adequacy of immersive virtual reality for the perception of daylit spaces: comparison of real and virtual environments. *Leukos*, 15(2-3):203–226, 2019.

[178] L. Maffei, M. Masullo, A. Pascale, G. Ruggiero, and V. P. Romero. Immersive virtual reality in community planning: Acoustic and visual congruence of simulated vs real world. *Sustainable Cities and Society*, 27:338–345, 2016.

[179] G. M. E. Sanchez, T. Van Renterghem, K. Sun, B. De Coensel, and D. Botteldooren. Using virtual reality for assessing the role of noise in the audio-visual design of an urban public space. *Landscape and Urban Planning*, 167:98–107, 2017.

[180] K. Sun, B. De Coensel, K. Filipan, F. Aletta, T. Van Renterghem, T. De Pessemier, W. Joseph, and D. Botteldooren. Classification of soundscapes of urban public open spaces. *Landscape and Urban Planning*, 189:139–155, 2019.

[181] J. Y. Jeon and H. I. Jo. Effects of audio-visual interactions on soundscape and landscape perception and their influence on satisfaction with the urban environment. *Building and Environment*, 169:106544, 2020.

[182] Y. Hao, J. Kang, and H. Wörtche. Assessment of the masking effects of birdsong on the road traffic noise environment. *The Journal of the Acoustical Society of America*, 140(2):978–987, 2016.

[183] Y. Meng and J. Kang. Fast simulation of sound fields for urban square animation. In *Proceedings of the 33rd International Congress on Noise Control*

Engineering, Prague, Czech Republic, volume 2004, pages 1888–1895. Institute of Noise Control Engineering, 2004.

[184] J. Kang, Y. Meng, and Y. Smyrnova. Objective and subjective evaluation of urban acoustic modelling and auralisation. In *Audio Engineering Society Convention 124, Amsterdam, Netherlands*, page 7325. Audio Engineering Society, AES E-LIBRARY, 2008.

[185] C. Schissler, A. Nicholls, and R. Mehra. Efficient HRTF-based spatial audio for area and volumetric sources. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1356–1366, 2016.

[186] E. Aromataris and A. Pearson. The systematic review: an overview. *AJN The American Journal of Nursing*, 114(3):53–58, 2014.

[187] H. J. Eysenck. Systematic reviews: Meta-analysis and its problems. *The BMJ*, 309(6957):789–792, 1994.

[188] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, and D. Moher. The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology*, 62(10):e1–e34, 2009.

[189] M. Lionello, F. Aletta, and J. Kang. A systematic review of prediction models for the experience of urban soundscapes. *Applied Acoustics*, 170:107479, 2020.

[190] Y. Hasegawa and S. K. Lau. Audiovisual bimodal and interactive effects for soundscape design of the indoor environments: A systematic review. *Sustainability*, 13(1):339, 2021.

[191] J. Kang. Sound propagation in interconnected urban streets: a parametric study. *Environment and Planning B: Planning and Design*, 28(2):281–294, 2001.

[192] J. Kang and Z. Li. Numerical simulation of an acoustic window system using finite element method. *Acta Acustica united with acustica*, 93(1):152–163, 2007.

[193] H. Fasti, A. Jaroszewski, E. Schorer, and E. Zwicker. Equal loudness contours between 100 and 1000 hz for 30, 50, and 70 phon. *Acta Acustica united with Acustica*, 70(3):197–201, 1990.

[194] A. Axelsson and A. Sandh. Tinnitus in noise-induced hearing loss. *British Journal of Audiology*, 19(4):271–276, 1985.

[195] J. Y. Jeon and H. I. Jo. Three-dimensional virtual reality-based subjective evaluation of road traffic noise heard in urban high-rise residential buildings. *Building and Environment*, 148:468–477, 2019.

[196] J. Y. Hong, B. Lam, Z. T. Ong, K. Ooi, W. S. Gan, J. Kang, J. Feng, and S. T. Tan. Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications. *Building and Environment*, 149:1–14, 2019.

[197] M. Cornock. How the writers of case reports need to consider and address consent and the general data protection regulation (GDPR). *Case Reports in Women's Health*, 19:e00060, 2018.

[198] M. Goddard. The eu general data protection regulation (GDPR): European regulation that has a global impact. *International Journal of Market Research*, 59(6):703–705, 2017.

[199] C. Xu, T. Oberman, F. Aletta, H. Tong, and J. Kang. Ecological validity of immersive virtual reality (IVR) techniques for the perception of urban sound environments. *Acoustics*, 3(1):11–24, 2021.

[200] J. J. Gibson. *The ecological approach to visual perception: classic edition*. Psychology Press, 2014.

[201] R. I. Naugle and G. J. Chelune. Integrating neuropsychological and "real-life" data: A neuropsychological model for assessing everyday functioning. In *The neuropsychology of everyday life: Assessment and basic competencies*, pages 57–73. Springer, 1990.

[202] T. D. Parsons. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, 9:660, 2015.

[203] J. M. Loomis, J. J. Blascovich, and A. C. Beall. Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers*, 31(4):557–564, 1999.

[204] I. D. Bishop, J. R. Wherrett, and D. R. Miller. Assessment of path choices on a country walk using a virtual environment. *Landscape and Urban Planning*, 52(4):225–237, 2001.

[205] D. M. Spooner and N. A. Pachana. Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of Clinical Neuropsychology*, 21(4):327–337, 2006.

[206] R. J. Sbordone and C. Long. *Ecological validity of neuropsychological testing*. CRC Press, 1996.

[207] N. Chaytor and M. Schmitter-Edgecombe. The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, 13(4):181–197, 2003.

[208] L. Kenworthy, B. E. Yerys, L. G. Anthony, and G. L. Wallace. Understanding executive control in autism spectrum disorders in the lab and in the real world. *Neuropsychology Review*, 18(4):320–338, 2008.

[209] R. L. Wood and C. Liossi. The ecological validity of executive tests in

a severely brain injured sample. *Archives of Clinical Neuropsychology*, 21(5):429–437, 2006.

[210] J. Jallouli and G. Moreau. An immersive path-based study of wind turbines' landscape: A French case in Plouguin. *Renewable Energy*, 34(3):597–607, 2009.

[211] T. Oberman, B. B. O. Šćitaroci, and K. Jambrošić. Towards a virtual soundwalk. In *Handbook of Research on Perception-Driven Approaches to Urban Assessment and Design*, pages 317–343. IGI Global, 2018.

[212] J. Diemer, G. W. Alpers, H. M. Peperkorn, Y. Shiban, and A. Mühlberger. The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in Psychology*, 6:26, 2015.

[213] J. L. Higuera-Trujillo, J. L. T. Maldonado, and C. L. Millán. Psychological and physiological human responses to simulated and real environments: A comparison between photographs, 360 panoramas, and virtual reality. *Applied Ergonomics*, 65:398–409, 2017.

[214] J. Carles, F. Bernáldez, and J. d. Lucio. Audio-visual interactions and soundscape preferences. *Landscape Research*, 17(2):52–56, 1992.

[215] L. Galbrun and F. M. Calarco. Audio-visual interaction and perceptual assessment of water features used over road traffic noise. *The Journal of the Acoustical Society of America*, 136(5):2609–2620, 2014.

[216] F. Liu and J. Kang. Relationship between street scale and subjective assessment of audio-visual environment comfort based on 3d virtual reality and dual-channel acoustic tests. *Building and Environment*, 129:35–45, 2018.

[217] A. Preis, J. Kociński, H. Hafke-Dys, and M. Wrzosek. Audio-visual interactions in environment assessment. *Science of the Total Environment*, 523:191–200, 2015.

[218] S. Viollon, C. Lavandier, and C. Drake. Influence of visual setting on sound ratings in an urban environment. *Applied Acoustics*, 63(5):493–511, 2002.

[219] J. Kang and L. Yu. Modelling subjective evaluation of soundscape: Towards soundscape standardization. *The Journal of the Acoustical Society of America*, 129(4):2570–2570, 2011.

[220] F. Aletta, M. Masullo, L. Maffei, and J. Kang. The effect of vision on the perception of the noise produced by a chiller in a common living environment. *Noise Control Engineering Journal*, 64(3):363–378, 2016.

[221] T. Yu, H. Behm, R. Bill, and J. Kang. Audio-visual perception of new wind parks. *Landscape and Urban Planning*, 165:1–10, 2017.

[222] C. W. Thompson. Next-generation virtual worlds: architecture, status, and directions. *IEEE Internet Computing*, 15(1):60–65, 2011.

[223] J. Du, Y. Shi, C. Mei, J. Quarles, and W. Yan. Communication by interaction: A multiplayer vr environment for building walkthroughs. In *Construction Research Congress 2016, San Juan, Puerto Rico*, pages 2281–2290, 2016.

[224] M. T. Nguyen, H. K. Nguyen, K. D. Vo Lam, X. G. Nguyen, and M. T. Tran. Applying virtual reality in city planning. In *International Conference on Virtual, Augmented and Mixed Reality, Toronto, Canada*, pages 724–735. Springer, 2016.

[225] L. Jiang, M. Masullo, L. Maffei, F. Meng, and M. Vorländer. A demonstrator tool of web-based virtual reality for participatory evaluation of urban sound environment. *Landscape and Urban Planning*, 170:276–282, 2018.

[226] T. Yu, H. Behm, R. Bill, and J. Kang. Validity of VR technology on the smartphone for the study of wind park soundscapes. *ISPRS International Journal of Geo-Information*, 7(4):152, 2018.

[227] I. Muhammad, M. Vorländer, and S. J. Schlittmeier. Audio-video virtual reality environments in building acoustics: An exemplary study reproducing

performance results and subjective ratings of a laboratory listening experiment. *The Journal of the Acoustical Society of America*, 146(3):EL310–EL316, 2019.

[228] M. Basner, W. Babisch, A. Davis, M. Brink, C. Clark, S. Janssen, and S. Stansfeld. Auditory and non-auditory effects of noise on health. *The Lancet*, 383(9925):1325–1332, 2014.

[229] M. Annerstedt, P. Jönsson, M. Wallergård, G. Johansson, B. Karlson, P. Grahn, Å. M. Hansen, and P. Währborg. Inducing physiological stress recovery with sounds of nature in a virtual reality forest-results from a pilot study. *Physiology & Behavior*, 118:240–250, 2013.

[230] M. Hedblom, B. Gunnarsson, M. Schaefer, I. Knez, P. Thorsson, and J. N. Lundström. Sounds of nature in the city: no evidence of bird song improving stress recovery. *International Journal of Environmental Research and Public Health*, 16(8):1390, 2019.

[231] Z. Abdalrahman and L. Galbrun. Audio-visual preferences, perception, and use of water features in open-plan offices. *The Journal of the Acoustical Society of America*, 147(3):1661–1672, 2020.

[232] H. Jahncke, K. Eriksson, and S. Naula. The effects of auditive and visual settings on perceived restoration likelihood. *Noise & Health*, 17(74):1, 2015.

[233] C. Guastavino and B. F. Katz. Perceptual evaluation of multi-dimensional spatial audio reproduction. *The Journal of the Acoustical Society of America*, 116(2):1105–1115, 2004.

[234] L. Jiang and J. Kang. Effect of traffic noise on perceived visual impact of motorway traffic. *Landscape and Urban Planning*, 150:50–59, 2016.

[235] C. K. Chau, T. M. Leung, J. M. Xu, and S. K. Tang. Modelling noise annoyance responses to combined sound sources and views of sea, road traffic,

and mountain greenery. *The Journal of the Acoustical Society of America*, 144(6):3503–3513, 2018.

[236] D. Thery, D. Poirier-Quinot, B. N. Postma, and B. F. Katz. Impact of the visual rendering system on subjective auralization assessment in VR. In *International Conference on Virtual Reality and Augmented Reality*, pages 105–118. Springer, 2017.

[237] T. Lokki, L. McLeod, and A. Kuusinen. Perception of loudness and envelopment for different orchestral dynamics. *The Journal of the Acoustical Society of America*, 148(4):2137–2145, 2020.

[238] D. Poirier-Quinot, B. N. Postma, and B. F. Katz. Augmented auralization: Complimenting auralizations with immersive virtual reality technologies. In *International Symposium on Music and Room Acoustics (ISMRA), La Plata, Argentina*, volume 14, pages 1–10, 2016.

[239] S. Serafin, M. Geronazzo, C. Erkut, N. C. Nilsson, and R. Nordahl. Sonic interactions in virtual reality: state of the art, current challenges, and future directions. *IEEE Computer Graphics and Applications*, 38(2):31–43, 2018.

[240] L. S. Simon, N. Zacharov, and B. F. Katz. Perceptual attributes for the comparison of head-related transfer functions. *The Journal of the Acoustical Society of America*, 140(5):3623–3632, 2016.

[241] T. Oberman, K. Jambrošic, F. Aletta, and J. Kang. Towards a soundscape surround index. In *Proceedings of the 23rd International Congress on Acoustics*, pages 829–835. Deutsche Gesellschaft für Akustik (DEGA), 2019.

[242] M. Narbutt, S. O'Leary, A. Allen, J. Skoglund, and A. Hines. Streaming VR for immersion: Quality aspects of compressed spatial audio. In *Proceedings of the 23rd International Conference on Virtual System & Multimedia (VSMM), Dublin, Ireland*, pages 1–6. IEEE, 2017.

[243] F. Pausch, L. Aspöck, M. Vorländer, and J. Fels. An extended binaural real-time auralization system with an interface to research hearing aids for experiments on subjects with hearing loss. *Trends in Hearing*, 22:2331216518800871, 2018.

[244] S. Favrot and J. M. Buchholz. Lora: A loudspeaker-based room auralization system. *Acta Acustica united with Acustica*, 96(2):364–375, 2010.

[245] L. Perotin, R. Serizel, E. Vincent, and A. Guérin. Multichannel speech separation with recurrent neural networks from high-order ambisonics recordings. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada*, pages 36–40. IEEE, 2018.

[246] I. Pollack. Monaural and binaural threshold sensitivity for tones and for white noise. *The Journal of the Acoustical Society of America*, 20(1):52–57, 1948.

[247] I. J. Hirsh. Masking of speech and auditory localization. *Audiology*, 10(2):110–114, 1971.

[248] A. K. Nábělek and J. M. Pickett. Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners. *Journal of Speech, Language, and Hearing Research*, 17(4):724–739, 1974.

[249] A. K. Nábělek and J. Pickett. Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing aids. *The Journal of the Acoustical Society of America*, 56(2):628–639, 1974.

[250] J. Allen, D. Berkley, and J. Blauert. Multimicrophone signal-processing technique to remove room reverberation from speech signals. *The Journal of the Acoustical Society of America*, 62(4):912–915, 1977.

[251] A. K. Nábělek and P. K. Robinson. Monaural and binaural speech perception

in reverberation for listeners of various ages. *The Journal of the Acoustical Society of America*, 71(5):1242–1248, 1982.

[252] L. E. Marks. Binaural versus monaural loudness: Supersummation of tone partially masked by noise. *The Journal of the Acoustical Society of America*, 81(1):122–128, 1987.

[253] K. Genuit and A. Fiebig. Psychoacoustics and its benefit for the soundscape approach. *Acta Acustica united with Acustica*, 92(6):952–958, 2006.

[254] B. Berglund and M. E. Nilsson. Soundscape perceived in built environments. In *Proceedings of the 18th International Congress on Acoustics, Kyoto, Japan*, pages 259–260. International Commission for Acoustics, 2004.

[255] B. Berglund and M. E. Nilsson. On a tool for measuring soundscape quality in urban residential areas. *Acta Acustica united with Acustica*, 92(6):938–944, 2006.

[256] B. De Coensel, S. Vanwetswinkel, and D. Botteldooren. Effects of natural sounds on the perception of road traffic noise. *The Journal of the Acoustical Society of America*, 129(4):EL148–EL153, 2011.

[257] H. I. Jo and J. Y. Jeon. Downstairs resident classification characteristics for upstairs walking vibration noise in an apartment building under virtual reality environment. *Building and Environment*, 150:21–32, 2019.

[258] M. Rychtáriková and G. Vermeir. Soundscape categorization on the basis of objective acoustical parameters. *Applied Acoustics*, 74(2):240–247, 2013.

[259] G. Brambilla, V. Gallo, F. Asdrubali, and F. D'Alessandro. The perceived quality of soundscape in three urban parks in Rome. *The Journal of the Acoustical Society of America*, 134(1):832–839, 2013.

[260] K. Jambrošić, M. Horvat, and H. Domitrović. Assessment of urban soundscapes with the focus on an architectural installation with musical features. *The Journal of the Acoustical Society of America*, 134(1):869–879, 2013.

[261] J. Kang. Acoustic comfort in non-acoustic spaces: a review of recent work in Sheffield. In *Proceedings of the Institute of Acoustics*, volume 25, pages 125–132, 2003.

[262] ISO 9613-2:1996. *Acoustics — Attenuation of sound during propagation outdoors — Part 2: General method of calculation.* International Organization for Standardization, Geneva, Switzerland, 1996.

[263] E. A. Lehmann, A. M. Johansson, and S. Nordholm. Reverberation-time prediction method for room impulse responses simulated with the image-source model. In *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 159–162. IEEE, 2007.

[264] Y. Smyrnova and J. Kang. Determination of perceptual auditory attributes for the auralization of urban soundscapes. *Noise Control Engineering Journal*, 58(5):508–523, 2010.

[265] F. Aletta, Ö. Axelsson, and J. Kang. Towards acoustic indicators for soundscape design. In *Proceedings of the Forum Acusticum 2014 Conference, Krakow, Poland*, pages 1–6. European Acoustics Association, 2014.

[266] M. Zhang and J. Kang. Towards the evaluation, description, and creation of soundscapes in urban open spaces. *Environment and Planning B: Planning and Design*, 34(1):68–86, 2007.

[267] J. Y. Jeon, P. J. Lee, J. Y. Hong, and D. Cabrera. Non-auditory factors affecting urban soundscape evaluation. *The Journal of the Acoustical Society of America*, 130(6):3761–3770, 2011.

[268] C. Lavandier and B. Defréville. The contribution of sound source characteristics in the assessment of urban soundscapes. *Acta Acustica united with Acustica*, 92(6):912–921, 2006.

[269] J. Y. Hong and J. Y. Jeon. The effects of audio–visual factors on perceptions

of environmental noise barrier performance. *Landscape and Urban Planning*, 125:28–37, 2014.

[270] W. G. Gardner and K. D. Martin. HRTF measurements of a KEMAR. *The Journal of the Acoustical Society of America*, 97(6):3907–3908, 1995.

[271] J. S. Bradley and G. A. Soulodre. The influence of late arriving energy on spatial impression. *The Journal of the Acoustical Society of America*, 97(4):2263–2271, 1995.

[272] P. J. Lee, J. Y. Hong, and J. Y. Jeon. Assessment of rural soundscapes with high-speed train noise. *Science of The Total Environment*, 482:432–439, 2014.

[273] R. S. Shankland. Architectural acoustics in America to 1930. *The Journal of the Acoustical Society of America*, 61(2):250–254, 1977.

[274] D. Griesinger. Room impression, reverberance, and warmth in rooms and halls. In *Audio Engineering Society Convention 93, San Francisco, CA, USA*, page 3383. Audio Engineering Society, 1992.

[275] A. Warzybok, J. Rennies, T. Brand, S. Doclo, and B. Kollmeier. Effects of spatial and temporal integration of a single early reflection on speech intelligibility. *The Journal of the Acoustical Society of America*, 133(1):269–282, 2013.

[276] J. S. Bradley, H. Sato, and M. Picard. On the importance of early reflections for speech in rooms. *The Journal of the Acoustical Society of America*, 113(6):3233–3244, 2003.

[277] F. M. Wiener, C. I. Malme, and C. M. Gogos. Sound propagation in urban areas. *The Journal of the Acoustical Society of America*, 37(4):738–747, 1965.

[278] J. Kang. *Acoustics of Long Spaces: theory and design guidance.* Thomas Telford, 2002.

[279] P. Steenackers, H. Myncke, and A. Cops. Reverberation in town streets. *Acta Acustica united with Acustica*, 40(2):115–119, 1978.

[280] Y. Meng, J. Kang, R. Huerta, and M. Zhang. Soundscape animation of urban open spaces in an interactive virtual environment. In *Proceedings of the 12th International Congress of Sound and Vibration, Lisbon, Portugal*, pages 129–135, 2005.

[281] H. Onaga and J. H. Rindel. Acoustic characteristics of urban streets in relation to scattering caused by building facades. *Applied Acoustics*, 68(3):310–325, 2007.

[282] P. Richmond, Y. Smyrnova, S. C. Maddock, and J. Kang. Audio-visual animation of urban space. In *Theory and Practice of Computer Graphics*, pages 183–190, 2010.

[283] J. Y. Hong, B. Lam, Z. T. Ong, K. Ooi, W. S. Gan, J. Kang, S. Yeong, I. Lee, and S. T. Tan. The effects of spatial separations between water sound and traffic noise sources on soundscape assessment. *Building and Environment*, 167:106423, 2020.

[284] C. Calleri, L. Shtrepi, A. Armando, and A. Astolfi. Evaluation of the influence of building façade design on the acoustic characteristics and auditory perception of urban spaces. *Building Acoustics*, 25(1):77–95, 2018.

[285] H. Kuttruff. *Room acoustics*. CRC Press, 2016.

[286] C.-H. Jeong. Kurtosis of room impulse responses as a diffuseness measure for reverberation chambers. *The Journal of the Acoustical Society of America*, 139(5):2833–2841, 2016.

[287] M. Hodgson, N. York, W. Yang, and M. Bliss. Comparison of predicted, measured and auralized sound fields with respect to speech intelligibility in classrooms using catt-acoustic and odeon. *Acta Acustica united with Acustica*, 94(6):883–890, 2008.

[288] M. Gomes, S. Gerges, and R. Tenenbaum. On the accuracy of the assessment of room acoustics parameters using mls technique and numerical simulation. *Acta Acustica united with Acustica*, 86(5):891–895, 2000.

[289] J. B. Fritz, M. Elhilali, S. V. David, and S. A. Shamma. Auditory attention-focusing the searchlight on sound. *Current Opinion in Neurobiology*, 17(4):437–455, 2007.

[290] C. Xu, H. Tong, and J. Kang. Perceived width evaluation on interpolated line sources in a virtual urban square. In *Proceedings of the International Conference on Immersive and 3D Audio, Bologna, Italy*, pages 1–4. IEEE, 2021.

[291] R. Mehra, L. Antani, S. Kim, and D. Manocha. Source and listener directivity for interactive wave-based sound propagation. *IEEE Transactions on Visualization and Computer Graphics*, 20(4):495–503, 2014.

[292] U. Kurze and G. Anderson. Sound attenuation by barriers. *Applied Acoustics*, 4(1):35–53, 1971.

[293] S. Martin and D. Hothersall. Numerical modelling of median road traffic noise barriers. *Journal of Sound and Vibration*, 251(4):671–681, 2002.

[294] C. Steele. A critical review of some traffic noise prediction models. *Applied Acoustics*, 62(3):271–287, 2001.

[295] J. Y. Jeon, P. J. Lee, J. You, and J. Kang. Acoustical characteristics of water sounds for soundscape enhancement in urban open spaces. *The Journal of the Acoustical Society of America*, 131(3):2101–2109, 2012.

[296] H. Tong and J. Kang. Relationship between urban development patterns and noise complaints in England. *Environment and Planning B: Urban Analytics and City Science*, page 2399808320930247, 2020.

[297] G. C. Stecker, J. A. Beim, H. Bharadwaj, A. K. Bosen, E. Buss, M. Braza, A. C. Diedesch, C. M. Dorey, A. R. Dykstra, R. Freyman, F. J. Gallun, R. L.

Goldsworthy, L. Gray, E. C. Hoover, A. Ihlefeld, T. Koelewijn, J. G. Kopun, J. Mesik, E. Peng, V. M. Richards, Y. Shen, D. E. Shub, J. H. Venezia, and S. Waz. Remote testing for psychological and physiological acoustics: Initial report of the asa p&p task force on remote testing. *The Journal of the Acoustical Society of America*, 148(4):2713–2713, 2020.

Appendix A

Copyright Licences

In this appendix, we provide the copyright licences of the two figures, Fig. 2.1 and Fig. 2.2, in Chapter 2.

Copyright Clearance Center Marketplace™

British Standards Institution - License Terms and Conditions

This is a License Agreement between Chunyang Xu ("You") and British Standards Institution ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by British Standards Institution, and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	06-Apr-2021	Type of Use	Republish in a thesis/dissertation
Order license ID	1109902-1	Publisher	BSI British Standards
Standard ID	BS ISO 12913-1:2014	Portion	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	Acoustics. Soundscape : Definition and conceptual framework	Rightsholder	British Standards Institution
Date	09/30/2014	Publication Type	Standard
Language	English		

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	Distribution	Worldwide
Number of charts / graphs / tables / figures requested	1	Translation	Original language of publication
Format (select all that apply)	Print, Electronic	Copies for the disabled?	Yes
Who will republish the content?	Academic institution	Minor editing privileges?	No
Duration of Use	Current edition and up to 5 years	Incidental promotional use?	No
Lifetime Unit Quantity	Up to 19,999	Currency	GBP
Rights Requested	Main product		

NEW WORK DETAILS

Title	Mr	Institution name	University College London
Instructor name	Chunyang Xu	Expected presentation date	2021-09-01

ADDITIONAL DETAILS

Order reference number	N/A	The requesting person / organization to appear on the license	Chunyang Xu

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure 1 — Elements in the perceptual construct of soundscape	Title of the article/chapter the portion is from	Acoustics — Soundscape — Part 1: Definition and conceptual framework
Editor of portion(s)	N/A	Author of portion(s)	International Organization for Standardization
Volume of serial or monograph	N/A	Issue, if republishing an article from a serial	N/A
Page or page range of portion	2	Publication date of portion	2014-09-30

Figure A.1: Copyright licence for Fig. 2.1.

Copyright Clearance Center Marketplace™

ACM (Association for Computing Machinery) - License Terms and Conditions

Order Date	26-Mar-2021	**Type of Use**	Republish in a thesis/dissertation
Order license ID	1107457-1	**Publisher**	ACM
ISBN-13	978-0-89791-228-0	**Portion**	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	Proceedings of the 1986 workshop on Interactive 3D graphic	**Country**	United States of America
		Rightsholder	ACM (Association for Computing Machinery)
Date	01/01/1987	**Publication Type**	Conference Proceeding
Language	English		

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	**Distribution**	Worldwide
Number of charts / graphs / tables / figures requested	1	**Translation**	Original language of publication
Format (select all that apply)	Print, Electronic	**Copies for the disabled?**	Yes
		Minor editing privileges?	No
Who will republish the content?	Academic institution	**Incidental promotional use?**	No
Duration of Use	Current edition and up to 5 years	**Currency**	GBP
Lifetime Unit Quantity	Up to 99,999		
Rights Requested	Main product		

NEW WORK DETAILS

Title	Mr	**Institution name**	University College London
Instructor name	Chunyang Xu	**Expected presentation date**	2021-09-01

ADDITIONAL DETAILS

Order reference number	N/A	**The requesting person / organization to appear on the license**	Chunyang Xu

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Fig.5 Light-weight, head-mounted display unit with wide-angle optics and LCD displays.	**Title of the article/chapter the portion is from**	Virtual environment display system
		Author of portion(s)	Fisher, S. S. and McGreevy, M. and Humphries, J. and Robinett, W.
Editor of portion(s)	N/A		
Volume of serial or monograph	N/A	**Issue, if republishing an article from a serial**	N/A
Page or page range of portion	80	**Publication date of portion**	1987-01-01

Figure A.2: Copyright licence for Fig. 2.2.

Appendix B

Questions for Subjective Evaluation

We introduce the questions used in the subjective tests for Chapters 5, 6 and 7.

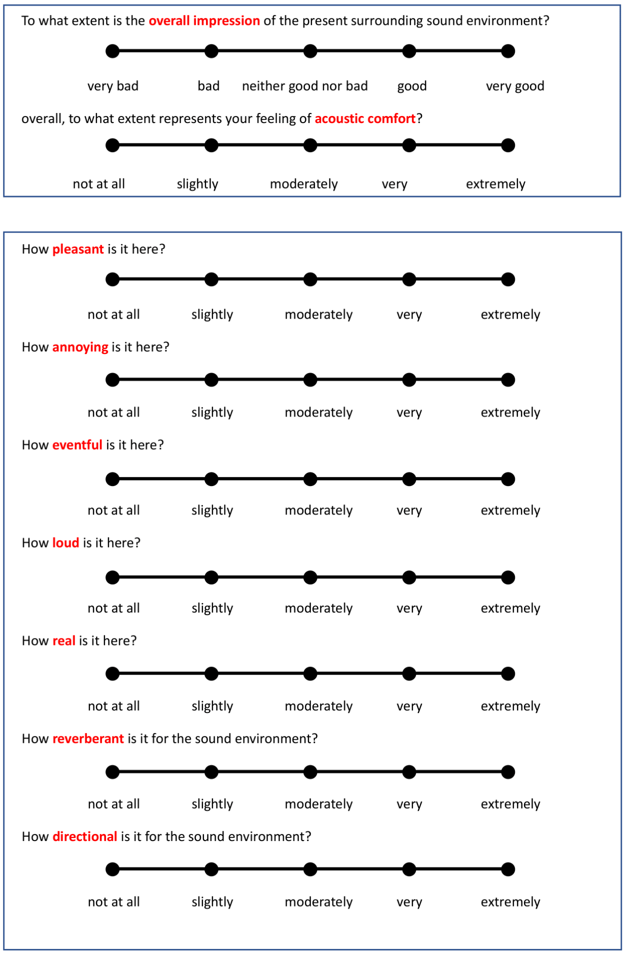


Figure B.1: Questions used in the subjective test for Chapter 5.

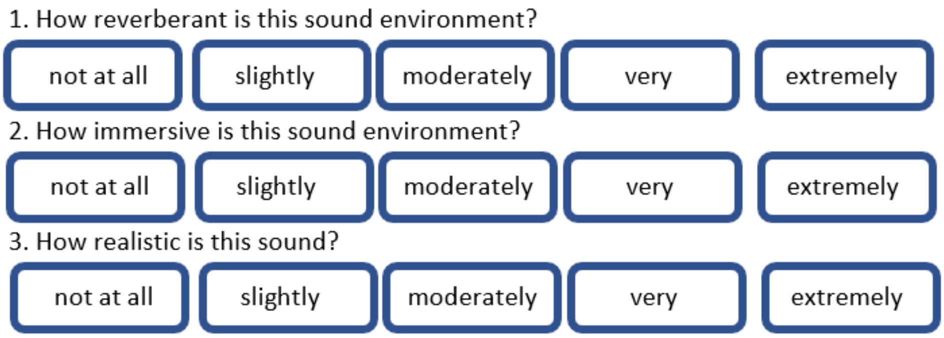


Figure B.2: Questions used in the subjective test for Chapter 6.

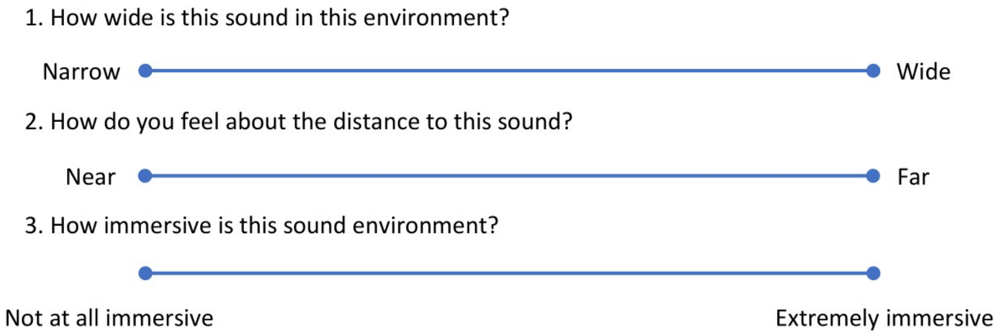


Figure B.3: Questions used in the subjective test for Chapter 7.

Appendix C

Participant Information Sheet 1

We introduce the participant information sheet used in the subjective test in Chapter 5.

Participant Information Sheet

You are being invited to take part in a research project. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

1. Research Project Title:

Soundscape evaluation by auditory experiments

2. What is the project's purpose?

The project is to study the subjective sound comfort in urban public spaces. People in different public spaces (e.g. parks, railway station, commercial streets and fountain squares) will perceive sounds differently. Twelve public sites in Sheffield were selected including the Peace Garden, Western Park, City Hall, etc. For each site, you will hear a sound lasting 30 seconds, and according to the sounds give your subjective sensation to (1) overall acoustic impression and comfort and (2) various perceived attributes (e.g. pleasant-unpleasant, dry-everywhere and quiet-noisy).

3. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep (and be asked to sign a consent form). You do not have to give a reason.

4. What will happen to me if I take part?

Twelve public sites were chosen in Sheffield and the sounds were recorded. Before the formal test, you will hear a short sound cut through a headphone including several typical site sounds to experience the overall variation in sound environments. During the test, you will hear a series of sounds in order, and according to each sound give your subjective sensation based on the test sheet. Each sound lasts 30 seconds, and the total test does not exceed one hour.

You should answer each question in the test sheet following your actual sensation and give you feedback thoughtfully.

5. What are the possible disadvantages and risks of taking part?

During the auditory test, if there are any unexpected discomforts with your headphones or playing sounds, please inform us immediately.

6. What are the possible benefits of taking part?

Your participating in the project will benefit the future urban soundscape planning leading to a more pleasant acoustic environment.

7. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any reports or publications.

8. What will happen to the results of the research project?

The results are likely to be published, and you will not be identified in any report or publications. Due to the nature of this research it is very likely that other researchers may find the data collected to be useful in answering future research questions. We will ask for your explicit consent for your data to be shared in this way and if you agree, we will ensure that the data collected about you is untraceable back to you before allowing others to use it. If the results are published in a journal, we will be glad to offer a copy of the published results to you, and please contact the leader researcher below.

9. Who has ethically reviewed the project?

This project has been ethically approved via the ethics review procedure of the University's Research Ethics Committee monitors the application and delivery of the University's Ethics Review Procedure across the university.

10. Contact for further information

Chunyang Xu (Leader Researcher)

Prof Jian Kang

Thank you for considering to take part in this project!

Appendix D

Participant Information Sheet 2

We introduce the participant information sheet used in the subjective test in Chapter 6.

Participant Information Sheet

Title of Study: Soundscape evaluation based on virtual reality technology

Department: Institute for Environmental Design and Engineering, the Bartlett

Name and Contact Details of the Researcher(s): Chunyang Xu

Name and Contact Details of the Principal Researcher: Prof Jian Kang

1. You are being invited to take part in a research project, and it is a PhD thesis research study. Before you decided it is important for you to understand why the research us being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

2. What is the project's purpose?

The project aims to investigate subjective perception on different urban sound environment. Combined with virtual reality technologies, the study focuses on audio-visual interaction. Through the human subjective perception, identify different acoustic conditions in urban sound environments.

3. Why have I been chosen?

You should be aged over 18 with normal vision and hearing. Other participants less than 30 with normal vision and hearing will also participate this subjective evaluation.

4. Do I have to take part? It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. There will be no benefits after the engagement, and you can withdraw at any time without giving a reason. If you decide to withdraw you will be asked what you wish to happen to the data you have provided up that point.

5. What will happen to me if I take part? You will hear some conventional sounds in cities (e.g. fountain, birdsong and clapping), and you will also wear a pair of VR glasses during the experiment to combine what you see and what you hear. For each video, you will hear a sound lasting 10 seconds. After the VR experience or only hearing experience, you will answer some questions according to what you have heard, e.g., how immersive is this sound environment? You only need take part in once, and the subjective evaluation will be less than one hour. Because this study occur in the campus, no travel expense will be reimbursed.

6. What are the possible disadvantages and risks of taking part?

There is no foreseeable risk during the whole evaluation. If you feel any discomfort during the evaluation owing to the VR headset and headphones, you can immediately tell the data collector and withdraw at any time.

7. What are the possible benefits of taking part?

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will help the city establish a more comfortable sound environment in the futuRL

8. What if something goes wrong?

If you have any complaints during the evaluation, you can contact the principal researcher listed below (Prof Jian Kang) or Chair of the UCL Research Ethics Committee - ethics@ucl.ac.uk

9. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications.

10. Limits to confidentiality

Please note that assurances on confidentiality will be strictly adhered to unless evidence of wrongdoing or potential harm is uncovered. In such cases the University may be obliged to contact relevant statutory bodies/agencies.

Please note that confidentiality will be maintained as far as it is possible, unless during our conversation I hear anything which makes me worried that someone might be in danger of harm, I might have to inform relevant agencies of this.

Please note that confidentiality may not be guaranteed; due to the limited size of the participant sample.

Confidentiality will be respected subject to legal constraints and professional guidelines.

Confidentiality will be respected unless there are compelling and legitimate reasons for this to be breached. If this was the case we would inform you of any decisions that might limit your confidentiality.

Confidentiality may be limited and conditional and the researcher has a duty of care to report to the relevant authorities possible harm/danger to the participant or others.

11. What will happen to the results of the research project?

When the results are likely to be published, you can obtain a copy of the published results from the researchers listed below, and you will not be identified in any report or publication.

The results of the project will be also the part of the PhD thesis research, and they might be used for subsequent research relevant to soundscapes and urban sound environments.

12. Local Data Protection Privacy Notice

Notice: The controller for this project will be University College London (UCL). The UCL Data Protection Officer provides oversight of UCL activities involving the processing of personal data, and can be contacted at data-protection@ucl.ac.uk

This 'local' privacy notice sets out the information that applies to this particular study. Further information on how UCL uses participant information can be found in our 'general' privacy notice: individuals whose data may be processed indirectly as part of research conducted by UCL. The information that is required to be provided to participants under data protection legislation (GDPR and DPA 2018) is provided across both the 'local' and 'general' privacy notices. No sensitive categories of personal data will be collected including racial or ethnic origin, political opinions, etc. The lawful basis that would be used to process your personal data will be performance of a task in the public interest. The lawful basis used to process special category personal data will be for scientific and historical research or statistical purposes. Your personal data will be processed so long as it is required for the research project. If we are able to anonymise or pseudonymise the personal data you provide we will undertake this, and will endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, or if you would like to contact us about your rights, please contact UCL in the first instance at data-protection@ucl.ac.uk.

You will be given a copy of the information sheet and a signed consent form to keep.

Thank you for reading this and considering to take part in this study.

Appendix E

Supplementary Statistical Results

We introduce the supplementary statistical results in the subjective test for the Fig. 6.5 in Chapter 6.

Table E.1: Pairwise comparisons between different orders for different indicators, sounds in Site 1. (RV—Reverberance, IM—Immersion, RL—Realism)

Site	Sound	Indicator	Order	Mean Difference (I-J)	Std. Error	Sig.b	95% Confidence Interval for Differenceb	
							Lower Bound	Upper Bound
1	B	RV	20*5	.633*	0.182	0.002	0.260	1.007
			20*1	.600*	0.149	0.000	0.296	0.904
			5*1	-0.033	0.148	0.823	-0.335	0.269
		IM	20*5	0.467	0.234	0.055	-0.011	0.944
			20*1	.433*	0.177	0.021	0.071	0.796
			5*1	-0.033	0.227	0.884	-0.498	0.432
		RL	20*5	0.267	0.203	0.199	-0.149	0.682
			20*1	0.367	0.237	0.133	-0.119	0.852
			5*1	0.100	0.264	0.708	-0.440	0.640
	C	RV	20*5	0.100	0.200	0.620	-0.308	0.508
			20*1	0.067	0.197	0.738	-0.337	0.470
			5*1	-0.033	0.148	0.823	-0.335	0.269
		IM	20*5	-0.100	0.188	0.599	-0.484	0.284
			20*1	0.200	0.162	0.227	-0.131	0.531
			5*1	.300*	0.145	0.048	0.003	0.597
		RL	20*5	0.067	0.159	0.677	-0.258	0.391
			20*1	0.100	0.188	0.599	-0.284	0.484
			5*1	0.033	0.155	0.831	-0.284	0.351
	F	RV	20*5	0.033	0.155	0.831	-0.284	0.351
			20*1	0.267	0.135	0.058	-0.010	0.543
			5*1	0.233	0.133	0.090	-0.038	0.505
		IM	20*5	0.000	0.179	1.000	-0.367	0.367
			20*1	-0.067	0.151	0.662	-0.376	0.242
			5*1	-0.067	0.159	0.677	-0.391	0.258
		RL	20*5	-0.033	0.227	0.884	-0.498	0.432
			20*1	-0.033	0.206	0.873	-0.455	0.388
			5*1	0.000	0.214	1.000	-0.439	0.439

Table E.2: Pairwise comparisons between different orders for different indicators, sounds in Site 2, 3 and 4. (RV—Reverberance, IM—Immersion, RL—Realism)

Site	Sound	Indicator	Order	Mean Difference (I-J)	Std. Error	Sig.b	95% Confidence Interval for Differenceb	
							Lower Bound	Upper Bound
2	B	RV	20*5	0.100	0.168	0.557	-0.245	0.445
			20*1	0.300	0.174	0.095	-0.056	0.656
			5*1	0.200	0.256	0.440	-0.323	0.723
		IM	20*5	0.167	0.180	0.362	-0.201	0.535
			20*1	0.233	0.196	0.243	-0.167	0.634
			5*1	0.067	0.239	0.783	-0.423	0.556
		RL	20*5	0.267	0.166	0.118	-0.072	0.605
			20*1	0.300	0.193	0.130	-0.094	0.694
			5*1	0.033	0.182	0.856	-0.340	0.407
	C	RV	20*5	0.067	0.172	0.702	-0.286	0.419
			20*1	0.267	0.235	0.265	-0.213	0.746
			5*1	0.200	0.139	0.161	-0.084	0.484
		IM	20*5	0.100	0.121	0.415	-0.147	0.347
			20*1	0.133	0.184	0.475	-0.243	0.510
			5*1	0.033	0.176	0.851	-0.327	0.393
		RL	20*5	-0.167	0.128	0.202	-0.428	0.094
			20*1	-0.100	0.232	0.669	-0.574	0.374
			5*1	0.067	0.214	0.758	-0.371	0.504
	F	RV	20*5	0.033	0.200	0.869	-0.377	0.443
			20*1	0.400	0.306	0.201	-0.225	1.025
			5*1	0.367	0.182	0.054	-0.007	0.740
		IM	20*5	0.200	0.130	0.136	-0.067	0.467
			20*1	0.133	0.257	0.608	-0.392	0.659
			5*1	-0.067	0.262	0.801	-0.603	0.470
		RL	20*5	-0.100	0.182	0.586	-0.471	0.271
			20*1	0.100	0.251	0.693	-0.413	0.613
			5*1	0.200	0.305	0.517	-0.423	0.823
3	B	RV	20*5	-0.033	0.140	0.813	-0.319	0.252
			20*1	0.233	0.149	0.129	-0.072	0.538
			5*1	0.267	0.159	0.103	-0.058	0.591
		IM	20*5	0.000	0.192	1.000	-0.392	0.392
			20*1	.567*	0.184	0.004	0.191	0.942
			5*1	.567*	0.184	0.004	0.191	0.942
		RL	20*5	-0.033	0.195	0.865	-0.431	0.365
			20*1	0.267	0.279	0.348	-0.305	0.838
			5*1	0.300	0.221	0.184	-0.151	0.751
	C	RV	20*5	.533*	0.184	0.007	0.157	0.910
			20*1	.933*	0.235	0.000	0.454	1.413
			5*1	.400*	0.156	0.016	0.081	0.719
		IM	20*5	0.400	0.212	0.070	-0.034	0.834
			20*1	.633*	0.273	0.028	0.074	1.192
			5*1	0.233	0.149	0.129	-0.072	0.538
		RL	20*5	0.167	0.235	0.484	-0.315	0.648
			20*1	0.200	0.293	0.501	-0.400	0.800
			5*1	0.033	0.155	0.831	-0.284	0.351
	F	RV	20*5	-0.033	0.200	0.869	-0.443	0.377
			20*1	0.233	0.252	0.363	-0.283	0.749
			5*1	0.267	0.179	0.147	-0.099	0.633
		IM	20*5	-0.200	0.176	0.264	-0.559	0.159
			20*1	.767*	0.184	0.000	0.391	1.142
			5*1	.967*	0.176	0.000	0.607	1.327
		RL	20*5	-0.133	0.190	0.489	-0.522	0.256
			20*1	.733*	0.230	0.003	0.264	1.203
			5*1	.867*	0.196	0.000	0.466	1.268
4	B	RV	20*5	.333*	0.161	0.048	0.003	0.663
			20*1	0.333	0.227	0.152	-0.130	0.797
			5*1	0.000	0.166	1.000	-0.340	0.340
		IM	20*5	0.100	0.188	0.599	-0.284	0.484
			20*1	0.100	0.246	0.687	-0.403	0.603
			5*1	0.000	0.203	1.000	-0.416	0.416
		RL	20*5	0.367	0.232	0.125	-0.109	0.842
			20*1	0.333	0.241	0.178	-0.160	0.827
			5*1	-0.033	0.242	0.891	-0.528	0.462
	C	RV	20*5	.733*	0.244	0.005	0.234	1.233
			20*1	1.367*	0.273	0.000	0.808	1.926
			5*1	.633*	0.169	0.001	0.287	0.980
		IM	20*5	-0.033	0.155	0.831	-0.351	0.284
			20*1	0.367	0.212	0.094	-0.066	0.800
			5*1	.400*	0.156	0.016	0.081	0.719
		RL	20*5	-0.200	0.182	0.281	-0.572	0.172
			20*1	0.067	0.253	0.794	-0.452	0.585
			5*1	0.267	0.166	0.118	-0.072	0.605
	F	RV	20*5	0.333	0.200	0.106	-0.075	0.742
			20*1	0.500	0.287	0.092	-0.086	1.086
			5*1	0.167	0.167	0.326	-0.174	0.508
		IM	20*5	0.267	0.185	0.161	-0.112	0.646
			20*1	.767*	0.184	0.000	0.391	1.142
			5*1	.500*	0.239	0.045	0.012	0.988
		RL	20*5	-0.033	0.269	0.902	-0.584	0.517
			20*1	.600*	0.274	0.037	0.040	1.160
			5*1	.633*	0.200	0.004	0.223	1.043

Appendix F

Participant Information Sheet 3

We introduce the participant information sheet used in the subjective test in Chapter 7.

Participant Information Sheet

Title of Study: Sound environment evaluation under VR

Department: Institute for Environmental Design and Engineering

Name and Contact Details of the Researcher: Chunyang Xu

Name and Contact Details of the Principal Researcher: Prof Jian Kang

You are being invited to take part in a research project, and it is a PhD thesis research study. Before you decided it is important for you to understand why the research us being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

2. What is the project's purpose?

The project aims to study subjective perception on different urban sound environment. Combined with audio-video technologies, the study focuses on audio-visual interaction. Through the human subjective perception, identify different sound conditions in urban sound environments.

3. Why have I been chosen?

You should be aged over 18 with normal vision and hearing.

4. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. There will be no benefits after the engagement, and you can withdraw at any time without giving a reason before submitting your data. If you decide to withdraw before the submission, the platform will not record your data. Meanwhile, if you submitted your data successfully online, due to the anonymity of the platform, we cannot recognise your data within our database.

5. What will happen to me if I take part?

You will first look through the participant information sheet. You will hear some common sounds in cities (e.g. fountain, road noise and talking), and you will use your smart phone and earphone during the experiment to combine what you see and what you hear. For each video, you will hear a sound lasting 10 seconds. After seeing videos, you will answer some questions according to what you have heard, e.g., how immersive is this sound environment? You only need take part in once, and the subjective evaluation will be less than one hour. No future research will be informed.

6. What are the possible disadvantages and risks of taking part?

There is no foreseeable risk during the whole evaluation. If you feel any discomfort during the evaluation owing to seeing your phone, you can immediately withdraw at any time.

7. What are the possible benefits of taking part?

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will help the city establish a more comfortable sound environment in the future.

8. What if something goes wrong?

If you have any complaints during the evaluation, you can contact the principal researcher listed below (Prof Jian Kang) or Chair of the UCL Research Ethics Committee (ethics@ucl.ac.uk)

9. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications.

10. Limits to confidentiality

Please note that assurances on confidentiality will be strictly adhered to unless evidence of wrongdoing or potential harm is uncovered. In such cases the University may be obliged to contact relevant statutory bodies/agencies.

Please note that confidentiality will be maintained as far as it is possible, unless during our conversation I hear anything which makes me worried that someone might be in danger of harm, I might have to inform relevant agencies of this.

Please note that confidentiality may not be guaranteed; due to the limited size of the participant sample.

Confidentiality will be respected subject to legal constraints and professional guidelines.

Confidentiality will be respected unless there are compelling and legitimate reasons for this to be breached. If this was the case we would inform you of any decisions that might limit your confidentiality.

Confidentiality may be limited and conditional and the researcher has a duty of care to report to the relevant authorities possible harm/danger to the participant or others.

11. What will happen to the results of the research project?

When the results are likely to be published, you can obtain a copy of the published results from the researchers listed below, and you will not be identified in any report or publication.

The results of the project will be also the part of the PhD thesis research, and they might be used for subsequent research relevant to soundscapes and urban sound environments.

12. Local Data Protection Privacy Notice

The controller for this project will be University College London (UCL). The UCL Data Protection Officer provides oversight of UCL activities involving the processing of personal data, and can be contacted at data-protection@ucl.ac.uk This 'local' privacy notice sets out the information that applies to this particular study. Further information on how UCL uses participant information can be found in our 'general' privacy notice:

The information that is required to be provided to participants under data protection legislation (GDPR and DPA 2018) is provided across both the 'local' and 'general' privacy notices.

No sensitive categories of personal data will be collected including racial or ethnic origin, political opinions, etc. The lawful basis that would be used to process your personal data will be performance of a task in the public interest. Your personal data will be processed so long as it is required for the research project. If we are able to anonymise or pseudonymise the personal data you provide we will undertake this, and will endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, or if you would like to contact us about your rights, please contact UCL in the first instance at data-protection@ucl.ac.uk.

Thank you for reading this information sheet and for considering to take part in this study.