Chapter 1.

Introduction

1.1. Background

Urban noise pollution affects 80 million EU citizens (approx. 20% of the population) with substantial impacts on public health which are not well addressed by conventional noise control methods (EEA, 2020). Concerns about noise pollution have recently received increased attention both as a global environmental issue (Aletta, 2022) and as a necessary component of the UN Sustainable Development Goals (King, 2022). Noise pollution has been recognised as the second most impactful environmental health concern in cities, behind air pollution. The WHO found that, among other vectors, transport noise accounted for a loss of 903,000 disability-adjusted life years (DALYs) due to sleep disturbance and 587,000 DALYs due to annoyance in the EU (WHO Centers of Disease Control, 2011).

Traditional noise control methods have typically limited their focus to the reduction of unwanted noise, ignoring the potential benefits of increasing positive sounds and remaining restricted by practical limitations of noise reduction. Modern approaches to achieve improved health outcomes and public satisfaction aim to incorporate a person's perception of an acoustic environment, an approach known as 'Soundscape'[1].

The soundscape concept represents a positive approach to understanding society's relationship with urban sound. In particular, it stands in contrast to the negative, reactive approach taken in existing noise control regulations. In a recent editorial, Ö. Axelsson (2020) stated:

> In practice, noise abatement is a reactive approach to sound. First, a member of the public must submit a complaint to the competent authority, which must verify that the complaint is valid and may then take actions. It is a common view among noise and health inspectors that they have no mandate to act, unless there is a complaint, the validity of which is verified. This makes noise abatement comparable to waste management. Sound is deemed a harmful waste product of human activity that must be removed.

By contrast, soundscape studies view sound as a resource which both needs to be appropriately managed, but can also contribute positively. Towards this, soundscape studies strive to understand the perception of a sound environment, in context, including acoustic, (non-acoustic) environmental, contextual, and personal factors. These factors combine together to form a person's soundscape in complex interacting ways (Berglund & Nilsson, 2006). In order to predict how people would perceive an acoustic environment, it is essential to identify the underlying acoustic and non-acoustic properties of soundscape.

As the future moves toward a more holistic soundscape focus, the ability to affect change at large scales and in a wide range of projects will require that familiar engineering tools and approaches can apply to soundscape design. When attempting to apply soundscape in practical applications in the built environment, it becomes apparent that a predictive model of the users' perceptual response to the acoustic environment is necessary. Whether to determine the impact of a design change, or to integrate a large scale data at neighbourhood and city levels, a mathematical model of the interacting factors will form a vital component of the implementation of the soundscape approach.

1.2. Research Aims & Questions

This work is intended to identify methods for incorporating contextual and objective information into a useable and interpretable predictive model of urban soundscapes and to develop tools for documenting, analysing, and visualising soundscape assessments. In order to achieve this, a protocol for collecting the multi-level, multi-factor perceptual assessment data has been developed and implemented, resulting in a large soundscape database. Several avenues of investigation are then drawn from the database and addressed throughout this thesis. The primary research questions are:

1. What are the primary acoustic features involved in soundscape perception and what are the driving interactions between acoustic features and soundscape assessment?
2. To what extent can a predictive model be used to investigate changes in likely soundscape perception in situations where the actual soundscape cannot be assessed?
3. What are the design requirements of a predictive model of soundscape assessments and how can future work move towards achieving these?
4. How does the sound source composition in a complex sound environment mediate this interaction and how can this effect be simplified and modelled?
5. What are the non-acoustic, personal factors which influence an individual's perception of the sound environment and to what degree do these factors explain the variance in soundscape assessments?

6. How can the inherent variation in soundscape assessments best be represented and in what ways and to what extent can this analysis of soundscapes be applied to address future urban design challenges?

Throughout this thesis, a Multi-Level Model (MLM) approach has been developed and progressively improved. Although the key chapters may make use of separate datasets or be focussed on different aspects of the multi-dimensional perception of urban soundscapes, underlying each of them is an analysis based on MLM and a goal towards integrating each of their findings into a final, cohesive model.

1.3. Thesis Structure

This thesis presents the results of several studies which develop the conceptual and statistical frameworks to enable the prediction and presentation of the soundscape analysis of urban spaces. Portions of Chapters 3, 5 and 7 to 9 have been published in peer-reviewed academic journals. Chapters 7 and 8, although written in heavy collaboration with coauthors are based primarily on the MLM analysis developed in this thesis.

Chapter 2 begins by reviewing the engineering approach taken in traditional noise control and highlights the improvements offered by the soundscape approach. The current state-of-the-art of soundscape for design is reviewed and the arguments for why predictive modelling is necessary are presented. Finally, the pre-existing framework for predictive soundscape models developed by Aletta, Kang, and Axelsson (2016) and previous examples of predictive models are reviewed.

After the introduction and literature review, the body of the text containing the original work is separated into three parts: **Part I** contains the methodology; **Part II** presents an initial effort to develop and apply a predictive soundscape model; and **Part III** further develops this initial model with two related modelling studies with discussions of how each of these developments can be integrated into a single, general model. Fig. 1.1 presents an overview of the chapter structure of this thesis and how each published study and research question are integrated into the structure.

The methods section is split into two chapters: the first (**Chapter 3**) provides an in-depth explanation of the unique data collection protocol developed for the Soundscape Indices (SSID) Project and this thesis. The collection and organisation methods included in this protocol were directed toward the creation of a large scale database suitable for training a soundscape prediction model which includes both acoustical and contextual information. The second part of the methods section (**Chapter 4**) focusses on how this database is analysed, which includes the questionnaire analysis methods and the multi-level modelling strategy employed throughout the rest of the thesis.

Chapter 5 details the development of a predictive soundscape model based on a multivariate psychoacoustic analysis of the collected database. To demonstrate the usefulness of predictive models,

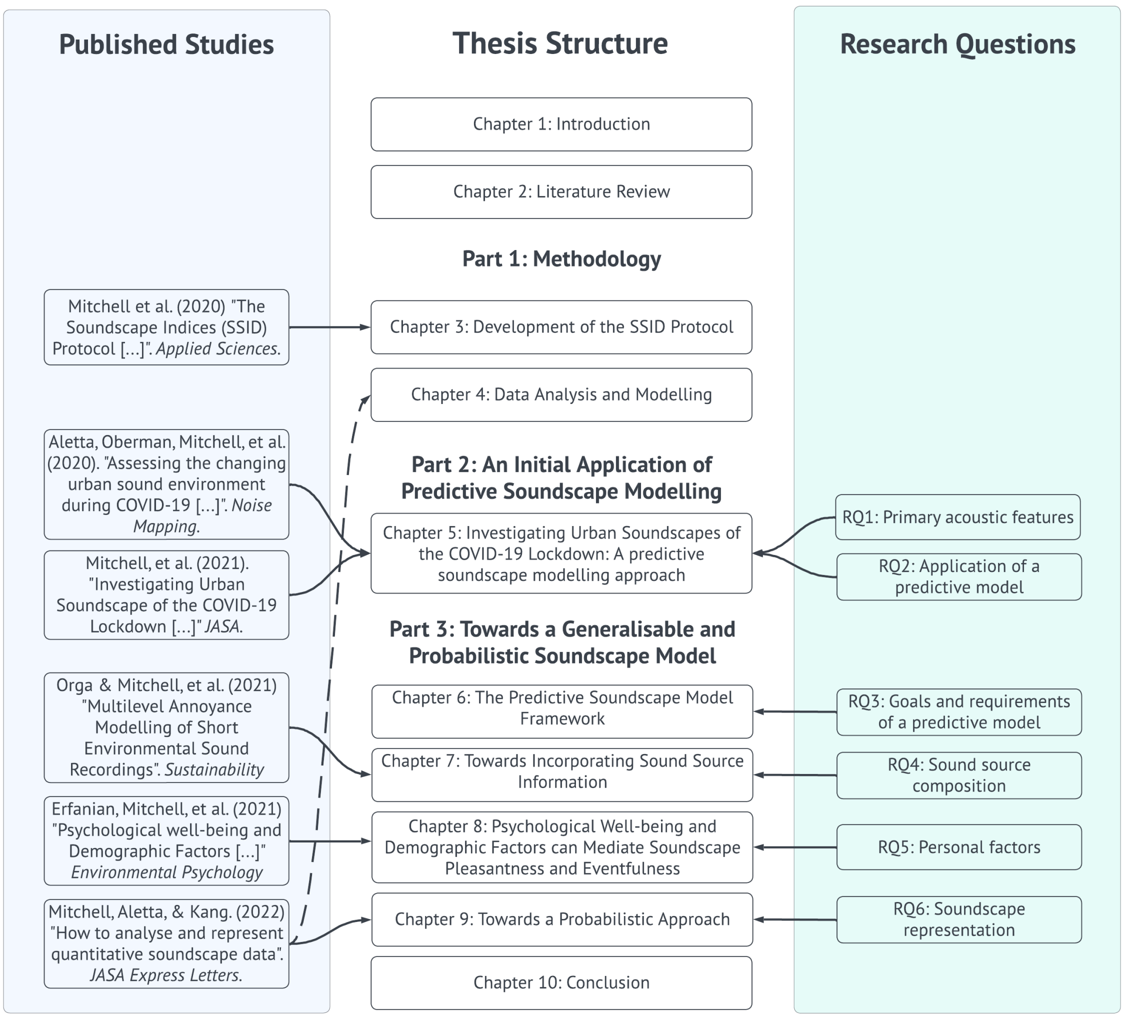


Figure 1.1.: The overall thesis structure showing how each Research Question (RQ) and published study relates to each chapter.

this initial model was applied to answer how the soundscape perception of urban public spaces was likely affected by the COVID-19 lockdowns in 2020. This chapter begins with an analysis of the sound environment impacts of the lockdowns, as revealed by the psychoacoustic analysis of binaural recordings. An initial MLM to predict soundscape pleasantness and eventfulness was trained on data collected before the lockdowns (in 2019) according to the protocol set out in Chapter 3, then applied to recordings taken in the same spaces during the height of the 2020 lockdowns (where traditional soundscape assessment methods are impractical) to investigate how these changes to the sound environment would likely have been perceived.

Following this initial application of soundscape prediction, Part III presents several case studies which attempt to expand on the initial model. **Chapter 6** begins by laying out a framework of the goals and constraints on a general soundscape model, which was derived from the experience gained in Chapter 5. These goals and constraints establish a series of improvements which can be made to the initial model to move it towards being a useful tool for soundscape design in engineering contexts.

Chapter 7 is the first case study and attempts to integrate sound source information into a psychoacoustic model of annoyance. By collaborating with a separate dataset (from the DYNAMAP project) which contains sound-source-labelled recordings from a Wireless Acoustic Sensor Network (WASN), we investigated the potential for incorporating sound source information in predictive soundscape models and further discuss how this should contribute to a general model. **Chapter 8** is the second case study and investigates to what degree personal factors (demographics and psychological well-being) influence soundscape perception. This study provides the basis for a consideration of how or whether these factors are necessary for predictive modelling.

Finally, **Chapter 9** reflects on the analysis methods for soundscape assessment data provided by ISO/TS 12913-2:2018 (2018) and presented in Chapter 4. Building on the lessons learned in the preceding chapters, I discuss the utility of these methods for describing the soundscapes of public spaces and propose a new analysis and visualisation method that better reflects the variety of experiences that people can have to the same soundscape. The implications of this new conception of a 'collective perception' are discussed and proposals for how predictive models should consider this collective perception for the design of public soundscapes are developed.

In all, the results of six peer-reviewed studies are presented. These studies represent a series of work to

1. Advance the conceptual development and practice of soundscape studies as applied to public spaces
2. Develop a transparent and useful method of predicting soundscape assessments
3. Investigate the various components which influence soundscape perception, including personal factors like psychological well-being, acoustical factors, and sound source specifics and

to integrate these components into the predictive modelling methods.

4. Propose a future framework for predictive soundscape models and provide proposals for how a general model should be developed that can be put to use in engineering approaches to designing the soundscapes of public urban spaces.

[1]**A note on terminology: Soundscape Perception?** According to the definition of soundscape provided in ISO 12913-1:2014 (2014), the soundscape is 'the acoustic environment as perceived or experienced and/or understood by people'. Both in the standard and elsewhere, this has commonly been taken to mean that the soundscape is the perception itself, while the factors which lead to the soundscape are separate entitites. In this definition, the soundscape is not made up of sound sources, the visual environment, etc. but instead is the perception formed by them. This definition was proposed by A. L. Brown (2012) where the author made this distinction very clear in a section titled **'Soundscape is perception of the acoustic environment of a place'**: 'Thus, a soundscape exists through human perception [...] the soundscape of a place is thus a percieved entity'.

Given this definition, speaking about the 'soundscape perception' would be redundant; the soundscape already is the perception. By extension, saying 'the soundscape is perceived as pleasant' also would not make sense; we should rather say 'the soundscape is pleasant'. However, even among the foundational modern soundscape literature these uses are relatively widespread; Ö. Axelsson, Nilsson, and Berglund (2010); Liu, Kang, Behm, and Luo (2014) both refer to soundscape perception within the title.

This definition also conflicts with other popular definitions of soundscape. The term soundscape is commonly used in acoustic ecology and underwater acoustics – see titles such as 'The soundscape of bat swarms' (Kloepper et al., 2017), 'An integrated underwater soundscape analysis in the Bering Strait region' (McKenna, Southall, Chou, Robards, & Rosenbaum, 2021), 'Soundscape analysis and acoustic monitoring document impacts of natural gas exploration on biodiversity in a tropical forest' (Deichmann, Hernández-Serna, C., Campos-Cerqueira, & Aide, 2017), and 'Identification and quantification of soundscape components in the Marginal Ice Zone' (Geyer, Sagen, Hope, Babiker, & Worcester, 2016). Several analysis packages have also been developed for the purpose of soundscape analysis, whether for urban-, underwater-, or bio-acoustics, which include no aspect of human perception in context (see e.g. Soundscape Viewer (Y.-J. Sun & Lin, 2020) and `scikit-maad` (Ulloa, Haupert, Latorre, Aubin, & Sueur, 2021)).

These fields appear to use the term *soundscape* more broadly, without a reference to human perception, to refer to either a broad consideration of the entire sound environment or to a focus on the sound environment as perceived by all creatures, not just humans. This first definition comes from Pijanowski et al. (2011) where the authors state that 'soundscape ecology focuses mostly on macro or community acoustics [...] the composition of all sounds heard at a location that are biological, geological, or anthropogenic' to differentiate it from previous acoustic ecology studies which 'focus on a single species or a comparison of species'. Within the ISO 12913 framework, this would more accurately be described as the *acoustic environment* ('sound at the receiver from all sound sources as modified by the environment'). In the end, all of these conflicting and overlapping definitions can make cross-disciplinary communication more difficult and prone to disagreements and misunderstandings. **draft** *cont. If you want to keep this in.*

Chapter 2.

Literature Review

Although soundscape studies have seen increased attention over the last decade (Kang et al., 2016), engineering practice is still dominated by a noise control approach. For this reason, this review of the literature will begin with some examples of how urban sound is assessed by standard noise control methods before moving on to discuss how the soundscape approach represents an improvement over these methods. From here, I will present the conceptual framework for predictive soundscape models (Aletta et al., 2016) from which the work in this thesis began and discuss why predictive models are necessary to enable an engineering approach to soundscape design. Finally, some tools and previous predictive models are reviewed.

2.1. The importance of perception and experience

Despite being the dominant focus of urban noise mitigation, the reduction of sound levels has been proven to not necessarily correlate with perception or lead to improved health outcomes (Andringa et al., 2013; Asdrubali, 2014; Kang, 2006; Kang et al., 2016; van Kempen, Devilee, Swart, & van Kamp, 2014). Research from the early 2000s demonstrated that reducing the sound level does not necessarily lead to better acoustic comfort in urban areas (De Ruiter, 2000; Schulte-Fortkamp, 2001). W. Yang and Kang (2005a) assessed the acoustic comfort of people in 14 urban spaces of five European countries (Greece, Italy, UK, Germany, and Switzerland). For this study, users of the space were randomly selected in the spaces and asked to evaluate the subjective sound level on a scale from 1 (very quiet) to 5 (very noisy), while an additional measure of acoustic comfort was assessed (from 1 [very comfortable] to 5 [very uncomfortable]) in the 2 case study sites in Sheffield, UK. While each participant was interviewed, the researchers measured the 1-minute L_{Aeq} as well as additional microclimate indices.

By examining the relationship between the subjective sound level and the measured sound level within each site separately, their results indicated an inconsistent relationship across sites, with correlation values ranging from $R = 0.373$ for Sesto San Giovanni, Italy to $R = 0.941$ for Karaisakaki square, Greece. This indicates that although L_{Aeq} can be a good indicator for the subjective sound level, the strength of this relationship and in particular the slope of the relationship depends on other factors not captured by the dB(A). This is further reinforced by the acoustic comfort results. Fig. 2.1 shows the acoustic comfort and subjective sound level ratings for the two Sheffield case study sites reported in W. Yang and Kang (2005a). Although in these sites there is a strong correlation between the L_{Aeq} and the subjective sound level, the correlation with acoustic comfort is much weaker. In addition, Fig. 2.1(a) in particular shows a nonlinear relationship between dB(A) and acoustic comfort. It can be seen that above $\sim$70 dB(A) the acoustic comfort decreases as the sound level increases, however below 70 dB(A) there is no significant change in acoustic comfort.

Parallels in visual perception Similar results showing the disconnect between commonly used metrics of the physical environment and the environment's impact on the users have also been demonstrated by studies looking at visual perception. Kruize et al. (2019) demonstrated that the experience and use of natural urban spaces were strongly related to health outcomes associated with those spaces, while strictly physical characteristics of the space were not. Their study made use of a cross-sectional design to investigate the relationships between several factors related to the experience of natural outdoor environments and key health outcome indicators. The outcome indicators investigated included physical activity, social contact, and mental health. As input indicators, they considered a set of GIS-derived quantitative indicators (i.e. Normalized Difference Vegetation Index (NDVI)) (Smith et al., 2017) and a selection of metrics describing the use and experience of the space, derived from surveys of the study participants. These survey-derived experience metrics included *perceived greenness*, *satisfaction with the natural environment*, and *importance of the natural environment*. Through multilevel regression analyses, the authors found that, in general, NDVI was not statistically significantly related to increased physical activity or improved mental health, while perceived greenness, satisfaction with the space, and importance of the space were. A one point increase in perceived greenness (ranging from 0-12) was associated with an additional 10 min of physical activity per week and an increased mental well-being score (MHI-5) of 0.331 (range 8-100). What is particularly interesting in these results is the difference in the findings between the objective measure of greenness and the perceived greenness. A study which used only the GIS-derived metric would have concluded that there was no relationship between greenness and the outcome factors. By including a perceptual attribute, Kruize et al. (2019) were able to demonstrate its importance. This suggests that what really matters to people's use of and the health and well-being impacts of these spaces is *how they are perceived* more so than just what their physical characteristics are.

Across both the visual and the auditory domain, research has suggested that a disconnect exists between the physical metrics used to describe urban environments and how they are perceived. In addition, this disconnect can be extended further into how these environments influence the health and well-being of their users. To gain a better understanding of these spaces and their impacts on

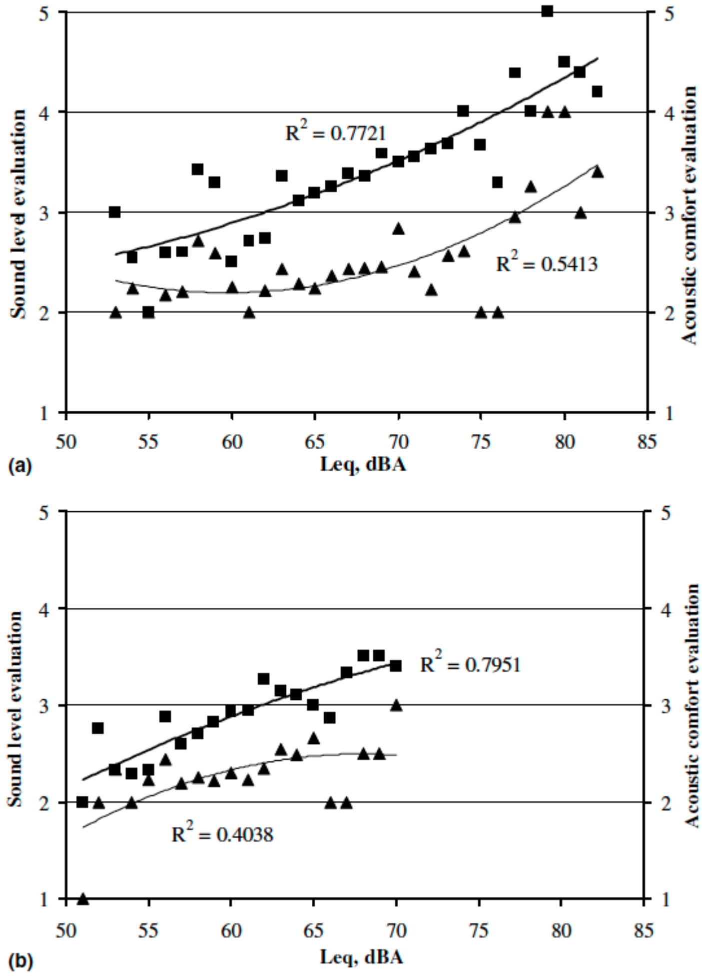


Figure 2.1.: Reproduced with permission from (W. Yang & Kang, 2005a, Fig. 2) showing subjective responses and measured sound levels in urban spaces in Sheffield, UK. Relationships between the measured sound level, the mean subjective evaluation of the sound level and the mean acoustic comfort evaluation, with binomial regressions and correlation coefficients squared R^2. (a) The Peace Gardens. (b) The Barkers Pool. ■ – subjective evaluation of sound level (1 [very quiet]; 2 [quiet]; 3 [neither quiet nor noisy]; 4 [noisy]; 5 [very noisy]). ▲ – acoustic comfort evaluation (1 [very comfortable]; 2 [comfortable]; 3 [neither comfortable nor uncomfortable]; 4 [uncomfortable]; 5 [very uncomfortable]).

people who work and live in cities, we must create assessment methods and metrics which go beyond merely characterising the physical environment and instead translate through the users' perception. In order to make the case for why a new approach to perception-focussed assessment and design is necessary, we must first understand how the existing assessment methods have attempted to consider perception.

2.2. Attempts to reconcile dB-focussed noise control with human perception

In this section I will begin with a brief discussion of the noise approach to annoyance and review some examples of noise assessment and mitigation methods and how they have attempted to reconcile the noted disconnect between the dB and the sound perception. This will provide the existing context for how traditional noise control approaches are targeted.

2.2.1. Assessing noise

In the UK, BS 4142:2014+A1:2019 (2019) is the current reference document for assessing and addressing noise impacts in outdoor environments. In particular, BS 4142 is intended to assess the impact of a specific noise source when introduced to a given background level. BS 4142 makes use of a 'rating level' based on a comparison between the sound which is being assessed and the background sound which would exist without it. Within the standard, a series of noise metrics are defined (BS 4142:2014+A1:2019, 2019, Sec. 3):

- **ambient sound level,** $L_a = L_{Aeq,T}$ - equivalent continuous A-weighted sound pressure level of the totally encompassing sound in a given situation at a given time, usually from many sources near and far, at the assessment location over a given time interval, T.
- **background sound level,** $L_{A90,T}$ - A-weighted sound pressure level that is exceeded by the residual sound at the assessment location for 90% of a given time interval, T, measured using time weighting, F, and quoted to the nearest whole number of decibels.
- **equivalent continuous A-weighted sound pressure level,** $L_{Aeq,T}$ - value of the A-weighted sound pressure level in decibels of continuous steady sound that, within a specificed time interval, $T = t_2 - t_1$, has the same mean-squared sound pressure as a sound that varies with time, and is given by the following equation:

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

where:
p_0 is the reference sound pressure (20 μPa); and
$p_a t$ is the instantaneous A-weighted sound pressure (Pa) at time *t*

- **residual sound level,** $L_r = L_{Aeq,T}$ - equivalent continuous A-weighted sound pressure level of the residual sound[1] at the assessment location over a given time interval, T.
- **specific sound level,** $L_s = L_{Aeq,Tr}$ - equivalent continuous A-weighted sound pressure level produced by the specific sound source at the assessment location over a given reference time interval, T_R.

In each of these metrics, the primary sonic feature which is assessed is the sound level, with some consideration for frequency content by using A-weighting. The standard then sets out the procedure to be used to measure the existing background sound level, measure or estimate the level of the specific sound, and calculate the margin between the specific sound and the background sound level. Throughout this process, BS 4142 notes 'certain acoustic features can increase the significance of impact over that expected from a basic comparison between the specific sound level and the background sound level.' To address this, it introduces certain methods to add a character correction to the specific sound level, resulting in the rating level:

- **rating level,** $L_{Ar,Tr}$ - specified sound level plus any adjustment for the characteristic features of the sound

The sonic characteristics included for these rating level adjustments are tonality, impulsivity, intermittency, and 'other sound characteristics' (described as 'otherwise readily distinctive against the residual acoustic environment'). As an example of how these adjustments are applied, I will quote the guidance to adjust for tonality:

> **Tonality**
> For sound ranging from not tonal to prominently tonal the Joint Nordic Method (ISO 1996-1:2016, 2016) gives a correction of between 0 dB and + 6 dB for tonality. Subjectively, this can be converted to a penalty of 2 dB for a tone which is just perceptible at the noise receptor, 4 dB where it is clearly perceptible, and 6 dB where it is highly perceptible.
>
> BS 4142:2014+A1:2019 (2019, pg. 13)

The goal of these rating level adjustments is to incorporate aspects of the specific perception of sound, which may make a sound more disturbing, more noticeable, or generally more impactful

[1] Ambient sound remaining at the assessment location when the specific sound source is suppressed to such a degree that it does not contribute to the ambient sound.

than the dBA value alone would suggest. This is an important part of rating the impact of these sounds and necessary to achieve the goals of the standard. However, by implementing this as adjustments in terms of dB, it still centres the sound level in the assessment and enables only a one-dimensional approach to assessing impact.

While BS 4142 is targeted towards assessing the impact of a specific sound, and primarily in an industrial and commercial context, ISO 1996-1:2016 (2016) is more general, including provisions for assessing 'community noise'. In the Introduction to the standard, several acknowledgements are raised about the importance of human perception, but the primary focus on the sound level is confirmed:

> To be of practical use, any method of description, measurement, and assessment of environmental noise is intended to be related in some way to what is known about human response to noise. [...] The methods and procedures described in this part of ISO 1996 are intended to applicable to noise from various sources, individually or in combination, which contribute to the total exposure at a site. At the stage of technology at the time of publication of this part of ISO 1996, the evaluation of long-term noise annoyance seems to be best met by adopting the adjusted A-weighted equivalent continuous sound pressure level, which is termed a "rating level".
>
> ISO 1996-1:2016 (2016)

Despite the nods toward a perception-focussed approach, ISO 1996 re-emphasises the focus on sound pressure level when it comes to discussing community noise annoyance:

> If the sound has special characteristics, then the rating equivalent continuous sound pressure level shall be the primary measure used to describe the sound. [...] research has shown that different transportation sounds or industrial sounds evoke different community annoyance responses for the same A-weighted equivalent continuous sound pressure level.
>
> ISO 1996-1:2016 (2016, Sec. 6.1)

What this review of some of the relevant standards and guidance demonstrates is that the decibel (specifically the equivalent continuous sound pressure level) is the dominant rating metric for all types of environmental noise and that, despite attempts to incorporate adjustments for 'special' characteristics of the sound, the only sonic characteristic really being considered is the sound level. In addition, what my review of this guidance reveals is that the rating level adjustments are impractical to apply to complex sound environments for general evaluation. ISO 1996-1:2016 (2016, Eq. 4) provides a formula for calculating the rating level of combined sources:

$$L_{Req,T} = 10lg\left(\frac{1}{T}\sum_{n}\sum_{j}T_{nj} * 10^{0,1L_{Reqj,Tnj}}\right) dB$$

where

$$T = \sum_{n} T_{nj}$$

for each source *j*.

However, ISO 1996 notes that 'As a practical matter, Formula (4) is typically evaluated one source at a time.' This makes it clearly impractical as a method for assessing a general sound environment, with multiple competing sound sources, or for automated monitoring. In this way, these standards preclude the possibility of making use of the rating level to provide a more nuanced view of a sound environment which accounts for the sonic characteristics.

2.2.2. Community noise annoyance

There is an existing methodology to address community noise annoyance, which began to develop in the 1940s following an increase in community complaints, primarily in response to aircraft noise (Kryter, 1994)[2]. From this, definitions of noise annoyance and single-number assessment indices were developed which focussed on the aircraft and transport noise impacts on residential areas. ISO/TS 15666:2021 (2021) defines annoyance (specifically 'noise-induced annoyance') as 'one person's individual reaction to noise.' This is assessed through socio-acoustic surveys using questions with either verbal or numerical rating scales. In contrast to many studies from the soundscape literature, the noise annoyance scales in ISO/TS 15666:2021 (2021) refer to long time scales ('Thinking about the last (12 months or so) ...') whereas soundscape studies have tended to focus on shorter and more immediate time scales (Rychtáriková & Vermeir, 2013; M. Yang & Kang, 2013). This is perhaps due to the more complex nature of the perception under investigation in soundscape.

ISO/TS 15666:2021 (2021) then extends to *community noise annoyance*, defined as 'the prevalence rate of this individual reaction in a community as measured by the responses to questions specified in Clause 4 and expressed in appropriate statistical terms.' This approach to community noise inherently recognises that 1) noise annoyance is an individual response which will vary among people and 2) the most appropriate way to discuss this impact on a broader scale is to describe the aggregate response statistically (e.g. 80% prevalence of 'highly annoyed' individuals).

A commonly used metric in noise annoyance studies in this context is the L_{dn}, which is the average of the sum of the A-weighted sound energy over 24 hours, with a penalty of 10 dB added for the hours from 10:00 pm to 7:00 am, on an annualised basis (Kryter, 1994, pg. 571):

$$L_{dn} = 10log((1/54,000(10^{L_A1s,7am..L_A1s,10pm/10}) \\ + (10(1/32,400(10^{L_A1s,10pm..L_A1s,7am/10})) \tag{2.1}$$

In the contexts for which it was developed, L_{dn} provides a good correlation with 'the cumulative percentage of people' being moderately, very, or highly annoyed for specific sources of noise, such as aircraft noise (0.89, 0.89, and 0.87, respectively) (Kryter, 1994). A predictive trend curve can be derived from this:

$$\%\text{Highly annoyed} = 110.091 + (-5.023 \times L_{dn}) = (0.058 \times L_{dn}^2)$$

However, as noted, this is employed for assessing the annualised annoyance from a particularly noted source averaged over 24 hours in a residential setting. The residential setting is particularly noted in the inclusion of the 10 dB penalty during night-time hours, to consider sleep disturbance impacts. The long time-scales and the and lack of consideration of either more complex sonic characteristics or the potential positive impacts of sounds make these methods unsuitable for assessing the soundscape of public spaces. However, noise annoyance methodology does provide a valuable advantage for our purposes when compared to the environmental noise assessment methods reviewed earlier:

> The procedures for estimating subjective annoyance and complaints about environmental noise are intended for the assessment of the reactions of large groups or neighborhoods of people, and not specific individuals within a group.
>
> Kryter (1994, pg. 571)

This concept will be further explored and addressed in Chapter 9.

Traditional noise control methods face several challenges in decreasing noise pollution in modern cities. In many cases, these challenges stem from an approach primarily focussed strictly on decreasing the noise levels or noise exposure in a given space. Part of this approach stems from traditional assessment methods which centre the sound level as the sole metric and which struggle to account for additional sonic characteristics such as tonality or the meaning associated with

[2]Kryter (1994) opts not to refer to annoyance when referring to a single sound, in order to 'avoid some of the ambiguity possible with the word *annoyance*. Instead, the author prefers to use the phrase *perceived noisiness*, defined as 'the subjective unwantedness felt from a sound, independently of any meanings or effects it may have.' I find this a somewhat strange definition, due to the specific attempt to define it independent of any meaning or effects. Whether someone considers a particular sound to be *noise* could be entirely dependent on the meaning they associate with it, and independent of the acoustical characteristics of the sound.

a sound. In particular, the metrics used, particularly when attempting to adjust for these characteristics, are impractical to apply for complex sound environments[3] with several overlapping sound sources. This approach can often prove impractical in situations where a problematic noise source cannot be moved or decreased, or where mitigation methods such as building a sound wall to block the sound transmission are expensive, infeasible, or merely undesirable (Ekici & Bougdah, 2003). This can result in many urban spaces which are intended to provide a restorative space in the city being unpleasant due to the unwanted noise and going underutilised with little way to address the issue. Where noise and acoustics is considered by planners and architects, their concern is typically with compliance of ordinances and regulations, or with maintaining the existing environmental conditions. Noise mitigation efforts frequently fail to centre human perception within the design (Coelho, 2016).

As can be seen in the structure and guidance of these standards documents, the goal is to maintain the existing sound environment and mitigate noise impacts from newly introduced sounds, with the dB, in terms of a rating level, as the metric of assessment. Attempts have been made across the various standards to include adjustments for how sonic characteristics such as tonality, impulsivity, and spectral content influence human perception in a useful and practical way. It should also be clear that, as expressed by Axelsson, these standards approach sound reactively, addressing it only as something to be managed, reduced, or tracked. Although the goals laid out in the Environmental Noise Directive (END) have shown a shift to considering sound as a potential benefit to be protected and promoted, the assessment methods available are still rooted in the waste management approach.

2.2.3. Soundscape: A perception approach

Soundscape studies strive to understand the perception of a sound environment, in context, including acoustic, (non-acoustic) environmental, contextual, and personal factors. These factors combine together to form a person's soundscape perception in complex interacting ways (Berglund & Nilsson, 2006). Humans and soundscapes have a dynamic bidirectional relationship – while humans and their behaviour directly influence their soundscape, humans and their behaviour are in turn influenced by their soundscape (Erfanian, Mitchell, Kang, & Aletta, 2019). Researchers in the areas of acoustics, environmental psychology, and auditory neuroscience outline the adverse impact of noise or negative sounds on well-being in an attempt to improve modern living standards (Hao, Kang, & Wörtche, 2016; Ising & Kruppa, 2004; Lawton & Fujiwara, 2016; Pedersen & Waye, 2007). In this regard, evidence indicates that positively perceived sounds (e.g. natural sounds) are tied with a high quality of life and enhanced psychological and physical health (Aletta, Oberman, & Kang,

[3] Throughout this thesis, I will use the phrase 'complex sound environment' or 'complex soundscape' to refer to a real-world environment with overlapping and competing sound sources which consist of sonic characteristics that any single metric currently struggles to encapsulate.

2018; Alvarsson, Wiens, & Nilsson, 2010; Jeon, Lee, You, & Kang, 2010; Shepherd, Welch, Dirks, & McBride, 2013).

When applied to urban sound and specifically to noise pollution, the soundscape approach introduces three key considerations beyond traditional noise control methods:

1. considering all aspects of the environment which may influence perception, not just the sound level and spectral content;
2. an increased and integrated consideration of the varying impacts which different sound sources and sonic characteristics have on perception; and
3. a consideration of both the positive and negative dimensions of soundscape perception.

This approach can enable better outcomes by identifying existing positive soundscapes (in line with the END's mandate to 'preserve environmental noise quality where it is good' (European Union, 2002)), better identify specific sources of noise which impact soundscape quality and pinpoint the characteristics which may need to be decreased, and illuminate alternative methods which could be introduced to improve a soundscape where a reduction of noise is impractical (Fiebig, 2018; Kang & Aletta, 2018). These can all lead to more opportunities to truly improve a space by identifying the causes of positive soundscapes, while also potentially decreasing the costs of noise mitigation by offering more targeted techniques and alternative approaches.

2.3. Soundscape Studies

This section will review the terms used for defining soundscapes, some existing methods for characterising urban soundscapes, and the previous attempts and frameworks for connecting the physical environment with soundscape perception.

Soundscape, conceived as the acoustic equivalent of landscape, is defined as the human's perception of the acoustic environment, in context (ISO 12913-1:2014, 2014; Kang, 2010; Schafer, 1977). The soundscape can be the result of a single sound or a combination of sounds that arises from an engaging environment. The Canadian composer and naturalist R. Murray Schafer led much of the original work to advance research in the area (Schafer, 1969), borrowing the term originally from work carried out by city planner Michael Southworth (Southworth, 1969). Since Schafer, there have been several multi-dimensional classifications for soundscapes. However, according to Schafer, the main components of the soundscape consist of keynote sounds, sound signals, and soundmarks. The soundscape ecologist Bernie Krause characterised soundscapes into three main domains based on the source of the sound. According to his classification, the soundscape refers to a wide spectrum of sounds, encompassing natural sounds relating to non-organic elements of nature such as waterfalls (geophony), organic but non-human sources such as animals' copulatory vocalisations (known as biophony), and all environmental sounds generated by human sources (anthrophony) such as human voices or human activity-related sounds (Kang & Schulte-Fortkamp, 2016; Krause, 1987). From this starting point in music and soundscape ecology, urban soundscape studies have advanced over the last two decades (Kang, 2006; Kang & Aletta, 2018). Fiebig (2018) noted that the standardization of soundscape methods was necessary to provide 'minimum measurement requirements leading to a (minimal) guaranteed level of reliability'. The next section will review the history and outcome of the resulting standard.

2.3.1. Standardising Soundscape: The ISO 12913 series

The soundscape community is undergoing a period of increased methodological standardization in order to better coordinate and communicate the findings of the field. This process has resulted in many operational tools designed to assess and understand how sound environments are perceived and apply this to shape modern noise control engineering approaches. Important topics which have been identified throughout this process are soundscape 'descriptors', 'indicators', and 'indices'. Aletta et al. (2016) defined soundscape descriptors as 'measures of how people perceive the acoustic environment'; soundscape indicators as 'measures used to predict the value of a soundscape descriptor'; soundscape indices can then be defined as 'single value scales derived from either descriptors or indicators that allow for comparison across soundscapes' (Kang et al., 2019).

This conception has recently been formalized and expanded upon with the adoption of the recent ISO 12813 set of standards (ISO 12913-1:2014, 2014; ISO/TS 12913-2:2018, 2018; ISO/TS 12913-3:2019, 2019). ISO 12913 Part 1 sets out the definition and conception of Soundscape, defining it as the 'acoustic environment as perceived or experienced and/or understood by a person or people, in context'. Here, the soundscape is separated from the idea of an acoustic environment, which encompasses all of the sound which is experienced by the receiver, including any acoustically modifying effects of the environment. In contrast, the soundscape considers the acoustic environment, but also considers the impact of non-acoustic elements, such as the listener's context and the visual setting, and how these interact with the acoustic environment to influence the listener's perception.

2.3.2. Soundscape data collection methods

Methods for collecting data on how people experience acoustic environments have been at the forefront of the debate in soundscape studies for the past 20 years. While the soundscape research field as we understand it today dates back to the late 1960s with the pioneering work of authors like M. Southworth (Southworth, 1969), R.M. Schafer (Schafer, 1977), and H. Westerkamp (Westerkamp, 2002), the theme of data collection methods for soundscape assessment emerged more prominently only recently (Kang et al., 2016). There is a general consensus in the research community that standardised tools to gather and report individual responses on the perception of urban acoustic environments are indeed desirable, to provide comparable datasets and soundscape characterisations across different locations, times, and samples of people, as well as allowing for replicability studies and offering inputs for modelling algorithms in soundscape prediction and design tasks. These were among the main drivers for the establishment of a Working Group at the International Organization for Standardization (ISO) back in 2008, which was named "Perceptual assessment of soundscape quality" (ISO/TC 43/SC1/WG 54) that has so far published three documents within the ISO 12913 series on soundscape. Part 1 (ISO 12913-1:s014) is a full standard and provides a general framework and definitions of soundscape concepts (ISO 12913-1:2014, 2014), while Part 2 (ISO/TS 12913-2:2018) and Part 3 (ISO/TS 12913-3:2019) are technical specifications and offer guidance on how data should be collected and analysed, accordingly (ISO/TS 12913-2:2018, 2018; ISO/TS 12913-3:2019, 2019) (Part 4, on soundscape design interventions, is currently under development by the working group, also registered as a technical specifications document). Specifically, Part 3 presents the proposed methods for analysing and representing the data collected by the soundscape surveys. Since the development of these standards, the focus has shifted from understanding individual perception to characterising the collective perception of increasingly large groups.

The ISO/TS 12913-2:2018 is the current reference document addressing data collection and reporting requirements in soundscape studies. In terms of methods, the ISO document covers two main approaches, namely: soundwalks combined with questionnaires (Methods A and B) and narrative interviews (Method C) (ISO/TS 12913-2:2018, 2018), which relate to on-site and off-site data collection, accordingly. Part 3 of the ISO 12913 series builds on Part 2 and provides guidelines for analysing data gathered using only those methods (ISO/TS 12913-3:2019, 2019). However, the range of possible methodological approaches to soundscape data collection is much broader and it includes, for instance, laboratory experiments (Aletta et al., 2016; Oberman, Šćitaroci, & Jambrošić, 2018; K. Sun et al., 2019), pseudo-randomized experience sampling (Craig, Moore, & Knox, 2017), and even non-participatory studies (Lavia et al., 2018).

2.4. Soundscape Descriptors, Indicators, and Perceptual Mapping

Aletta et al. (2016) provides a review of the soundscape descriptors and indicators commonly used in soundscape research and outlines an initial framework for developing predictive soundscape models. In their review, the authors identified 8 potential soundscape descriptors:

1. Noise annoyance
2. Pleasantness
3. Quietness or tranquility

4. Music-likeness
5. Perceived affective quality
6. Restorativeness
7. Soundscape quality
8. Appropriateness

To this list, I would add 'acoustic comfort' as used in W. Yang and Kang (2005a) as a ninth potential descriptor. Similarly, the authors identified a range of potential indicators used to characterise the acoustic environment:

- L_{Aeq}
- statistical levels ($L_x - L_{100-x}$)
- proportion of low-frequency sounds ($L_{Ceq} - L_{Aeq}$)
- Loudness (N_5)
- Sharpness (S)
- Roughness (R)
- Fluctuation Strength (FS)

However, it is noted that several studies show that no single psychoacoustic indicator alone can explain the variation in soundscape responses (as expressed via the descriptors) (e.g. (Persson Waye & Öhrström, 2002)). The goal of statistical modelling, therefore is to create a more complex and complete representation of the relationship between soundscape indicators and descriptors, beyond what any single indicator could achieve.

Fig. 2.2 shows a conceptual view of this relationship. We start with **soundscape indicators**, describing the physical environment to which a listener is exposed. Soundscape indicators characterise the physical and contextual environment to which the listener is exposed. This can be broken down into **sonic features** (e.g. the acoustical features listed above) and **characteristics of the space** itself (e.g. the amount of visible sky, the intended use-case of the space, how crowded the space is, etc.). In order to translate from the physical inputs to an expressed description of the soundscape perception, we introduce the concept of a **perceptual mapping** (Lionello, 2021). This mapping represents a simplified idea of how each individual's brain processes the inputs from the soundscape which they experience, forms a perception, and finally expresses that perception through their description of the soundscape. For our purposes, this perceptual mapping is treated as essentially a black box mapping inputs to outputs. It can be conceived of as a network of weights in which certain characteristics of the sound may have different weights and directions depending on the context, through which all

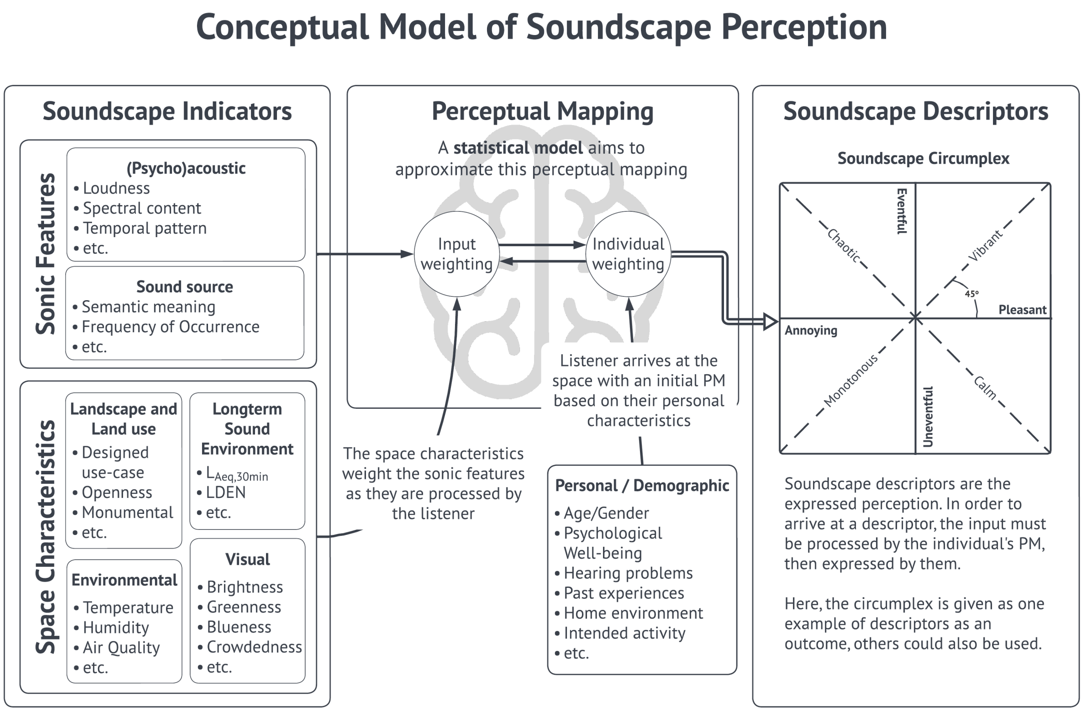


Figure 2.2.: The conceptual model of soundscape perception, illustrating the perceptual mapping from physical inputs, through personal experience, to soundscape descriptors. The role of the statistical model is to attempt to approximate or reflect this perceptual mapping.

of the inputs are processed, resulting in the soundscape rating. In general, this perceptual mapping is formed prior to an individual's exposure to the soundscape in question.

We then break the perceptual mapping into two parts: how the inputs are weighted relative to each other (which is relatively consistent across participants) and the particular variation in each person's perception based on their own experiences and background. In this conceptual model, the weighting of the sonic features (both the acoustic features and the sound source information) are mediated by the space characteristics as they are processed by the listener. The individual weighting represent the effects due to the listener's particular personal characteristics (their age, gender, psychological well-being, etc.) as well as the inherent unpredictable randomness inherent in each individual's experience of the soundscape.

It should be made clear that this represents a very simplified view of how a soundscape perception is formed, however for the purposes of understanding and modelling how someone's perception forms in response to their exposure to a space, it provides a useful conceptual framework. One way to consider the function of a statistical model of soundscape perception is as replicating the perceptual mapping between soundscape indicators and descriptors (Lionello, 2021). As a person experiences an urban space, they are exposed to an array of physical inputs, these are then processed by the listener through their own personal experience and mapped to their perception of that space. This perception is then expressed through their description of this experience of the soundscape. It is this mapping of physical inputs to perceptual description which the statistical model aims to reflect. The most successful model would then accurately replicate the general perceptual mapping across the population.

2.4.1. Perceived Affective Quality

Based on the work in Aletta et al. (2016) and a recent review of predictive soundscape models (Lionello, Aletta, & Kang, 2020), among the potential soundscape descriptors which can be used, I have selected the soundscape circumplex (Ö. Axelsson et al., 2010), specifically the version in Method A of ISO/TS 12913-2:2018 (2018) as the most appropriate for predictive modelling.

Method A is built on a series of descriptors referred to as the Perceived Affective Quality (PAQ), proposed by Ö. Axelsson et al. (2010). These PAQs are based on the pleasantness-activity paradigm present in research on emotions and environmental psychology, in particular Russell's circumplex model of affect (Russell, 1980). As summarised by Axelsson: 'Russell's model identifies two dimensions related to the perceived pleasantness of environments and how activating or arousing the environment is.' This circumplex model is formed of two dimensions, pleasantness (often referred to as valence) and activity (or arousal), which are orthogonal to each other. In their study, three primary dimensions of soundscape perception were extracted from participants' responses to complex sound samples measured on 116 attributes, using Principal Components Analysis. The first compo-

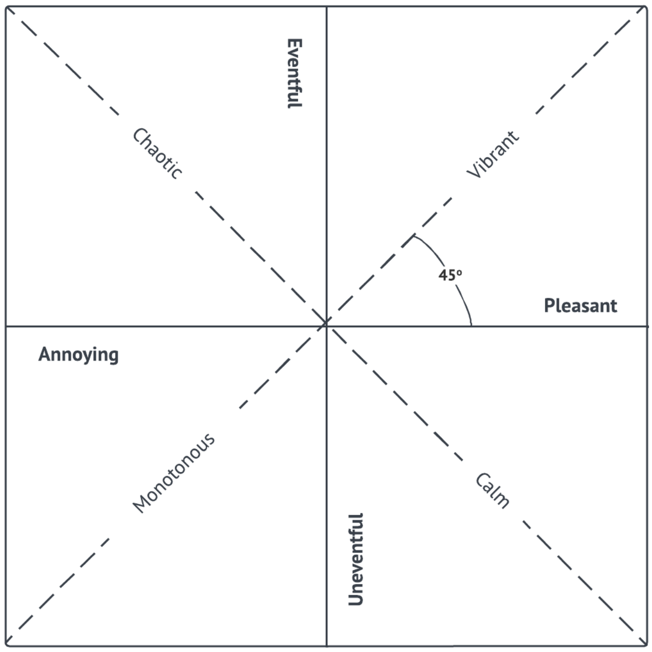


Figure 2.3.: The soundscape circumplex, as originally derived by Ö. Axelsson et al. (2010) and updated in ISO/TS 12913-2:2018 (2018).

nent was found to represent pleasantness (aligning with attributes such as comfortable, appealing, uncomfortable, disagreeable, and inviting) and explained 50% of the variance in the dataset. The second component was found to represent eventfulness (eventful, lively, uneventful, full of life, and mobile) and explained 18% of the variance. The third component was found to represent familiarity (commonplace, common, and familiar) and explained 6% of the variance. When applied to soundscape, Axelsson re-termed these main axes as 'Pleasant' and 'Eventful', and also identified a set of additional axes which are rotated 45° from the main axes. This rotated axis contains additional attributes which represent various mixtures of the pleasant and eventful attributes: 'Exciting', 'Chaotic', 'Monotonous', and 'Calm'. This circumplex model of soundscape can be seen in Fig. 2.3. In Method A, these PAQs are collected through a series of questions with 5-point Likert-type responses where participants are asked to what extent they agree or disagree that the present surrounding sound environment is pleasant, exciting, etc. for each of the 8 descriptors. Method A also includes questions on: the sound source composition of the space, broken down into 'Traffic noise', 'Other noise', 'Sounds from human beings', and 'Natural sounds'; overall soundscape quality; appropriateness of the sound environment to the place. The circumplex model, along with the sound source and general soundscape questions represent a relatively comprehensive method for assessing the soundscape of a space.

One benefit of the circumplex model is that, as a whole, it encapsulates several of the other proposed soundscape descriptors - in particular, annoyance, pleasantness, tranquility, and possibly restorativeness (Aletta et al., 2016). According to Ö. Axelsson (2015), the two-dimensional circumplex model of perceived affective quality provides the most comprehensive information for soundscape assessment. It is also possible that the overall soundscape quality could itself be derived from the pleasant-eventful scores derived for a soundscape.

The circumplex also lends itself well to questionnaire-based methods of data collection, as proposed in ISO/TS 12913-2:2018 (2018). In contrast to methods such as soundwalks, interviews, and lab experiments, in-situ questionnaires are able to provide the quality and amount of data which is necessary for statistical modelling. Large-scale, in-situ questionnaires are therefore considered the most appropriate data collection approach for generating a soundscape assessment database intended for predictive modelling.

2.5. Statistical Models of Soundscape Perception

Several studies prior to the formalization of the ISO standards on soundscape demonstrated the general, but inadequate, relationship between traditional acoustic metrics, such as L_{Aeq}, with the subjective evaluation of the soundscape (R. M. Alsina-Pagès, Freixes, Orga, Foraster, & Labairu-Trenchs, 2021; Aumond et al., 2017; Berglund & Nilsson, 2006; Rychtáriková & Vermeir, 2013; W. Yang & Kang, 2005a). These have typically aimed to address the existing gap between traditional environmental acoustics metrics and the experience of the sound environment. Yang and Kang (2005) showed that, when the sound level is 'lower than a certain value, say 70 dBA', there is no longer a significant change in the evaluation of acoustic comfort as the sound level changes. However, the perceived sound level does continue to change along with the measured sound level, showing that (1) measured sound level is not enough to predict soundscape descriptors such as 'acoustic comfort', and (2) there is a complex relationship between perceived sound level and soundscape descriptors which is mediated by other factors.

Contrary to the hopes expressed by Aletta, Axelsson, and Kang (2014), that 'ideally there should be one acoustic indicator per dimension', the evidence from subsequent investigations and modelling attempts (Lionello et al., 2020) indicates this to be unlikely. There appears to be no reason we should think the perceptual dimensions should be reduced to a single acoustic indicator. The dimensions of soundscape represent complex perceptual concepts which we should expect to be composed of a multi-factor interaction between the input features. This necessary complexity highlights the need for a more sophisticated machine learning approach in order to handle and interpret the interactions between the many input features which contribute to the formation of a soundscape perception. (Aletta et al., 2016)

According to a recent review of predictive soundscape models from Lionello et al. (2020), the degree of employing auditory and non-auditory factors in soundscape prediction varies, with some studies relying on contextual, personal/demographic (Erfanian, Mitchell, Aletta, & Kang, 2021; Tarlao, Steffens, & Guastavino, 2020) or social media (Aiello, Schifanella, Quercia, & Aletta, 2016) data entirely to predict and generate soundscape features. Some methods also incorporate perceptually-derived features, such as subjective sound level and visual pleasantness as predictors (Lionello et al., 2020). In general, these methods which incorporate perceptually-derived inputs achieve better accuracy rates than those which don't, however this information must also be obtained from people via a survey and therefore are unsuitable for predictive modelling where surveys are not possible. For example, Ricciardi, Delaitre, Lavandier, Torchia, and Aumond (2015) proposed two models based on data collected from a smartphone application to predict urban sound quality indicators based on linear regressions. The first model which incorporated perceptually-derived input features (visual quality and familiarity) achieved an R^2 of 0.72, while a second model without these features achieved an R^2 of 0.58. This indicates the necessity for considering and accounting for the influence which contextual factors in a space have on the relationship between the sound environment itself and the listener's perception of it (i.e. the soundscape) while also highlighting the challenges associated with a predictive model which depends only on measurable features.

These previous studies have generally been limited by one or many of the following factors:

- limited number or types of locations;
- limited responses sample size;
- no non-acoustic factors.

These factors generally limit the generalizability of their results beyond the investigated locations.

Psychoacoustic Annoyance Models for the prediction of annoyance based solely on a combination of psychoacoustic metrics have been previously proposed, with the most notable model based on psychoacoustic metrics proposed by Zwicker and Fastl (2007). The authors provide definitions and the empirical basis behind a series of psychoacoustic metrics (loudness, roughness, sharpness, fluctuation strength), the specifics of which will be further expanded upon in Chapter 4. Briefly, these metrics relate to specific psychophysical sensations which move beyond the strictly physical descriptions of sounds. Acoustical metrics such as L_{Zeq} describe the physical characteristics of a sound, derived from the magnitude of the pressure changes induced by the sound. By contrast, psychoacoustical metrics attempt to relate these physical characteristics to the sensation they induce in humans. They therefore provide a more direct insight into how sounds are perceived and interpreted by a listener.

Each of the proposed psychoacoustic metrics therefore attempts to describe one aspect of the sonic quality of the sound such that a sound can be broken down and described through some combination of these metrics. Zwicker and Fastl then propose a model which combines these metrics to quantitatively describe annoyance ratings obtained in psychoacoustic experiments. From Zwicker and Fastl (2007, p. 327):

> Basically, psychoacoustic annoyance depends on the loudness, the tone colour, and the temporal structure of sounds. The following relation between psychoacoustic annoyance, PA and the hearing sensations loudness, N, sharpness, S, fluctuation strength, F, and roughness, R can be given:
>
> $$PA \sim N(1 + \sqrt{[g_1(S)]^2 + [g_2(F, R)]^2}) \qquad (2.2)$$

Based on the results of psychoacoustic experiments, the authors expand on this theory to provide the following general model of psychoacoustic annoyance:

$$PA = N_5(1 + \sqrt{w_s^2 + w_{FR}^2}) \qquad (2.3)$$

with

- N_5 percentile loudness in sone
- $w_s = (\frac{S}{acum} - 1.75) * 0.25 \log{(\frac{N_5}{sone} + 10)}$ for $S > 1.75$ acum

describing the effects of sharpness S and

- $w_{FR} = \frac{2.18}{(N_5/sone)^{0.4}}(0.4 * \frac{F}{vacil} + 0.6 * \frac{R}{asper})$

describing the influence of fluctuation strength F and roughness R.

Demographic differences

Several studies have attempted to study the degree to which personal and demographic factors influence a person's soundscape perception. In some conceptions (Erfanian et al., 2019; Kou, Kwan, & Chai, 2020) these personal factors are classed as 'contextual' soundscape indicators - features which influence or, in a modelling context, be used as independent variables to predict the value of a soundscape descriptor. The personal factors help to create a personal soundscape interpretation model which is individual to each person.

In this way, a person's individual state-of-mind, ethnic identity, educational background, gender identity, etc. form a pseudo-deterministic framework through which the physical inputs from their environment are filtered. Clearly, many of these personal factors could never be measured and even those which are measurable will have wide ranges of legitimate effects, however estimating the degree and type of effect they may have can both help us better predict individual soundscape assessments and understand how group identities influence sound perception.

2.6. An Engineering Approach: The need for predictive soundscape models

The existing methods for soundscape assessment and measurement, such as those given in the ISO 12913 series, have been focussed primarily at determining the *status quo* of an environment. That is, they are able to determine how the space is *currently* perceived, but offer little insight into hypothetical environments. As such, they are less relevant for design purposes, where a key goal is to determine how a space *will be* perceived, not just how an existing space is perceived. The methods for assessment outlined in ISO/TS 12913-2:2018 (2018) and for analysis given in ISO/TS 12913-3:2019 (2019) are inherently limited to post hoc assessments of an existing space. Since they are focussed on surveying people on their experience of the environment, it stands that the space must already exist for people to be able to experience. How then would an urban planner, architect, or other designer estimate how a potential user would react to a space which is under design and not available to be assessed. One approach to this challenge has been through the use of virtual reality and auralization (Oberman et al., 2018). Toward this, and following from the combination of perceptual and objective data collection encouraged in ISO/TS 12913-2:2018 (2018), the natural push from the design perspective is towards 'predictive modelling' In this context, predictive modelling involves predicting how physical acoustic environments would likely be perceived or assessed by the users of the space.

The soundscape approach faces several challenges in practical applications which are unaddressed by current assessment methods, but which may be solved through the development of a predictive modelling framework. The first of these challenges is predicting how a change in an existing sound environment will be reflected in the soundscape. While it is possible in this scenario to measure the existing soundscape via questionnaire surveys, if a change is then introduced to the acoustic environment, it is so far impossible to say what the resulting soundscape change would be. This question relates strongly to the idea of soundscape interventions; where a particular noise pollution challenge is addressed by introducing more pleasant sounds (e.g. a water feature), following the soundscape principle of treating sound as a resource (Lavia, Dixon, Witchel, & Goldsmith, 2016). Predicting how much a particular intervention would improve the soundscape (or, indeed whether it would improve at all) is not yet possible with the retrospective methods available. This question is also addressed in Chapter 5 of this thesis which uses a predictive model to look at how the changes in the acoustic environment due to the COVID-19 lockdowns resulted in changes in the soundscapes of the spaces.

The ability to predict the likely soundscape assessment of a space is crucial to implementing the soundscape concept in practical design. Current methods of assessing soundscapes are generally limited to a post-hoc assessment of the existing environment, where users of the space in question are surveyed regarding their experience of the acoustic environment (Engel, Fiebig, Pfaffenbach, &

Fels, 2018; X. Zhang, Ba, Kang, & Meng, 2018). While this approach has proved useful in identifying the impacts of an existing environment, designers require the ability to predict how a change or proposed design will impact the soundscape of the space. To this end, a model that is built upon measurable or estimate-able quantities of the environment would represent a leap forward in the ability to design soundscapes.

2.6.1. Soundscape mapping

Similarly, a move towards modelling methods based on objective and/or measurable factors would facilitate the application of mapping in soundscape. While noise maps have become common in urban noise research and legislation (EEA, 2020; Gasco et al., 2020), they can be difficult to translate into a soundscape approach. The Environmental Noise Directive (END) (European Union, 2002), first implemented in 2002, is the main EU instrument to identify noise pollution impacts and track urban noise levels across the EU. Its goals were to determine the population's exposure to environmental noise, make information on environmental noise available to the public, and prevent and reduce environmental noise and its effects. In general, noise maps are based on modelled traffic flows, from which decibel levels are extrapolated and mapped, although interpolation and mobile measurement methods also been recently developed (Aumond, Can, et al., 2018). Alternatively, they can be produced using longterm SLMs or sensor networks. While these methods have significant utility for tracking increases in urban noise levels and are important for determining the health and societal impacts of noise on a large scale, their restricted focus on noise levels alone limits their scope and reduces the potential for identifying more nuanced health and psychological effects of urban sound.

Several studies have attempted to bring soundscape to urban noise mapping. The most notable of these attempts (Aletta & Kang, 2015; Aumond, Jacquesson, & Can, 2018; Hong & Jeon, 2017; Kang, Aletta, Margaritis, & Yang, 2018) bring new, more sophisticated methods for mapping urban sound (not just noise levels). For instance, all four present methods which map the relative level of various sound sources, producing maps of the spatial distribution of bird sounds, human voices, water sounds, etc. In Aletta and Kang (2015); Hong and Jeon (2017) the mapping relied on soundscape surveys conducted in public spaces, then used interpolation methods and basic relationships to the measured noise levels to generate a map of the perceived soundscape over the entire study space. Kang et al. (2018), after starting with survey responses attempted to create a prediction method which relied only on the audio recordings made in the space to create visual maps of the predicted soundscape perception (i.e. the PAQs 'pleasant', 'calm', 'eventful', 'annoying', 'chaotic', 'monotonous'). According to the authors, the prediction and mapping model would follow three steps: (1) sound sources recognition and profiling, (2) prediction of the soundscape's perceptual attributes, and (3) implementation of soundscape maps. Unfortunately, from the paper, it appears that the prediction model results were not actually used for the mapping and, again, the survey responses from 21 respondents were interpolated to create the soundscape map. Their results indicated to how a predictive model could have been slotted into a mapping use-case, but this was limited by (1) the relatively poor predictive performance for several of the attributes, (2) the inability to automatically recognise sound sources, and (3) a very limited dataset in terms of sample size and variety of locations.

While the connection is not made to perception, Aumond, Jacquesson, and Can (2018) focussed on creating sound maps which can reflect the pattern of sound source emergences over time within a city. By stochastically activating varying sound sources across their map, they could map the percentage of time when a sound source emerges from the overall complex sound environment. If a predictive soundscape model which incorporates sound source information can be developed, then the same procedure which led to their sound source emergence maps could also feed the soundscape model, resulting in a map of predicted perception over time.

The broader use-case and need for such soundscape models and maps was recently highlighted by Jiang et al. (2022), which opens the discussion for how the value and impact of soundscapes should be measured and what tools are needed to enable the valuation of policy interventions for soundscapes. In response to Question 5, the authors make the necessity of predictive soundscape models quite clear:

> **Question 5: What soundscape metrics and data will be needed?**
>
> Answer: Quantitative soundscape metrics that link subjective perceptions to objective acoustic and contextual factors will be needed, to enable monetisation while at the same time account for the perception-based nature of soundscape.
>
> ...
>
> Despite the varied requirements for soundscape metrics and data between and even within valuation methods, a standardised metric or set of metrics, such as dB in noise valuation [...] will allow comparison and integration of different studies and building compatible evidence bases. In this respect, standardised soundscape data collection, reporting and analysis methods have been developed and suggested (ISO, 2018; 2019), and the data outputs, such as the two soundscape dimensions based on affective quality ratings, have the potential to be used as standardised soundscape metrics for valuation purposes.
>
> Jiang et al. (2022)

It was a combination of research highlighting the health impact of noise (Ising & Kruppa, 2004), economic impacts of urban noise (Bristow, Wardman, & Chintakayala, 2014; Galilea & de Dios Ortúzar, 2005) and international-level noise monitoring and mapping efforts which led to the eye-opening statistics showing the true impact of urban noise with which I opened this thesis (EEA, 2020). By creating improved methods and tools which enable the same scale and type of evidence, we can allow research to investigate the full impact of urban sound beyond just its negative, noise-focussed impact, and do so on city- and country-level scales.

2.6.2. Soundscape prediction for design

Soundscape perception, while primarily driven by sound level, is mediated heavily by non-acoustic factors which interact with the sound level, spectral information, and temporal acoustic behaviour in complex ways. The soundscape is influenced by several levels of factors: the immediate and long-term acoustic environment, other environmental factors (e.g. temperature, air quality), the physical / visual characteristics of the space, the type of architectural space, and even cultural and country-level expectations. When approached in a predictive model context, the acoustic data must form the core components, but a coherent framework for describing how the influence of the acoustic factors is affected by the non-acoustic factors is required.

Simpler analyses have taken a fragmented approach, for instance where separate acoustic-factor models are built independently for each type of architectural space considered in the data set and, separately, statistical models are built to investigate another non-acoustic factor, e.g. visual greenness vs lack of greenness. In order to properly extract the influences of all of these levels of factors as well as to build a generalisable model which can be used in practice, this fragmented approach should be combined into a single multi-level model.

The first key step for this approach is the creation of a coherent, large-scale, multi-factor database of objective environmental measurements and subjective perceptual responses. My research makes use of in-person field questionnaires, long-term manned questionnaires, and multi-factor characterisation of the environment as part of the ERC-funded project Soundscape Indices (SSID) and in further collaboration with the DYNAMAP project to collect this database across a wide range of locations and soundscape types.

Part I.

Methodology

Introduction

The first key step toward developing a predictive soundscape model is the creation of a coherent, large-scale, multi-factor database of objective environmental measurements and subjective perceptual responses. My research makes use of in-person field questionnaires, long-term manned questionnaires, and multi-factor characterisation of the environment. Conducting urban soundscape studies on a scale large enough to form a machine learning dataset presents a unique challenge. The standardised methods of conducting soundscape surveys (ISO/TS 12913-2:2018, 2018) are labour-intensive, time-consuming, and provide limited information about the acoustical and environmental context. Towards this, we developed an in depth soundscape assessment protocol based on ISO/TS 12913-2:2018 (2018).

Due to the unique nature of this data collection method as a novel contribution submitted as its own journal paper, the methodology section has been split into two sections: Chapter 3 presents the data collection protocol as its own, stand-alone contribution, and Chapter 4 then presents the data analysis and modelling methods made use of throughout the rest of the thesis.

Chapter 5 makes use of the paired binaural recordings and soundscape assessments to model the relationship between (psycho)acoustic features and the soundscape descriptors., including information related to the context of the assessment location. Chapter 8 relies on the accompanying demographic information. The visual and environmental information and the Recording Stage information have not been used in the models presented in this thesis, but are considered necessary for future studies and for the ability to generalise to new locations. Once the data was collected via the protocol in Chapter 3, the questionnaire data was then analysed and modelled according to the methods presented in Chapter 4.

Chapter 3.

Development of The Soundscape Indices (SSID) Protocol and The International Soundscape Database (ISD)

This chapter presents the data collection method used to build the dataset used throughout this thesis. Published as *The Soundscape Indices (SSID) Protocol* (Mitchell et al., 2020), this protocol gives detailed instructions for carrying out soundscape assessments, how this data is organised, and descriptions of each data type collected. This protocol is presented both to document the data collection methods used throughout the SSID project and to provide comprehensive instructions for future researchers hoping to make use of the protocol. The protocol consists of two stages: (1) a Recording Stage to collect audio-visual recordings for further analysis and for use in laboratory experiments, and (2) a Questionnaire Stage to collect in-situ soundscape assessments via a questionnaire method paired with acoustic data collection. Key adjustments and improvements have been made to enable the collation of data gathered from research groups around the world. The data collected under this protocol has been compiled into the International Soundscape Database (Mitchell, Oberman, Aletta, Erfanian, et al., 2021).

3.1. Purpose

The SSID Protocol was designed to achieve two primary goals:

1. gather in-situ soundscape assessments from the public, which can be further analysed and utilised in designing a soundscape index;
2. conduct recordings needed to reproduce the audio-visual environment of a location in a laboratory setting for conducting controlled experiments on soundscape.

These two goals represent two levels of data required for developing a general soundscape model. The first enables large scale data collection, resulting in a database with thousands of perceptual responses and their corresponding quantitative data which can be statistically analysed on a large scale, or used for training in machine learning modelling. In-situ assessments also represent the most holistic assessment, ensuring all factors that influence the soundscape are present, including those which cannot be reproduced elsewhere.

However, there are questions that cannot be practically addressed in-situ, such as soundscape assessment of less- or un-populated areas, the influence of mismatched acoustic and visual cues, physiological and neural responses to soundscapes, and so on (Kogan, Turra, Arenas, & Hinalaf, 2017). Laboratory experiments with controlled environments are required to address these aspects. Toward the development of a coherent SSID, therefore, it is important that these two forms of data are collected simultaneously and with compatible methods, such that the results of the two approaches can be confidently combined and compared. In addition, since this protocol is intended to be used for the creation of a large-scale international database with additions carried out by several different and remote teams, it has been designed for efficiency, scalability, and information redundancy.

3.2. Protocol Design and Equipment

The first goal is achieved by conducting in-situ questionnaires using a slightly altered version of Method A (questionnaire) from Annex C of the ISO/TS 12913-2:2018 technical specification (ISO/TS 12913-2:2018, 2018) collected either via handheld tablets or paper copies of the questionnaire. Typically, a minimum of 100 responses are collected at each location during multiple 2-5 hr sessions over several days. During the survey sessions, acoustic data are collected via a stationary class 1 or class 2 Sound Level Meter (SLM) (as defined in IEC 61672-1:2013 (IEC 61672-1:2013, 2013)) running throughout the survey period and through binaural recordings taken next to each respondent. These acoustic and response data are linked through an indexing system so that features of the acoustic environment can be correlated with individual responses or with the overall assessment of the soundscape, as required by researchers.

The second goal is achieved by making First-Order (or higher) Ambisonic recordings simultaneously with 360°video which can be reproduced in a virtual reality environment. It has been shown that head-tracked binaural and multi-speaker ambisonic reproduction of recorded acoustic environments recorded in this way have high ecological validity (Davies, Bruce, & Murphy, 2014), particularly when paired with simultaneous head-tracked virtual reality video (De Coensel, Sun, & Botteldooren, 2017; Hong et al., 2018).

The on-site procedure to collect these data are separated into two stages, which will be outlined in detail in Section 3.4. The stage during which the audio-visual recordings are made for lab experiments is called the **Recording Stage**, while the stage during which questionnaires and environmental data are captured is called the **Questionnaire Stage**.

The procedure has been designed to include multiple levels of data and metadata redundancy, making it robust to on-site issues and human error. The most crucial aspect of the redundancy is ensuring the perceptual responses can be matched with the appropriate corresponding environmental and acoustic data even when some information is lost or forgotten.

3.2.1. Labelling and Data Organisation

In order to be able to identify all of the many data components of the Recording and Questionnaire Stages and to associate these with their various corresponding data, the following labelling system is suggested. This system is focussed on (1) relating all of the separate recordings and factors to specific questionnaire responses and (2) efficiency and consistency on site. A recent paper by Aumond et al. (2017) demonstrated the importance of addressing multiple levels of factors which influence perception, from individual-, to session-, to location-level. The successful pleasantness models built incorporating these information levels showed a marked improvement over the equivalent individual-level or location-level only models. The data organisation system proposed here was designed in order to maintain this important information, and the levels of information for the data collected on site are shown in Table 3.1.

At the top level is the **Location** information. This includes information about the location which does not change day-to-day, and generally characterises the architectural character of the space, or typical climate conditions for the area. As described in Section 3.2.2, each 'environmental unit' should be considered a new location. Therefore, if researchers want to investigate the differences in soundscape assessment in the middle of a small urban park and along the road next to the same park, these would be considered different locations since they would (typically) have different environmental factors and should be given different names. The name chosen should be concise, but it should be obvious what location is referred to.

The next level is information which is specific to each session, labelled with a **SessionID**. This SessionID should contain the name of the location and a numerical index which will increase with each repeated session at that location. The SessionID is associated with the data collected during the Recording Stage, and with the data which are continuous throughout the Questionnaire Stage, SLM, and ENV data. For easy automatic processing, correct spelling and consistency with the format is crucial so that data can be filtered according to the SessionID or the location, as is often necessary. In addition, for ease of automatic processing, it is recommended not to include spaces in the SessionID to avoid string splitting issues in analysis code.

Underneath each SessionID will be a set of **GroupID**s. One GroupID is assigned for *each group of participants*. This should correspond to a single binaural recording and a single 360°photo[1]. This

[1]Note that for the data used throughout this thesis – which was generally the first round of data collection –

will be used to (1) relate multiple surveys taken simultaneously and (2) link the recording and photo with the surveys. The GroupID is particularly crucial as it allows commonly missing data to be shared across multiple collection methods. For instance, occasionally paper questionnaires will be missing start and end time information. In this case, this information can be pulled directly from other questionnaires with the same GroupID. Where no questionnaires have the times, it is possible to extract an approximate start time from the binaural recordings or 360°photo and then estimate an average end time.

The GroupID should have the following format: [a set of letters representing the location name][the SessionID index number][an incrementing index for each group]. For example, for the second session at Regent's Park Japanese Garden, the location name is 'RegentsParkJapan', the GroupID letters might be 'RPJ'; the SessionID would be 'RegentsParkJapan2', so the GroupIDs for that session would start at '201'. Therefore, for example, the tenth group of participants for that session would be labelled 'RPJ210'. This format ensures that, if the location or SessionID are not recorded for a questionnaire, it is still obvious which session it belongs to.

Table 3.1.: Labelling system for on-site data collection. Regents Park Japanese Garden is used as an example location. Abbreviations as defined in Table 3.3 - SLM: Sound Level Meter (acoustical factors); ENV: Environmental factors; QUE: Questionnaires; PIC: Site pictures.

<table>
<tr><th>Level of information</th><th colspan="5">Example Label</th><th>Factors measured at this level</th></tr>
<tr><td>Location</td><td colspan="5">RegentsParkJapan</td><td>GPS, Architectural typology, visual openness, etc.</td></tr>
<tr><td>SessionID</td><td colspan="3">RegentsParkJapan1</td><td colspan="2">RegentsParkJapan2</td><td>SLM, session notes, ENV</td></tr>
<tr><td>GroupID</td><td>RPJ101</td><td>RPJ102</td><td>...</td><td>RPJ201</td><td>...</td><td>BIN, PIC</td></tr>
<tr><td>Questionnaire</td><td>1, 2, 3</td><td>4, 5</td><td>...</td><td>25, 26</td><td>...</td><td>QUE, Start & End time</td></tr>
</table>

3.2.2. Location and Measurement Point Selection

To select the appropriate measurement point, it should be ensured that the following contextual factors representative of the site are present in the spatial recording: openness, greenness, presence of landmarks, dominant use (walking, staying), and social presence (related to the dominant use). These are identified as objective metrics often used in urban and landscape research (Ewing & Clemente, 2013; Joglekar et al., 2020; Kaplan, 1989; Lynch, 1964; Quercia, O'Hare, & Cramer, 2014), possibly contributing to soundscape assessment (Aletta, Astolfi, et al., 2019; Pheasant, Fisher, Watts, Whitaker, & Horoshenkov, 2010). This relies on the researcher's opinion-driven assessment – it is advised to observe the location for a moment and then choose the point representative of the context and the first-person user experience. For instance, in a park, it would probably be near a bench in the

360°photos were not collected for each GroupID. This was an adjustment made to this protocol after the experience gained in this first round.

central area near the fountain; in a busy square, it would be a place where most people gather and have the best view of the landmark. While doing so, the placement too near the prominent vertical objects such as a statue, a wall, or a mast should be avoided as it might cause issues in later handling the visual data (3m is considered a safe distance from these features). Similar concerns are also true for the audio data and careful attention should be paid to avoid placing the recording equipment near extraneous noisy equipment or acoustic shadows. Further guidance on this is given in Point 4 of Section 3.4. It is important to avoid placing the recording equipment at a position where no users are expected (i.e. avoid putting the equipment in the middle of a flower bed or a grass area that nobody uses).

For the purposes of this protocol, a single location was considered to be an 'environmental unit' wherein the environmental factors are consistent and is typically perceived to constitute a single distinct area. The exact dimensions and delineation of the environmental unit will vary depending on the characteristics of the space, so it is ultimately up to the judgement of the researchers on site to select an appropriate measurement point to best capture the character of the environmental unit.

3.2.3. Equipment

The equipment listed in Table 3.2 is designed to facilitate both the audio-visual recording of the location and the collection of objective environmental factors, as given in Table 3.3. What equipment is brought on site should be adjusted depending on availability, needs of the researchers, and whether only one of the protocol stages will be carried out, or both. The equipment selected should be neutral and not noticeable. In general, this means dark or neutral colours as opposed to high-visibility colours and selecting compact equipment.

The use of class 1 or class 2 SLMs has been stipulated to maintain verifiable consistency and quality of data across all soundscape studies which make use of this protocol, as well as with data collected under various other environmental acoustics purposes. As the accuracy of acoustic information gathered at the site is the most vital in the discussion of soundscape indices, specific requirements have only been set out for the acoustic equipment. Class 1 is highly preferred, but consideration is made for cost and availability of equipment. It should be noted what standard of SLM was used in the data collection and appropriate consideration of the precision and tolerances of the equipment should be taken during the data analysis.

[2] The recommended acoustic data settings are given here in order of importance. In cases where researchers do not have access to a meter capable of spectral logging, L_{Aeq} logging should be prioritised over spectral analysis. During both stages, spectral data can typically be extracted from the audio recordings, but accurately tracking the sound level is crucial.

[3] The recommended environmental factors are given here in order of importance. more flexibility is allowed in selecting which factors to record and investigate (compared to the acoustic data) as it is still unclear how and to what extent environmental factors influence soundscape assessment. However, previous studies have indicated visual (i.e. lighting level) and temperature are significant factors (Jeon, Lee, Hong, & Cabrera, 2011).

Table 3.2.: Recommended equipment for implementing the SSID protocol. SLM: Sound Level Meter; AMB: Ambisonics; BIN: Binaural; QUE: Questionnaires.

Equipment	**Requirements**
Tripod stand	With add-on hooks/holders for AMB microphone, SLM, environmental meter(s) and 360 camera with suitable suspension for microphones
360 camera	4K, 5.1K or better resolution video, with suitable battery life and optional remote control
Spatial audio/Ambisonics (AMB) microphone system	Min. quality should be First-order Ambisonics (FOA) capability, however systems which achieve higher-order ambisonics would be preferred where available.
Multi-channel field recorder	Min. inputs to accomodate output from AMB microphone
Windshield(s) for AMB and SLM microphones	This can a single large windshield which can accomodate both microphones or separate windscreens for each microphone
Sound Level Meter (SLM)	Class 1 (preferred) or class 2 with omnidirectional pattern measurement microphone
Binaural recording system	Portable, worn by the researcher, or with a mounted binaural head
Sound calibrator for SLM, AMB microphones and binaural system	According to IEC 60942: 2017 Electroacoustics – Sound calibrators
Environmental meter(s)	See Table 3.3 for the recommended metrics
Tables and/or printed questionnaires	Internet connectivity or offline app to submit the questionnaires on site.

3.3. Techniques for Field Data Collection

There are several methods available for characterising the physical environment and collecting soundscape assessments. Here, I will address the techniques employed in this protocol and general best practice for each of them.

3.3.1. Questionnaire Surveys

As stated above, the questionnaire is primarily based on Method A of ISO/TS 12913-2:2018. This method begins with a set of questions relating to the sound environment which are assess on a 5-point Likert scale, coded from 1 to 5. A sample codebook to demonstrate the recommended variable naming and response coding is included in Appendix B.

The first section includes four questions relating to sound source identification, where the sound sources are divided into four categories: Traffic noise, Other noise, Sounds from human beings, and Natural sounds (labelled SSI01 through SSI04, respectively). These taxonomic categories of environmental sounds are based on the work done by Guastavino (2007) and A. L. Brown, Kang, and Gjestland (2011).

Table 3.3.: Table of recommended context and acoustic measurement factors.

Factor category	**Category code**	**Factors collected**	**Protocol stage**	**Measurement Duration**
Spatial Audio	AMB	Ambisonics A format 44.1 kHz, 24 bit resolution Min. first-order ambisonics (FOA)	Recording Stage	15 minutes
360 Video	VID	4K, 5.1K or better resolution video	Recording Stage	15 minutes
360 Photos	PIC	4K, 5.1K or better resolution still photos	Questionnaire Stage	Captured with each GroupID
Binaural Audio	BIN	Binaural audio recording. Note down the corresponding GroupID in recording metadata	Questionnaire Stage	30s of clean audio captured with each GroupID
Sound Level Meter Acoustic Data and Audio	SLM	Acoustic data[2]: (a) 1-second logging period (b) L_{Aeq}, L_{AFmax} , 1/3rd Octave Band L_{Aeq}, Octave Band L_{Aeq}, Full statistics, and Full Spectral Statistics. Recording: (a) .wav audio recordings (b) 44.1 kHz, 24 bit resolution	Both	Span or survey (approx. 3-4 hours)
Environmental Data[3]	ENV	10-second logging period: (a) Temperature (C) (b) Lighting Intensity, Lux (LI) (c) Air Quality (CO_2) (d) Relative Humidity (RH) (e) Dew Point (C)	Both	Span of survey (approx. 3-4 hours)
Questionnaires	QUE	SSID Questionnaire given in Appendix C Additional data: (a) GroupID for each group of participants (b) SessionID (c) Start and End time for each participant (if electronic) or each group (if paper) (d) GPS Location (if electronic)	Questionnaire Stage	On average, questionnaires last 5-10 minutes per GroupID

Next are the 8 scales which make up the circumplex model of the Swedish Soundscape Quality Protocol (SSQP) (Ö. Axelsson, Nilsson, & Berglund, 2012), describing the Perceived Affective Quality (PAQ). These are assessed on a 5-point Likert scale from 'Strongly Disagree (1)' to 'Strongly Agree (5)'. These are included as follows: Pleasant, Chaotic, Vibrant, Uneventful, Calm, Annoying, Eventful, and Monotonous (labelled PAQ01 through PAQ08, respectively).

Following this are five questions addressing the participant's overall assessment of the surrounding sound environment, addressing overall acoustic quality, the appropriateness of the sound environment to the location, perceived loudness, and how often the participant visits the place and how often they would like to visit again (labelled SSS01 through SSS05, respectively).

The fourth section comprises the WHO-5 Well-being Index (WHO-5), asking how the participants have been feeling over the last two weeks, such as 'I have felt calm and relaxed'. The WHO-5 index is constructed to constitute an integrated scale in which the items add up related information about the level of the individual's general psychological well-being (Hall, Krahn, Horner-Johnson, & Lamb, 2011; Topp, Østergaard, Søndergaard, & Bech, 2015). This information can provide additional insight into how exposure to pleasant or annoying soundscapes may impact psychological well-being as was investigated by Aletta, Oberman, et al. (2019) or, alternatively, how a person's current psychological status may influence their perception of the sound environment as recently investigated by Erfanian et al. (2021). Each of the five WHO-5 questions (labelled WHO01 to WHO05) are assessed on a 6-point scale coded from 0 to 5.

The final section of the participant-facing questionnaire comprises five questions on the participant's demographic information (age [AGE00], gender [GEN00], occupational status [OCC00], education level [EDU00], ethnicity [ETH00], and local vs. tourist [MISC03]) and a free response for the participant to provide any additional comments they would like to make on the sound environment [MISC01]. It is important to note that the section on ethnicity, and to a lesser extent education level, will need to be adjusted to ensure the available responses are appropriate for the location where the survey is being conducted.

At the end of the questionnaire are a set of spaces available for the researcher conducting the survey to fill out, adding additional information about the observed behaviour of the participants, indexing and labelling metadata, and space for any additional notes. More information and guidance on this information is included below.

This questionnaire is intended to collect a consistent core set of perceptual responses and information about the participant, with space to add additional questions as required by specific research goals. Some examples of this which have been implemented by the various research groups are specific questions calling attention to water sounds and features, the perception of visual features, and an open response for identifying the dominant sound source. Given the proper labelling and coding, these additional questions can be fully integrated into the overall dataset, allowing the researchers the freedom to pursue their own research interests while maintaining consistency and compatibility with the overall database.

General notes for conducting the questionnaires:

- The core questionnaire is reported in Appendix A. The labels and corresponding scales are also reported. Ideally, the form should be submitted and filled on a tablet via a survey app (e.g. REDCap, Qualtrics, KoBoToolbox, or similar) so that data can then be easily downloaded in an .xlsx or .csv file. Using paper forms is also acceptable; however, researchers on site will need to take more careful note of information such as the time of response and the information will need to be manually input after the session is completed. If using an electronic version, the system should be set up to record the start and end times and GPS coordinates for each survey.

- If using an electronic version, be sure to have enough tablets with internet connectivity (if required by the survey system) and sufficient battery life; if using the paper version, be sure to print enough copies. Even if using the electronic version, it is recommended to also print a number of paper versions as a backup or if a large group agrees to participate at once.
- Regardless of the translation of the items, it is important that the label (e.g. SSI01) is kept, as well as the size and direction of the scales (1-5, etc.) to maintain data consistency.

3.3.2. Contextual and Environmental Factor Data Collection

During each survey, the equipment listed in Section 3.2.3 is set up to capture the contextual and environmental data for the location. Table 3.3 lists the factors to be collected and at what stage they should be collected.

Spatial Audio-Visual Recordings

In order to capture the acoustic and visual information in the space for replication in a laboratory setting, 360°video and AMB audio are recorded to be used in Virtual Reality (VR) playback. The goal of this is two-fold: first, to enable researchers to document and replicate the in-situ environment of the space as it was during a questionnaire survey session for lab experiments and, second, to capture environments in which performing a questionnaire survey is not feasible.

Typically, questionnaire surveys are carried out over a period of several days at the same location. The goal of these multiple sessions is to capture as many questionnaire responses as needed (100 for a particular soundscape is typically recommended (Engel et al., 2018)), which, in the experience of the author is prohibitively difficult to achieve in a single session in most locations. It is recommended that the repeated sessions are conducted under similar circumstances and environmental conditions. As such, it is not entirely necessary to repeat the spatial recordings each time a questionnaire survey is conducted. Instead, it may be useful to use the spatial recording as a chance to gain a different perspective on the space under investigation. For instance, if the questionnaires are conducted in the middle of a large urban park, the first session could collect a spatial recording within the environmental unit of the questionnaire site, but the subsequent returns to the site could collect spatial recordings in a different environmental unit, say, along a road bounding the park, or in a space in the park which does not typically have many people. This enables the simultaneous expansion of the questionnaire database and the gathering of additional environments to investigate in a laboratory setting.

General notes for spatial recordings:

- The audio-video recordings can be done before or after the questionnaire survey.
- The purpose of the audio-video recordings is to capture representative recordings which can be reproduced in a laboratory setting. During the first time at a location, the focus should be on capturing the environment as experienced by the respondents to the questionnaires at that location. Therefore, the recordings should be performed in nearly the same spot, with similar lighting and environmental conditions. For further survey sessions, provided the conditions are similar, other recordings could be taken which provide additional perspectives around the space for reproducing in the lab.
- These recordings can be performed entirely separately from the questionnaire survey, ,if desired. Reasons for doing this may be (but are not limited to): location is not populated, making questionnaires impossible; specific locations or conditions are required for a lab experiment; time limitations require many sites in an area to be captured and in-situ questionnaires could not be completed in time.
- The 360°video will take a significant amount of storage space. Researchers should ensure that there is ample free space on the camera SD cards prior to going out on site. If conducting multiple surveys away from their home institute (i.e. in another city), teams are recommended to bring a large external hard drive so that videos can be offloaded after each session.

Reference Recordings

A soundscape index, or any investigation of the impact of the physical environment on the soundscape, requires consistent and accurate measurement of the environment, most importantly calibrated measurement and recording of the acoustic environment. For this protocol, this has been achieved through the use of separate calibrated binaural recordings and measurements made with a calibrated SLM.

3.4. Procedure

Fig. 3.1 shows the whole process of the on-site soundscape protocol. The relevant equipment in each row should be operating when the row is coloured in, such that when multiple rows are shaded this means that multiple pieces of equipment should be running during that time period. The following section prepares step-by-step instructions for conducting the in-situ surveys, including the Recording Stage and Questionnaire Stage. Fig. 3.2 shows an example of the recommended equipment setup.

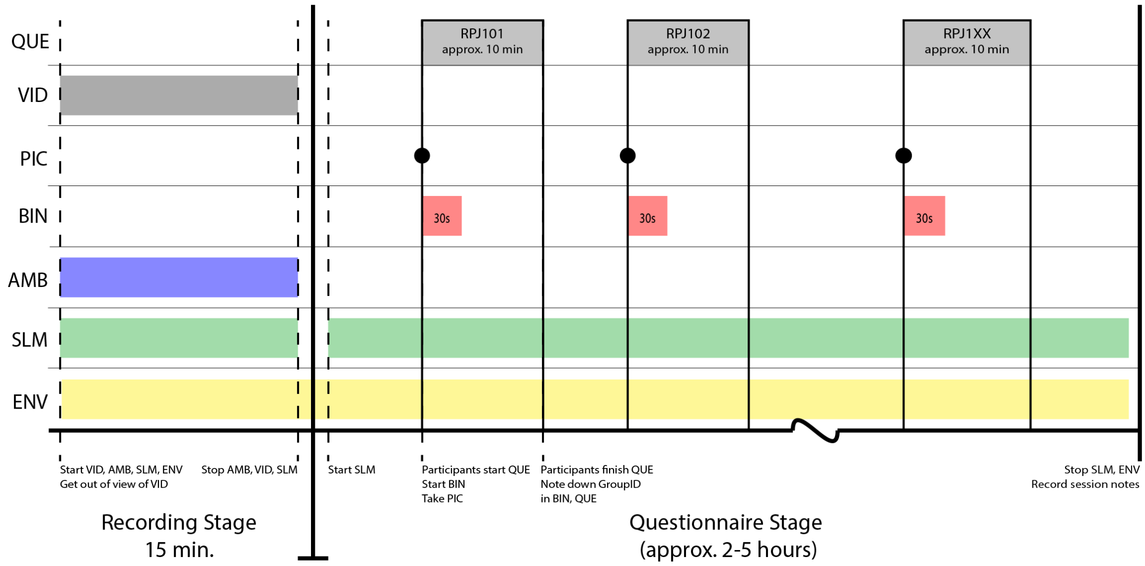


Figure 3.1.: Timeline of the on-site soundscape protocol. RegentsParkJapan (RPJ) is used as an example. Abbreviations as defined in Table 3.3 – QUE: Questionnaires; VID: 360°video; PIC: Site pictures; BIN: Binaural Recording; AMB: Ambisonic recording; SLM: Sound Level Meter (acoustical factors); ENV: Environmental factors.

Setup & Calibration The equipment should be assembled, checked, and calibrated prior to arriving at the measurement location. Calibrate the equipment according to the manufacturer's instructions. All SLMs should have built-in methods to calibrate using a standard 94 dB 1 kHz tone calibrator. If a similar method is available for the ambisonic microphone, this should be used. If a built-in method is not available, but a calibrator can be fitted to the microphone capsules, then the ambisonic microphone should be calibrated by recording the 1 kHz signal through the system for each microphone capsule after the gain settings have been finalised on site (see below). If it is not possible to calibrate the ambisonic microphone, then the levels recorded will need to be compared to the levels taken simultaneously with the SLM. This is why it is crucial to have an appropriate quality, calibrated SLM included within the same setup as the AMB recordings.

3.4.1. Assembling the Equipment

1. Set up the equipment by prioritising the position of the 360°camera and position the lens at the average eye level 160–180 cm, as shown in Fig. 3.2.

   It is advisable to test the setup for video stitching issues and reconfigure if needed (e.g. the

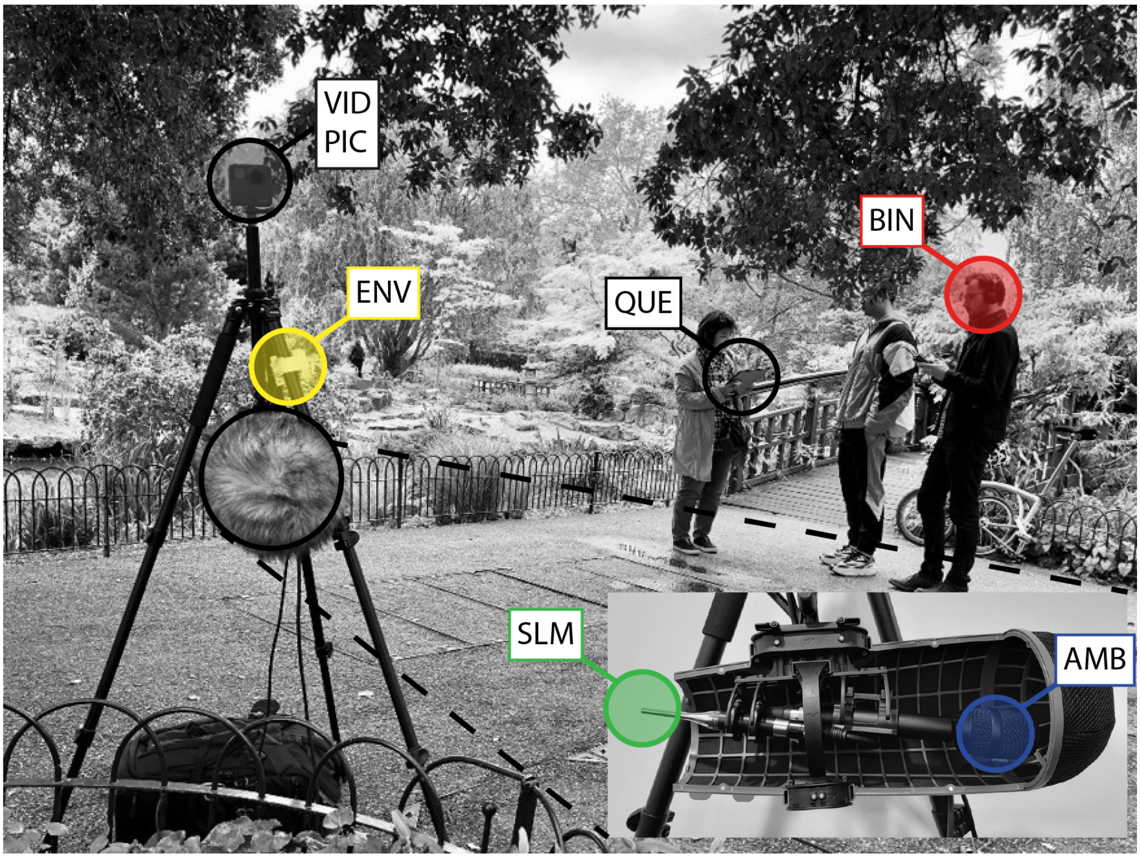


Figure 3.2.: Photo of a full survey carried out in a park in London during the Questionnaire Stage. To the left is the equipment (colour-coded to match Fig. 3.1), with the ambisonic microphone and SLM microphone in the windscreen, with the 360°camera on top of the tripod and to the right are one researcher interacting with the participant while the second researcher conducts the binaural recording. The body of the SLM and the multichannel recorder are stored in a bag under the tripod which can contain all of the pieces of equipment for easy transport.

equipment will be partially visible in the raw video recording, so you need to test if the chosen setup allows for efficient erasing/hiding/patching of the exposed parts in the post-processing). Companies selling 360°cameras usually offer free software for basic editing and previewing. It is advisable to position the camera as the highest item in the set to avoid the need for editing both the sky and the ground.

2. Carefully position the AMB microphone so its axes are aligned with the axes of the 360°camera; the microphone's front (usually marked by the logo) and the camera's front should be looking in the same direction. Many AMB microphones allow them to be oriented vertically or horizontally (end-fire), this should be noted and adjusted in the relevant software settings.

This is essential for informed post-processing. It is advisable to position the capsules of the AMB microphone and the capsule of the SLM as near to each other as possible, without introducing scattering effects. It can usually be done within the same windshield unit, but it is not essential to do so and depends on the available clamps and stands.

3. The gain settings for the four ambisonic audio channels should be set to the same level. In some devices (such as the MixPre10), this can be set by locking the channel gain settings to a single channel. Many devices also offer ambisonic plugins which simplify these settings and automatically link the gain settings – these should be used where available.
4. Set the SLM to log sound levels and simultaneously record .wav audio. The recommended logging settings are given in Table 3.2. The SLM should be mounted and positioned according to standard guidance for environmental noise measurements, like that given in Section 9 of ISO 1996-1:2016 (2016) or Section 5 of ANSI/ASA S12.9-2013/Part 1 (2013). Generally, the microphone should be a minimum of 1.2m above the ground and a minimum of 1m from any vertical reflecting surfaces.
5. Attach the environmental meter(s) to the tripod. Care should be taken when positioning the environmental monitor. Most units will include guidance on their use from the manufacturer – these should be followed where available. Some general items to keep in mind include not accidentally covering air quality sensor holes, not positioning light sensors in the shade of the other equipment, and not positioning temperature sensors in direct sunlight unless this is how they are intended to be positioned.

3.4.2. Recording Stage

The following section prepares step-by-step instructions for conducting the Recording Stage of the on-site protocol, as shown in Fig. 3.1.

1. Double check all settings and file save locations on the recording equipment.
2. Adjust gain settings to ensure there is no clipping. Good practice is to listen for what is expected to be the loudest sound event during the recording period (e.g. sirens) and set the gain such that the level is comfortably under clipping during this event.
3. Start recording on all devices, including the ambisonic microphone, 360°camera, SLM, and environmental meter.
4. Stand at the front of the camera/ambisonic microphone and clap. The clap can help synchronise the audio with the video, if necessary, and ensuring you are standing in line with the front of the 360°video can help with lining up the directionality of the two, if necessary.

5. Retreat out of view of the camera, blending into the surrounding crowd, or otherwise make sure not to be obvious to someone watching the video.
6. Record at least 5 min of consistent and representative audio and video. It is recommended to record for 15 min to give the best chance of being able to extract a solid 5 min of useful video and audio.
7. Stop recording on all devices and ensure all files are saved properly.

3.4.3. Questionnaire Stage

The following section prepares step-by-step instructions for conducting the in-situ questionnaires and their accompanying reference recordings as part of the Questionnaire Stage. Typically these are performed during the same working session as the Recording Stage, using the same set of equipment. The selection of an appropriate location and setup of the equipment should follow the guidance given in Section 3.2.2, while making sure the location selected is representative of where the respondents will be stopped. Wherever possible, the equipment should be assembled and located so as not to draw the attention of the respondents and particularly to avoid influencing their perception of the space.

1. Double check all settings and file save locations on the recording equipment. If starting this stage immediately after the Recording Stage, make sure to rename or advance the index of the filenames for the SLM and ENV meters.
2. Start recording on the SLM and ENV meter (or leave running from preceding Recording Stage). These will continue running until the end of the Questionnaire Stage.
3. Gather the tablets and/or paper questionnaires and prepare to approach potential participants.
4. Approach participants and ask if they would be willing to take part in a research study. If the participants are in a group, they can participate at the same time, but should each fill out a separate questionnaire. When approaching participants, you should identify yourself as a researcher or student researching urban sound. We advise avoiding phrases such as 'noise', 'noise pollution', 'noise disturbance' or other terms which carry a negative connotation. In general, explanations and answers to questions should strive to be as neutral as possible regarding the nature of the soundscape.
5. Once the participant has consented to participate, hand them the questionnaire or tablet and provide them with basic instructions for answering the questionnaire. Emphasise that they should be responding and assessing the current sound environment, in the current place. Note that this is a common misunderstanding – many participants assume the questionnaire is focussed on the sound environment at their home, or in the city in general. Where a

mix of tablets and paper questionnaires are being used, each group should have at least one participant using a tablet such that start and end times and precise GPS coordinates can be pulled from the accompanying electronic questionnaire. While one researcher is interacting with the participants, the second should arrange the equipment for taking the BIN recordings and 360°photo (PIC).

6. Once the participant has started answering the questionnaire, start recording the BIN audio. If the participants are in a group and all are taking the survey at the same time, only one binaural recording is needed for the whole group. The researcher conducting the recording should strive to keep their head as stationary as possible and to avoid making any extraneous noise.

   Make sure that at least 30s of consistent audio is recorded while the participant is filling in the questionnaire. This should not include talking either from the researcher or the participant. If talking or other intrusive (non-representative) sound occurs, extend the recording period to end up with a solid 30s of good audio. The goal is to capture the sound environment which the participant is exposed to while filling out their questionnaire, but to exclude sounds which the participant is not likely considering as part of their assessment. Most commonly, this would be the researcher talking, or the participant themselves talking. Any other sounds which the participant was 'naturally' exposed to should be included.

   When taking the BIN recording, attempt to orient the head (artificial or researcher wearing a headset) in the same direction as the participants. This is not crucial as it is often impossible to achieve, but it is preferable. Be careful not to move the head during the recording.

7. Note the GroupID in the metadata for the BIN recording, or make a manual note of the BIN recording file name and the GroupID separately.

8. Take one 360°photo (PIC) with the camera to capture the general setting. This can also be done at regular intervals during the survey session.

9. When the participant has finished filling in the questionnaire, thank them for their participation and fill in the additional research questions at the end of the questionnaire. These help to both track the data collected and to document the conditions on site. The most important of these are:

   - (For paper version) Start and End time. If a Start time was not noted, at minimum, the End Time must be recorded and an average survey duration can be subtracted to estimate the Start Time.
   - GroupID
   - SessionID

10. Repeat steps 4–9 for the remainder of the session, incrementing the GroupID by one with each new group of participants. If there are more than two researchers on site, the additional researchers can stop new groups of participants simultaneously. The researcher operating the BIN equipment can then shift between the groups once they have finished the 30s recording. This researcher should also have the responsibility of keeping track of the GroupID numbers for each group. Experience has shown this is possible up to about three groups at a time, with four researchers on site.
11. Once the session is finished, stop the equipment and ensure all files are saved properly.
12. After each session, make note of the character of the site and the environmental conditions during the survey. This might include, but is not limited to:
    - Site typology and intended use (e.g. urban park, transit station, urban square, etc.)
    - Weather
    - Crowdedness (i.e. how many people are present in the space)
    - Dominant sound sources and any key soundmarks
    - Visual character (e.g. amount of greenness, enclosed vs. open, etc.)

3.5. Lessons from International Data Collection

As this protocol has already been implemented by several research groups across four countries, it has undergone a rigorous testing and development process. Throughout this process, adjustments have been made which resulted in the final protocol presented here. However, no process is perfect or applicable in all situations. As such, after consultation with the research groups involved, I have compiled the most common feedback and guidance to keep in mind when implementing this protocol.

3.5.1. Sampling

The research groups were instructed to try keeping the structure of respondents well-balanced. This often led to longer times and larger sample sizes required as most comments from five research groups addressed age and type of location as the most influential factors for participant sampling. However, while some reported higher response rates from younger (student) members of the public, others reported higher response rates in case of older highly-educated people. A common observation was that public parks are the locations with the highest response rates, most likely due to a high number of people taking part in activities that allow enough time to take part in a survey. The type of space was also reflected in the sense of privacy. In locations that were more public, people in groups were more likely to take part in the survey, while in the more private locations it was the opposite. Amongst other comments, whether a participant was a tourist or a local also had an influence on the response rate. Tourists seemed more likely to participate in the survey.

Several groups reported extremely hot and cold weather to negatively affect the response rates. One research group, which conducted the survey also in a residential area, distinguished privacy/ownership of the survey site as a major factor.

3.5.2. Data Collection

A group of three researchers seems to be the minimum number needed to conduct the survey, as observed by the partner research groups. A group of nine researchers on site proved to be the most effective number. The time needed to complete the survey varied greatly depending on the location.

Although the questions are written in a manner that emphasises the focus on the actual acoustic environment perceived at the moment, additional care should be made to ensure a proper understanding of that concept while approaching the participants. Researchers' comments are invaluable here to keep track of the outliers if a researcher feels similar issues or other factors (i.e. wearing headphones) lead to collecting invalid or misleading data.

3.5.3. Equipment

Some partners had previous experience in soundscape research, but for all this was the first study that featured surveying a large number of public participants around a single measurement point. All the research groups found it very important to delegate one researcher or technician to care exclusively about the equipment and the quality of the recordings.

The intention of the recording stage is to record a first-person experience most representative of the location. Therefore, the researchers are instructed to 'make themselves invisible' in the recording. However, at some locations, various research groups decided to put out a sign asking members of the public not to touch or come near the measurement point as they experienced passers-by touching the windshield out of curiosity.

The equipment setup has been designed to be as compact and unobtrusive as possible so as to limit any intrusion on the participant's experience of the space. From our experience, most participants do not end up with the equipment within their field of view during the questionnaire and often do not notice the presence of the stationary equipment. In some locations, this is not possible and participants may comment on its presence; however, over the thousands of surveys collected, only a small number of respondents have commented on the equipment as noticeably impacting their experience.

3.5.4. Translation

Regarding the onsite soundscape survey, the translation of the questionnaires (and in particular the perceptual adjectives used for the soundscape appraisal) is a key point to consider when using the protocol in regions where English is not the local language. Indeed, while the ISO/TS 12913-2:2018 document from which the soundscape-related questions of this protocol are derived aims at providing standardised scales, it does not provide official translations in languages other than English. Some perceptual constructs are difficult to render in different languages and people might assign different meanings to them (e.g. (Almagro Pastor, Vida Manzano, & García Quesada, 2019; Jeon et al., 2018; Nagahata, 2019; Tarlao, Steele, Fernandez, & Guastavino, 2016)). For this reason, in the soundscape research community, there is a growing interest in testing and validating reliable translation of the ISO soundscape adjectives (Aletta, Oberman, Axelsson, et al., 2020), which will hopefully lead to a wide-spread use of this soundscape tool. It is expected that these validated translations could simply be substituted for their English counterparts in this protocol, when they become available.

3.6. The Database

The implementation of this protocol within the SSID project has resulted in the creation of a large dataset of urban soundscape assessments. This database forms the basis for the majority of the work reported in this thesis. Fig. 3.3 shows how the various portions of the database have been used throughout this thesis for constructing and using the models.

On 8 November, 2022, the first publicly available version, v0.2.0, was published on zenodo.org to coincide with the publication of Mitchell, Oberman, Aletta, Kachlicka, et al. (2021) (i.e. the study reported in Chapter 5). As of 26 May, 2022, the published version of the database is v0.2.4. It is the intention that this dataset be added to and augmented with new locations, cities, and contexts in the future, with contributions from the SSID team at UCL, our collaborators at other institutions around the world, and via submissions from similar research groups. If a soundscape assessment is collected according to the SSID Protocol, it can be integrated with the rest of the database to form a large, cohesive, and ever-growing database of soundscape assessments. As v0 was intended as a pre-release dataset, it included only the 13 locations in London and Venice used throughout this thesis and the results of the psychoacoustic analysis of the binaural recordings, but not the recordings themselves. In the near future, the intention is for this to be updated to V1.0, which will include more cities in Europe and China with thousands more surveys and will include the binaural recordings themselves. By following a consistent protocol designed to follow the ISO 12913 data collection standards, The International Soundscape Database (ISD) (Mitchell, Oberman, Aletta, Erfanian, et al., 2021) will hopefully form a new standard database of soundscape assessments and documentation to be used by soundscape researchers and practitioners around the world.

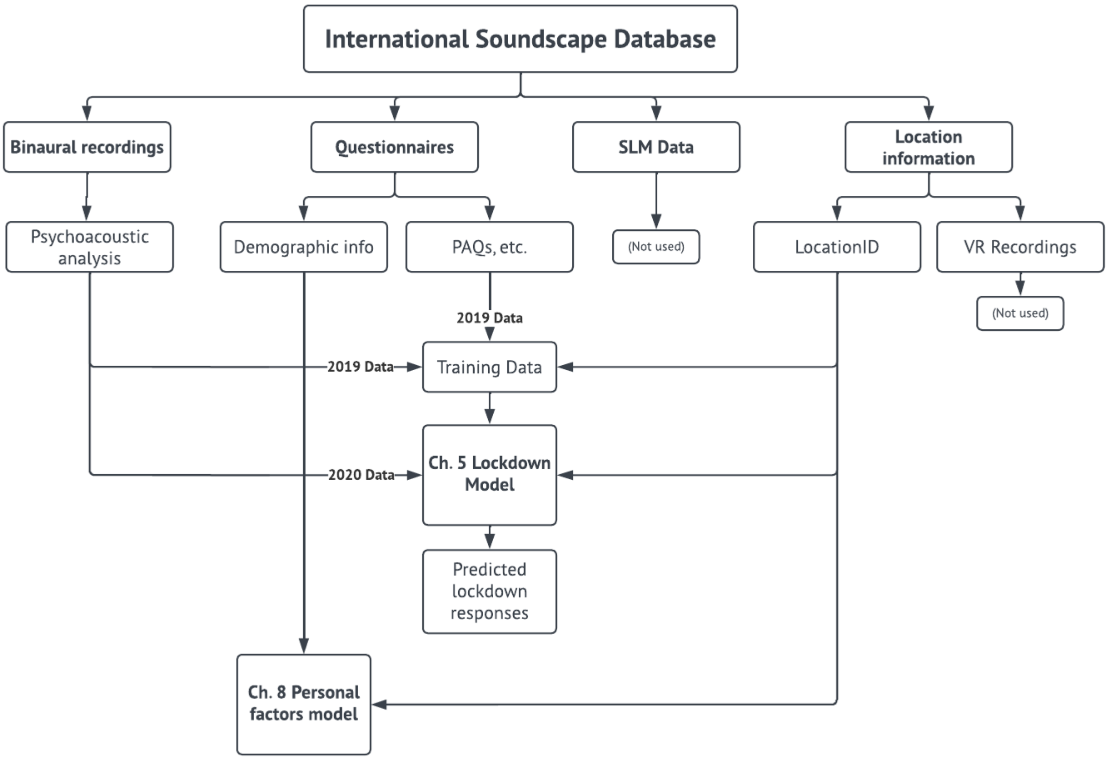


Figure 3.3.: Diagram of how the data from the ISD has been used in the models throughout this thesis.

Chapter 4.

Data Analysis and Modelling

The full protocol developed for this thesis is outlined in Chapter 3. The development and presentation of this protocol involved a substantial development and testing phase, and represents a novel advancement in soundscape survey methodology. Therefore it was submitted and published as a peer-reviewed journal article in MDPI Applied Sciences as Mitchell et al. (2020) and is presented as a stand-alone chapter within this thesis. This section therefore presents those methods not associated with the data collection procedure, i.e. the analysis and statistical methods used.

This chapter begins by reviewing the methods presented in ISO/TS 12913-2:2018 (2018) for analysing soundscape assessment data which makes use of the soundscape circumplex, such as that collected with the SSID Protocol. Next, a review of machine learning prediction and regression methods is presented and an in-depth review of the Multi-Level Model (MLM) method used throughout this thesis is given.

4.1. The current ISO 12913 framework

Although different methods are proposed for data collection in ISO12913 Part 2 (ISO/TS 12913-2:2018, 2018), in the context of this thesis I focus on the questionnaire-based soundscape assessment (Method A), because it is underpinned by a theoretical relationship among the items of the questionnaire that compose it. The core of this questionnaire is the 8 perceptual attributes (PA) originally derived in Ö. Axelsson et al. (2010): pleasant, vibrant (or exciting), eventful, chaotic, annoying, monotonous, uneventful, and calm. In the questionnaire procedure, these PAs are assessed independently of each other, however they are conceptually considered to form a two-dimensional circumplex with *Pleasantness* and *Eventfulness* on the x- and y-axis, respectively, where all regions of the space are equally likely to accommodate a given soundscape assessment (Aletta et al., 2016). In Ö. Axelsson et al. (2010), a third primary dimension, *Familiarity* was also found, however this only accounted for 8% of the variance and is typically disregarded as part of the standard circumplex. As will be made clear throughout, the circumplex model has several aspects which make it useful for representing the soundscape perception of a space as a whole.

4.1.1. Coordinate transformation into the two primary dimensions

To facilitate the analysis of the PA responses, the Likert scale responses are coded from 1 (Strongly disagree) to 5 (Strongly agree) as ordinal variables. In order to reduce the 8 PA values into a pair of coordinates which can be plotted on the Pleasant-Eventful axes, Part 3 of ISO 12913 (ISO/TS 12913-3:2019, 2019) provides a trigonometric transformation, based on the 45°-relationship between the diagonal axes and the pleasant and eventful axes. This transformation projects the coded values from the individual PAs down onto the primary Pleasantness and Eventfulness dimensions, then adds them together to form a single coordinate pair. In theory, this coordinate pair then encapsulates information from all 8 PA dimensions onto a more easily understandable and analysable two dimensions. The ISO coordinates are thus calculated by:

$$\begin{aligned} ISOPleasant = {} & [(pleasant - annoying) + \cos 45° * (calm - chaotic) \\ & + \cos 45° * (vibrant - monotonous)] * 1/(4 + \sqrt{32}) \end{aligned} \tag{4.1}$$

$$\begin{aligned} ISOEventful = {} & [(eventful - uneventful) + \cos 45° * (chaotic - calm) \\ & + \cos 45° * (vibrant - monotonous)] * 1/(4 + \sqrt{32}) \end{aligned} \tag{4.2}$$

where the PAs are arranged around the circumplex as shown in Fig. 4.1. The $\cos 45°$ term operates to project the diagonal terms down onto the x and y axes, and the $1/(4+\sqrt{32})$ scales the resulting coordinates to the range (-1, 1). The result of this transformation is demonstrated in Fig. 4.1. This treatment of the 8 PAs makes several assumptions and inferences about the relationships between the dimensions. As stated in the standard (ISO/TS 12913-3:2019, 2019, p. 5):

> According to the two-dimensional model, vibrant soundscapes are both pleasant and eventful, chaotic soundscapes are both eventful and unpleasant, monotonous soundscapes are both unpleasant and uneventful, and finally calm soundscapes are both uneventful and pleasant.

4.1.2. Interpreting the Soundscape Circumplex

The circumplex model of soundscape, as originally defined by Ö. Axelsson et al. (2010), is commonly understood to be a two-dimensional space (its main orthogonal components being annoying-pleasant and uneventful-eventful) where all regions of the space are equally likely to accommodate a given soundscape assessment (Aletta et al., 2016). For instance, in theory, an extremely vibrant soundscape

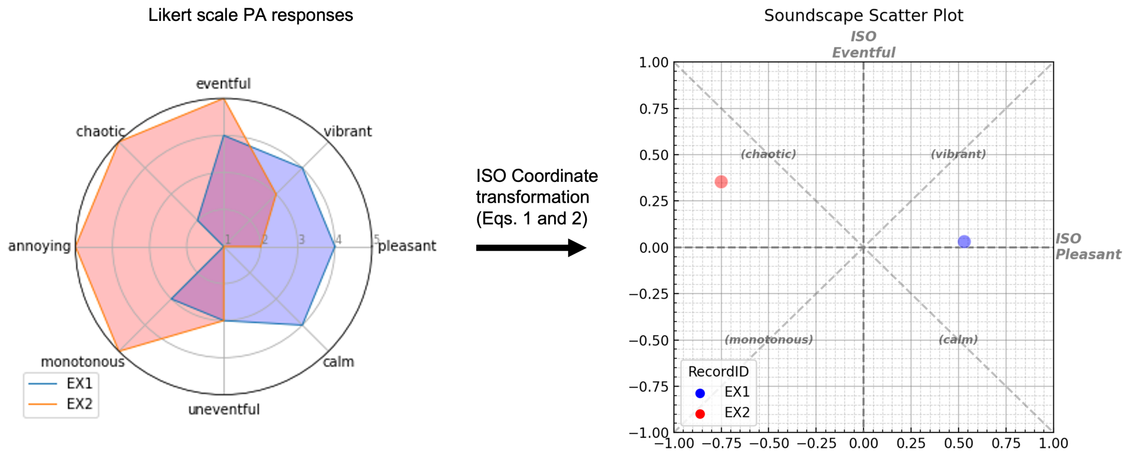


Figure 4.1.: Example of representations of two soundscape assessments. Left: Radar plot of two example perceptual attribute (PA) ratings on the Likert scales (1 to 5). Right: Scatter plot of the same assessments on the soundscape circumplex, transformed according to ISO 12913 Part 3.

(e.g., with a score of 1) should be as likely to occur as an extremely annoying one, as well as one neutral on all dimensions (e.g., with a score of 0). However, a recent work by Lionello et al. (Lionello, Aletta, Mitchell, & Kang, 2021) incidentally highlighted a possible issue with the process for representing soundscape assessments with the current ISO protocols. More specifically, when considering big numbers, soundscape assessments seem to have a bivariate normal distribution around the origin of the circumplex model. This would imply that not the whole space of the model is equally accessible to any given soundscape[1]. Studies in the field show that data collection campaigns rarely return extreme values for soundscape dimensions (Mancini, Mascolo, Graziuso, & Guarnaccia, 2021) and so far the general interpretation has been that some soundscapes (e.g., extremely monotonous) may simply be difficult to find and detect with people in urban contexts (K. Sun et al., 2019).

4.1.3. Usefulness for predictive modelling

This trigonometric projection method enables us to transform the 8 Likert scale PAQ values into a pair of coordinate values. This transformation has a few beneficial effects for applying standard modelling techniques to soundscape data. First, it simplifies and reduces the target problem; rather than needing to model eight separate responses, we are now focussed on only two. Second, it transforms the data from ordinal responses on a 1 to 5 scale into continuous values between -1 and +1. While it is clearly possible to model ordinal outputs through classification, the methods are often less familiar and more complicated than dealing with a more standard regression problem. For those

[1] **??** offers a more in-depth critical critique and discussion of the specific consequences of this projection method in more detail than is appropriate to include here.

outside of machine learning (i.e. designers, engineers, etc.) regression methods, especially linear regression, are already familiar and interpretable while methods of classification and ordinal modelling are typically less familiar. By applying the ISO projection to each individual's soundscape assessment, we generate a vector of output values which can be matched up to physical data measured for each individual. This creates the sort of input-output pair vector necessary for supervised regression learning.

4.2. Machine Learning and Regression Techniques

Machine learning approaches are typically divided into three broad categories: supervised, unsupervised, and reinforcement learning. In supervised learning, the training data consists of input-output pairs and learns a model which can map from the inputs to the outputs. In unsupervised learning, no corresponding output data is available to the training model, thus it learns patterns in the input without feedback. Reinforcement learning does not necessarily begin with training data, instead the learning agent is given a series of reinforcements in the form of rewards and punishments (Stuart Russell, 2021). Reinforcement learning will not be used in this thesis. Unsupervised learning has been applied to a limited degree to the acoustic data collected in several sound environments which will be expanded upon later.

The majority of this thesis is therefore focussed on creating a supervised learning model wherein the input data are the result of measurements and the output data are the perceptual assessments of the soundscapes. In the context of this thesis, there are two primary types of supervised machine learning models - regression and classification. Regression is applied when the output is a continuous number (e.g. temperature) whereas classification is used when the output is a finite set of values.

4.2.1. Multi-level Linear Regression

Multi-level regression modelling is a technique commonly used in fields such as psychology (Quené & van den Bergh, 2004; Volpert-Esmond, Page-Gould, & Bartholow, 2021), for applied prediction models (Frees & Kim, 2006; Gelman, 2006), and for a small number of previous soundscape studies (Aumond et al., 2017). MLMs are particularly useful when data is organised at one or more levels or groups. As noted in Table 3.1, the International Soundscape Database (ISD) forms a hierarchical structure with several groups nested within each other: Questionnaires within GroupIDs within SessionIDs within Locations, making a MLM especially well-suited. The concept behind MLMs can be built up starting from simple linear regression, as given by:

$$y_i = \alpha + \beta x_i + \epsilon \tag{4.3}$$

For a classical multiple linear regression, we expand the k coefficients out as so:

$$y_i = \alpha + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \epsilon_i \tag{4.4}$$

where:

- *i* indexes each unit, the smallest items of measurement. This is often a measurement per individual within a group or, in the ISD this can be for each recording.
- $y_i = (y_1, \ldots, y_n)$ the modelled outcome for each unit *i*;
- $k = 1, \ldots, K$ denotes each of the multiple predictors;
- β_k is the slope coefficient for the k^{th} predictor;

Their primary feature of an MLM is the ability to have coefficients and intercepts which are allowed to vary depending on the group (Gelman, 2006). This can take three forms:

1. Random intercepts
2. Random slopes
3. Random intercept and random slopes

In a random intercept structure, the intercept for each input feature is allowed to vary according to the second level. This structure assumes that the linear relationship between each input feature and the output is consistent across the second level groups, but that the zero point (the intercept) is different. This is expressed mathematically (Gelman, 2006) as:

$$y_i = \alpha_{j[i]} + \beta x + \epsilon i \tag{4.5}$$

where groups are indexed by $j =, \ldots, J$, and *j[i]* is the i^{th} individual *i* in group *j*. This results in a vector of length *J* containing one intercept result per group.

In the context of auditory perception studies, this is most appropriate for repeated measures experimental designs (as will be demonstrated in Chapter 7). A repeated measures study is one in which all participants experience all levels of the independent variables and provide some response in terms of the output variable. In other words, each participant constitutes a group in the model and they respond to all of the input variables (Kristjansson, Kircher, & Webb, 2007). In this case, the MLM framework is used to account for starting differences between respondents; populations are expected to demonstrate similar behaviours in response to a given stimulus, but may have differing initial starting points, i.e. different intercepts for each analysed feature. The MLM framework using a varying intercept for each participant allows this initial difference among individuals to be accounted for while also highlighting the overall relationship between e.g. acoustic features and annoyance ratings for a given sound.

Random slope structures take the opposite assumption; each level shares the same intercept, while the coefficients for each feature are allowed to vary depending on the group. This assumes that different groups will have a different relationship between the input features and the output, but that these relationships may begin at a different threshold. This structure appears to be less commonly used than random intercept models. This can be mathematically described as:

$$y_i = \alpha + \beta_{j[i]} x_i + \epsilon_i \tag{4.6}$$

where $\beta_{j[i]}$ is the slope coefficient for individual *i* in group *j*.

With multiple predictors, we write:

$$y_i = X_i B + \epsilon_i \tag{4.7}$$

where B is a matrix of slope coefficients of size $J \times K$ with one coefficient for each predictor (*k*) for each group (*j*):

$$B_{J \times K} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1K} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{J1} & \beta_{J2} & \cdots & \beta_{JK} \end{bmatrix}$$

Finally, a random-slope, random-intercept model allows both the slope and intercept of the coefficients to vary for each group in the second level:

$$y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i \tag{4.8}$$

Wilkinson-Rogers notation The analysis package used for constructing these models is `lme4` (Bates, Mächler, Bolker, & Walker, 2015) in the R statistical language (R Core Team, 2018). This package makes use of a style of writing MLMs called Wilkinson-Rogers notation (Wilkinson & Rogers, 1973). Wilkinson-Rogers notation provides a way to specify MLMs without the need to specify coefficient values in a straightforward and easily readable way. For these models, the important operators to be familiar with are: $\sim$ indicates that a model regresses upon the dependent variable; $+$ sums the model terms;· indicates interaction terms; `(var|grp)` specifies a grouping variable or random effects term for an MLM[2]. Table 4.1 shows some example models written in Wilkinson-Rogers notation:

The structure inherent within the ISD means that this approach is particularly appropriate. In order to further demonstrate the structure and use of an MLM, I'll further describe it in terms

[2]It appears this notation was an extension of Wilkinson-Rogers introduced into the `nlme` R package by Pinheiro and Bates (1997).

Table 4.1.: Mathematical and Wilkinson-Rogers notation for several example models to demonstrate how to translate from one to the other. `var` is used to denote the independent variables to demonstrate that the variable name (e.g. *loudness*) can be used directly in the notation.

Description	Model	Wilkinson-Rogers Notation
Two predictors	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$	`y ~ var1 + var2`
Two predictors and no intercept	$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$	`y ~ var1 + var2 - 1`
Two predictors with interaction	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \epsilon_i$	`y ~ var1 · var2`
Random intercept	$y_i = \alpha_{j[i]} + \beta_{x_i} + \epsilon_i$	`y ~ var1 + (1|grp)`
Random slope	$y_i = \alpha + \beta_{j[i]x_i} + \epsilon_i$	`y ~ 1 + (var1|grp)`
Random slope Random intercept	$y_i = \alpha_{j[i]} + \beta_{j[i]x_i} + \epsilon_i$	`y ~ var1 + (var1|grp)`

of the ISD data, where the most obvious second level for this MLM is the location (a categorical variable defined by the LocationID). To demonstrate this, we'll specify a simple random-intercept random-slope MLM which has the loudness (N_5) and sharpness (S) as the two input variables, the recordings (taken per group in the ISD and indexed by the GroupID) make up the units at level 1, and the LocationID is level 2.

$$\texttt{ISOpl} \sim \texttt{loudness + sharpness + (loudness + sharpness | LocationID)} \tag{4.9}$$

which would also be written as:

$$ISOPl_i = \alpha_{j[i]} + \beta_{j[i]} N_{5_i} + \beta_{j[i]} S_i + \epsilon_i \tag{4.10}$$

where $\alpha_{j[i]}$ is the mean ISOPleasant score for LocationID *j* where recording *i* was taken and $\beta_{j[i]} N_5$ and $\beta_{j[i]} S$ are the loudness and sharpness slope coefficients for location *j*.

Mixed effects: fixed and random effects In some applications and fields, it is more common to refer to a MLM as a Linear Mixed-Effects Regression (LMER), however the two are simply different ways of speaking about the same mathematical concept (Gelman, 2006; Pinheiro & Bates, 2000). Although I typically refer to MLMs, I sometimes find it helpful to use the random-effects/fixed-effects terminology and Chapter 8 primarily refers to the model as an Linear Mixed-Effects Regression (LMER) since that work was targeted at a psychology audience, where mixed-effects is the more common term. In a random-intercept random-slope model, where certain features can be specified only at the unit level and others vary at the group level, the features at the unit level are considered *fixed*: i.e. the relationship between independent and dependent variable remains fixed across all groups. In an LMER the second level of effects are then termed the *random effects*. This is useful – particularly in a random-intercept model – where the effects in this second level are somewhat unexplainable. Consider a repeated measure study using a random-intercept model: the grouping factor is the participant who has been exposed to multiple inputs. The second level operates to account for that participants consistent difference from the other participants, but that difference can be considered to be random for each participant, hence it is a *random* effect when the goal is to elucidate the effects which are consistently seen across the sample. However, this use of the word random to refer to the second-level effects is not always useful, as highlighted by Gelman (2006, pg. 2). When we expect the grouping factor to have some explainable impact on the relationship between the independent and dependent variables (e.g. the context of the location has a complex, but non-random effect on the soundscape perception), then it is not appropriate to refer to it as 'random'.

In all, this concept helps to highlight the specific benefit of a MLM approach. Conceptually, it would be possible to achieve a similar goal by treating the grouped data as 'fully un-pooled' and to fit a separate linear regression model for each group. In the case of Eq. (4.9), we would treat the data from each location as a separate dataset and train a model on them each independently. This would reveal, within each location, the relationship between the loudness and sharpness of a sound and its perceived pleasantness. However, it would ignore the fact that there is a general, *fixed*, relationship between loudness and pleasantness, regardless of the effects of the location.Alternatively, we could treat the data as 'fully-pooled', creating a linear model with the form given in Eq. (4.4) and only considers the fixed relationship across the entire dataset. A MLM, which treats the data as 'partially-pooled', and includes effects which are both fixed and can vary according to the location, enables us to investigate both the degree to which loudness generally impacts on pleasantness *and* how this relationship changes according to the context of the location.

Structural Equation Models

Structural Equation Modelling (SEM) is a statistical approach to testing hypotheses about relationships between variables and latent features which has been used by several studies in soundscape (Hong & Jeon, 2015; Tarlao et al., 2020). SEM is a flexible approach in that it can include one or more independent or dependent variables and the variables can be continuous or discrete, factors or measured. As a collection of statistical methods, SEM is focussed on causal inference and is fundamentally built on a combination of path diagrams and regression modelling techniques (Ullman & Bentler, 2012). Most frequently in an SEM, the researcher constructs a path diagram which expresses the hypothesised relationships between the variables of interest. These path diagrams can be quite complex, including covariance relationships, latent variables, residuals, and distinctions between factors and measured variables. The model is then fit to the data through a selected estimation method (most typically Maximum likelihood (ML) as in regression modelling) and evaluated. The model may then be further reduced or modified if necessary. SEM also allows for multilevel modelling where separate models are developed for different levels of a nested hierarchy. This is conceptually equivalent to the MLM or LMER discussed throughout this thesis. Several studies in the soundscape literature have made use of SEM.

Part II.

An Initial Application of Predictive Soundscape Modelling

Introduction

The ability to predict the likely soundscape assessment of urban spaces has great potential for bringing a soundscape focus to design, but it also unlocks new opportunities to track how urban spaces change. In order to demonstrate this potential, and to give a practical example of the principles I have outlined in this thesis, this part presents a novel application of predictive soundscape modelling used to track how the COVID-19 lockdowns impacted urban soundscapes. The model presented in Chapter 5[3] represents a first-step approach which will be further built upon and developed throughout the following chapters in Part III.

[3]The study presented in Chapter 5 has been published as (Mitchell, Oberman, Aletta, Kachlicka, et al., 2021) in a special issue on COVID-19 Pandemic Acoustic Effects.

Chapter 5.

Investigating Urban Soundscapes of the COVID-19 Lockdown: A predictive soundscape modelling approach

The engineering soundscape approach developed in this thesis strives to make certain applications of soundscape possible. As highlighted earlier, one of these applications is the ability to assess the soundscape perception in situations where surveys are impractical as well as to be able to track changes in the soundscape of public spaces. This chapter therefore presents a unique application of predictive modelling towards these goals as well as the development of the predictive modelling method itself. First, I will review the identified impacts of the COVID-19 lockdowns on the sound environment in cities around the world, investigates these changes in detail within London and Venice, and finally presents the impact of these changes on the likely perception of the soundscape of public spaces through the results of the predictive model.

5.1. Review of the impacts of COVID-19

The global emergency caused by COVID-19 in early 2020 required national lockdown measures across the world, primarily targeting human activity. In the United Kingdom, construction and transport were allowed to continue, but a decrease in activity was observed (Hadjidemetriou, Sasidharan, Kouyialis, & Parlikad, 2020). In other countries, such as Italy, the restrictions were more severe and even included limiting people's movement to a certain radius from their place of residence (Ren, 2020). The explorations in environmental acoustics of lockdown conditions across the world have revealed various degrees of impact on the acoustic environment, with researchers reporting reductions in noise levels affecting the population at the scale of urban agglomerations such as the Ruhr Area in Germany (Hornberg et al., 2021) and conurbations in the south of France (Munoz et al., 2020). Impacts have also been reported at a scale of a multimillion city such as Madrid (Asensio, Pavón, & de Arcas, 2020) or Barcelona (Bonet-Solà, Martínez-Suquía, Alsina-Pagès, & Bergadà, 2021) as well as at a more local, city-centre or even public space-scale in cities such as Stockholm (Rumpler, Venkataraman, & Göransson, 2021), London (Aletta, Oberman, Mitchell, Tong, & Kang, 2020), Girona (R. M. Alsina-Pagès, Bergadà, & Martínez-Suquía, 2021), or Granada (Vida Manzano et al., 2021). In general, these studies have demonstrated a decrease in urban noise levels and indicated a difference in the amount of decrease depending on the type of space investigated (e.g. parks, urban squares, etc.) and the type of human activity characteristic for the space, with higher reductions in places typically associated with human sounds and activities such as shopping and tourism.

5.1.1. The lockdown measures in London and Venice

The lockdown measures implemented in the UK to contain the spread of the SARS-CoV-2 virus were not particularly strict if compared with other countries. In general, lockdown measures involved 'stay at home' recommendations, social distancing, stopping non-essential commercial activities, banning public gatherings, limiting traffic mobility and alike. Specifically, the UK Government passed the Health Protection (Coronavirus, Restrictions) (England) Regulations 2020, which were put into place at 1:00 pm on 26th March 2020 (Public Health England, 2020). Under these restrictions, the public were only allowed to leave their homes once per day for essential activities and exercise. All offices and shops selling non-essential goods were told to close, gatherings of more than two people in public were banned, and individuals were advised to only interact with members of their own household. These restrictions were set to be reviewed by the Secretary of State at least once every 21 days and would continue indefinitely until they were no longer necessary to prevent the spread of infection in England. In practice the lockdown continued through the spring of 2020 and was first partially eased on the 1st of June, with school children in England returning to school, but the broader lockdown continued throughout the summer (Tong, Aletta, Mitchell, Oberman, & Kang, 2021). During this period of lockdown, noise complaints increased by 48% compared to the same period during the preceding year, with an immediate uptick seen once lockdown measures were implemented. Fig. 5.1 shows the timeline of these restrictions and the accompanying noise complaints received by 22 boroughs in London (Tong et al., 2021).

5.1.2. Perceptual changes

Those studies were mostly focussed around the L_{Aeq}, as well as a standardization approach to reporting subsequent changes in soundscape proposed by Asensio, Aumond, et al. (2020). They were

[1]Figure originally created for Tong et al. (2021).

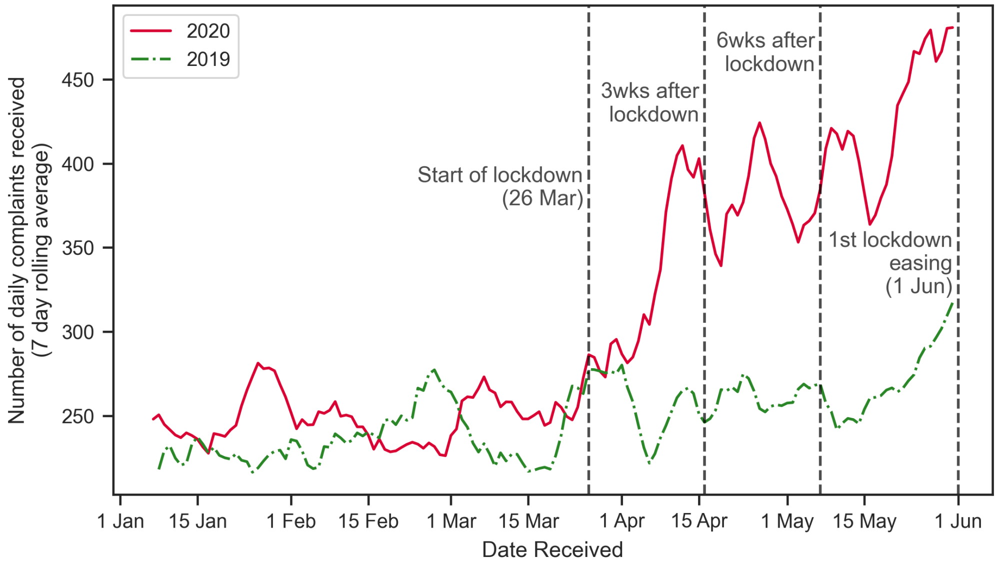


Figure 5.1.: Time series of the number of noise complaints received in the first half year of 2019 and 2020. A 7-day rolling average window is applied to account for weekly patterns in noise complaint reporting[1].

not able to reveal the perceptual impact of such conditions in public spaces as well because of: 1) the lack of subjective data for the exact or comparable locations in previous years; and 2) the lack of participants present in public spaces during the lockdown, hence the inability to collect soundscape data *in-situ*. Attempts have been made to bridge this gap by using social networks to source subjective data, but this resulted in a focus on indoor conditions following the shift in the citizens' behaviour, i.e. spending more time indoors (Bartalucci, Bellomini, Luzzi, Pulella, & Torelli, 2021; Lee & Jeong, 2021). Garrido-Cumbrera et al. (2021) relied on an online survey deployed in England, Ireland, and Spain to explore the perceived change in natural environments in particular. They observed a consistent increase in the perceived presence of natural sounds across all major cities and rural areas respectively in these three countries. A very similar trend was observed in Argentina, also based on an online questionnaire without a listening task (Maggi et al., 2021).

Munoz et al. (2020) combined noise measurements with an online questionnaire deployed to residents, some of which were residing in the areas covered by the noise monitoring network available. The participants were asked to recall how their lockdown area sounded before and during the first lockdown in 2020 and to describe the perceived change. They observed a consistent reduction in levels, followed by the perceived reduction of transport sounds (air and road) and an increase of natural sounds, while the resulting environment was described as pleasant, calm, and peaceful. By combining field recordings and focus groups, Sakagami (2020) and Lenzi, Sádaba, and Lindborg

(2021) observed changes in the sound source composition and the affective quality of soundscape in a residential area in Kobe, Japan and a public space in Getxa, Spain, respectively, during the different stages of the lockdown period. Following the easing of lockdown measures, a decrease in animal and traffic sounds was observed in Kobe, while an increase in eventfulness, loudness, and presence of human sound sources, followed by a decrease in pleasantness, was shown in Getxa.

This metric- and , by necessity, indoor-focussed approach left the following research questions unanswered:

1. How would people have perceived these outdoor urban spaces as a result of this change in acoustic environment? (RQ1)
2. Would these sound level reductions result in improvements to the soundscape of the spaces? (RQ2)
3. What are the key features needed for a soundscape prediction model based on comprehensive acoustic on site measurements to be used for assessing locations with low social presence or in situations where conducting surveys is impractical (RQ3)?

The 1st research question (RQ1), addressing the perceptual effect of the change in urban soundscape induced by the lockdowns, can be further broken down into the following questions:

- How was the sound source composition influenced by the change?
- How would the affective response to the acoustic environment in lockdowns change?
- Could this demonstrate the effect of human activities on the perception of an acoustic environment in general?

5.2. Materials and Methods

This study was conducted via initial onsite data collection campaigns in Central London and Venice in 2019 before the outbreak of COVID-19 as part of the SSID project (Mitchell et al., 2020) and in 2020 during the strictest part of the lockdowns (Aletta, Oberman, Mitchell, et al., 2020), including objective acoustic data (2019 and 2020) and subjective responses (2019 only).

Using both 2019 and 2020 binaural recordings, an online listening experiment was conducted to provide an understanding about the change in sound source composition. The 2019 onsite questionnaire data were used to define the dominant sound source at each location as a starting point for interpreting the soundscape change. A predictive model was developed to reveal the change in the perceived pleasantness and eventfulness using objective acoustic data and location to predict subjective responses. Although the initial (2019) dataset contains additional locations (specifically, in Spain, the Netherlands, and China), due to the nature of this study as a reaction to the strict movement and activity restrictions, the sites which could be included in the lockdown (2020) measurement campaigns were limited to locations where staff and equipment had access and where recordings could be undertaken during the spring of 2020.

The sites were selected to provide a mixture of sizes and uses, varying in typology ranging from paved squares to small and large parks to waterside spaces across both cities. Throughout the text they are indexed via a LocationID based on the location's name (e.g. CamdenTown, SanMarco), while a more in-depth overview of each is given in Appendix C. London is taken as an example of a large, typically noisy city while the Venice sample provides a unique look at spaces with typically very high human activity levels and no road traffic activity. In particular, the 2019 Venice surveys were taken to coincide with the yearly Carnevale festival in order to capture its distinct soundscape.

ISO/TS 12913-2:2018 (2018) was consulted for reporting on soundscape data. A detailed description of the 2019 survey campaigns is featured through the paper and in the public database. This study was approved by departmental UCL IEDE Ethics Committee on 17th July 2018 for onsite data collection and on the 2nd of June 2020 for the online listening experiment and is conducted in adherence to the ethical requirements of the Declaration of Helsinki (World Medical Association, 2013).

5.2.1. Onsite data: Questionnaires, binaural measurements, and recordings

The initial onsite data collection featured both questionnaire data collected from the general public and acoustic measurements, conducted across thirteen urban locations (in London $N = 11$, in Venice $N = 2$) between the 28th of February and the 21st of June 2019, with additional sessions in July and October 2019. Although the total survey period in 2019 extended over several seasons, the surveys at any individual location did not extend over seasons with different occupancy patterns. A total of 1,318 questionnaire responses were collected from the general population across the measurement points during 1 – 3 hour-long campaigns in both cities in 2019, accompanied by 693 approximately 30-second long 24-bit 44.1 kHz binaural recordings. After data cleaning, each of the 13 locations was characterised by between 14 to 80 recordings and between 24 to 147 questionnaire responses. Mean age of the participants was 33.9, with a standard deviation of 14.57 (45% male, 53.8% female, 0.4% non-conforming, 0.9% prefer-not-to-say).

Although recent results from both Tarlao et al. (2020) and Erfanian et al. (2021) indicate the important influence of personal and demographic factors – in particular age and gender – on soundscape perception, these factors were not included as potential features in the modelling process[2]. Given the nature of this study as addressing a scenario when people could not be surveyed, no ad-

[2] See Chapter 8 for the results of (Erfanian et al., 2021) and an in depth discussion of how these factors should or could be integrated into the predictive modelling process.

ditional demographic information is available in the lockdown case to be fed into the model and is therefore not useful to include for the development and application of this specific predictive model. This information is reported throughout the study simply to provide further context to the data collection.

The subsequent measurement campaign in 2020 mimicked the binaural recording strategy applied in the initial campaign and was performed between the 6th and the 25th of April 2020 in both cities, this time excluding the questionnaire. An additional 571 binaural recordings were collected on site in 2020.

Data collection

The 2019 data collection was performed across all the locations using the protocol based on the Method A of the ISO/TS 12913-2:2018 (ISO/TS 12913-2:2018, 2018), as described in Chapter 3, collected either via handheld tablets or paper copies of the questionnaire. The full questionnaire and data collection procedure are given in Mitchell et al. (2020), however the key parts used for this study are those addressing sound source dominance and Perceived Affective Quality (PAQ).

In order to simplify the results and allow for modelling the responses as continuous values, the 8 PAQs undergo a trigonometric projection to reduce them onto the two primary dimensions of pleasant and eventful, according to the procedure outlined in Part 3 of the ISO 12913 series (ISO/TS 12913-3:2019, 2019). In order to distinguish the projected values from the Likert-scale PAQ responses, the projected values will be referred to as ISOPleasant and ISOEventful and can be considered to form an x-y coordinate point (x = ISOPleasant, y = ISOEventful) as explained in detail in Chapter 4 (Lionello et al., 2021).

Data cleaning

The cleaning of the samples was conducted using the ArtemiS SUITE 11. I discarded or cropped whole recordings, or its parts affected by wind gusts or containing noises and speech generated by the recording operator by accident or for the purpose of explaining the questionnaire to a participant. This resulted in 1,291 binaural recordings which were then processed further, as described in Section 5.2.1. Psychoacoustic analyses are shown in the publicly available database.

In order to maintain data quality and exclude cases where respondents either clearly did not understand the PAQ adjectives or intentionally misrepresented their answers, surveys for which the same response was given for every PAQ (e.g. 'Strongly agree' to all 8 attributes) were excluded prior to calculating the ISO projected values. This is justified as no reasonable respondent who understood the questions would answer that they 'strongly agree' that a soundscape is pleasant and annoying, calm and chaotic, etc. Cases where respondents answered 'Neutral' to all PAQs are not excluded in this way, as a neutral response to all attributes is not necessarily contradictory. In addition, surveys were discarded as incomplete if more than 50% of the PAQ and sound source questions were not completed. The site characterisation per ISO/TS 12913-2:2018 (2018) is available in Appendix C, featuring the address, overall psychoacoustic characteristics of the location, typical use of each location, and pictures taken during the survey sessions.

Psychoacoustic analyses

The binaural recordings were analysed in ArtemiS SUITE 11 to calculate the suite of 11 acoustic and psychoacoustic features given in Table 5.1 to be used as initial predictors.

Table 5.1.: Psychoacoustic features considered for inclusion in the predictive models. All metrics are calculated for the full length of the recording (30s). As recommended by ISO 532-1:2017 (2017) and ISO/TS 12913-2:2018 (2018), the 5th percentile of Loudness is used rather than the average.

Feature	**Symbol**	**Unit**	**Calculation Method**
Loudness (fifth percentile)	N_5	sones	ISO 532-1:2017 (2017)
Sharpness	S	acum	ISO 532-1:2017 (2017)
Roughness	R	asper	ECMA-418-2 (2020)
Impulsiveness	I	iu	ECMA-418-2 (2020)
Fluctuation Strength	FS	vacil	ECMA-418-2 (2020)
Tonality	T	tuHMS	Sottek (2016)
Psychoacoustic Annoyance	PA	–	Zwicker and Fastl (2007)
L_{Aeq}	L_{Aeq}	dB	IEC 61672-1:2013 (2013)
$L_{A10} - L_{A90}$	$L_{A10} - L_{A90}$	dB	ISO 1996-1:2016 (2016)
$L_{Ceq} - L_{Aeq}$	$L_{Ceq} - L_{Aeq}$	dB	ISO 1996-1:2016 (2016)
Relative Approach	RA	cPA	Sottek and Genuit (2005)

The (psycho)acoustic predictors investigated were selected in order to describe many aspects of the recorded sound – in particular, the goal was to move beyond a focus on sound level, which currently dominates the existing literature on the acoustic effects of lockdowns noted in 2. In all, they are expected to reflect the sound level (L_{Aeq}), perceived sound level (N_5), spectral content (S, $L_{Ceq} - L_{Aeq}$, T), temporal character or predictability (I, FS, RA), and overall annoyance (PA). These metrics have been proposed as indicators to predict perceptual constructs of the soundscape (Aletta, Axelsson, & Kang, 2017; Aletta et al., 2016) and have shown promise when combined together to form a more comprehensive model applied to real-world sounds (Orga et al., 2021). The maximum value from the left and right channels of the binaural recording are used, as suggested in ISO/TS 12913-3:2019 (ISO/TS 12913-3:2019, 2019).

Table 5.2.: Pearson correlation coefficients between candidate acoustic features and ISOPleasant and ISOEventful across all 13 locations. Only statistically significant ($p < 0.01$) coefficients are shown.

Parameter	ISOPl	ISOEv	PA	N_5	S	R	I	FS	T	L_{Aeq}	$L_{A10} - L_{A90}$	$L_{Ceq} - L_{Aeq}$
ISOPleasant												
ISOEventful	-0.24											
PA	-0.28	0.24										
N_5	-0.37	0.33	0.94									
S			0.71	0.56								
R	-0.36	0.32	0.63	0.74	0.11							
I			-0.10		-0.37	0.24						
FS	-0.11	0.14	0.37	0.43		0.46	0.55					
T	-0.21	0.30	0.58	0.63	0.12	0.54	0.16	0.52				
L_{Aeq}	-0.34	0.37	0.84	0.93	0.56	0.72	-0.09	0.37	0.57			
$L_{A10} - L_{A90}$	-0.18	0.15	0.21	0.33	-0.20	0.31	0.36	0.44	0.40	0.23		
$L_{Ceq} - L_{Aeq}$		-0.20	-0.49	-0.49	-0.54	-0.31		-0.27	-0.28	-0.61	-0.22	
RA	-0.34	0.31	0.60	0.74	0.18	0.71	0.31	0.63	0.58	0.73	0.23	-0.14

Table 5.2 shows the Pearson correlation coefficient between each of the candidate acoustic features and the outcome pleasantness and eventfulness. For ISOPleasant ($ISOPl$), we can see three tiers of correlations:

1. The more highly correlated tier ($|r| > 0.28$) consists of RA, L_{Aeq}, R, N_5, and PA
2. The low correlation tier consists of $L_{A10} - L_{A90}$, T, and I
3. $L_{Ceq} - L_{Aeq}$, I, and S show no correlation

For ISOEventful ($ISOEv$), these tiers are:

1. The more highly correlated tier ($|r| > 0.30$) consists of RA, L_{Aeq}, T, R, and N_5
2. The low correlation tier consists of $L_{Ceq} - L_{Aeq}$, $L_{A10} - L_{A90}$, FS, and PA
3. I and S show no correlation

Among the correlations for the psychoacoustic metrics considered for inclusion as input features, we can see several highly inter-correlated features. As expected, PA, L_{Aeq}, and N_5 are highly correlated, meaning that careful consideration is paid to these features to ensure they do not contribute to multicollinearity in the final model.

5.2.2. Modelling

Two linear multi-level models (MLM) were computed to predict: 1) ISOPleasant, and 2) ISOEventful. The inherent grouped structure of the ISD necessitates a modelling and analysis approach which considers the differing relationships between the objective acoustic features and the soundscape's perceived affective quality ratings across the various locations and contexts. The individual-level of the models is made up of the acoustic features calculated from the binaural recordings made during each respondent's survey period, while the group-level includes the categorical 'LocationID' variable indicating the location in which the survey was taken, acting as a non-auditory contextual factor.

A separate backwards-step feature selection was performed for each of the outcome models in order to identify the minimal feature set to be used for predicting each outcome. In this feature selection process, an initial model containing all of the candidate features was fit. Each feature was then removed from the model one at a time, then the best-performing model is selected and the procedure continues step-wise until no improvement is seen by removing more features. This process is carried out first on the location-level features (including the potential to remove all features including LocationID, resulting in a 'flat' or standard multivariate linear regression model), then on the individual-level features. The performance criterion used for this process was the Akaike

Information Criterion (AIC) (Akaike, 1974). To check for multicollinearity among the selected features, the Variance Inflation Factor (VIF) was calculated and a threshold of $VIF < 5$ was set. Any features which remained after the backwards step-wise selection and which exceeded this threshold were investigated and removed if they were highly collinear with the other features.

All of the input features are numeric values, in the units described above. Before conducting feature selection, the input features are z-scaled to enable proper comparison of their effect sizes. After the feature selection, the scaled coefficients are used in the text when reporting the final fitted models to facilitate discussion and comparison between the features. The unscaled model coefficients are reported in Appendix D to enable the models to be applied to new data. In order to properly assess the predictive performance of the model, an 80/20 train-test split with a balanced shuffle across LocationIDs was used. The z-scaling and feature selection were performed on the training set only, in order to prevent data leakage. To score the performance of the model on the training and testing sets, I use the Mean Absolute Error (MAE), which is in the scale of the response feature – for ISOPleasant and ISOEventful this means our response can range from -1 to $+1$. However, since the end-goal of the model is to predict the soundscape assessment of the location as a whole, rather than the individual responses, I also assess the performance of the model in predicting the average response in each location. To do this, the mean response value for each location is calculated, and the R^2 accuracy across LocationIDs is reported for both the training and testing sets.

The model fitting and feature selection was performed using the `step` function from `lmerTest` (v3.1.3) (Kuznetsova, Brockhoff, & Christensen, 2017) in R statistical software (v.4.0.3) (R Core Team, 2018). The summaries and plots were created using the `sjPlot` package (v.2.8.6) (Lüdecke, 2021) and `seaborn` (v.0.11.1) (Waskom, 2021).

5.2.3. Online Survey

A online listening test was conducted using the Gorilla Experiment Builder (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2019). The participants were exposed to a random selection of 78 binaural recordings (39 from 2019 and 39 from 2020, 6 recordings per location). Each participant had the option to evaluate either 1 or 2 sets of 6 recordings randomly assigned between 13 stimuli sets. Mp3 files, converted at 256 kBps were used due to the requirements of the Gorilla platform.

No visual stimuli were used in the experiment. The experiment consisted of:

1. an initial exercise to enhance the chances of participants complying with the instructions and wearing headphones
2. a training set using two randomly chosen binaural recordings (then not used in the main task) from the dataset
3. a soundscape characterisation questionnaire starting with an open-ended question about

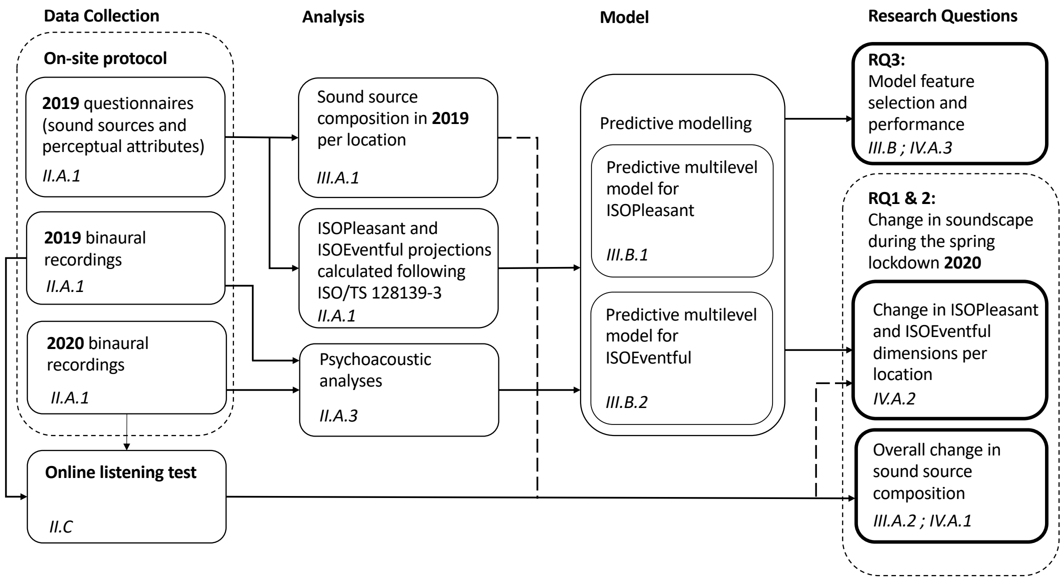


Figure 5.2.: The study flowchart indicating the data collection, analysis, modelling, and discussion throughout the study. draft *The subsections in the text to which each box refers are indicated in italics.*

perceived sound sources and featuring the same questions as the one used in-situ, looking into the perceived sound source dominance of the following four types: traffic noise, other noise, human sounds, and natural sounds

4. a questionnaire on the basic demographic factors.

The questionnaire used in Part 3 of the online experiment is reported in Table D.1.

Having in mind the remote nature of the study and to ensure a minimum level of robustness for reliable sound source recognition, an initial exercise was performed consisting of a headphone screening test (Woods, Siegel, Traer, & McDermott, 2017) and a headphone reproduction level adjustment test (Gontier, Lavandier, Aumond, Lagrange, & Petiot, 2019). The level adjustment was performed using an 11-second-long pink noise sample matched to the lowest and the highest L_{A90} values from the experimental set. Participants were asked to adjust their listening level to clearly hear the quieter sample while keeping the level low enough, so they don't find the louder sample disturbing. The headphone screening test followed, featuring a stereo signal of 1-second-long 100 Hz sin tone, generated with Izotope RX6 application, played at a 3 dB difference where one of the equally loud pairs had its phase inverted. A 100 Hz sin was used because the pilot tests revealed that the 200 Hz sin tone proposed by Woods et al. (2017) created a higher uncertainty varying across different laptop models and would likely contribute to the chances of a participant fooling the test. It was expected that participants using speakers would not be able to either hear the sin wave or would be fooled by the inverted phase effect and therefore not able to pass the trials, unless they were indeed using headphones. The participant needed to recognise the quietest of the 3 samples in a trial of 6 attempts. Only participants correctly answering 5 or more out of 6 trials were allowed to proceed with the experiment. Participants were asked not to change their audio output settings during the rest of the experiment. This was introduced to ensure that a participant is using a headphone playback system which allows a listener to clearly recognise a 3 dB difference at 100 Hz as a proxy for sufficient audio quality playback.

Online questionnaire data was collected between the 9th of June and the 9th of August 2020. Within the Gorilla Experiment Builder, a total of 250 attempts to complete the experiment were recorded, where 165 participants were excluded either on the basis of not passing the headphone screening ($N = 79$) or for not completing the experiment, usually before engaging into the screening ($N = 83$). Out of a total of 88 participants who completed the test, 2 participants were excluded as outliers as they provided uniform answers across all the questions and commented on not being able to properly hear the stimuli, despite their successful completion of the training tests. The participants of the online experiment were of mean age 32.42, 45.1% male, 54.9% female.

Fig. 5.2 illustrates and summarises the framework and sections described above.

5.3. Results

The results of the onsite surveys, online experiment, and the model development are reported here. They are reported following the structure of the ISO/TS 12913 series, revealing the perceived sound source dominance, key perceptual attributes (ISOPleasant and ISOEventful) and the lockdown-related changes.

5.3.1. The sound environment impacts of the lockdown in London and Venice

Before continuing to the predictive model and the impacts on the soundscape perception, I will first review results indicating the change in the sound environment. To summarise these impacts, I present the location-level changes in the L_{Aeq}, N_5, and S from the 2019 condition to the 2020 lockdown condition. The original analysis in this section was presented in Aletta, Oberman, Mitchell, et al. (2020). This analysis has been updated to include the Venice data and to correct slight discrepancies between the datasets and psychoacoustic analyses used in the two papers ((Aletta, Oberman, Mitchell, et al., 2020) and (Mitchell, Oberman, Aletta, Kachlicka, et al., 2021)).

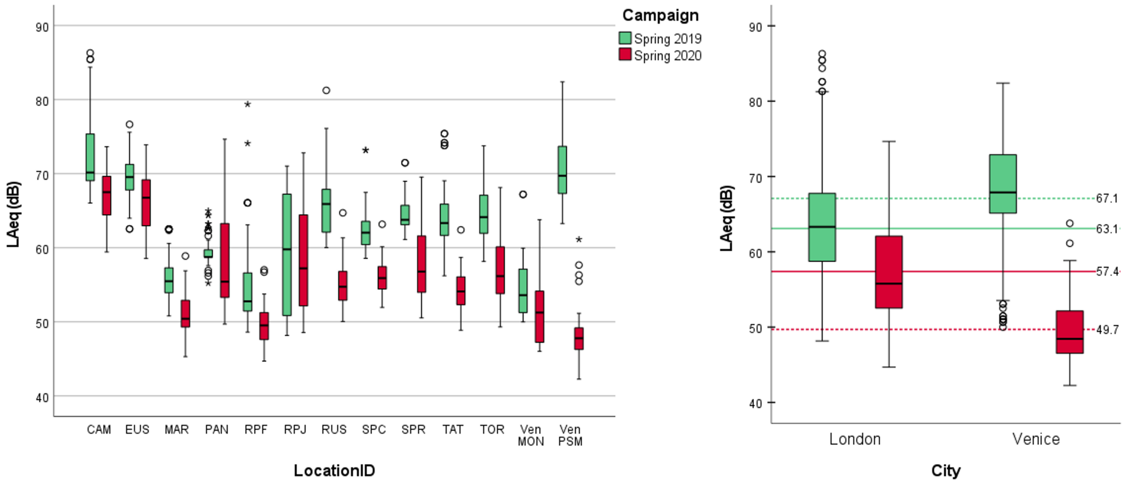


Figure 5.3.: On the left: Sound levels distributions at the 11 London locations before and during the lockdown measures implementation; on the right: Sound levels distributions in each city (aggregated across locations) and corresponding mean values before and during the lockdown measures implementation (solid line: London; dashed line: Venice).

Fig. 5.3 presents the distributions of L_{Aeq} values measured in each of the 13 locations in 2019 and during the lockdown in 2020. To determine whether there is a statistically significant difference between the two campaigns in each location, a one-tailed t-test was used. The results and the mean and standard deviation for each location in both campaigns are presented in Table 5.3.

The t-test results indicate that all locations except Regents Park Japan demonstrate a significant change in the sound environment. Averaged across all locations, sound levels in London and Venice decreased by 5.66 dB and 17.45 dB, respectively, and an overall decrease of 7.08 dB for both cities. In London, the amount of reduction ranges from 10.49 dB (Russell Square) to 1.27 dB (Regents Park Japan). The largest reduction (22.03 dB) is seen in Piazza San Marco which is to be expected given the drastic contextual difference from Carnevale in 2019 to a deserted square in 2020.

Fig. 5.4 presents the distributions of N_5 values measured in each of the 13 locations in 2019 and during the lockdown in 2020. The results here closely mirror the L_{Aeq} results. In London the N_5 reduction ranges from 11.03 sones (Russell Square) to 0.35 sones (Pancras Lock). Fig. 5.5 presents the distributions of S values measured in each of the 13 locations in 2019 and during the lockdown in 2020. An interesting point here is that, unlike the sound level and loudness results, there was not a universal reduction in the sharpness levels across all locations.

Effect of the urban setting on sound levels reduction In addition to strictly documenting the changes in the sound environment in London, we also aimed to investigate whether the lockdown measures would result in different sound level reductions depending on the urban

Table 5.3.: Mean and standard deviation values for 2019 and 2020 (Lockdown) measurement campaigns. The difference is the 2020 mean value subtracted from the 2019 mean value to demonstrate the level of change due to the lockdown. *p*-values calculated for one-tailed t-test.

		2019				2020 (Lockdown)				Difference and t test p-value		
LocationID		**Samples**	**N5 (sone)**	**S (acum)**	L_{Aeq} (dB)	**Samples**	**N5 (sone)**	**S (acum)**	L_{Aeq} (dB)	**N5 (sone)**	**S (acum)**	L_{Aeq} (dB)
CAM	Mean	90	39.29	2.28	72.41	44	30.50	1.87	67.19	-8.79**	-0.41**	-5.21**
	Std. Deviation		13.74	0.47	5.08		8.43	0.17	3.72	0.00	0.00	0.00
EUS	Mean	99	31.16	2.25	69.55	38	24.72	1.88	66.03	-6.44**	-0.38**	-3.52**
	Std. Deviation		5.62	0.13	2.76		7.34	0.21	4.25	0.00	0.00	0.00
MAR	Mean	92	12.83	1.67	55.85	41	9.34	1.53	51.27	-3.48**	-0.15**	-4.58**
	Std. Deviation		3.15	0.15	2.66		2.62	0.23	2.86	0.00	0.01	0.00
PAN	Mean	78	15.32	1.64	59.42	80	14.97	2.04	58.11	-0.35	0.40**	-1.30**
	Std. Deviation		2.28	0.11	1.82		6.80	0.51	6.06	0.36	0.00	0.06
RPF	Mean	107	11.98	1.67	54.47	43	8.18	1.40	49.61	-3.79*	-0.28**	-4.86**
	Std. Deviation		7.31	0.17	5.01		2.19	0.16	2.76	0.04	0.00	0.00
RPJ	Mean	82	16.90	2.72	59.72	35	15.38	2.64	58.45	-1.52	-0.08	-1.27
	Std. Deviation		8.20	0.65	8.21		7.15	0.59	7.35	0.26	0.33	0.30
RUS	Mean	149	23.07	2.62	65.49	40	12.04	1.47	55.00	-11.03**	-1.15**	-10.49**
	Std. Deviation		5.10	0.48	3.58		3.35	0.10	3.00	0.00	0.00	0.00
SPC	Mean	60	18.22	1.81	62.40	27	13.18	1.46	56.33	-5.04**	-0.35**	-6.07**
	Std. Deviation		3.81	0.12	2.88		3.21	0.10	2.58	0.00	0.00	0.00
SPR	Mean	57	19.72	1.73	64.59	48	13.93	1.45	57.77	-5.78**	-0.28**	-6.82**
	Std. Deviation		3.02	0.12	2.22		4.62	0.28	4.71	0.00	0.00	0.00
TAT	Mean	129	19.86	1.76	63.61	41	11.33	1.25	54.37	-8.54**	-0.50**	-9.24**
	Std. Deviation		4.95	0.22	3.80		2.50	0.15	2.71	0.00	0.00	0.00
TOR	Mean	115	22.69	2.03	64.77	41	15.20	1.47	56.79	-7.49**	-0.56**	-7.98**
	Std. Deviation		5.60	0.23	3.79		6.38	0.21	4.37	0.00	0.00	0.00
VenMON	Mean	24	11.26	1.60	54.91	33	8.33	1.70	51.28	-2.93**	0.10	-3.63*
	Std. Deviation		3.54	0.19	4.84		2.71	0.20	4.46	0.01	0.06	0.02
VenPSM	Mean	90	28.64	2.04	70.38	40	7.09	1.11	48.36	-21.55**	-0.92**	-22.03**
	Std. Deviation		7.06	0.18	4.33		2.32	0.12	3.66	0.00	0.00	0.00
London	Mean	1058	21.26	2.05	63.08	478	15.37	1.70	57.42	-5.89**	-0.35**	-5.66**
	Std. Deviation		9.96	0.50	6.62		8.31	0.48	6.67	0.00	0.00	0.00
Venice	Mean	114	24.98	1.94	67.13	73	7.65	1.38	49.68	-17.33**	-0.56**	-17.45**
	Std. Deviation		9.62	0.26	7.73		2.56	0.34	4.27	0.00	0.00	0.00
Total	Mean	1172	21.62	2.04	63.47	551	14.35	1.66	56.40	-7.27**	-0.39**	-7.08**
	Std. Deviation		9.99	0.48	6.84		8.22	0.47	6.92	0.00	0.00	0.00

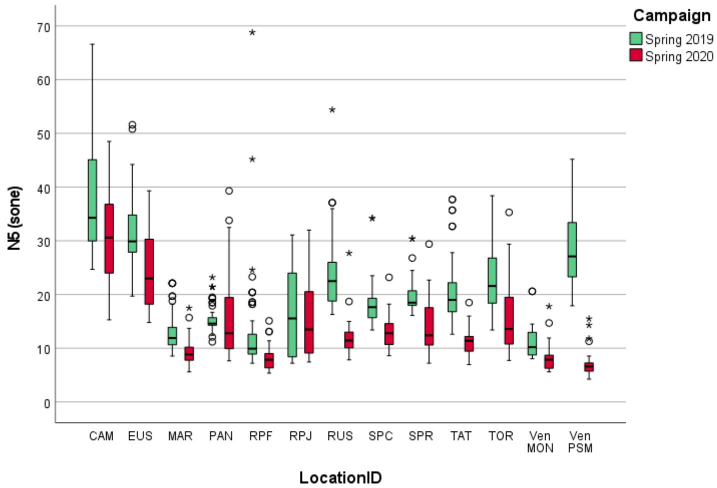


Figure 5.4.: Loudness distributions at the 11 London locations before and during the lockdown measures implementation.

scenario (and its composition of sound sources). For this purpose, it was decided to define an 'Area type' variable that would serve as a proxy for urban (acoustic) context: a *k*-means cluster analysis was performed on the mean values of L_{Aeq}, L_{A10}, L_{A90}, N_5, T, FS, and S of the 2019 measurements campaign for the 11 locations, after those have been *z*-score standardized to meet the algorithm criteria. The rationale was that clustering urban areas *a priori* based on their 'typical' acoustic climate (hence using only data from 2019) would allow us to see whether there was an association between area type and noise reduction. The algorithm was set to a three-cluster solution, based on visual inspection of the scree plot as reported in Fig. 5.6a ('elbow method') (Ketchen Jr. & Shook, 1996). The analysis was conducted in R (R Core Team, 2018) and figures were produced using the package `factoextra` (Kassambara & Mundt, 2020).

Fig. 5.6b shows a plot of clustered data based on the two most relevant underlying dimensions for the three-cluster solution. Dimension 1 seems to describe a pattern related to sound level and associated metrics, whilst Dimension 2 is related to Sharpness. This is consistent with previous findings in literature where it was observed that when it comes to categorization and classification of urban acoustic environments based on objective features, most solutions are reduced to intensity- and spectral-related parameters (Aletta et al., 2017; De Coensel & Botteldooren, 2006).

Table 5.4 shows the basic descriptive statistics of the psychoacoustic features for the 11 locations according to cluster membership; when combining those patterns with information about dominant sound sources as derived from data from Mitchell et al. (2020), the three clusters could be labelled as: *Traffic/Noise-dominated areas* (locations: Camden Town, Euston Tap, Piazza San Marco), *Active Areas* (locations: Regents Park Japan, Russell Square, St Pauls Cross, St Pauls Row, Tate Modern, Torrington Square) and *Quiet Areas* (locations: Marchmont Garden, Pancras Lock, Regents Park

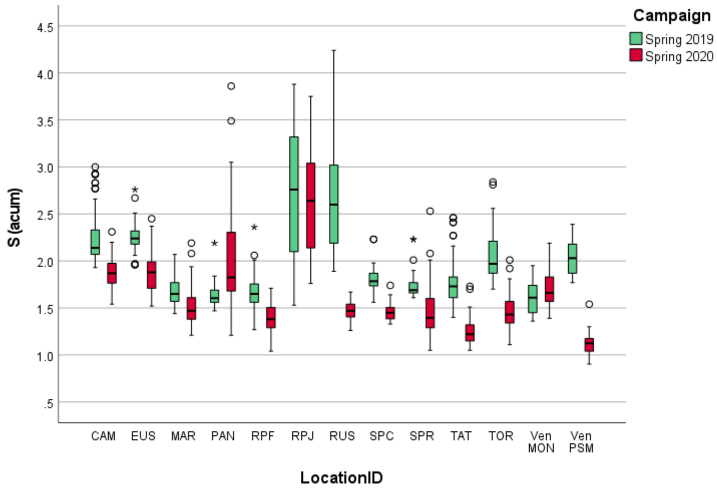


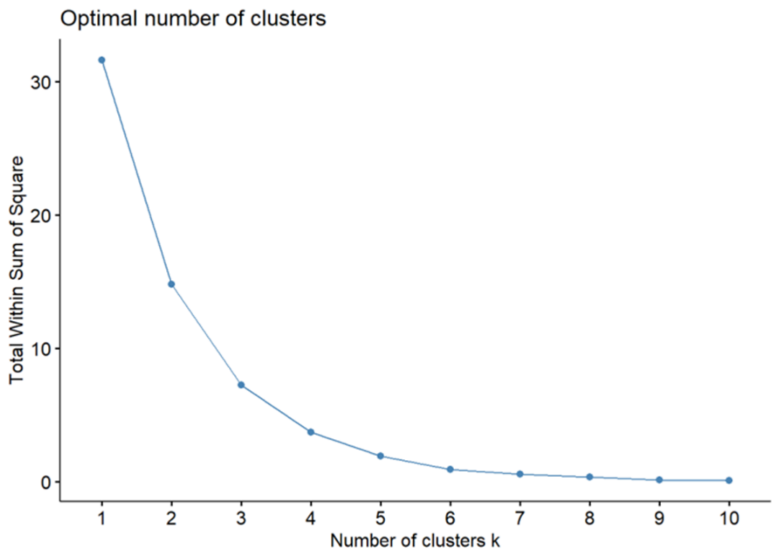
Figure 5.5.: Sharpness distributions at the 11 London locations before and during the lockdown measures implementation.

Table 5.4.: Descriptive statistics of the psychoacoustic metrics for the three identified clusters.

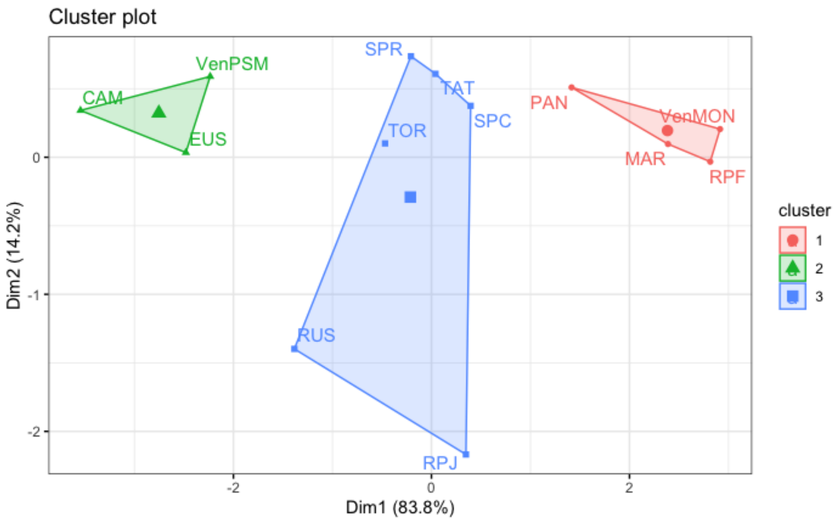
Cluster	Parameter	Mean values				
		L_{Aeq}	L_{A10}	L_{A90}	S	N_5
1 – (Quiet areas)	Mean	55.9	58.1	52.4	1.7	12.9
[N=3]	Std. deviation	2.6	2.5	3.1	0.0	1.8
	Variance	7.0	6.3	9.3	0.0	3.1
2 – (Active areas)	Mean	62.5	64.3	59.6	2.1	18.9
[N=6]	Std. deviation	2.3	2.5	2.3	0.4	2.5
	Variance	5.3	6.4	5.5	0.1	6.3
3 – (Traffic-dominated areas)	Mean	70.4	72.9	66.1	2.4	34.1
[N=2]	Std. deviation	1.6	1.9	0.3	0.0	4.4
	Variance	2.4	3.6	0.1	0.0	19.8

Fields, Monumento Garibaldi). Traffic/Noise-dominated (with the exception of San Marco) areas with an exceptionally high noise level; typically this is because they are on major roads, where traffic noise is the dominant sound source. However Piazza San Marco also clusters here due to its unique use-case. Active areas are locations where the human activity (also combined with traffic) is the main contributor to the acoustic environment. Quiet areas are generally parks or areas with greenery that tend to have a relatively low background noise (lack of traffic sources).

When considering the mean L_{Aeq} reductions between 2019 and 2020 as a function of Area type, it can be observed that they vary across the three clusters, as shown in Fig. 5.7. The biggest reductions are for Active areas (M = 6.6 dB; SD = 3.2 dB), followed by Traffic-dominated areas (M = 4.5 dB; SD = 0.8 dB), and Quiet areas (M = 3.6 dB; SD = 1.9 dB). A possible explanation for this is that road



(a) 'Scree' plot used to identify the optimal number of clusters to use in the *k*-means clustering algorithm where an 'elbow' can be identified for a three-cluster solution.



(b) Bi-dimensional plot for the three-cluster solution. The clusters have been labelled as: Cluster 1 – Quiet Areas; Cluster 2 – Active Areas; Cluster 3 – Traffic/Noise-dominated Areas.

Figure 5.6.: Clustering analysis of 2019 prelockdown data.

traffic at the selected locations in London is still sustained to some extent (e.g. circulation of public transport, key workers, etc.), while the most significant variation in Active areas is possibly due to the complete lack of (non-motorized) human activity on site. The locations in the cluster labelled as Quiet areas were already not particularly noisy even before the lockdown, thus the small changes observed are probably once again due to the absence of people.

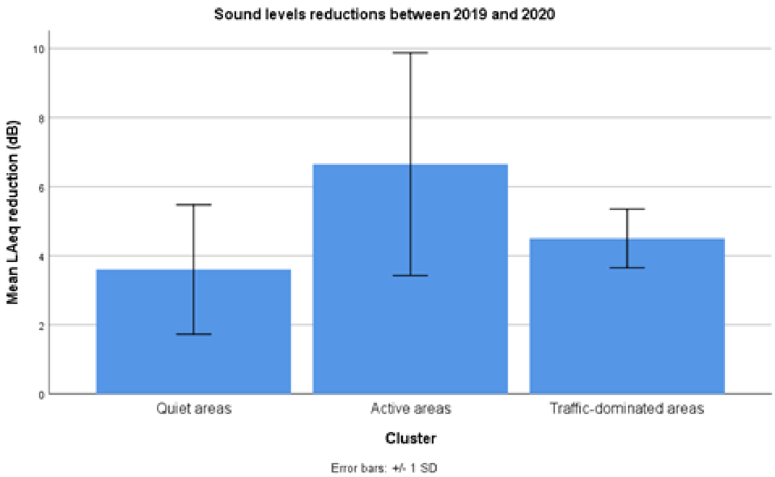


Figure 5.7.: Mean A-weighted equivalent sound level reductions between the pre- and during-lockdown conditions as a function of cluster membership (i.e. Area type).

5.3.2. Perceived sound source dominance

2019 sound source composition per location

While the previous clustering was focussed on differentiating locations according to their strictly acoustic parameters, the listening test allowed us to also differentiate locations according to their sound source profiles. Based on the results derived from the source-dominance questions in the online listening test, this section presents the differences between the locations. Questionnaire data was collected English, Italian, and Spanish in both cities. Data presented here was aggregated per LocationID.

According to the highest scored mean value of the dominant sound source type, as shown in Fig. 5.8, the locations can be grouped into: natural sounds dominated (RegentsParkJapan, RegentsParkFields, RussellSq), human sounds dominated (SanMarco, TateModern, StPaulsRow, StPaulsCross, MonumentoGaribaldi), noise (traffic and other noise) sounds dominated (CamdenTown, EustonTap, TorringtonSq, PancrasLock).

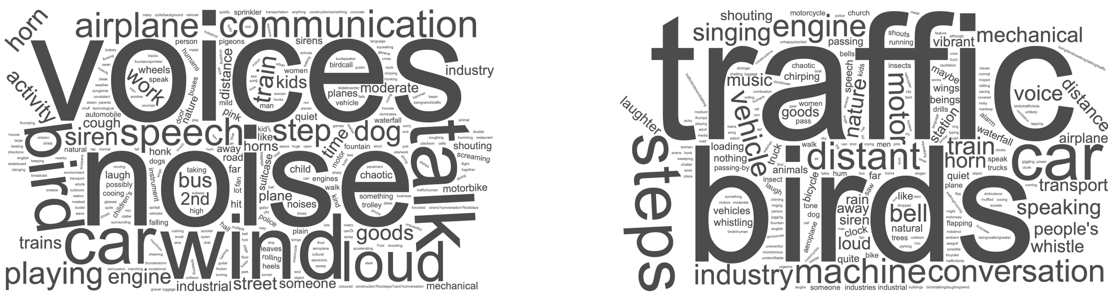
Overall change in the perceived sound source dominance during lockdown

1,803 words describing the sound sources present in the 2019 recordings and 1,395 words related to the 2020 recordings were input by participants in response to the open-ended question Q1 (see Table D.1). The frequency of occurrence, generated using the Word-Clouds web app, is shown in

Locations	Traffic	Human	Natural	Other
CamdenTown	3.8	3.3	1.3	2.7
EustonTap	3.7	2.6	1.7	3.0
MarchmontGarden	2.7	2.7	2.6	2.5
MonumentoGaribaldi	1.9	3.4	3.0	2.0
PancrasLock	2.4	2.5	2.4	3.3
RegentsParkFields	2.4	2.9	3.1	1.9
RegentsParkJapan	1.9	2.5	4.0	1.5
RussellSq	2.8	3.0	3.3	2.1
SanMarco	1.4	4.0	2.2	1.9
StPaulsCross	2.6	3.3	2.3	2.1
StPaulsRow	2.5	3.4	1.7	2.3
TateModern	2.5	3.6	2.6	2.1
TorringtonSq	3.2	3.3	1.9	2.8

Figure 5.8.: Mean values per LocationID for the perceived dominance of the sound source types, for the 2019 on-site campaign.

Fig. 5.9, for the 2019 and the 2020 sets respectively. The highest frequency words from both 2019 and 2020 groups are: noise, car/traffic, bird/birds, talk/voice, and (foot)steps.



(a) 2019 (b) 2020

Figure 5.9.: A graphic illustrating the frequency of occurrence of the sound sources reported by the participants of the online study across all locations, shown for recordings from 2019 (a) and 2020 (b).

The results from the listening tests deployed online were analysed using SPSS Statistics v. 25. Levene's test for equality of variances resulted in highly statistically significant values for all 4 sound sources investigated (< 0.001). Therefore, a Mann-Whitney U-test was used as a non-parametric equivalent to the t-test to investigate the change in the perceived dominance of the four sound source types (McKnight & Najab, 2010). The results for human sounds indicated that the perceived dominance was greater for the 2019 sample ($M = 3.82$) than for the 2020 sample ($M = 2.62, U =$

$41,656, p < 0.001$). The results for natural sounds indicated the perceived dominance increased from 2019 ($M = 2.00$) to 2020 ($M = 2.54, U = 63,797, p < 0.001$). However, the differences for the noise sources (traffic and other) were not statistically significant.

Table 5.5.: Mean values and standard deviation for the perceived dominance of sound sources (rated from 1 - 5), assessed via an online survey.

Sound source type	Campaign	N	Mean	Std. Dev.	Std. Error Mean
Traffic	2019	422	2.51	1.369	.067
	2020	383	2.56	1.525	.078
Other	2019	422	2.00	1.182	.058
	2020	382	2.23	1.333	.068
Human	2019	423	3.82	1.143	.056
	2020	382	2.62	1.346	.069
Natural	2019	424	2.00	1.307	.063
	2020	380	2.54	1.441	.074

5.3.3. Model selection, performance, and application

ISOPleasant model selected

Following the feature selection, the ISOPleasant model (given in Table 5.6) has N_5 as the fixed effect with a scaled coefficient of -0.06, and L_{Aeq}, $L_{A10} - L_{A90}$, and $L_{Ceq} - L_{Aeq}$ as coefficients which vary depending on the LocationID. The training and testing MAE are very similar, indicating that the model is neither over- nor under-fitting to the training data ($MAE_{train} = 0.259$, $MAE_{test} = 0.259$). The model performs very well at predicting the average soundscape assessment of the locations ($R^2_{train} = 0.998$, $R^2_{test} = 0.85$).

The high intraclass correlation ($ICC = 0.90$) demonstrates that the location-level effects are highly important in predicting the pleasantness dimension. Within this random-intercept random-slope model structure, these effects include both the specific context of the location (i.e. the LocationID factor), but also the L_{Aeq}, $L_{A10} - L_{A90}$, and $L_{Ceq} - L_{Aeq}$ features whose effects vary across locations. These slopes are given in Fig. 5.10. This point highlights the need to consider how the context of a location will influence the relationship between the acoustic features and the perceived pleasantness.

ISOEventful model selected

Through the group-level feature selection, all of the group-level coefficients were removed, including the LocationID factor itself. Therefore the final ISOEventful model is a 'flat' multi-variate linear

Table 5.6.: Scaled linear regression models of ISOPleasant and ISOEventful for 13 locations in London and Venice. The ISOPleasant model is a multi-level regression model with one level for individual effects and a second level for LocationID effects, while the ISOEventful model is a 'flat' multi-variate linear regression with no location effects.

	ISOPleasant			**ISOEventful**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.24	0.15 - 0.33	**<0.001**	0.14	0.12 - 0.16	**<0.001**
N_5	-0.06	-0.10 - -0.02	**<0.001**			
S				-0.08	-0.11 - -0.06	**<0.001**
FS				-0.02	-0.05 - -0.00	**0.033**
T				0.04	0.01 - 0.07	**0.002**
L_{Aeq}				0.14	0.11 - 0.17	**<0.001**
$L_{Ceq} - L_{Aeq}$				-0.03	-0.05 - 0.00	0.052
Random Effects						
σ^2	0.11					
τ_{00}	$0.03_{LocationID}$					
τ_{11}	$0.02_{LocationID.L_{Aeq}}$					
	$0.00_{LocationID.L_{A10}-L_{A90}}$					
	$0.00_{LocationID.L_{Ceq}-L_{Aeq}}$					
ICC	0.90					
N	$13_{LocationID}$					
Observations	914			914		
MAE Test, Train	0.258	0.259		0.233	0.231	

regression model, rather than a multi-level model. The ISOEventful is a linear combination of S, FS, T, L_{Aeq}, and $L_{Ceq} - L_{Aeq}$. The training and testing MAE are very similar, indicating that the model is not over-fit to the training data ($MAE_{train} = 0.233$; $MAE_{test} = 0.231$). The model performs slightly worse than the ISOPleasant at predicting the mean location responses, but still performs well ($R^2_{train} = 0.873$; $R^2_{test} = 0.715$).

Application to lockdown data

Once the two models were built and assessed, they were then applied to the lockdown recording data in order to predict the new soundscape ISO coordinates. Fig. 5.11a shows the pre-lockdown ISO coordinates for each location and Fig. 5.11b shows how the soundscapes are predicted to have been assessed during the lockdown period. As in the model assessment process, the predicted responses are calculated for each recording individually, then the mean for each location is calculated and plotted on the circumplex.

In 2019 the majority of locations in the dataset fall within the 'vibrant' quadrant of the circumplex, particularly those which are primarily dominated by human activity (e.g. SanMarco, TateModern). CamdenTown and EustonTap, which are both in general visually and acoustically dominated by traffic, are the only two to be rated as 'chaotic', while no locations are overall considered

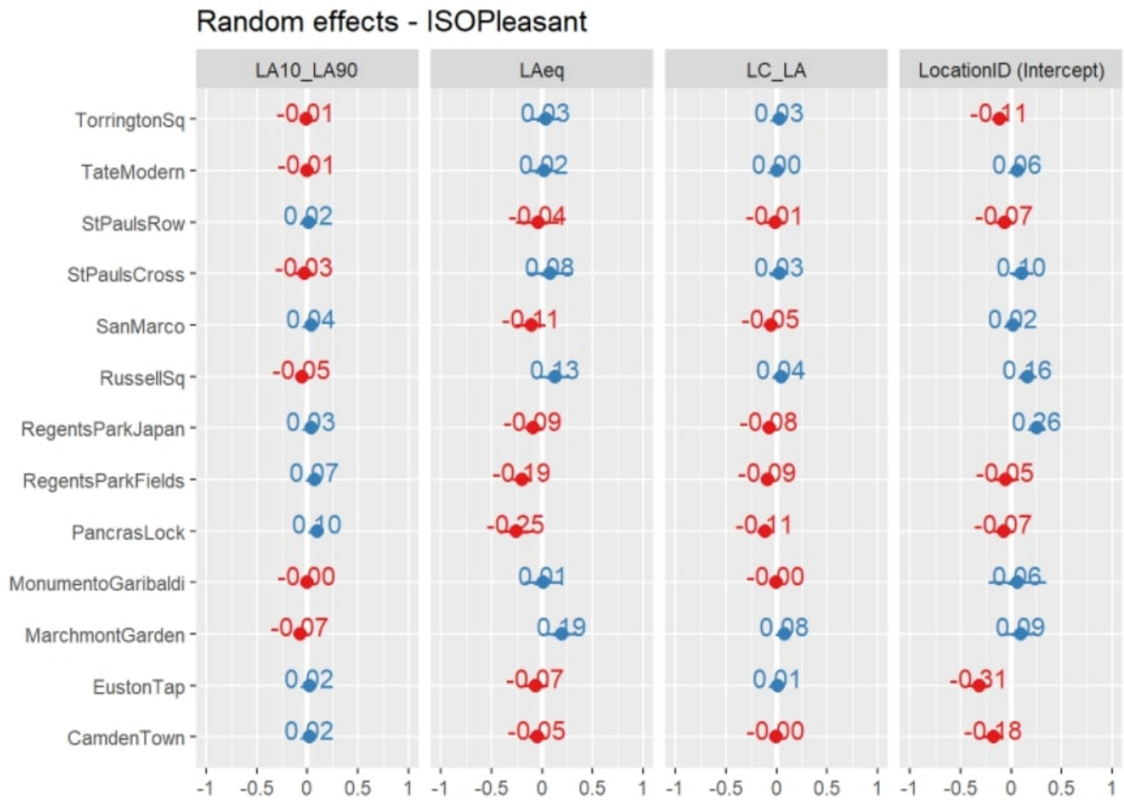


Figure 5.10.: Location-level scaled coefficients for the ISOPleasant model.

to be 'monotonous'. During the 2020 lockdown, there is a general positive move along the 'pleasant' dimension and a general negative move along the 'eventful' dimension, but several patterns of movement can be noted. These are investigated further in the Discussion section below.

5.4. Discussion

5.4.1. Interpretation of the results

To interpret the results addressing RQ1 and RQ2, it is necessary to separately look into the overall change in sound source composition, and the change in the affective quality of soundscapes per location.

Change in the sound source composition

The open-ended question about sound sources in the online survey did not reveal a change in sound source types but rather confirmed that all types were still present in both conditions. The sound

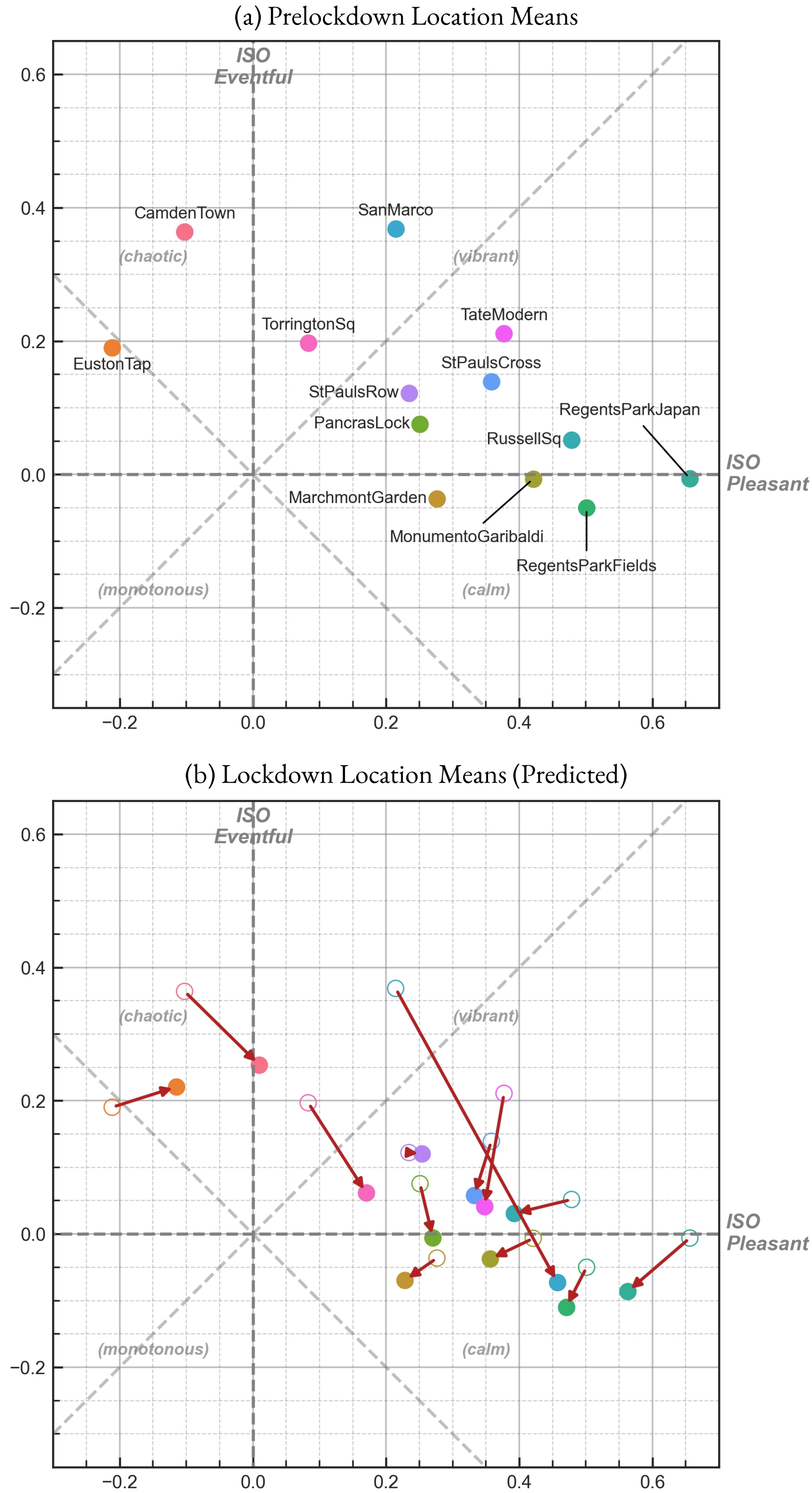


Figure 5.11.: Soundscape circumplex coordinates for (a) the mean ISOPleasant and ISOEventful responses for each location; and (b) the mean predicted responses based on recordings made during the lockdown and the location's movement in the circumplex.

source composition question taken from the Method A of the ISO/TS 12913-2:2018 (ISO/TS 12913-2:2018, 2018) revealed a statistically significant reduction in human sound sources and a significant increase in the perceived dominance of natural sound sources.

The most frequent sound sources detected from the open-ended question correspond to the main four sound source types investigated, which indicated that all types remained present in the lockdown condition (at all the locations). While traffic intensity might have gone down, where the results of the Mann-Whitney U-test were inconclusive, but supported by the psychoacoustic measurements according to Aletta, Oberman, Mitchell, et al. (2020), traffic-related sound sources were still clearly present.

The sound source composition of an outdoor acoustic environment is extremely complex. Removing one component, such as human sounds, has implications on the whole (Gordo, Brotons, Herrando, & Gargallo, 2021). Testing the effects of this *in-situ* is not straightforward and interpreting this study in line with 'what is the impact of human sounds' must be taken within the broader context of the range of conditions which changed within the acoustic environment. However, looking at the overarching picture, the lockdown condition was a useful and unique case study to understand the impact which human activities – and the human sound source type in particular – can have on soundscape perception of urban open spaces.

Movement of soundscapes

In order to interpret how the change of the acoustic environment at the locations examined would have been perceived, and to answer RQ2, movement vectors within the circumplex space are shown in Fig. 5.12. This clearly shows a few different patterns of movement due to the effects of the 2020 lockdown. These can be further looked into depending:

1. the magnitude of change,
2. the direction of change,
3. shift between the quadrants shown in Fig. 5.11,
4. sound source composition.

SanMarco The largest change is seen in Piazza San Marco, with a predicted increase in pleasantness of 0.24 and a decrease in eventfulness of 0.44, enough to move the soundscape out of the 'vibrant' quadrant and into 'calm'. This extreme change (relative to the rest of the locations) is exactly what would be expected given the unique context of the measurements taken in 2019 – the measurement campaign corresponded with Carnevale, a yearly festival which centres around the square. By contrast, due to the particularly strict measures imposed in Italy, during the lockdown measurement period the square was almost entirely devoid of people. What is promising is that,

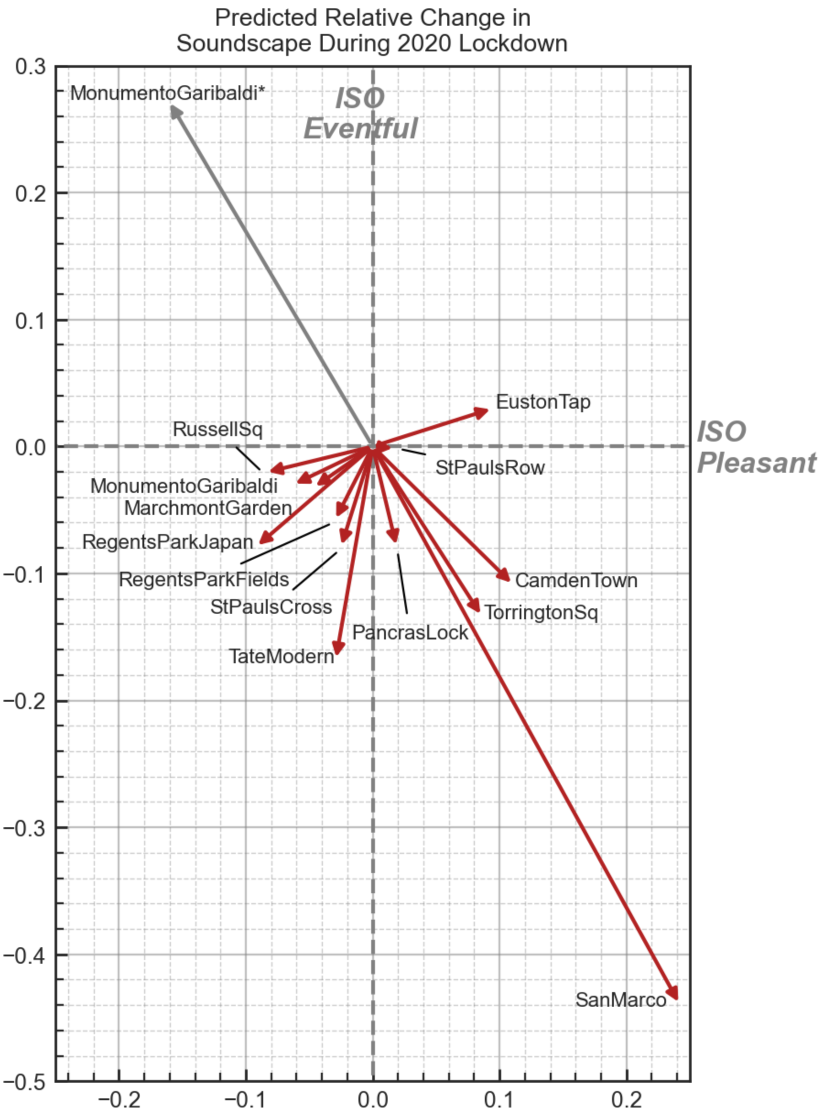


Figure 5.12.: The relative movement of soundscape perception in the circumplex due to the COVID-19 lockdowns, represented as vectors centred on the origin. *The lawn-works dominated session is shown separately as MonumentoGaribaldi* with a grey arrow to indicate that this is distinct from the effects of the lockdown changes.

without any of this contextual information about the presence or absence of people, my model is able to capture and reflect what may be considered a reasonable and expected direction and scale of movement within the soundscape circumplex.

EustonTap, CamdenTown, TorringtonSq, PancrasLock The next locations of interest are those which, in the 2019 survey data, were rated as being dominated by traffic noise: Euston-

Tap, CamdenTown, TorringtonSq, and PancrasLock. These are the only locations (besides SanMarco) which show a predicted increase in pleasantness. Of these traffic-dominated spaces, the two which were most heavily dominated by traffic noise (CamdenTown and EustonTap) showed the most increase in pleasantness, with TorringtonSq having slightly less of an increase. PancrasLock, which was also rated as having high levels of both Human and Natural sounds shows only a modest improvement in pleasantness.

StPaulsCross, TateModern, RegentsParkJapan, RegentsParkFields, RussellSq Among the locations which are predicted to experience a negative effect on pleasantness we see a mix of spaces which were assessed as being dominated by Human (StPaulsCross and TateModern) and Natural (RegentsParkJapan, RegentsParkFields, RussellSq) sounds before the lockdown. It is hard to discern a pattern of difference between these two groups, although it appears that the Human-dominated spaces saw a greater reduction in eventfulness, compared to the Natural-dominated spaces.

In general, we note that most of the spaces experience some degree of reduction in eventfulness. This pattern is particularly consistent with what would be expected from a reduction in human presence in these spaces (Aletta & Kang, 2018), as reflected by the observation that, in general, those spaces which had the most human sounds prior to the lockdown showed the greatest reduction in eventfulness during the lockdown.

EustonTap An unexpected result is that EustonTap is predicted to experience an increase in eventfulness and it is unclear whether this accurately reflects the real experience people would have had in the space. Normally, EustonTap is a mostly-outdoor drinking venue located at the entrance to the Euston Train station and is situated directly along a very busy central London road. During the 2020 survey, the researchers noted that the music and chatter of people from the pub was noticeably missing, but that the perceived reduction in road traffic was minimal. Based on the theory of vibrancy which would suggest it is driven by human presence and sounds (Aletta & Kang, 2018), we would not therefore expect a shift in the vibrant direction as indicated here. This discrepancy may reveal a weakness in the context-independent ISOEventful model, or it may in fact be indicating that, at certain thresholds of traffic noise, a reduction in level – and therefore a reduction in energetic masking – will allow other aspects of the sound to influence the perception.

MonumentoGaribaldi Finally, special attention should be paid to the results shown for Monumento Garibaldi, which in 2019 was perceived as a pleasant and slightly calm green space featuring a gravel walkway. During the first measurement session during the lockdown in 2020, the researcher noted that the soundscape was dominated by landscaping works, in particular noise from strimmers (or weed whackers). In order to gain a sample which was more representative of the impact of the lockdowns, the researcher returned another day to repeat the measurements without interference from the works.

To examine the impact of these two scenarios separately, the prediction model was applied to the data from the two sessions independently and the session which was impacted by the landscaping works is shown in Fig. 5.12 in grey and labelled MonumentoGaribaldi*, while the unaffected session is shown in red. In the latter case, the predicted change in soundscape as a result of the lockdowns fits neatly into what would be expected and closely matches the predicted behaviour of similar locations in London (i.e. MarchmontGarden and RussellSq). On the other hand, the session which was dominated by noise from the strimmers is predicted to have become much more chaotic, with a decrease in pleasantness of 0.16 and an increase in eventfulness of 0.27. This indicates that, although the model has no contextual information about the type of sound and in fact the training data never included sounds from similar equipment, just based on the psychoacoustic features of the sound it is able to reasonably predict the expected change in soundscape.

As a whole, the primary impact of the 2020 lockdowns on the soundscapes in London and Venice was an overall decrease in eventfulness. With the exception of EustonTap, all of the sessions show some degree of reduction in eventfulness, reflecting the general decrease in sound levels and human sound sources across the locations. The impact of the lockdowns on pleasantness is more mixed and seems to be driven by the previous dominance of traffic noise in the space. However, it could also be noted that, while all locations experienced a reduction in sound level, those which are predicted to become more pleasant had an average L_{Aeq} above 60 dB in 2019. By contrast, the locations which were predicted to experience a decrease in pleasantness generally had sound levels below 60 dB(A) in 2019. This may indicate that reductions in sound level can improve pleasantness when the sound level exceeds some threshold of around 60 - 65 dB(A) but are ineffective when sound levels are below this threshold. Similarly, W. Yang and Kang (2005a) showed that, when the sound level is 'lower than a certain a certain value, say 70 dB' there is no longer a significant improvement in the evaluation of acoustic comfort as the sound level reduces. It is unclear at this point where this threshold would lie for pleasantness/annoyance, how strict it may be, or how it is impacted by the sound source composition of the acoustic environment, therefore further research is needed in this area.

Model selection results

The most immediately interesting result of the model-building and feature selection process, answering to RQ3, is the apparent irrelevance of location context to the ISOEventful dimension. The multilevel model structure was chosen since the starting assumption was that soundscape perception is heavily influenced by contextual factors, such as expectations of the space and visual context. For this modelling, these factors can be considered as location-level latent variables at least partially accounted for by the inclusion of the LocationID as the second-level factor. While this assumption certainly held true for ISOPleasant, my results indicate that these types of contextual factors are not significant for ISOEventful, and do not affect the relationship between the acoustic features of the sound and the perception.

In particular this result may herald a shift in modelling approach for soundscapes – where previous methods, in both the soundscape and noise paradigms, have mostly focused on deriving acoustic models of annoyance (in other words have focussed on the ISOPleasant dimension) perhaps they should instead consider the acoustic models as primarily describing the eventfulness dimension when considered *in-situ*. In addition this study takes the approach of modelling responses at an individual level in order to derive the soundscape assessment of the location. Rather than either attempting to represent the predicted response of an individual person – which is less useful in this sort of practical application – or to base the model on average metrics of the location, the goal is instead to characterise the location itself, through the aggregated predicted responses of individuals. I believe this modelling approach better addresses the practical goal of predictive soundscape modelling and reflects the structure of the data collection.

5.4.2. Limitations of the study

The onsite sampling method was initially not intended as the ultimate characterisation of a location's soundscape but rather as a tool for model development. Therefore, the change observed does not necessarily represent the ground truth about the site's soundscape, if such a thing exists. Further, the online listening tests took a relatively small but random sample from the available database and did not include any contextual information. This proved to be sufficient for the purpose of detecting a change in sound source composition, however, the relatively small sample of recordings included in the online study does limit how representative they are of the location's sound environment as a whole.

The surveys and recordings taken represent only a snapshot of the soundscape or sound environment for a short period in time. This is a flaw in most soundscape sampling methods presented both in the literature and in ISO/TS 12913-2. To truly be said to characterise the soundscape of a space, long-term monitoring and survey methods will need to be developed in order to capture the changing environmental and contextual conditions in the space. Models of the sort presented here, which are based on measurable quantities, could prove to be useful in this sort of long-term monitoring as they could take continuous inputs from sensors and generate the likely soundscape assessment over time.

Further, the lockdown condition is likely to cause distortions of the circumplex soundscape perception model. Therefore, it is important to acknowledge that all the predictions were made for the people with no experience of the pandemic and its psychological effects. Conceptually, this model captured the perceptual mapping (i.e. the relationship between the acoustic indicator inputs and the soundscape descriptor outputs) of people in 2019, but this perceptual mapping is likely to have been affected by the psychological and contextual impacts of the lockdown itself, independent of its changes on the sound environment. Future research might look into potential perception changes in the post-pandemic world.

The model presented in this chapter is also not generalisable. Specifically, since the multi-level structure relies on the LocationID as a categorical variable, the model cannot be applied to other locations not in the original training set. This was suitable for the specific research questions of this study, focussed on how these particular locations changed during the lockdown period, but will need to be solved for a practical method of prediction for engineering and design. Chapter 6 will discuss how this can be addressed. Finally, since this model is dependent only on the (psycho)acoustic indicators, it does not yet represent a holistic representation of the perceptual mapping, for which we also need to consider semantic information, visual and contextual factors, and personal factors. This model represents only a starting point for further model development, which will be presented in Part III.

5.5. Conclusion

This chapter demonstrates an application of predictive modelling to the field of soundscape studies. The model building results reveal that, within this dataset, an approach based on psychoacoustics can achieve $R^2 = 0.85$ for predicting the pleasantness of locations and $R^2 = 0.715$ for predicting the eventfulness. A modelling-focussed method of this sort is a key component to the potential scalability of the soundscape approach to applications such as smart city sensing, urban planning, and cost-effective, sustainable design. To demonstrate the usefulness and feasibility of such an approach, I applied my predictive model to a unique case study in which traditional soundscape survey methods were impossible.

By applying this predictive model to recordings collected during the 2020 lockdown, the change in perception of the urban soundscapes is revealed. In general, soundscapes became less eventful, and those locations which were previously dominated by traffic noise became more pleasant. By contrast, previously human- and natural-dominated locations are in fact predicted to become less pleasant despite the decrease in sound levels. Although the results are limited in that they present one snapshot of the soundscape of the spaces, the success of the model in responding to new and disturbing sound events demonstrates its potential usefulness in long-term monitoring of urban soundscapes.

Starting from the success and unique application of this model, Part III will present developed framework for creating a generalisable predictive model, and a series of case studies illustrating how additional soundscape indicators could be integrated into the model to improve its performance and usefulness.

Part III.

Towards a Generalisable and Probabilistic Soundscape Model

Introduction

Chapter 5 has demonstrated a novel application of predictive soundscape modelling. Given the conditions of the COVID-19 lockdowns, existing soundscape assessment methods were impractical (or illegal) to implement. As outlined in Chapter 2, this is an example of one of the scenarios where predictive modelling is necessary, as the existing *post-hoc* survey methods cannot be used. Other applicable scenarios include: soundscape mapping, predicting soundscapes of not-yet-existing spaces, unattended sensor networks, and long-term soundscape monitoring.

The model in Chapter 5 represents an advance for current soundscape methods and provided valuable insight into the likely soundscape during the COVID-19 lockdown. However, it does not yet represent a generalisable soundscape model which would be necessary to address these other applications of predictive modelling. Towards this goal, I therefore present a forward-looking proposal for expanding this model to include the additional types of soundscape indicators introduced in Chapter 2.

In Chapter 6 I develop the overall goals of a generalised predictive soundscape model and the development constraints that these requirements place on such a model. I describe how we can begin to build on the model from Chapter 5, incorporating architectural and visual information, and make it generalisable to new locations, contexts, and applications.

Chapter 7 addresses the necessity of incorporating semantic information about the sound source type into a model for predicting soundscape perception. I illustrate how this information can be integrated by presenting the results of my study on predicting noise annoyance with a MLM incorporating sound source information (Orga et al., 2021)[3]. Finally, I discuss how sound source information could be integrated into the general model in an automated way.

In Chapter 8 I investigate to what extent personal factors (e.g. age, gender, psychological well-being) influence soundscape perception. I constructed a third MLM and performed feature selection to determine the relative importance of these personal factors (Erfanian et al., 2021)[4]. Several approaches for how this information could be integrated into a practical predictive model are proposed, following the constraints laid out in Chapter 6.

Finally, in Chapter 9 I present a new method for analysing and representing soundscape assessments, published as Mitchell, Aletta, and Kang (2022). This method emphasises the necessity of considering the variation in people's responses to a soundscape. Given this variation, the soundscape of a space cannot be represented by an single person's assessment, but is better thought of as a

[3]The content in this study and chapter are my own work, conducted as part of this thesis. I shared first-authorship with Dr. Ferran Orga for the published version and his contributions to the text included in the chapter are clearly marked and referenced.

[4]The content in this study and chapter are my own work, conducted as part of this thesis. I shared co-authorship with Ms. Mercede Erfanian for the published version. The text included in this chapter represents my portion of the original work. Where Ms. Erfanian drafted the original text, this has been clearly indicated and referenced.

collective perception. As such, the method I have developed represents a soundscape as a distribution of soundscape assessments. This chapter introduces this visualisation method and its implications. I also provide a proof-of-concept demonstration for how a predictive soundscape model can reflect this consideration of the natural variation in perception.

Chapter 6.

The predictive soundscape model framework

6.1. Goals

Before defining what form a general practical predictive model should take, we first need to make clear what the goals of such a model is. First, that it to a reasonable extent is successful in predicting the collective perception of a soundscape. As described in detail in Chapter 9, it should succeed at both indicating the central tendency of the soundscape perception, but importantly it should also inform the likely spread of perception among the population. The outcome of the predictive model should not be focussed on predicting an individual assessment; the goal is not to predict the perception of any specific individual, but to reflect the public's perception of a public space. In other words, ideally the model will result in an accurate distribution of soundscape perceptions for the target population.

Second, that it can be implemented automatically. Once an initial setup is performed, such as identifying what location the measurements are conducted in, the model should be capable of moving from recorded information to predicted soundscape distribution without human intervention. We need soundscape to be able to be performed instrumentally. This enables it to be applied to unmanned uses, such as smart city sensors and soundscape mapping. It is impractical to conduct soundscape surveys or soundwalks in every location we wish to map and certainly not when we wish to see how these locations change over longer periods of time. A predictive model should allow us to survey these soundscapes remotely in order to extend soundscape to city-scale assessments.

Third, the model should enable us to test and score proposed interventions. In a design context, it is crucial that various design strategies and interventions can be tested and that the influencing factors can be identified. The model should assist the user in highlighting what factor is limiting the success of a soundscape, spark ideas for how to address it, and allow these ideas to be tested. Several other useful features of predictive soundscape models arise out of these goals and will be discussed later, but these form the core goals of the framework.

6.2. Constraints

If we accept that predictive models are necessary to advance a more holistic approach to urban sound in smart cities, we must then define the constraints of such a model. The goal here is to define a framework for what is needed from a future model intended to be used in a smart city sensors, soundscape mapping, or urban design context.

The first constraint is that the model must be based on measurable factors. By this, I mean the data which eventually feeds into the predictive model should be collected via sensor measurements of one sort or another; this could be acoustic sound level measurements or recordings, environmental measurements, video recordings, or GIS measurements, etc.. What it certainly cannot include is perceptual data. This is strictly a practical constraint - for a predictive model designed to be used in practice, there is no justification to include other perceptual factors derived from surveys but not whichever factor you desire to predict. If the goal is to predict soundscape pleasantness and it is necessary to survey people about the visual pleasantness, why not just also survey them about the soundscape pleasantness directly? Certainly this mix of perceptual data is useful in research and can elucidate the relationships between the sonic and visual environments, but it is not useful in a practical predictive context. Any results which arise from research combining this sort of perceptual information must eventually be translated into a component which can itself be measured or modelled.

The second constraint is that any analysis of the measured data can be done automatically, without human intervention. If the eventual goal is to deploy the model on continuously-running, unmanned sensor nodes or to enable practical large-scale measurements, the predictive model should be capable of operating with minimal human input. This means, for instance, that if the model includes information about the sound source, this identification of the source should be possible to do automatically (i.e. through environmental sound recognition). Towards this goal, and given the current practical limitations of environmental sound recognition, a model using a manually-labelled sound-source data is used in Chapter 7 to investigate the future potential of sound-source aware prediction models.

The third constraint is for the model to be generalisable to new locations. Ideally, it will be generalisable to new and (to it) unfamiliar soundscape types, but the minimum requirement should be that it can be applied to new locations which are otherwise similar to those in the training data. This means that any factors which are used to characterise the context provided by the location should be distinguished from a simple label of the location and should instead be derived from measurements of the location. In practice this could be geographical or architectural characteristics of the space, a proposed use-case of the space, or consistent visual characteristics of the space such as the proportion of pavement to green elements. This is in contrast to the model created for Chapter 5 which was constrained to be used only on those locations included in the training data since it made use of a location label.

For this third point, some aspects of the first and second constraints can be relaxed. Since this would only need to be defined once for a location, definitions such as the use case of the space could be defined by the person using the model. What is necessary is that the model and its component location-context factors can be set up ahead of time by the user, then the recording-level effects are able to be calculated automatically. In the MLM context this essentially amounts to choosing the appropriate location-level coefficients ahead of time then automatically calculating the features which are fed into those coefficients (per constraint 1 & 2).

A potential constraint for some applications is related to computation time. Since one proposed application of a predictive soundscape model is to embed the model on a WASN node, the model would then need to be able to run on relatively low-power hardware such as a Raspberry Pi with a reasonable latency. This would especially present an issue for a model which relies on the combination of several psychoacoustic features, such as that in Chapter 5 since these features are computationally intensive to calculate and several of them may need to be computed for each time step of the model. Although this is a real practical concern that should be addressed in the future, for the sake of this initial definition of a general predictive model used across many applications, I have not considered this as a strict constraint. The model being developed here would primarily be intended to operate as an off-line design tool operating on standard desktop hardware and not necessarily requiring real-time calculations. In the case of a WASN, a model of this sort could still be used by sending recording information from the node to a central computer for further computation and analysis. Further efficiency improvements and a specific algorithm for embedding on the node is left as a future development.

Finally, the model should be robust to missing components. If the original or full construction of the model depends on demographic information of the population using the space, in cases where this information is not available, it should be possible to omit it and still obtain a reasonable result. Here we may define potential 'must-have' and 'optional' factors. Given the amount of variance explained by the various factors explored in this thesis, in-depth acoustic information is a must-have, while demographic and personal factors are an optional factor where the trade-off of losing 3% of the explained variance in eventfulness (as will be shown in Chapter 8) is accepted as reasonable. Based on the results of Chapter 5, it would appear that location-context is crucial for modelling the pleasantness, but not for modelling the eventfulness. In order to determine the must-have factors for characterising the location-context, more work will need to be done to determine the appropriate input factors and their relative importance. The first step of this is to bring the model in line with constraint 3.

6.3. Expansions and advancements for future predictive models

To illustrate how a model which fits this proposed framework could be developed, I will start with the model presented in Chapter 5. As it is, this model most obviously violates constraint 3 - it is not generalisable to new locations. Since the structure of the MLM has the 'LocationID' as the categorical feature used in the second level, any new data must be able to conform to one of the original 13 locations. Technically, it would be possible to select the location in the training data which is considered most similar to the new location and use the coefficients derived for it, but this is a poor design for a generalisable model. Therefore the first stage to generalise this starting point model would be to replace the location-label variable with a more general categorical description of the location-context.

6.3.1. Incorporating architectural and visual information

The simplest version of this replacement would be to sort the locations into a predefined architectural or landscape location type. Suligowski, Ciupa, and Cudny (2021) presents a definition in which urban spaces can be classed as 'green' (unsealed, permeable, biologically active areas), 'blue' (water areas), or 'grey' (human made, predominantly formed by sealed, impermeable, hard surfaces built from concrete or tarmac). Thus we could sort the 13 locations used in the model according to whether they would be classed as green, blue, or grey and reconstruct the multilevel model using these categories in place of the location labels. In this way, new locations could then be similarly identified and fed into the model.

However, this simplified method has a few potential drawbacks. The green-blue-grey classification likely would not capture a wide enough array of potential landscape or architectural types and therefore would limit the ability of the model to differentiate the varying relationships between the acoustic features and the soundscape perception. In addition, the green-grey-blue paradigm does not provide an indication of the visual quality of the space. Although it might be assumed that green spaces are visually pleasant and grey spaces less so, there would presumably be some spectrum of quality within each of these categories which may provide additional information for the prediction of soundscape quality.

An alternative method is to make use of the visual information about the locations which can be capture as part of the SSID protocol. The most straightforward method for this is to make use of a clustering algorithm to analyse measured features of the spaces and sort a given location into one of several location types. From a visual analysis model such as the FaceLift model presented by Joglekar et al. (2020), it is possible to extract elements such as the percent of visible sky (i.e. openness), the percent of greenness, and an overall visual quality rating from photos and videos taken in the space.

By applying an unsupervised clustering algorithm to this visual data, many more categories of the architectural and visual characteristics of the space can be derived and measurements of new spaces can likewise be assigned to these categories.

With this method, we would then have a two-stage model, where the first stage is to measure the visual characteristics of the space and, using the clustering algorithm, sort it into one of the identified categories. This category would be assigned to all subsequent acoustic measurements taken in the space as they are fed into the MLM to predict the likely soundscape assessment.

6.3.2. Additional acoustic, psychoacoustic, and bioacoustic metrics

Although the model presented in Chapter 5 began its feature selection with a relatively wide array of potential psychoacoustic features, many aspects of the sound were not captured and many potential metrics were not included. This includes additional statistical breakdowns of the included features - for instance, L_{A90} (as a measure of the background level), $N_{10} - N_{90}$, $L_{A,max}$ and $L_{A,min}$, etc. There are also a host of bioacoustic metrics which were not considered, such as those presented in Devos (2016), including the Acoustic Complexity Index (ACI), Normalized Difference Soundscape Index (NDSI), Bioacoustic Index (BIO), and so on.

An important sonic feature which is underutilised is the temporal behaviour of the sound. A few metrics are able to capture some aspects of how the sound changes over time, such as FS looking at amplitude modulations in the sub-audible range, but this is still limited. One approach to characterising the temporal structure of complex acoustic scenes is through $1/f$ analysis (De Coensel & Botteldooren, 2006; De Coensel, Botteldooren, & de Muer, 2003; M. Yang, De Coensel, & Kang, 2015).

Future work on expanding the predictive model should begin by considering these additional metrics and exploring their potential as new and better-performing input features for predicting soundscape perception.

With this increased slate of potential input parameters, a more efficient feature selection method will need to be employed, compared to the stepwise feature selection used throughout this thesis. The multicollinearity between the candidate features, the increased training time, and the ratio between independent variables and sample size make it infeasible to apply a stepwise selection with a large number of candidate features.

In addition to a more sophisticated analysis of the acoustic characteristics of the sound, addressing the semantic meaning which listeners attach to certain sources is an important advance in predicting soundscape perception. The next chapter presents how I have begun to address this goal using data collected via a WASN.

Chapter 7.

Towards incorporating sound source information

[1] In L. Brown (2009), the author proposes that one of the underutilised concepts in making use of sound as a resource is the disaggregation of sound sources. He states that 'the type of sound sources present is critical in judgements about outdoor sound quality'. The goal is to move away from the straightforward use of aggregate sound metrics, which attempt to summarise the sound environment as a whole, through various acoustic metrics. Given the various semantic meanings that listeners associate with certain sound sources, the annoyance elicited by particular sounds will vary as will the relationship between the acoustic features of that sound and the annoyance (Lafay, Rossignol, Misdariis, Lagrange, & Petiot, 2018).

Given the nature of the ISD as a large dataset containing hundreds of *in-situ* recordings, identifying sound sources manually was impractical at this stage. In order to progress towards a sound-source aware model, I therefore partnered with researchers from the LIFE DYNAMAP project who had curated a set of labelled recordings selected from a WASN installed in Milan (Italy). For this study, the DYNAMAP researchers asked more than 100 people to conduct three different perceptual tests through an online survey (R. M. Alsina-Pagès, Freixes, et al., 2021).

The perceptual tests were designed to measure the annoyance in people relating to different urban sounds and their characteristics (R. M. Alsina-Pagès, Freixes, et al., 2021; Labairu-Trenchs, Alsina-Pagès, Orga, & Foraster, 2018), by means of short excerpts of raw acoustic audio obtained from the DYNAMAP project (Sevillano et al., 2016). The audio excerpts which were most representative of the site were selected, using a wide range of sound types (sirens, airplanes, people talking, dogs barking, etc.) (Alías, Orga, Alsina-Pagès, & Socoró, 2020; Alías, Socoró, & Alsina-Pagès, 2020).

[1] The content of this chapter was originally published as Orga et al. (2021), a collaborative work between myself from the SSID team at UCL and Dr Ferran Orga, a researcher at Grup de Recerca en Tecnologies Média, La Salle-URL. Dr Orga and I shared first authorship on this paper. Original data collection was performed by the team at La Salle-URL while the data analysis and modelling strategy was conceived by the team at UCL and implemented by me. Dr Orga and myself drafted the original manuscript, with my work focussing on the analysis method section, results, and discussion of the modelling results.

Sound annoyance depends on the acoustic characterisation of each sample, and it is possible to classify the acoustic excerpts depending on their sound source characterisation, which can be the basis to ask participants about their perceptions. The psychoacoustic characterisation is based on the psychoacoustic measurements of loudness, sharpness, and others defined by Zwicker and Fastl (2007).

Based on the data collected by the DYNAMAP team, I aim to determine the psychoacoustic parameters that have an effect in the individual annoyance scores, and how the relationships between these parameters and annoyance may vary according to the dominant sound source. For this reason, a multilevel psychoacoustic model is trained using the results of the MUSHRA test (Rec. ITU-R BS.1534, 2015), focused on annoyance evaluation by the participants over several different types of sound. The results show that sound source identification provides valuable information for a predictive model and that sharpness is a primary predictor for annoyance which is independent of the sound source.

7.1. Methods

7.1.1. Dataset

This study makes use of a dataset collected in collaboration with the LIFE DYNAMAP[2] project conducted in Milan (Italy) (Alías, Socoró, & Alsina-Pagès, 2020; Sevillano et al., 2016). This project makes use of a WASN, enabling the collection of data over a longer period of time than was possible with the SSID protocol outlined in Chapter 3. A WASN enables a broader characterisation of the acoustic events present in a location, as recording conditions can be made consistent across the nodes and data can be retrieved at any time of the day.

The dataset used in this study has been obtained by homogeneously sampling several hours, in both weekday and weekend, with 24 sensors distributed along the urban District 9 of Milan (R. Alsina-Pagès, Alías, Socoró, & Orga, 2018). After that, experts from the DYNAMAP development team labelled the acoustic events of the recordings manually to obtain a 151-h dataset (Alías, Socoró, & Alsina-Pagès, 2020). Due to the nature of the project, this consisted in removing events not related to traffic noise from the noise map computation, events were grouped in Road Traffic Noise (RTN) that belongs to the 83.7% of the total time of the dataset, and Anomalous Noise Event (ANE) with the 8.7% of the total time. Another class was used to include overlapping and unidentified events: complex (COMPLX) with 7.6% of the total time (Alías, Orga, et al., 2020). During the labelling process, the DYNAMAP developers found up to 26 types of anomalous events, which

[2]The data collection (both the collection of the recordings and the online survey) was performed by the DYNAMAP team. To maintain consistency with the published version of this study and to provide the appropriate context, the text on data collection (Sections 7.1.1 and 7.1.2) has been reproduced verbatim from our study (Orga et al., 2021) and was initially drafted by Dr. Orga, the other first author.

they decided to group in the following classes: airplane, alarm, bell, bike, bird, blind, brake, bus door, construction, dog, door, glass, horn, interference, music, people, rain, rubbish service, siren, squeak, step, thunder, tramway, train, trolley, wind, works (construction) (Alías & Socoró, 2017).

The most common sound classes were picked to evaluate the relationship between the event measurements and the citizens' perception of annoyance. These selected events used in the study belong to the following 9 classes: airplane, bird, brake, construction, dog, door, horn, people, and siren (Orga, Alías, & Alsina-Pagès, 2017). As the selected events are the most common, those are the ones that contain the widest variety of recording conditions, including different sensor locations and recording hours (Labairu-Trenchs et al., 2018). The reason for that choice was two-fold:

1. the availability of a wide range of examples of each type of sound to choose for the design of the tests, including the possibility of finding different samples that keep similar psychoacoustic values,
2. the fact that the most common sounds are the most reasonable to evaluate with people, as they best summarise the character of the soundscape around each sensor.

More details about the event selection process and the availability of the study sensors are detailed in (Labairu-Trenchs et al., 2018), and the time of each event in the sensors is depicted in (R. M. Alsina-Pagès, Freixes, et al., 2021).

7.1.2. Design of the Perceptual Tests

In order to assess the degree of annoyance produced by the aforementioned classes of sounds, an online test was conducted using the Web Audio Evaluation Tool (Jillings, Man, Moffat, Reiss, et al., 2015). Specifically, the MUSHRA test method (Rec. ITU-R BS.1534, 2015) – which was originally designed for the evaluation of audio codecs – has been adapted for this purpose. Participants were given a clear explanation of what they were going to be asked, including detailed instructions on the operation of the test. No training phase was therefore considered. A demographic survey was included at the beginning of the test for all 100 participants, asking for them to identify their age, gender, and a subjective rating of the participant's residential area (zr1 - very quiet, zr2 - quiet, zr3 - bit noisy, zr4 - noisy, zr5 - very noisy).

The second part of the test consists of five sets. Each set presents a group of short acoustic events with similar values of loudness and sharpness but from different classes, and recorded in the same sensor, in order to maintain the recording conditions and location of the sounds under comparison. For each set, the participants were asked to evaluate the annoyance produced by the presented recordings, ordering them in a $0 - 10$ scale, where zero corresponds to *not at all* and 10 corresponds to *extremely disturbing* following the ICBEN recommendation. The interface was customised including a colour scale to help the participants place the stimuli according to the degree of annoy-

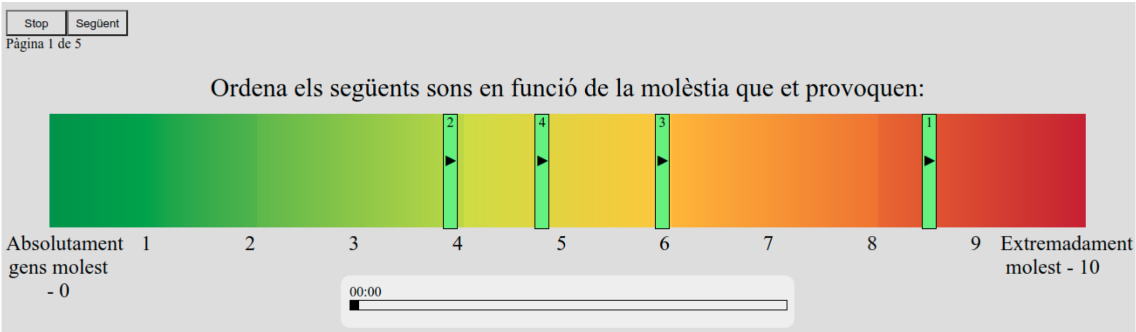


Figure 7.1.: Screenshot of the MUSHRA test conducted to assess the annoyance provoked by different sounds. Title: sort the following sounds according to the cause annoyance. The scale ranges from *not annoying at all* to *extremely annoying*.

ance that they perceive. Each audio is represented with a green bar with a 'play' icon on it and the recordings are sorted randomly along the MUSHRA scale (see Figure ??). An audio recording is reproduced when the corresponding bar is clicked. The system ensures the participant listens to all the recordings and moves all the bars before they jump to the next set of recordings. The sets were presented in a random order to prevent learning biases. MUSHRA tests usually include hidden reference stimuli, which in audio or speech quality evaluation corresponds to the highest quality samples and that are used to remove outlier responses.Since stimuli pertaining to different classes are compared, no audio reference was included, thus avoiding biases towards a certain audio class. The participants were asked to take the test using headphones and to keep the same volume during all the tests, to maintain the same conditions throughout the entire testing process. One hundred participants undertook this test, 59 men and 41 women, with a mean age of 33. Participants were volunteers, mainly from the university and also gathered via social networks. The distribution according to residential area is the following: 9 in zr1, 37 in zr2, 35 in zr3, 18 in zr4, and 1 in zr5. The MUSHRA test allows us to:

1. obtain an individual score of annoyance for each audio,
2. carry out comparisons among the different types of events contained in a set.

The detail of the stimuli included in each of the five sets of the test can be found in Table 7.1.

7.1.3. Psychoacoustic Data Analysis

The dataset resulted in 27 audio recordings of identified sound events with durations ranging between 1.01 and 2.69 s. The calibrated audio files were imported into the ArtemiS Suite software (v. 11.5, Head Acoustics GmbH) and the following psychoacoustic parameters were computed: *loudness, sharpness, roughness, tonality* and *impulsiveness* (Zwicker & Fastl, 2007); values for these parameters are reported in Table 7.1. The rationale for selecting a relatively large set of psychoacoustic metrics is that they are often used as indicators to predict perceptual constructs (such as annoyance) in perceptual studies, as shown in recent soundscape literature (Aletta et al., 2017, 2016). Fluctuation Strength, which could otherwise be included in this list of psychoacoustic parameters, as in Zwicker's annoyance model, was not included as the length of the recordings are too short to obtain a valid value. Loudness was calculated according to the DIN 45631 / A1 standard for time-varying sounds, in a free-field (DIN 45631/A1, 2008). As recommended by the standard, in order to avoid the under-estimation of evaluated loudness which is seen when using the arithmetic average of the loudness curve, the N_5 value (the 5% percentile value of the time-dependent loudness curve) is used as the single value of loudness. Sharpness was calculated according to DIN 45692 (2009), in a free-field. With this sharpness method, the absolute loudness of the sound is not accounted for, so there should not be a duplication of information across the loudness and sharpness metrics. Roughness was calculated according to the hearing model by Sottek (2017), with the option to skip the first 0.5 s in order to not distort the single value. Impulsiveness was also calculated according to the hearing model by Sottek, with a 0.5 s skip interval. Finally, tonality was calculated according to ECMA-74 (17th Edition) (2019), which is based on the hearing model of Sottek, with a frequency range of 20 Hz to 20 kHz.

7.1.4. Multi-Level Linear Regression Modelling

The analysis for this study utilises a MLM, with a random intercept and a random slope, using backward step feature selection. MLMs are commonly used in psychological research for repeated measures studies (Quené & van den Bergh, 2004; Volpert-Esmond et al., 2021) and for applied prediction models (Frees & Kim, 2006; Gelman, 2006). MLMs allow for the incorporation of nested and non-nested groups effects within the structure of the model, where the coefficients and intercepts for the independent variables are allowed to vary across groups. For this study, the data are grouped into two non-nested sets to form a two-level model: by repeated measures per respondent (*user*) and by sound type (*label*). In order to take into account the repeated measures across participants, and to correct for the participant's mean annoyance level, the *user* variable is included in the second-level as a random intercept. We then include the psychoacoustic features as label effects, with coefficients which are allowed to vary across the sound type labels. The psychoacoustic features are also included as fixed effects in the first level, which do not vary across either the user or label groups.

The initial model structure, as written in Wilkinson-Rogers notation (Wilkinson & Rogers, 1973), is thus:

Table 7.1.: Psychoacoustic parameters calculated for the 27 stimuli used in the listening experiment.

Sensor	**Label**	**Psychoacoustic Parameters**				
		Loudness (N_5 sone)	**Sharpness (acum)**	**Roughness (asper)**	**Tonality (tuHMS)**	**Impulsiveness (iu)**
hb133	peop	15.1	1.46	0.032	0.204	0.270
hb133	door	16.8	1.43	0.029	0.113	0.354
hb133	dog	13.1	1.22	0.033	0.373	0.266
hb133	brak	16.0	1.76	0.030	0.326	0.241
hb133	bird	12.6	1.73	0.024	0.283	0.214
hb133	airp	13.0	1.27	0.060	0.438	0.231
hb127	sire	17.7	1.56	0.045	1.540	0.178
hb127	peop	16.1	1.62	0.035	0.410	0.417
hb127	horn	18.1	1.56	0.028	0.666	0.260
hb127	door	19.8	1.72	0.037	0.037	0.479
hb127	brak	19.0	1.95	0.034	0.251	0.281
hb127	sire	20.1	1.73	0.046	1.670	0.288
hb127	peop	22.0	1.96	0.036	0.322	0.452
hb127	horn	19.9	2.16	0.034	1.290	0.336
hb127	brak	21.0	1.81	0.030	1.170	0.285
hb127	airp	24.4	1.65	0.056	0.172	0.446
hb115	wrks	20.3	1.97	0.054	0.227	0.267
hb115	trck	24.4	1.60	0.033	0.040	0.276
hb115	sire	19.5	1.46	0.054	0.861	0.333
hb115	peop	25.1	1.79	0.032	0.411	0.331
hb115	horn	22.3	2.00	0.032	0.806	0.155
hb115	door	26.3	1.62	0.038	0.045	0.397
hb115	brak	20.6	1.93	0.034	0.216	0.313
hb115	wrks	24.6	1.92	0.064	0.447	0.317
hb115	sire	26.6	1.77	0.044	0.626	0.290
hb115	horn	29.5	2.35	0.039	0.486	0.262
hb115	door	31.3	1.88	0.048	0.223	0.402

$$Annoyance \sim N_5 + R + S + T + I + (1|user) + (1 + N_5 + R + S + T + I|label) \quad (7.1)$$

Feature Selection

The MLM is initially fitted with all of the potential features included within both levels. In order to reduce the complexity of the model, a backwards step features selection process is applied to both levels of the model. This process involves fitting the full model which includes all of the potential independent features (i.e. Eq. (7.1)). The feature with the highest *p*-value (least significant)[3] is then removed from the candidates and the model is refit. This process is repeated until all features meet the predefined significance threshold of $p < 0.05$. For a two-level model, first backward elimination of the second level is performed, followed by backward elimination of the first-level (or fixed) part.

If more than one feature is selected in the first-level, then the VIF is calculated in order to check for multicollinearity, with a pre-determined threshold of $VIF < 5$. Any features which remain after the backwards stepwise selection and exceeded this threshold were investigated and removed if they were highly collinear with the other features. Once the feature selection process is completed, the final model with only significant features of interest included is fit and the table of the model coefficients is printed along with plots of the random effects and standardised estimates terms. Finally, quantile plots of the residuals and random effects are examined to confirm they are normally distributed (Harrison et al., 2018).

The input and output features are z-scaled prior to the analysis and model building by subtracting the mean and dividing by the standard deviation in order to directly compare the coefficient values of independent variables measured on different scales (Harrison et al., 2018). The model fitting and feature selection was performed using the `step` function from `lmerTest` (v. 3.1.3) (Kuznetsova et al., 2017) in the R statistical software (v. 4.0.5) (R Core Team, 2018). The summaries and plots were created using the `sjPlot` package (v. 2.8.7) (Lüdecke, 2021) and the multi-level R^2 values were calculated using `MuMIn` (v. 1.43.17) (Barton, 2020).

[3]Note that this process differs slightly from that taken in Chapter 5 which used the AIC as an overall performance metric as opposed to a *p*-value per feature method used here. In this case, this change represents a development of my modelling expertise, as the study presented in this chapter was actually completed before (2021) the study presented in Chapter 5 (2022). While both are valid methods and may have different ideal applications, for a prediction-focussed approach, the AIC selection method is more approrpriate.

7.2. Results

7.2.1. Differences in Annoyance between Groups

[4]The average annoyance score of all users across all stimuli was $M = 0.58(SD = 0.05)$. Since some basic demographic information about the 100 participants of the perceptual test was known, it seemed logical to explore possible differences in annoyance scores between different groups/levels of stratification of the sample, mostly for descriptive purposes. Therefore, Areas of residence and Gender were considered as factors in this analysis. Gender was treated as a binary variable (F/M), while Areas of residence was treated as a five-level categorical variable based on people's self-reported character of the area where they typically reside (range: 1-5; very quiet-very noisy). One-way repeated measures ANalysis Of VAriance (ANOVA) was deemed to be the most appropriate approach to take into account the multiple responses that each of the 100 participants provided for the different recordings ($N = 27$). A first analysis was then conducted to determine whether there was a statistically significant difference in annoyance between Areas of residence: no statistically significant differences were observed in this case $F(4.95) = 1.374, p = 0.249$. Likewise, a second one-way repeated measures ANOVA was carried out to check whether statistically significant differences in annoyance existed between females and males: no statistically significant effect was observed in this case either $F(1.98) = 0.714, p = 0.400$. Such small differences between groups can indeed be observed in Figure **??**.

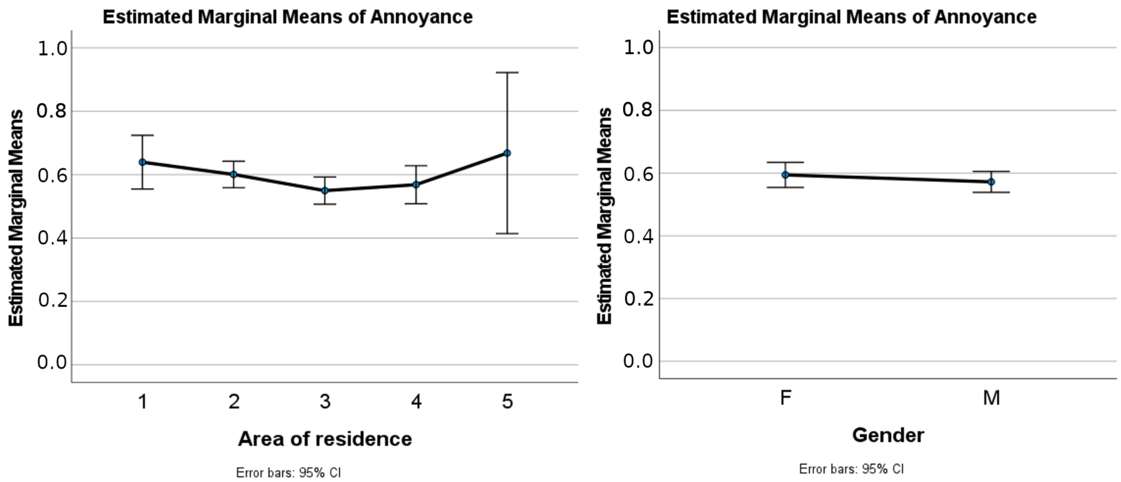


Figure 7.2.: Estimated Marginal Means for Annoyance as a function of Areas of residence (**left**) and Gender (**right**).

[4]The analysis carried out in Section 7.2.1 was performed by Dr. Francesco Aletta. These results are included here verbatim from the original published paper to provide context for the later discussion on the influence of demographic differences on soundscape perception.

7.2.2. Annoyance Model

In the context of the multi-level linear regression modelling, the included variables were assumed to have an effect at two levels: the first level (i.e. fixed effect(s)), and the second level, where annoyance score intercepts are allowed to vary as a function of users (i.e. the 100 participants), and where each feature of interest is allowed its own coefficient as a function of labels (i.e. the 7 types of sounds). Sharpness came up as the main predictor with a strong statistical significance in the fixed-effect level, as reported in Table 7.2. This implies that, regardless of any other factors, the sharper the sounds, the more annoying they are perceived to be.

Table 7.2.: Random-intercept random-slope multi-level model of psychaocoustic annoyance, accounting for repeated measures (user) and sound source type (label) within the second level. Coefficients and confidence intervals given are for z-scaled data.

	Annoyance		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.02	-0.13 – 0.16	0.811
Sharpness	0.33	0.25 – 0.40	**<0.001**
Random Effects			
σ^2	0.47		
τ_{00user}	0.28		
$\tau_{00label}$	0.02		
ICC	0.39		
N_{user}	100		
N_{label}	10		
Observations	2700		
Marginal R^2 / Conditional R^2	0.08 / 0.64		

The second-level effects presented in Figure **??** show that level- and loudness-based acoustic parameters do not play a significant role in predicting annoyance when considering other psychoacoustic factors and specific sound sources. However it should be noted that this may be influenced by the online data collection paradigm used in this study, which may struggle to control for the playback level. The variables selected by the feature selection algorithm within the type of sound (*label*) level include: impulsiveness, roughness, tonality, and type of sound are relatively small, while roughness appears to be more important. For instance, when other effects are controlled, the sound type 'horn' seems to be less annoying the rougher it is; while for the types of sound 'bird' and 'siren', higher roughness values will lead to higher annoyance scores. Looking at the model from the point of view of the types of sound, one could observe that 'horns' tend to be more annoying than other sounds if they are more impulsive, while 'people' or 'birds' or 'brakes' result in more annoying scores compared to other sounds if their tonal components are more prominent. Overall, for this model, the marginal and conditional R^2 values are 0.08 and 0.64, respectively. Marginal R^2 provides the variance explained by the fixed effects only, and conditional R^2 provides the variance explained by the whole model, i.e. both fixed effects and second-level effects. Thus, the majority of variance is explained by second-level factors, while a smaller portion (8%) is covered by sharpness alone.

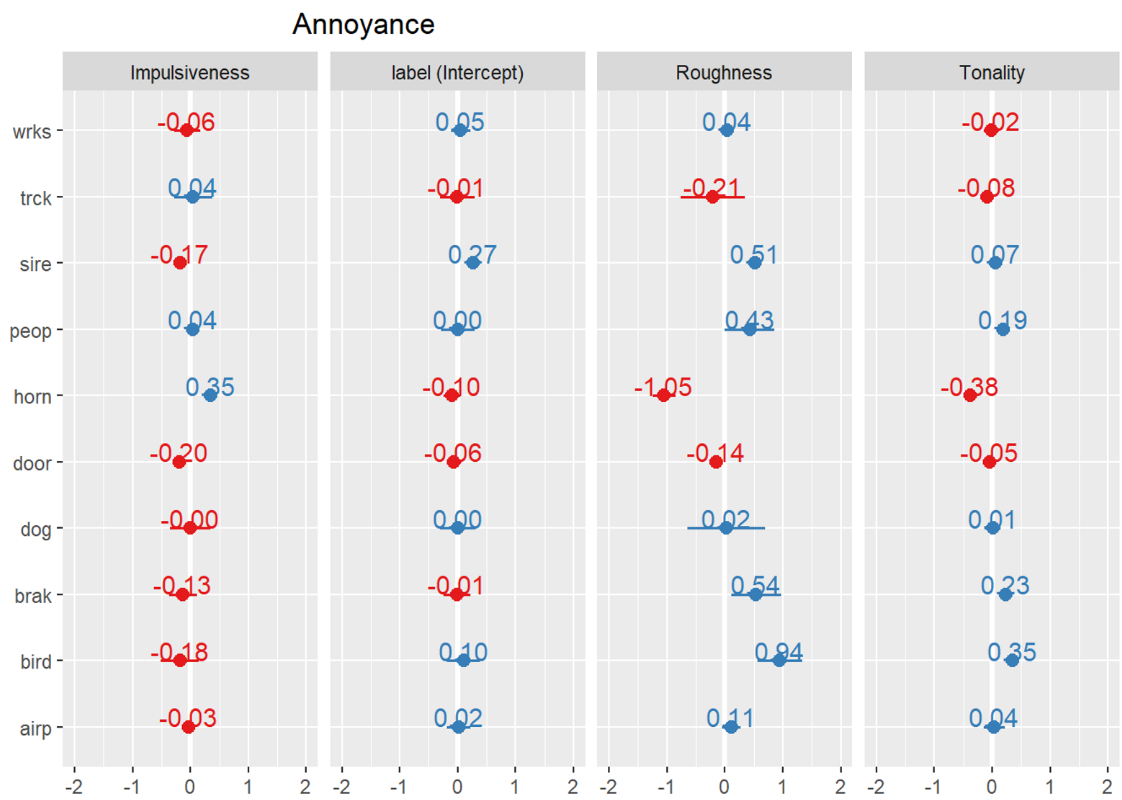


Figure 7.3.: Second-level effects figures representing the regression coefficients by types of sound (label) and for different psychoacoustic parameters.

7.3. Discussion

The fact that no significant differences in annoyance scores were observed between sample groups (i.e. gender or area of residence) is particularly interesting: it is common to assume in soundscape studies that personal and contextual factors play a strong role in how people respond to urban acoustic environments (Kang et al., 2016). However, this is probably more relevant when complex sound environments (e.g. multi-source) are being considered and when dealing with relatively longer duration of exposures (e.g. several minutes) as seen in in-situ surveys. For clearly identifiable sources of environmental noise, with signals of short duration (i.e. 1-3s) like those used for this experiment, it is likely it was easier for the sample to converge on similar annoyance scores, regardless of other demographic factors. This aspect will be further investigated in Chapter 8.

Regarding the noise annoyance scores, sharpness came up as an important predictor in the first level of the modelling stage (explaining up to 8% of the variance alone). It is important to highlight that the sharpness calculation method used in this study did not include any loudness correction; nor was any loudness-related parameter selected by the feature selection algorithm. To some extent, this is possibly due to the fact that, being an online experiment, it was not possible for the research team to actually calibrate the loudness playback level accurately for the remote participants. On the other hand, considering this aspect from the WASN implementation perspective, this could be seen as an encouraging finding, since calibrating a diffuse acoustic monitoring network may not be practical in real-world scenarios, so it is good to have models that can achieve up to 64% of variance explained regardless of actual levels. Furthermore, in complex acoustic environments, loudness would likely vary over time depending on the relative positions between sound sources and (human) listeners in ways in which the other psychoacoustic parameters such as sharpness and tonality are less likely to. This is something that is impossible for fixed sensors to take into account, so once again it is preferable not to rely on loudness as a predictor.

This result also appears to differ from some of the results in the model from Chapter 5, where sharpness was not selected as a final feature in the ISOPleasant model. There are a few potential explanations for this. First, although within the circumplex pleasantness is considered the opposite of annoying, the respondents' annoyance rating in this study may be focussed on more specific or different factors than what is captured in the combined ISOPleasant score. Second, since the ISD data is collected *in-situ*, the general pleasantness of the soundscape may emphasise different acoustic features than the online study procedure used in this study. Finally, the structure of this model effectively controls for the influence of sound source information, whereas the Chapter 5 model has no source information included. This could be interpreted as the sound-source-aware model more correctly identifying the acoustic feature which is important, independent of the sound source. Whereas the previous model's feature selection results may be more likely to select the features which help to differentiate sound sources, which is not necessary in the sound-source-aware model. This would indicate that the features selected in the Chapter 5 model were selected because they perform better at differentiating sound sources and accounting for the semantic meaning associated with sound sources, but when the semantic information itself is included in the model, other acoustic features are more important for determining annoyance.

Being able to predict noise annoyance from recorded sounds is particularly helpful from a public health perspective. In the context of a smart-city framework, one could imagine a WASN large enough to cover a whole urban area; having a noise annoyance prediction algorithm at the node position that can return live annoyance scores to a central server from sounds recorded locally by the sensor would make for a useful application for environmental protection officers and other stakeholders at community or local authority level (Kang & Aletta, 2018). A relevant issue to consider from the WASN perspective, is that in previous studies conducted in both urban (Alías, Socoró, &

Alsina-Pagès, 2020) and suburban (Alías, Orga, et al., 2020) environments, a clear influence of the type of environment around the sensor location on the types of noise was seen. Not all urban and suburban locations around the sensors have frequent sirens or horns. The presence of these sounds depends on the most common activities (e.g. leisure, hospitals, etc.), the type of road (wide, narrow), and the type of buildings and houses surrounding the location. The types of sounds and their relative frequency of occurence can vary widely given the combination of these architectural and landscape characteristics. In the design of a general model for quality of life, the approach for considering sound source information presented in this chapter should be combined with the proposal for incorporating this architectural and contextual information developed previously in Chapter 6.

7.3.1. Incorporating into the general model

Incorporating information about the sound source can greatly improve the prediction of perceived annoyance. The modelling structure used in this chapter effectively created separate, sound-source-dependent models of psychoacoustic annoyance, in contrast to the general annoyance model developed by Zwicker and Fastl (2007). Each sound source (traffic, horns, people, etc.) has its own linear combination of psychoacoustic values to demonstrate how the semantic meaning which the listener assigns to a sound mitigates the perceptual mapping from physical inputs to perceptual outcomes. Incorporating this information, such that this semantic meaning influences the outcome of the ideal predictive model, is crucial.

To demonstrate how this information could be integrated into the overall model, we can return to the model created for Chapter 5. For the purposes of this, we will assume that sound source labels have been derived for the ISD data in the same way as was done for the DYNAMAP dataset, resulting in one label per recording. This creates quite a complex series of relationships, where not only do we expect the relationship between a psychoacoustic feature and the perception to change depending on the sound source, but also according to the location-context, and in such a way that the location-context may change either the direct psychoacoustic→perception mapping or the source+psychoacoustic→perception mapping. To integrate this series of relationships into the Chapter 5 model we can include the sound source label as an interaction term at both the first and second (location) level[5]:

$$ISOPl \sim 1 + N_5 + N_5 \cdot label + (N_5 + N_5 \cdot label | LocationID) \tag{7.2}$$

where N_5 is used as a stand-in for the various psychoacoustic features we could consider; *label* is the sound source label, defined with the DYNAMAP ontology and taking *wrks* as the reference level.

[5]Here, I use N_5 as a stand-in for the various psychoacoustic features we could consider. This could be substituted with any of the other features discussed or with a set of features.

Unfortunately, since the sound source information is not (yet) available for the ISD, it is not feasible at this point to run an example of this model on the data to demonstrate its performance. This further development forms part of my future work. What is still to be determined is 1) how this process can be automated, with minimal manual input from the model's users; and 2) how to deal with complex scenes where multiple overlapping sound sources are present and how to integrate this into the model.

The first point is simpler, conceptually, but possibly more difficult in practice. An automated sound source recognition algorithm could be used, such as YAMNET (Hershey et al., 2017), based on the AudioSet ontology (Gemmeke et al., 2017) (available as a pre-trained model on MATLAB or for Python in Tensorflow Hub), or a model more specialised for urban environments, such as those presented in the Detection and Classification of Acoustic Scenes and Events (DCASE) challenges (Bello et al., 2019) using the UrbanSound 8k dataset (Salamon, Jacoby, & Bello, 2014). In general, these models slice the recording into short chunks (typically on the order of 1 second) and produce a relative weighting for the likely presence of each available sound class within that second. From this list of weightings, the top predicted sound sound can be extracted for that chunk. If necessary, these can then be summed to give the overall top predicted sound over the entire recording. A model such as this could be integrated into the data processing pipeline to first identify the sound source(s) for a given time step, before combining this with the psychoacoustic analyses for the same recording and fed into the above model. There is also a parallel field of research specifically into polyphonic sound event detection, which would likely suit this task more closely (Mesaros, Heittola, & Virtanen, 2016). This would enable a (theoretically) fully automated, sound-source aware, psychoacoustic model of soundscape perception. However, the lingering question – and the particular reason that this approach wasn't taken from the start of this research given the state-of-the-art at the time – is whether these environmental sound recognition models are accurate enough in the difficult and complex urban sound environments to be feasible for this purpose.

This second point relates to one of the key differences between the DYNAMAP dataset and the ISD data; although the recordings provided by the DYNAMAP team were real recordings of a complex sound environment, they were manually selected as having a single dominant sound source and were much shorter (only 1–2.7s). By contrast, the ISD data are more representative of the total sound environment, meant to reflect all that the respondent was exposed to while completing the questionnaire. This means it contains many overlapping sound sources over 30s to 1 minute, and presumably the respondents were responding to all of this. There may have been a single dominant sound source driving their response, or the soundscape may have been very heterogeneous. Effectively, this represents a weakly-labelled machine learning recognition task for the models, but how these labels are combined to give a usable label for the soundscape prediction model is less clear. Note that this is distinct from the recognition challenges for polyphonic sound event recognition task itself. What we are trying to solve is how to summarise a 30s recording containing many sound sources into a single label which best encapsulates the effect of the sound sources. This is the challenge facing a true soundscape prediction model based on the ISD which aims to reflect how people are exposed to a soundscape in a public space.

This challenge could be tackled in a variety of ways, which no doubt will be developed as the field progresses, but I will propose some possible solutions. Two options would directly use the output from the recognition models. The simplest is to identify the sound with the highest output weighting over the entire recording period and consider this the 'dominant' sound which then feeds into the soundscape model. Similarly, using the results from the recognition model, if the dominant sound in each time step is identified, then we could use the sound label which was identified for the plurality of the time steps, meaning it was the dominant sound for the most time. A similar, but perhaps more sophisticated approach takes inspiration from sleep disturbance. It is now common in sleep disturbance studies, particularly due to aircraft noise, to include information on the number of noise events, usually defined as the number of events with an $L_{A,max}$ which exceeds some threshold (e.g. > 60dB) (Janssen, Centen, Vos, & van Kamp, 2014). We could thus define a threshold level and define sound events as a contiguous period which exceeds the threshold; the recognition model would then be run on the longest contiguous period, giving some indication of the single sound which was noticeable for the longest time throughout the recording.

A more sophisticated approach is to define so-called *sound source profiles*. This would be a set of categorical labels which characterise some particular combination of sound sources; for example, the 'Traffic noise A profile' may be dominated by road traffic sounds with some amount of human voices and sirens present, while the 'Traffic noise B profile' might be dominated by road traffic with loading and unloading sounds and no sirens. Similar specific subprofiles would also be defined for human and natural sound dominated profiles. Since these are categorical, they can easily be incorporated into the multi-level model given in Eq. (7.2). Although a basic version of these profiles could be defined using the sound source dominance questions in the SSID protocol questionnaire (see Appendix A), defining and subsequently assigning recordings to these profiles using the results of the environmental sound recognition models would seem preferable. By calculating e.g. the top ten identified sound sources, an unsupervised clustering algorithm with an appropriate method of determining the optimal number of clusters could help us to define these profiles based on patterns in the sound source data. Once the clusters are defined based on the output of the recognition models, it would be trivial to then process new recordings and classify them into their correct profile, then subsequently feeding this into the soundscape model. In this way, we can define a fairly comprehensive set of possible profiles of sound sources that are often seen together in urban environments and train a predictive soundscape model which can alter its psychoacoustic coefficients depending on the combination of sound sources present. At all stages throughout this process (except the internal calculation of the source recognition model), the definitions used, and the coefficients used throughout the model are transparent, understandable, and traceable.

7.4. Conclusions

In this chapter, an online listening experiment was conducted with 100 participants to assess the noise annoyance induced by short recordings of individual environmental noise sources gathered via a wireless acoustic sensors network in Milan. The main conclusions of this study are:

- The acoustic samples gathered from selected sensors in Milan WASN of the DYNAMAP project led the DYNAMAP team to a structured MUSHRA test to evaluate the annoyance in an offline perceptual test.
- When considering short recordings of single-source environmental sounds, no significant differences in noise annoyance were observed as a function of demographic factors, such as gender and self-reported area of residence (i.e. from very quiet to very noisy).
- The multi-level linear regression model derived from this case study achieved an overall $R^2 = 0.64$, using sharpness as a fixed effect (the first level), and impulsiveness, roughness, and tonality as random effects allowed to vary according to the type of sound (the second level) as predictors for perceived noise annoyance.

By using a consistent MLM modelling strategy, the approach taken in this study highlights how a similar approach can be integrated into the general model. By incorporating sound source information along with the psychoacoustic metrics, the model can better reflect how listeners will respond to different sounds. The results given here demonstrate that the psychoacoustic features of a sound are most for determining annoyance not on their own, but in how they cause us to perceive a certain sound. By allowing the relationship between psychoacoustic features and annoyance to vary per sound source, we create a more representative analogue of the perceptual mapping from soundscape indicators to descriptors.

Given the somewhat unexpected result from this study that demographic factors had little difference on annoyance ratings, the next chapter will further investigate the influence of personal factors on soundscape perception making use of the larger and more diverse dataset from the ISD.

Chapter 8.

Psychological Well-being and Demographic Factors can Mediate Soundscape Pleasantness and Eventfulness

Soundscape studies aim to consider the holistic human perception of a sound environment, including both the physical phenomena and how these are mediated by internal factors. Within the context of the predictive modelling strategy presented in this thesis, the inclusion of personal factors presents a particular challenge. Referring back to our conceptual model of soundscape perception shown in Fig. 2.2, the form of the perceptual mapping which translates the soundscape indicators into a listener's perception (expressed via soundscape descriptors) is influenced by the listener's own psychological state and background. Although we can find general patterns in people's responses to soundscape indicators – for instance, in general people have a strong annoyance reaction to roughness in sirens, but not in horns (see Section 7.2.2, but this association may not have been formed for everyone equally) – precisely how these responses are formed will differ from person to person. Our next step then, is to investigate to what extent people with similar demographic backgrounds, stages of development, or psychological states share a similar perceptual mapping. In short, we want to determine which personal factors influence the perceptual mapping and how much of the perceptual response can be explained by these factors.

The first step to exploring how we might account for the influence of personal factors is to establish which factors have the most influence and to what extent they can mitigate soundscape perception. Towards this, a first study was conducted on an early subset of the ISD data, making use of the demographic information (age, gender, educational status, ethnicity, and occupational status) and the psychological well-being included in the SSID questionnaire.

8.1. Introduction

Whilst advancements have been made in understanding soundscape determinants, there is a lack of consensus in the literature about the impact of demographic factors on soundscape perception. Additionally, much work in this area relies on limited case studies (Fang et al., 2021; Ismail, 2014; W. Yang & Kang, 2005b). There is also a parallel set of literature examining the effect of psychological well-being on sound perception. However, as previous research in this area has typically focused on simple tones rather than complex sounds (Laufer, Israeli, & Paz, 2016; Riskind, Kleiman, Seifritz, & Neuhoff, 2014), controlled laboratory-based experiments, or subsets of individuals, the extent to which psychological well-being affects soundscape remains under-researched. Therefore, this study has two aims:

1. to determine the associations between soundscape perception and demographic factors (i.e. age, gender, ethnicity, education level, occupation status);
2. to understand whether high levels of psychological well-being are associated with increased soundscape pleasantness and eventfulness
3. to determine if or how these personal factors should be integrated into a soundscape prediction strategy.

To achieve this we explore the association between personal factors (including psychological well-being and demographics) and the soundscape perception using the *in-situ* data collected in the ISD. In keeping with the methodology used throughout this thesis, this was investigated through a MLM, which incorporates the LocationID as a proxy for contextual information in order to demonstrate what degree of explanatory power these personal factors may have for a predictive model. Finally, I discuss how these results should be considered in the context of my expanded predictive model and what methods may be used to include them as predictors.

8.2. Methods

The study was approved by the local ethics committee of University College London (UCL), BSEER, Institute for Environmental Design & Engineering (IEDE) (Dated 11-10-2019).

8.2.1. Data Collection

This chapter made use of a subset of data from the ISD. This study was conducted and published during the first round of SSID data collection, prior to the first publication of the ISD. It includes 11 locations in London, with data collected from general members of the public. This chapter made use of the same SSID questionnaire presented in full in Appendix A, which is an adapted version of ISO/TS 12913-2:2018 (2018) Method 'A' (urban soundwalk method) and the WHO-5 Well-being Index (Hall et al., 2011), as well as demographic information. As this chapter focusses on the items related to psychological well-being, demographics, and personal factors, we used a subset of the variables available in the full ISD. Only the sections of the questionnaire which were examined within this study are reported in this chapter. Table 8.1 reports the demographic characteristics of the sample used.

Psychological well-being/WHO-5 well-being index

The WHO-5 Well-being Index (WHO-5) is a validated metric used to quantify general well-being. It is measured by asking how individuals have been feeling over the last two weeks through a series of questions such as 'I have felt cheerful and in good spirits'. The WHO-5 has been designed for multiple research and clinical purposes, covering a wide range of mental health domains, namely perinatal mental health, geriatrics mental health, endocrinology, clinical psychometrics, and psychiatric screening.

The WHO-5 has been shown to be a coherent measure of well-being, with good validity (Topp et al., 2015). For the purpose of analysis, a composite WHO-5 score is calculated by summing the responses to each of the 5 questions (coded from 0 [for 'at no time'] to 5 [for 'all of the time']), then multiplying by 4 to get a single score which ranges from 0 (the lowest level of well-being) to 100 (the highest level of well-being) (Topp et al., 2015). Blom, Bech, Högberg, Larsson, and Serlachius (2012) and Lucas-Carrasco, Allerup, and Bech (2012) have confirmed that the WHO-5 items constitute an integrated scale in which items add up related information about the level of general psychological well-being among both young people and the elderly.

Demographic characteristics

The demographic characteristics of each participant, including age, gender (male, female, nonconforming), education level (some high school, high school, trade/technical/vocational training, university, postgraduate), occupational status (employed, unemployed, retired, student, employed-student, other, rather not say), and ethnicity (Asian, Black/Caribbean, Middle Eastern, White, Mixed) were collected. Blank spaces were also provided if the participant wished to provide additional information. The demographic breakdown of the sample is presented in Table 8.1.

Outcome variables (ISOPleasant and ISOEventful)

The outcome variables used for this study are the ISOPleasant and ISOEventful coordinate values calculated according to Part 3 of ISO/TS 12913-3:2019 (2019).

Table 8.1.: The sample demographic characteristics

Demographic characteristics	N(%)
Total Samples	N = 1134
Gender	
Female	610 (53.79)
Male	524 (46.2)
Age	
Mean	34.67 years ± 15.11
18-30	627 (55.29)
31-40	195 (17.19)
41-50	112 (9.87)
51-60	97 (8.55)
61-70	72 (6.34)
71+	31 (2.73)
Educational Level	
Some high school	22 (1.2)
High school graduate	315 (17.3)
Trade/technical/vocational training	51 (2.8)
University (undergraduate/bachelor)	422 (32.1)
Postgraduate degree (master)	324 (17.8)
Occupation Status	
Employed	613 (54.05)
Unemployed	25 (2.2)
Retired	84 (7.4)
Student	348 (30.6)
Employed-Student	5 (0.4)
Other	44 (3.8)
Rather not say	15 (1.3)
Ethnicity	
White	806 (71.08)
Mixed/Multiple ethnic groups	63 (3.5)
Asian/Asian British	156 (8.6)
Black/African/Caribbean/Black British	31 (1.7)
Middle Eastern	23 (1.3)
Rather not say	55 (3)

8.2.2. Data analysis strategy

Data quality, missing data, checking for outliers, and data scaling

In order to maintain data quality and exclude cases where respondents either clearly did not understand the PAQ adjectives or intentionally misrepresented their answers, surveys for which the same response was given for every PAQ (e.g. 'Strongly agree' to all 8 attributes) were excluded. This is justified as no reasonable respondent who understood the questions would answer that they 'strongly agree' that a soundscape is pleasant and annoying, calm and chaotic, etc. Cases where respondents answered 'Neutral' to all PAQs are not excluded in this way, as a neutral response to all attributes is not necessarily contradictory. In addition, surveys were discarded as incomplete if more than 50% of the PAQ and sound source questions were not completed.

Prior to the data analysis, we imputed missing data and the imputed data was used across all analyses. Missing education values were imputed with the mode value (University). Missing values for age were imputed with the median age value (29). WHO-5 (psychological well-being) missing values were imputed with the median value (64). We excluded those who responded 'non-conforming' (N=4) or 'decline' (N=21) for gender, due to the very small sample size and to simplify the effects of gender in the model. The initial data sample size was N=1467; the data included in the analysis N=1134.

We took a lenient approach to outliers. Due to the nature of survey data, it was typically inappropriate to remove data solely because it represented a deviation from the typical response. However, we wanted to catch data which was incorrect, intentionally wrong, or a typo and then remove them. For the most part, this was handled with my data quality method implemented in REDCap, to ensure the SSQP/perceptual attribute values (N=8) were filled in such that they complied with the circumplex theory to a minimum degree. We were therefore, only looking for values which were extreme outliers or impossible.

Correlation between predictors and output variables

To establish the relationships between all pairs of variables including the predictors and outcome variables, the Pearson correlation coefficient, ANOVA, and Chi-square were performed (as appropriate depending on whether the feature was continuous, ordinal, or nominal) between psychological well-being, age, gender, ethnicity, education level, occupation status, and the circumplex coordinate values (ISOPleasant and ISOEventful). These results are given in Table 8.2.

8.2.3. Model specification (linear mixed-effects modelling)

Linear Mixed-Effects Regression (LMER) with random intercept and fixed slope, using backward stepwise feature selection was utilised to (a) identify the association of the features of interests (FOIs) including psychological well-being, age, gender, education level, ethnicity, occupation status, and their interaction terms with ISOPleasant and ISOEventful and (b) accommodate associations within participants among locations. In order to account for latent differences in the pleasantness and eventfulness ratings of various locations, the intercepts of each model are allowed to vary as a function of the location. Therefore, the model is constructed with two levels – the individual level (the random effects) and the location level (the fixed effects). Separate models were constructed for each ISOPleasant and ISOEventful and take the form:

$$ISOPleasant_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + \beta_n x_{nij} + \epsilon_{ij} \tag{8.1}$$

$$ISOEventful_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + \beta_n x_{nij} + \epsilon_{ij} \tag{8.2}$$

where $ISOPleasant_{ij}$ or $ISOEventful_{ij}$ are the dependent variable value for individual i in Location j; β_{0j} is the intercept for Location j; β_1 through β_n are the slopes relating the independent variables x_1 through x_n to the dependent variable; x_{1ij} through x_{nij} are the dependent variables for individual i in Location j; ϵ_{ij} is the random error for individual i in Location j. In turn, β_{0j} can be expressed as:

$$\beta_{0j} = \gamma_{00} + U_{0j} \tag{8.3}$$

where γ_{00} is the mean intercept across Locations; and U_{0j} is the unique effect of Location j on the intercept. In a random intercept model, the slope coefficients (B_n) are considered fixed across the locations (hence, labelled as the fixed effects) indicating that the relationship between the dependent variable (e.g. age, gender, etc.) and the independent variable (ISOPleasant or ISOEventful) is the same for all locations, while the general ISOPleasant of the location is accounted for by the varying intercept.

In order to identify the significant FOIs within the multi-level structure, I employed a stepwise feature selection on the fixed effects portion of the mixed-effects model, with an inclusion threshold of $p < 0.05$. Since this model includes only the LocationID at the random effects level, only the fixed effects are reduced in the feature selection process. To check for multicollinearity among the selected features, the Variance Inflation Factor (VIF) was calculated and a threshold of $VIF < 5$ was set. Any features which remained after the backwards stepwise selection which exceeded this threshold were investigated and removed if they were highly collinear with the other features. Once the feature selection process is completed, the final model with only significant FOIs included is fit and the table of the model coefficients is printed along with plots of the random effects and z-scaled and non-standardised estimates terms.

The model fitting and feature selection was performed using 'lme4' (version 1.1) and the 'step' function from 'lmerTest' (version 3.1.3) (Kuznetsova et al., 2017) in R statistical software (version 4.0.3) (R Core Team, 2018). The summaries and plots were created using the 'sjPlot' package (version 2.8.6) (Lüdecke, 2021).

8.3. Results[1]

8.3.1. Correlations

Table 8.2 presents a matrix of the correlation coefficients for the features of interest. It should be noted that these correlations are calculated across the entire pooled sample, and therefore do not account for the multi-level structure of the LocationID. Age, education, gender (male), and WHO-5 are all positively correlated with ISOPleasant. However, only age is directly (negatively) correlated with ISOEventful. Age, education, and WHO-5 are all similarly correlated with ISOPleasant, while gender has a lower effect. It is worth noting that, while occupation is not directly correlated with either of the outcome variables, it is significantly correlated with all of the other independent variables considered in the study and highly correlated with age. As will be noted in the modelling results, this means it can act as a proxy for several of these other features in certain circumstances.

Table 8.2.: Correlation coefficients for study variables. $^{**}p < 0.005$, $^{*}p > 0.05$

Factors	Age	Education	Ethnicity	Gender	Occupation	WHO-5	ISOPleasant
Age							
Education	0.32						
Ethnicity	0.23	0.04					
Gender	**-0.1****	0.05	**0.08***				
Occupation	**0.71****	**0.19****	**0.13****	**0.1***			
WHO-5	**0.12****	0.1	**0.1***	0.02	0.16		
ISOPleasant	**0.13****	**0.12****	0.11	**0.07***	0.16	**0.14****	
ISOEventful	**-0.08****	0.08	0.07	0.05	0.12	0.00	**-0.24****

8.3.2. Linear mixed-effects modelling

The linear mixed-effects regression derived regularised models of the soundscape pleasantness and eventfulness. This model was then reduced via backward stepwise feature selection. Table 8.3 presents the ISOPleasant and ISOEventful models, including non-standardised and standardised estimate values and confidence intervals (CIs) for the selected features that survived from the initial model. After the feature selection, age, education, and ethnicity were not found to be significant features in either the ISOPleasant or ISOEventful models. It should be noted, however, that the presence of one feature (e.g. occupation) which is highly correlated with another (e.g. age and gender) may cause one of the features to not meet the threshold of significance when both are included, causing

[1] This section closely resembles the Results section of the original paper (Erfanian et al., 2021) of which I was the second author. I contributed significantly to the drafting of the original paper and in particular to the analysis and results presented here.

Table 8.3.: Fixed and random effects in a linear mixed model explaining variations in ISOPleasant and ISOEventful while controlling for psychological well-being and demographic factors. The standardised estimates are calculated by refitting the model on standardised data scaled by subtracting the mean and dividing by 1 SD, allowing a comparison of all features. $^{**}p < 0.005$, $^{*}p > 0.05$

	ISOPleasant			ISOEventful		
Predictor	**Estimates**	**Std. Est.**	**95% CI**	**Estimates**	**Std. Est.**	**95% CI**
WHO-5	0.001**	**0.03**	0.01, 0.05	0.001	0.01	-0.02, 0.04
Gender (male)	-	-	-	**-0.08***	**-0.04**	-0.07, -0.00
Occupation (Rather not say)	-0.19*	**-0.19**	-0.36, -0.02	**0.7****	**0.02**	-0.13, 0.17
Occupation (Retired)	0.1**	0.10	0.03, 0.18	**-0.18****	**-0.11**	-0.18, -0.04
Occupation (Unemployed)	0.01	0.01	-0.13, 0.14	**0.01****	**0.18**	0.06, 0.3
WHO-5 x Gender (male)	-	-	-	**-0.001***	**-0.04**	-0.07, -0.00
WHO-5 x Occupation (Rather not say)	-	-	-	**-0.01****	**-0.21**	-0.33, -0.09
Random Effects						
σ^2	0.11			0.08		
$\tau_0 0$	0.06$_{Location}$			0.01$_{Location}$		
ICC	0.35			0.15		
N	11			11		
Observations	1134			1134		
Marginal R^2/ Conditional R^2	0.014 / 0.354			0.039 / 0.181		
AIC	779.125			451.351		

it to be removed during the stepwise feature selection. Nonetheless, it may be that, in a final model which included either of these features (but not both), they would each be considered significant. In this way, even though occupation was selected during this process, age may also have been considered significant, when not considering occupation. This behaviour is explored in more detail later.

Psychological well-being and its association with pleasantness and eventfulness

The final models found that a higher level of psychological well-being and retirement are associated with higher pleasantness, while individuals that prefer not to report their occupational status showed a negative association with pleasantness. Further analysis revealed that psychological well-being was negatively associated with eventfulness in men and individuals that did not report their occupational status. Additionally, we detected that eventfulness is positively associated with unemployment, whereas it is negatively associated with gender (male) and retirement (Table 8.3).

The marginal and conditional R^2 values are given for each model in Table 8.3. In a mixed effects model, the marginal R^2 represents the variance explained by the fixed effects (the individual-level independent variables) while the conditional R^2 represents the variance explained by both the fixed and random effects (Nakagawa & Schielzeth, 2012). From the conditional R^2, we can say that the full models explain 35.4% and 18.1% of the variance in ISOPleasant and ISOEventful, respectively (Fig. 8.1). While the majority of the variance is explained by location-level differences (as confirmed by the intraclass correlation coefficients (ICCs)), 1.4% of variance in ISOPleasant and 3.9% of variance in ISOEventful is explained by the FOIs (i.e. psychological well-being and age) included as fixed effects.

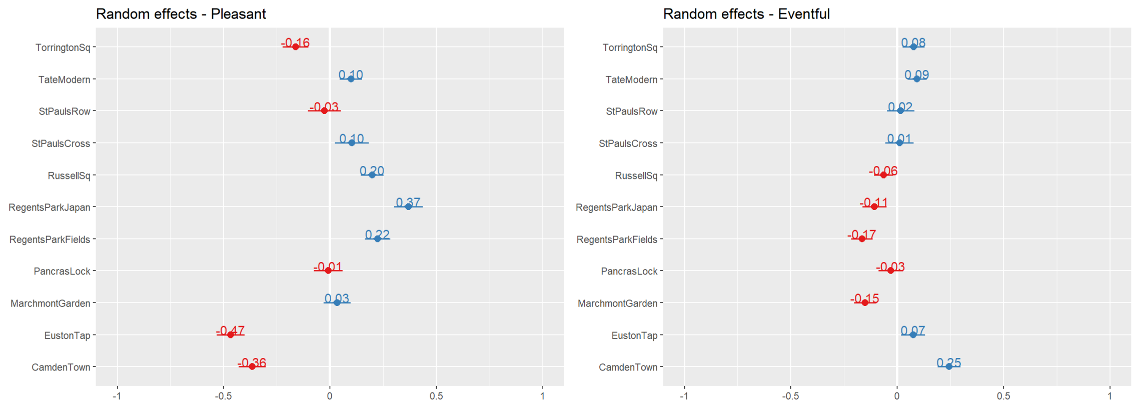


Figure 8.1.: The summary result demonstrated in the random-effects figures gives the average from the distribution of ISOPleasant across locations.

Occupation status

According to our findings, occupation status, in particular 'retirement' and to a lesser degree, gender (male) were important factors in the pattern of soundscape assessments. It is not clear why occupation (rather not say) demonstrates such a strong predictive relationship, either on its own or as an interaction term with WHO-5. A more detailed study or analysis will need to be performed to determine whether any other patterns around those who prefer not to state their occupation status can be found; it is possible that those who prefer not to say have some other characteristic which links them and somehow contributes to a change in their soundscape perception. We considered that 'rather not say' may be viewed as the default response for some people, so if they were confused about the question or didn't fit in some other category, they would elect not to response. However, this seems unlikely given that both 'other' and 'student' (possibly the most likely group to be unsure how to respond about their occupational status) were also options and did not reveal a strong relationship.

It is worthwhile to highlight that 'retirement' factor could potentially be a proxy for age (> 65) and gender (male). In order to further investigate the effect which the inclusion of occupational status had on the model building process, I re-ran the stepwise feature selection, this time without including occupation status in the initial model. This allowed me to determine whether other features (namely age and gender) would be finally selected and how they would interact within the

Table 8.4.: Linear mixed effects model resulting from the feature selection process when the initial model does not include occupational status. $^{**}p < 0.01$, $^{*}p > 0.05$

	ISOPleasant			ISOEventful		
Predictor	**Estimates**	**Std. Est.**	**95% CI**	**Estimates**	**Std. Est.**	**95% CI**
WHO-5	0.001**	**0.03**	0.01, 0.05	-	-	-
Age	**0.001***	**0.02**	0.001, 0.04	**-0.001****	**-0.03**	-0.05, -0.01
Gender (male)	-	-	-	**-0.04***	**-0.04**	-0.07, -0.001
Ethnicity	-	-	-	**-0.09****	**-0.09**	0.03, 0.14
Random Effects						
σ^2	0.11			0.08		
$\tau_0 0$	$0.06_{Location}$			$0.01_{Location}$		
ICC	0.34			0.14		
N	11			11		
Observations	1134			1134		
Marginal/Conditional R^2	0.009 / 0.345			0.023 / 0.165		
AIC	778.271			456.130		

model. The results of this process are given in Table 8.4.

Age (ISOPleasant: $\beta = 0.02, p = 0.05$; ISOEventful: $\beta = -0.03, p = 0.01$) and gender (ISOEventful: $\beta = -0.04, p = 0.05$) then came out as significant, as shown in Table 8.4. This would indicate that occupation status, particularly 'retirement', represents a group of older male individuals. Even though incorporation of occupation into the model complicates the interpretation of the outcome, it results in a slightly better fitting model (R_c^2 for ISOPleasant (0.354) and ISOEventful (0.181)) relative to 0.345 for ISOPleasant and 0.165 for ISOEventful in the model without occupation status, which is why it is selected by the feature selection process. These findings are in line with previous research, suggesting significant differences among age groups in the soundscape of different acoustic environments (Ren, Kang, & Liu, 2016; W. Yang & Kang, 2005a). These findings imply that an increase in age leads to an increase in the positive appraisal of the soundscape pleasantness. This is supported by a study by Aydın and Yılmaz (2016) in which they found that soundscape pleasantness reported by young individuals was significantly lower than the other age groups.

Age could potentially highlight the contextual role of the acoustic environment. Past experiences, memories, and even traumas give a particular context to our perception and shape the soundscape, making individual perception highly diverse, depending on the content of experience/memory. While the increase in age can lead to appreciating different sound elements, lower age seems to be related to more arousing and vibrant sounds (W. Yang & Kang, 2005a).

Like age, gender was found to be associated with the soundscape eventfulness. Past works have also reported that there are gender-related discrepancies in soundscape (Croome, 1977; W. Yang &

Kang, 2005a). These differences may be an indication of different auditory processing across genders.

Soundscape pleasantness and eventfulness differences among locations

The pleasantness and eventfulness were significantly different among locations. Pleasantness appeared to be highest in locations dominated by nature sounds (i.e. Regents Park Japan). In agreement with my results, Payne (2013) referred to the pleasantness dimension of the soundscape as the positive perception of natural places as well as the restorative capacity of the soundscape. Y. Zhang (2014) also reported a significant impact of natural soundscape on individuals' restorative experiences and boosting pleasantness. In the study by Ö. Axelsson et al. (2010) participants reported that the sound excerpts of natural components are more pleasant than human and technical sounds. Unlike pleasantness, the eventfulness increased the most in locations with dominant traffic and other sounds (i.e. Euston Tap). These findings are supported by previous research done by Bradley and Lang (2000) and Hume and Ahtamad (2013). In both studies, unnatural and urban sound-clips (i.e. fire engine siren and traffic noise), inherent in the traffic-dominant locations in this chapter, were rated highest in arousal and lowest in the pleasantness dimension. As formerly mentioned by Erfanian et al. (2019), throughout the soundscape literature, arousal has been applied as the equivalent of eventfulness and indicated on the Y-axis of the circumplex model (Ö. Axelsson et al., 2010; Erfanian et al., 2019).

These results insinuate the notion that there are multiple primary factors (Bradley & Lang, 2000) that contribute to the perception of the acoustic environment which should be considered important by urban designers and policymakers. It is expected that understanding these factors will provide multidimensional knowledge in guiding the implementation of the technological infrastructure of smart cities.

8.4. Discussion

The goal of this chapter was to determine to what extent secondary factors mediate soundscape perception, and to highlight which of these secondary factors are important to consider.

As expected, the majority of the total variance in the perceptual ratings is explained by the location-level differences (e.g. overall sound level) which represent primary contributing factors to the acoustic environment (see (McDermott, 2012)) and other non-acoustic factors. Approximately 3% of the variance is then explained by the combination of personal factors, which represent secondary contributing factors as defined by McDermott. Although the variance explained by these secondary factors is small compared to the primary factors, they are still found to contribute significantly.

8.4.1. Incorporating personal factors

Although, as Droumeva (2021) points out, each individual brings their own cultural and subjective aspects of listening to the stage of urban sound, when attempting to characterise the soundscape of a space, it is not a particular individual's aspects we should be concerned with. That individual forms a part of the collective perception of the space. Their cultural and subjective (i.e. personal) aspects mitigate their perception, but this perception then forms only a single component of the collective perception. How then should we consider these personal factors? Surely there is no suggestion to disregard their influence and importance within the soundscape approach? In my view, there are two approaches:

1. Incorporate these personal factors as demographic statistics of a location; or
2. An agent-based approach where each individual likely to use the space is simulated and modelled with their personal factors to then be included in the collective perception.

Let's look at how these two approaches would be implemented into the multilevel acoustics-based predictive model, such as those presented in Chapters 7 to 5.

Approach 1

In the first, the demographic breakdown of the space under investigation is estimated, either through a census or by the designers' desired use case. This demographic breakdown can then be compared to the results presented above (Erfanian et al., 2021) to derive weighting factors which adjust the predicted soundscape assessment. For instance, the results suggest that retired persons perceive the soundscape as 0.01 points more pleasant than others. If the particular space under investigation has a large proportion of retired persons, say 65% we could then apply an adjustment to the initial personal-factors-agnostic prediction to reflect this tendency. In this example, an initial location-level ISOPleasant prediction of 0.36, with a 65% retired population would be corrected by 0.0065 (0.65 x 0.01) for a final demographics-corrected ISOPleasant prediction of 0.3665.

Approach 2

In the second approach, rather than performing an overall estimation of the demographics and soundscape perception distribution, individual responses (with their own probability estimation) are modelled. To illustrate this, let's assume that the modelling is performed for a 30 second section of audio; for a given day, we would then model a single response to each 30s audio, where the individual response, including adjustments for that individual's demographic features would be modelled. For any given individual, their likely demographic profile would be randomly drawn from the estimated demographic breakdown of the space – i.e. if the space is expected to have 40% men and

60% women, there would be a 40% chance that the individual modelled for a randomly selected 30s audio is a man and the appropriate feature coefficients would be used. The multiple individual responses are then summed to give the overall distribution of responses. Again, the general demographics of the space would need to be estimated in order to ensure that a reasonable distribution is used. This is a more direct implementation of the modelling presented in this chapter, directly using the models derived in this chapter, as opposed to deriving weighting factors as in approach 1.

Benefits and downsides

Without knowing the implementation of a final model (i.e. exactly how the input data is measured and fed through the system and the structure of the model), it is difficult to know which approach would be more or less difficult to implement. Assuming a system which generates a predicted distribution of responses for each 30s recording, it seems likely that both approaches would be equally simple to implement. However, approach 2 seems to offer one useful advantage; by applying the demographic corrections to each individual response prediction, it would be easier to appropriately include estimates for intra-correlated demographic characteristics. For instance, as illustrated in Table 8.3, the ISOEventful has an interaction factor for psychological well-being (WHO-5) and gender (Male), so the proper breakdown both of the gender distribution and of the psychological well-being within the genders would be required. This seems much better suited to approach 2, where any individual's full personal profile could be created based on the factors which are likely to appear together (e.g. high psychological well-being and retired may be more prevalent in men than women and this would need to be reflected in the demographic statistics).

On the other hand, again depending on the particular implementation of the final model, approach 1 would seem to be easier to exclude personal factors. As noted in Section 6.2, a robust model should allow us to define 'optional' factors - ones which can be excluded from the model. Given their low impact and difficulty to estimate, personal and demographic factors will likely not always be available to include. If they are incorporated by applying weighting factors to the model predictions, then excluding them is as simple as not adding these weighting factors.

8.5. Conclusion

In this chapter, I conducted a linear mixed-effects model to show the associations of psychological well-being and demographic factors with soundscape pleasantness and eventfulness. The findings indicate that psychological well-being is positively associated with pleasantness and negatively associated with eventfulness in males and individuals that did not report their occupations. I further demonstrated that the occupational status, in particular retirement as a proxy of age and gender, was related to the perceptions of pleasantness and eventfulness. In total, these personal factors were shown to account for 1.4% of the variance for pleasantness and 3.9% of the variance for eventfulness.

These results confirm, to some degree, the results presented in Chapter 7, where gender was not found to be a significant predictor of annoyance. If we take annoyance to be the inverse of pleasantness, then both analyses reveal there is not a relationship between gender and the perception of pleasantness/annoyance. However, the results of this chapter do demonstrate that gender, both on its own and when paired with psychological well-being has a significant impact on a person's perception of the eventfulness of a soundscape. This further demonstrates the limitations of previous noise control studies which, at most, aimed to investigate only the annoyance dimension. It is important to ensure that we are not disregarding the secondary dimension of soundscape perception.

I then offer a series of proposals for how these personal factors can be included in a predictive modelling framework to enable soundscape predictions to account for differences in demographic patterns. At some point, it seems necessary that a truly holistic approach to soundscape design will need to account for these differences, particularly as we begin to make better comparisons across countries and cultures. However, given the level of uncertainty still present in soundscape predictions and the relatively low explanatory demonstrated for these demographic features, at this point it does not appear that further exploring personal factors in a predictive modelling context is the most necessary step. Other, more impactful features such as including sound sources and visual features, are more important for creating accurate and useful predictive models. That said, the inclusion of psychological well-being (as measured by WHO-5) does provide new empirical grounds for research to explore the influence of one's psychological state on their perception and experience of complex sound environments.

Chapter 9.

Towards a Probabilistic Approach

My goal in Part III is to demonstrate how the initial predictive model given in Chapter 5 can be further developed to form a practical and generalisable predictive soundscape model. So far, in Chapters 7 and 8, I have demonstrated the potential benefits of incorporating additional sound source information and secondary factors. At this point, after several years and studies working closely with the ISD data and the ISO 12913 specifications, it was important to further examine the analysis methods being used. Therefore, this chapter, originally published as Mitchell et al. (2022), interrogates the analysis methods presented in the ISO and further develops a new method of representing the soundscape of a space. Finally, I discuss how to bring the predictive model expanded throughout this thesis in line with the suggested improvements to the existing analysis and visualisation methods.

9.1. Summarising the soundscape assessment of a location

The studies presented so far have generally followed the analysis methods presented in ISO/TS 12913-3:2019 (2019). While the assessment methods available are able to record the soundscape perception of a single individual, and that person's perception is valid for themselves, it is not appropriate to then state that it is representative of the collective perception of that soundscape. In order to characterise the soundscape of a particular space or time, perceptual responses from multiple people must be collected and subsequently summarised or aggregated to describe the general soundscape of the location. The ISO guidelines stipulate a minimum of 20 participants for a soundwalk, with these broken up into sessions of no more than 5 participants at a time. Part 3 then provides the recommended methods for analysing this data.

Annex A.2 of ISO 12913 Part 3 provides the statistical measures to be used on the raw PA responses. The recommended measure of central tendency is the median, while the recommended measure of dispersion is the range. These are chosen as the data is ordinal by nature, however as will be demonstrated later, they have significant limitations. Although it is unclear, the implied intention is then that the median value of each PA is fed into Eqs. (4.1) and (4.2) presented in Chapter 4 to calculate the ISOPleasant and ISOEventful values, which can then be plotted in a two-dimensional scatter plot. Thus the standard suggests that (1) the projection method equations are not applied to individual responses and (2) only the median assessment of a location should be plotted.

9.2. Limitations of the ISO

How the ISO/TS 12913-3:2019 (2019) methods should be applied to represent the soundscape of a location has not been adequately discussed in previous literature, nor sufficiently in Part 3 of ISO 12913 itself. Indeed, in Section A.3, the technical specifications document state that (ISO/TS 12913-3:2019, 2019, p. 5):

> Results can be reported in a two-dimensional scatter plot with coordinates for the two dimensions 'pleasantness' and 'eventfulness'. The coordinates for 'pleasantness' are plotted on the X-axis, and the coordinates for 'eventfulness' on the Y-axis. Every data point in the scatter plot represents one investigated site.

However, it is not made clear whether this single point on the circumplex can be considered to be a realistic representation of the average perception of the acoustic environment. This is how I have so far represented the soundscape of a location (as in Figs. 5.11 and 5.12). Here, I will argue that this representation is incomplete. Effectively, there is no representation of dispersion in the soundscape assessment, nor a recommended use of the range that was calculated as part of the analysis recommend in Section A.2 of Part 3 of the ISO 12913. Absent a suggestion from the ISO 12913 for how the range should be used, I therefore apply this analysis to an existing real-world soundscape dataset to determine whether it provides a useful measure of dispersion. Here I use the data contained in the ISD (v0.2.4) (Mitchell, Oberman, Aletta, Erfanian, et al., 2021), which includes 1,300+ individual responses collected across 13 locations in London and Venice, according to the SSID Protocol.

For any large enough sample for a site, the range will always be from 1 to 5, the maximum and minimum available Likert-scale values. We would expect that collecting more data would result in more information or better precision, however the range will always increase as the sample size increases. As an example, within the ISD data, of the 8 PAs collected at 13 locations (for a total of 104 scales), 88% have a range from 1 to 5 and with larger sample sizes at each location, this percentage would only have increased. Using range to analyse the dispersion provides very limited information for comparing the soundscape assessments of different locations, or of a location under different conditions.

Although the range does not appear to be a useful measure of dispersion, the median does provide a useful measure and appropriately functions to describe the central tendency of the soundscape assessment of the sample. However, by stipulating that the median of each PA should be taken prior to applying the circumplex projection, the ISO procedure only allows for plotting a single scatter point in the circumplex for each location, and does not allow for plotting individual responses on the circumplex. This limits the possibilities for visualising the general trends in individual perception across the soundscape. Finally, no example or recommendation for how the circumplex scatter plot should be presented is given in the standard.

The instruments described in the ISO 12913 Part 2 (ISO/TS 12913-2:2018, 2018) were originally designed primarily for the context of individual or small group assessments. In these scenarios, the focus is on assessing the particular soundscape perception of the person in question. In order to develop this model to truly reflect the soundscape of a space, we must consider how these methods should be extended to analyse and represent the collective perception of that space. Recent advances in the soundscape approach since the development of the standards have shifted some focus from individual soundscapes to characterising the overall soundscape of public spaces (Mitchell et al., 2020) and to making comparisons between different groups of people (Jeon et al., 2018). In this context, a consideration of the natural variation in people's perception and the variation over time of a soundscape must be a core feature of how the soundscape is discussed. Reducing a public space which may have between tens and tens of thousands of people moving through it in a single day down to the mean (or median, or any other single metric) soundscape assessment often dismisses the reality of the space. Likewise, this overall soundscape of a public space cannot be determined through a ten person soundwalk, as there is no guarantee that the sample of people engaged in the soundwalk are representative of the users of the space (in fact it is very likely they would not be).

9.3. The Way Forward: Probabilistic Soundscape Representation

Given the identified issues with the recommended methods for statistical analysis and their shortcomings in representing the variety in perception of the soundscape in a space, how then should we discuss or present the results of these soundscape assessments? Ideally the method will: 1) take advantage of the circumplex coordinates and their ability to be displayed on a scatter plot and treated as continuous variables, 2) scale from a dataset of twenty responses to thousands of responses, 3) facilitate the comparison of the soundscapes of different locations, conditions, and groups, and 4) encapsulate the nuances and diversity of soundscape perception by representing the distribution of responses.

I therefore present a visualisation in Fig. 9.1 of the soundscape assessments of several urban spaces included in the ISD which reflects these goals. The specific locations selected from the ISD are chosen for demonstration only and these methods can be applied to any location. Rather than attempting to represent a single individual's soundscape or of describing a location's soundscape as a single average assessment (as in Chapter 5), this representation shows the whole range of perception of the users of the space. First, rather than calculating the median response to each PA in the location, then calculating the circumplex coordinates, the coordinates for each individual response are calculated. This results in a vector of ISOPleasant, ISOEventful values which are continuous variables from -1 to +1 and can be analysed statistically by calculating summary statistics (mean, standard deviation, quintiles, etc.) and through the use of regression modelling (as has been done throughout this thesis), which can often be simpler and more familiar than the recommended methods of analysing ordinal data. This also enables each individual's response to be placed within the pleasant-eventful space. All of the responses for a location can then be plotted, giving an overall scatter plot for a location, as demonstrated in Fig. 9.1(a).

Once these individual responses are plotted, we then overlay a heatmap of the bivariate distribution (with isodensity curves for each decile) and marginal distribution plots. In this way, three primary characteristics of the soundscape perception can be seen:

1. The distribution across both pleasantness and eventfulness, including the central tendency, the dispersion, and any skewness in the response;
2. The general shape of the soundscape within the space - in this case Russell Square is almost entirely in the pleasant half, but is split relatively evenly across the eventfulness space, meaning while it is perceived as generally pleasant, it is not strongly calm or vibrant;
3. The degree of agreement about the soundscape perception among the sample - there appears to be a relatively high agreement about the character of Russell Square, as demonstrated by the compactness of the distribution, but this is not the case for every location.

Fig. 9.1(a) includes several in-depth visualisations of the distribution of soundscape assessments, however the detail included can make further analysis difficult. In particular, a decile heatmap is so visually busy that, in my experience, it is not possible to plot more than one soundscape distribution at a time without the figure becoming overly busy. It also can make it difficult to truly grasp point 2, the general shape of the soundscape. To facilitate this, the soundscape can be represented by its 50th percentile contour, as demonstrated in Fig. 9.1(b) where the shaded portion contains 50% of the responses. This simplified view of the distribution presents several advantages, as is demonstrated in Fig. 9.1(c) and Fig. 9.1(d) and takes inspiration from the recommendation in the ISO standard to use the median as a summary statistic. In my testing, the 50th percentile contour has proved useful, clear, and compact, however this should not be taken as the definitive correct percentile cutoff. Further work will need to be done to validate the precise presentation.

When visualised this way, it is possible to identify outliers and responses which are the result of anomalous sound events. For instance if, during a survey session at a calm park, a fleet of helicopters

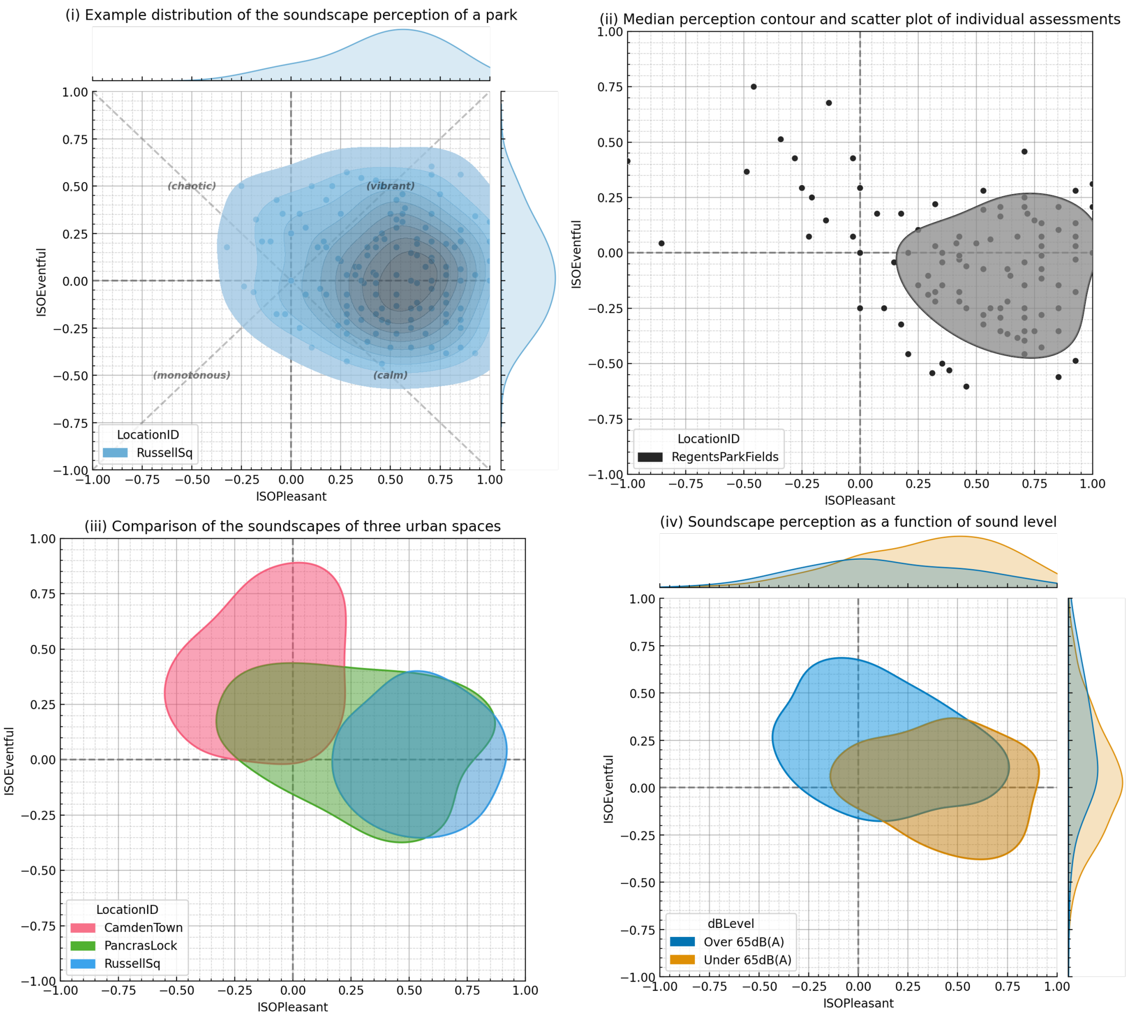


Figure 9.1.: A demonstration of some use cases of representing soundscape perception as probabilistic distributions. Data is drawn from the International Soundscape Database (ISD) and is used for demonstration only. (i) Demonstrates a high-level of detail for presenting the bivariate distribution of soundscape perception in a park (Russell Square in London). (ii) Simplified view of the distribution using the 50th percentile contour. The assessments impacted by a series of helicopter fly-overs are made obvious in the chaotic quadrant. (iii) A comparison of three popular public spaces in London. Their overlapping regions can reveal when and how their soundscapes may be similar. (iv) A comparison across the full ISD for soundscape perception at $< 65dBL_{Aeq}$ and $> 65dBA$. The introduction of other acoustic, environmental, and contextual data can reveal new and complex relationships with the soundscape perception.

flies overhead, driving the participants to respond that the soundscape is highly chaotic, we would see a group of scatter points in the chaotic quadrant which appear obviously outside the general pattern of responses. Often, these responses would be entirely discarded as outliers or the surveys and soundwalks would be halted entirely – ignoring what is in fact a significant impact on that location, its soundscape, and how useful it may be for the community. Alternatively, they would be naively included within the statistical analysis, significantly impacting the central tendency and dispersion metrics (i.e. median and range) without consideration for the context. This is the situation shown in Fig. 9.1(b) where it is obvious that there is strong agreement that Regents Park Fields is highly pleasant and calm, however we can see numerous responses which assessed it as highly chaotic. These responses were taken when a series of military helicopter fly overs drastically changed the sound environment of the space for several minutes.

Fig. 9.1(c) demonstrates how this simplified 50th percentile contour representation makes it possible to compare the soundscape of several locations in a sophisticated way. The soundscape assessments of three urban spaces, Camden Town, Pancras Lock, and Russell Square, are shown overlaid with each other. We can see that Camden Town, a busy and crowded street corner with high levels of traffic noise and amplified music, is generally perceived as chaotic, but the median contour shape which characterises it also crosses over into the vibrant quadrant. We can also see that, for a part of the sample, Russell Square and Pancras Lock are both perceived as similarly pleasant, however some portion of the responses perceived Pancras Lock as being somewhat chaotic and annoying. This kind of visualisation is able to highlight these similarities between the soundscapes in the locations and identify how they differ. From here, further investigation could lead us to answer what factors led to those people perceiving the location as unpleasant, and what similarities the soundscape of Pancras Lock has with Russell Square that could perhaps be enhanced to increase the proportion of people perceiving it as more pleasant.

In addition to solely analysing the distributions of the perceptual responses themselves, this method can also be combined with other acoustic, environmental, and contextual data. The final example, in Fig. 9.1(d) demonstrates how this method can better demonstrate the complex relationships between acoustic features of the sound environment and the soundscape perception. The data in the ISD includes approx. 30-s-long binaural audio recordings taken while each participant was responding to the soundscape survey, providing an indication of the exact sound environment they were exposed to. For Fig. 9.1(d) the entire dataset of 1,338 responses at all 13 locations has been split according to the analysis of these recordings giving a set of less than 63 dB L_{Aeq} and a set of more than 63 dB. The bivariate distribution of these two conditions are then plotted.

By presenting soundscape perception as a bivariate distributional shape on the circumplex, practitioners are obligated to address two key aspects of perception that are too often ignored: the distribution of potential responses and the eventful dimension. The array of potential responses to an environment is a crucial factor in assessing the successful design of a space and represents the reality of perception. There is no single perceptual outcome of an environment; it will always include some randomness inherent in human perception and this should be reflected in how we present soundscape assessments. Similarly, the eventful dimension is crucial to understanding how an environment is perceived and can have important impacts on the health and well-being of the users. Recent evidence also suggests that there is a more direct relationship between acoustic characteristics and the perception of eventfulness, while pleasantness is more dependent on context (Mitchell, Oberman, Aletta, Kachlicka, et al., 2021). Studies which explore the correlations between acoustic features and annoyance (or pleasantness) without considering eventfulness are perhaps missing the most direct effect of the acoustic features.

A python package called `Soundscapy` has been developed for performing the analysis and visualisations presented [making use of the `seaborn` plotting library (Waskom, 2021)] and is available for download from Github (`https://github.com/MitchellAcoustics/Soundscapy`. An interactive Jupyter notebook which provides a tutorial for using `Soundscapy`, working with the ISD data, and recreating these figures has also been included in the examples folder of the Github repository.

9.3.1. A sidenote on the proper distribution for the soundscape circumplex

The plots shown above make use of a kernel density estimation (Silverman, 2018) and assumes a normal distribution. Given that the ISOPleasant and ISOEventful values have a hard boundary at $[-1, +1]$, it is not in fact correct to consider the distribution of responses within the circumplex as a normal distribution. A normal distribution is defined as extending out to $(-\infty, \infty)$ with an area of 1 under the probability density. If the potential space of the responses is bounded, the assumption of them forming a normal distribution is violated, as part of the probability density function is unreachable, meaning the area under the probability density will not sum to 1.

If we assume the general shape of the responses to be normal, then they would instead form a truncated normal distribution (Barr & Sherrill, 1999; Burkardt, 2014). Briefly, a truncated normal distribution is estimated by first calculating the probability density function of the standard normal distribution. Then, the density function is truncated at the set boundary ($[a, \infty)$ or $(-\infty, b]$) or boundaries ($[a, b]$) and the portion of the density function which is truncated is redistributed within the boundary.

This redistribution means that the various parameters of a truncated distribution will be somewhat different than for a normal distribution, in particular the calculation of variance. This impacts the soundscape distribution plots demonstrated in Fig. 9.1 as the kernel density estimation performed by the underlying plotting library (`seaborn`) assumes a normal distribution with no boundary. It is possible that making use of a truncated normal distribution would change the shape of the distributions produced by `soundscapy`. Although at this point there does not seem to be a simple method of adapting the `soundscapy` code to make use of a truncated distribution, I chose to briefly test out how much of a change the truncated distribution is likely to make to the shape of the `soundscapy` plots through functions available in R.

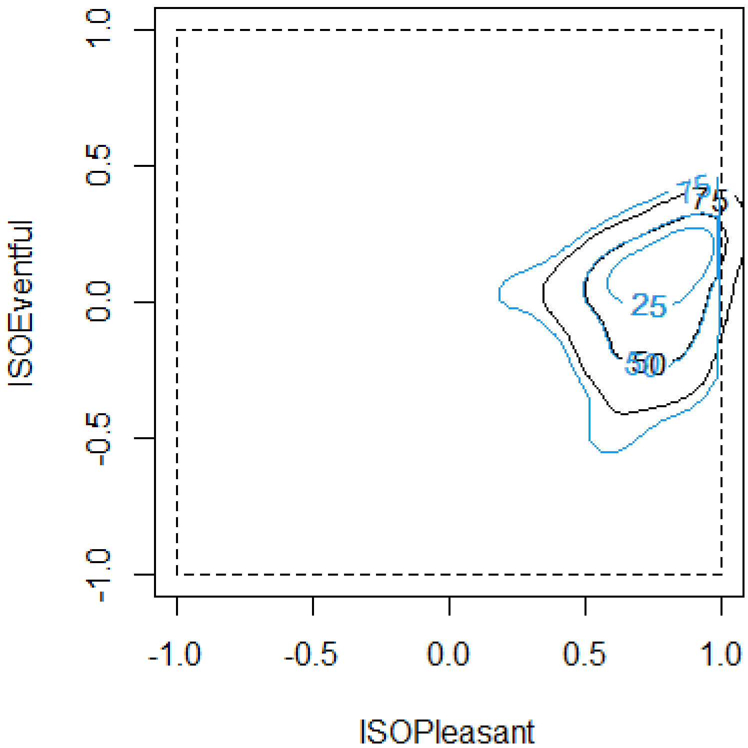


Figure 9.2.: Comparing a probability distribution in the soundscape circumplex (using Regents Park Japan as the worst case example) using a normal kernel density estimation method (black line) and a truncated KDE (blue line).

From Fig. 9.2, it appears that there would be some difference in the shape of the soundscape distribution when using a truncated distribution. However, I would note that Regents Park Japan was chosen as the worst case location in the whole ISD as the samples lie closest to the boundary and the density function estimated in `soundscapy` has the most area which lies outside the boundary. Most locations do not show any overlap with the boundary and would not be noticeably affected by the truncation. In addition, switching to the truncated normal distribution only affects those iso-density levels which overlap with the boundary. Therefore, the recommended simplified density curve given in Fig. 9.1(b) of 50% is effectively unchanged since it is very unlikely the 50th percentile curve would exceed the boundaries. For the time-being, therefore it appears that there is not a detriment to using a standard normal distribution as opposed to a truncated normal distribution, for the visualisations created by `soundscapy`.

However, as we move towards a probabilistic prediction framework, and even in the frequentist predictive models used throughout this thesis, it seems possible that the distinctions between these underlying distributions will become more important.

9.4. Making Use of the Soundscape Circumplex

There are various potential methods for integrating the probabilistic soundscape approach into a design and intervention setting. Representing the soundscape as a shape within the circumplex provides flexibility in setting design goals for a space. Not all spaces can or should have the same soundscape and soundscapes should be treated as dynamic, not static; identifying and creating an appropriate soundscape for the particular use case of a space is crucial to guiding its design. Proper forward-looking design of a soundscape would involve defining the desired shape and distribution of perceptions in the space. This can be achieved by drawing the desired shape in the circumplex and testing interventions which will bring the existing soundscape closer to the desired perception. A soundscape may need to be perceived as vibrant during the day and calm for some portion of the evening, meaning the desired shape should primarily sit within the vibrant quadrant but have some overlap into calm. This also enables designers to recognise the limitations of their environment and acknowledge that it is not always possible to transform a highly chaotic soundscape into a calm one. In these cases, instead the focus should be placed on shifting the distribution to some degree in a positive direction. The most sophisticated method of setting design goals is therefore to identify the desired shape which represents the variety of desired outcomes, and focus on designs and interventions which are most successful in matching the predicted outcome with that goal. This strategy of defining the optimal soundscape as an area or a shape within the 2-dimensional circumplex was previously illustrated by Cain et al. (2013). In Fig. 9.3, I have adapted Cain's Figure 6 to show how the shape of a target soundscape can be drawn and the shape of the existing soundscape compared to it. The work of a designer is then trialling intervention options which move the design soundscape closer to the target soundscape.

Although the visualisations shown in Fig. 9.1 are a powerful tool for viewing, analysing, and discussing the multi-dimensional aspects of soundscape perception, there are certainly cases where simpler metrics are needed to aid discussion and to set design goals. Taking inspiration from noise annoyance (ISO/TS 15666:2021, 2021), I propose a move toward discussing the 'percent of people likely to perceive' a soundscape as pleasant, vibrant, etc. when it is necessary to use numerical descriptions. In this way, a numerical design goal could also be set as e.g. 'the soundscape should be likely to be perceived as pleasant by at least 75% of users' or the result of an intervention presented as e.g. 'the likelihood of the soundscape being perceived as calm increased from 30% to 55%'. These numbers can be drawn from either actual surveys or from the results of predictive models.

Finally, although acknowledging the distribution of responses is crucial, it is sometimes necessary

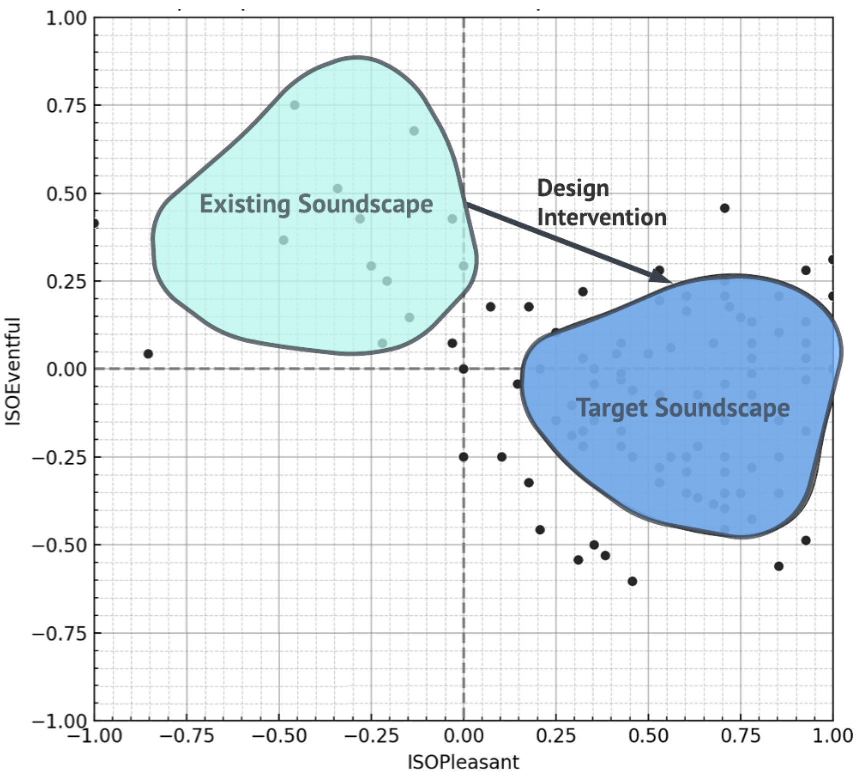


Figure 9.3.: Adapted from Cain et al. (2013, Fig. 6). Using the soundscape circumplex shape for target-setting for soundscape design.

to summarise locations down to a single point to compare many different locations and to easily investigate how the soundscape assessment has generally changed over time. For this purpose, the mean of the ISOPleasant and ISOEventful values across all respondents is calculated to result in a single coordinate point per location. This clearly mirrors the original intent of the coordinate transformation presented in the ISO, but by applying the transformation first to each individual assessment then calculating the mean value, it maintains a direct link to the distributions shown in Fig. 9.1. An example plot using the mean response of each location to compare many locations and to demonstrate change in soundscape perception can be found in Fig. 5.11 (Mitchell, Oberman, Aletta, Kachlicka, et al., 2021, Fig. 5). The key to all of these analysis methods, whether they be the distributional plots shown in Fig. 9.1, the numerical summaries, or the use of other standard statistical analyses is treating the soundscape of the space or group as a collective perception as expressed by a vector of individual circumplex coordinates.

Finally, the primary concern addressed by this method is the analysis of larger soundscape datasets, compared to what is suggested in the standard. This is necessary in order to statistically describe the groups or sub-groups being investigated, and is typically taken to need a minimum of 30 responses per group (although the full dataset, made up of many groups and locations may have many more responses in total, as in the ISD) [e.g. (Hong & Jeon, 2015; Puyana Romero, Maffei, Brambilla, & Ciaburro, 2016)]. It is unlikely that the bivariate distribution plots shown are appropriate for small datasets. However, the process of calculating the ISO coordinates for each individual response and treating this as a set of continuous values to subject to other statistical analyses holds for all sample sizes. Pleasant-eventful scatterplots are still useful for comparing differences in individual responses and appropriate methods of summarising small sample data should be explored (such as the univariate scatterplots described in Weissgerber, Milic, Winham, and Garovic (2015)).

9.4.1. Incorporating appropriateness

The discussion thus far has focussed on the two primary dimensions of soundscape perception - pleasantness and eventfulness. Our next goal is to somehow account for the third primary component identified by Ö. Axelsson et al. (2010) - Familiarity, sometimes also referred to as Appropriateness. In a later paper, Ö. Axelsson (2015) addresses critiques of the SSQP for its focus on perceptual attributes and its use of a Good-Bad scale, without considering the appropriateness of the soundscape. Given my proposal for representing soundscapes as a shape within the circumplex based on the distribution and for defining a target ideal soundscape shape, we could use additional information about the intended use of the location or the assessed appropriateness of the soundscapes to help identify the best shape for the ideal soundscape. This proposal effectively hinges on the idea that, for a given location and its intended use for relaxation, recreation, or commerce, there is some circumplex shape and placement which would be identified as most appropriate for that context. For a location meant to house an urban market, the most appropriate soundscape would primarily be vibrant, with perhaps some degree of chaotic-ness still being deemed appropriate and perhaps even desirable. For a pocket park meant to provide respite from a busy street, clearly a completely calm soundscape would be preferred, but some degree of monotony could be considered appropriate to the context. In this way, we could make use of appropriateness information to provide a more lenient and achievable definition of the 'ideal' distribution of a soundscape, against which the actual soundscape can be assessed. This method could also be further developed to provide a holistic single-value index of soundscape quality by defining standard 'ideal' soundscape shapes for different use cases and contexts, with the index indicating to what degree the assessed soundscape conforms to that ideal shape.

9.5. Probabilistic Predictions

This discussion on considering the distribution of responses leads us to my final proposal for improving the general prediction model. The model should be capable of reproducing the current tendency and size of the distribution of responses within the circumplex. This applies both to the prediction for an entire location, with many recordings feeding into the model, but also applies to the results from a single recording. For any single sound or recording, we should acknowledge that there will be a spread of potential responses from the sample population. Even given the exact same inputs, different people will have a range of different responses, all of which should be considered valid. Our goal, therefore, is not only to accurately predict the average response, but also to accurately reflect how much of a spread there will be and how it may be skewed.

As with many of my proposals for improving our initial model, there are several potential approaches to take – including a Bayesian approach, as suggested by Lionello (2021) – which should be developed as future work in the field. At this stage, I will offer a suggestion based on predicting the standard deviation of responses alongside the average response, which can be combined to generate an outcome distribution[1]. The first stage is to perform a deterministic prediction of the ISO coordinates, exactly as in Chapter 5. Then, two additional models will be created for separately predicting the standard deviation of ISOPleasant and ISOEventful responses to a single recording. We would then have four parallel models through which each recording would be processed: one to predict the centre of the ISOPleasant response, one to predict the standard deviation of the ISO-Pleasant response, and the same two for predicting the ISOEventful responses. By combining the centre and standard deviation predictions, and assuming a normal (or truncated normal) distribution, we can then generate a probability density function in the two dimensions, resulting in the circumplex shape for that recording. Fig. 9.4 demonstrates this workflow.

In order to train this model, we will need multiple responses per recording in order to calculate the standard deviations for the training set. Some degree of this exists in the ISD, where we have between an average of 1.57 responses per GroupID, meaning on average 1.57 people conducted their surveys at the same time and thus were exposed to the same sound environment, and indexed to the same binaural recording. However, it seems likely that only 3 responses would not be sufficient to calculate reasonable standard deviation values. In this case, there is a partner dataset to the ISD which has been created as part of the Soundscape Attributes Translation Project (SATP) (Aletta, Oberman, Axelsson, et al., 2020). This project aims to create validated translations of the circumplex PAQs into as many other languages as possible, such that soundscape assessments can be carried out across the world using a standardised metric. For this dataset, 27 binaural recordings, made using the same method and equipment as the ISD, were selected. The specific recordings were selected with the aim of providing a representative range of expected urban soundscapes which evenly cover the circumplex space. After a translation of the perceptual attributes was proposed by a partner research group, they conducted a lab experiment with at least 30 participants all listening to the 27 sounds and

[1] A version of this was recently done by Ooi, Watcharasupat, Lam, Ong, and Gan (2022) for predicting the probability distribution of ISOPleasant, however this idea was arrived at independently through the course of writing this chapter.

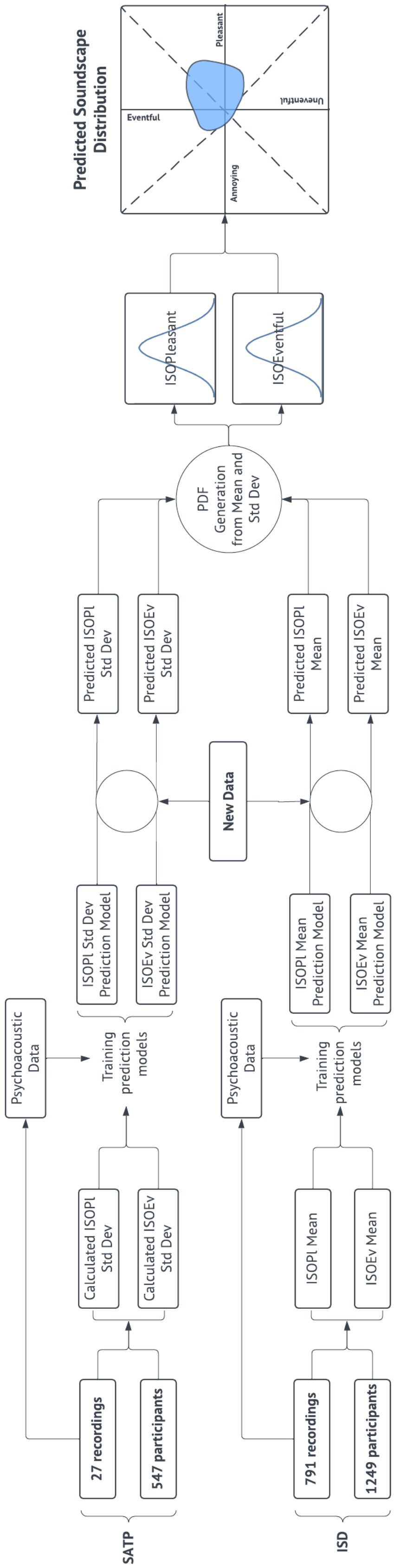


Figure 9.4.: A proposed structure and processing pipeline for probabilistic soundscape predictions.

completing the soundscape questionnaire. At this stage, this dataset contains 16 languages, with 546 participants each responding to the exact same 27 sounds. This dataset, in addition to its intended use for validating the perceptual attribute translations, also provides a rich dataset to investigate to what degree people's soundscape perception varies in response to the same soundscape, both across the entire dataset and within each language independently. We could thus use this data to train the models for predicting the standard deviation for each recording, and pair this with the centre predictions trained on the much broader ISD dataset. By combining these datasets and by predicting the expected distribution of responses we create a much more realistic representation of how a population perceives a given soundscape.

This model would create distribution predictions for each 30s recording which is input to it. To then predict the soundscape distribution of an entire location, we would make many recordings, preferably over several hours or days, and generate the predicted distribution to each of these recordings. These distributions can then be combined to result in an overall predicted distribution of the soundscape of the location, giving us a predicted soundscape shape.

9.6. Limitations of the circumplex and quantitative analysis

The method presented here is a solution for representing the soundscape of a space, which requires considering the perception of many people, but it is important to note that this is only one (very important) goal of the soundscape approach. Psychological and sociological investigations of people's relationship to their sound environment and the interactions between social contexts and individual perception are a crucial aspect of the field for which this approach would likely not be sufficient (Bild, Pfeffer, Coler, Rubin, & Bertolini, 2018). Open-response questions, structured interviews, and mixed-methods studies can provide additional insight into how people experience their environment and should be considered alongside or preceding this focus on how a space is likely to be perceived on a larger scale.

These other approaches are not in opposition to the methods proposed here, but instead further expand our view. The circumplex is a limited view of soundscape perception (this is made obvious by the fact that it excludes the third component, *familiarity*, identified in Ö. Axelsson et al. (2010)) but it is an exceptionally rich tool for dealing with the two primary aspects of soundscape perception which can readily expand the much more limited view provided by existing noise and annoyance assessment tools. Aspects of the psychological and sociological emphasis can also be integrated into a circumplex-focused approach, as demonstrated in Erfanian et al. (2021), where personal factors such as age, gender, and psychological well-being were analysed in terms of how they mediated the ISOPleasant and ISOEventful outcomes.

There has been some discussion regarding the interdependence of the PAs and the strict validity of the 90°and 45 °relationships between the attributes (Lionello et al., 2021). Further work has indicated that the scaling between the attributes may vary, but the underlying relationships hold. It is for this reason that I have taken the coordinate projection as the starting point of this critique. It should also be noted that the particular PA descriptors used in ISO 12913 are intended for outdoor environments and should not be directly applied to indoor spaces. However, a proposed set of descriptors for some indoor environments has been derived which further confirms the validity of the circumplex relationships (Torresin et al., 2020). The methods proposed here should be directly applicable to indoor spaces by using the comfort/content descriptors as well as to any other translations of soundscape descriptors into other languages (Aletta, Oberman, Axelsson, et al., 2020) as long as the dimensional relationships of the circumplex are maintained.

9.7. Conclusion

draft *need to adjust and add more general model discussion* Soundscape studies have been steadily growing as a research field over the past three decades. Their relevance for the planning and design of urban spaces is now generally acknowledged by both the academic and practitioners' communities. Yet, for their contribution in shaping better environments to be meaningful, it is necessary to agree on common methodological approaches and techniques to analyse and present standardised soundscape data. Therefore, the general goal of this work is to consider some of the questions that may still have been left unanswered by the ISO 12913 series when it comes to optimal ways to analyse and represent soundscape data coming from the ISO standardised protocols. As a result, I propose a method for presenting the results of standardised assessments as a distribution of soundscape perception within the circumplex space. This method provides an opportunity to conduct a nuanced discussion of soundscape perception which considers the variety of individual responses. The tools for generating these circumplex visualisation is made openly available as well. This shift is part of a move towards a more holistic approach to urban noise and to integrating the soundscape approach into urban design and regulations.

draft *======= Random bits to incorporate somewhere =======*

Sound perception as a chaotic system

Human perception of sound is a chaotic system. Given small changes in a large number of input conditions, a wide range of outcomes are possible **cit** *some chaos theory thing*. The same person, exposed to the same environment, could have very different responses to it depending on their state of mind, the route they took to get to the space, what they ate for breakfast, or even just some inherent randomness in their psychological response. This is further amplified by the innumerable differences between people which it will never be possible to fully capture. This is all the more true when discussing in situ perception, outside the controlled conditions of a laboratory.

However, the wide array of previous literature demonstrates that, when taken as a statistical group, conclusions can be made about the probable soundscape perception and its various causal factors.

> It means that in many situations, all we can say about a system's dynamics is of a statistical nature. -Gert van der Hejden

Chapter 10.

Conclusions

Urban soundscape studies have progressed a great deal over the last two decades. The soundscape approach has begun to make impacts in policy, such as with the Welsh *Noise and soundscape action plan (2018–2023)* (2018), and has achieved increased international recognition as a key tool in creating positive and restorative urban environments, as in the UN Environment Programme's *Frontiers 2022: Noise, Blazes and Mismatches* report (Aletta, 2022). However, this increased attention has highlighted some shortcomings with the tools currently available in soundscape.

Soundscape studies have been focussed for too long on retrospective post-hoc evaluations and on the individual or small group scale. If soundscape is to be effectively brought into assessment and legislation, data will be needed at the city scale. Predictive soundscape modelling thus provides a possibility for a more holistic approach to large scale urban sound investigations. Studies from outside of soundscape have demonstrated that a user's perception of a space is a much better predictor of how the use it – and of the benefits they derive from it – than the strict physical characteristics of the space (Kruize et al., 2019). It thus stands that a soundscape approach focussed on perception which can be generalised across a city-scale – rather than in isolated spaces – could provide more reliable metrics with which to investigate the health, social, and psychological effects of sound.

Society, designers, and engineers are interested in possibilities, in designing and improving future spaces.

- Soundscape has also been focussed on the local / individual scale, whereas assessment and legislation need data at the city-scale.
- Society (and engineers) are interested in possibilities, in designing and improving future spaces
- Because of this limited view, the methods available in soundscape studies are unsuitable for these challenges.
- If noise control wantss to progress beyond
- Psychoacoustics alone is not enough to model soundscape perception

- To be useful, Predictive models can't include perceptual inputs, this would be recursive and self-defeating.

10.1. Key Findings

The empirical and modelling work in this thesis represents a key step towards realising this application to soundscape mapping. When the predictive modelling approach is paired with data from, e.g. a large-scale acoustic sensor network, it could be used to produce a dynamic map of the likely perception of spaces across a city. Alternatively,

Chapter 5 demonstrated the human-level impact of a drastic change in urban transport. As a result of the COVID-19 lockdowns, an ideal implementation of noise reduction efforts was achieved through drastic reductions in traffic flows and substantial reductions in transport and delivery activity. Our question was then whether these changes in the common sources of urban noise actually resulted in the desired noise reduction and whether these noise reductions would have achieved an improved in the perceived soundscape of urban spaces. In the first case, how effective these traffic reductions were at reducing sound levels was heavily dependent on the type of space, although a general reduction was seen. However, the predictive modelling demonstrated that even large reductions in traffic noise levels at sites like Camden Town and Euston Tap were not enough to make those sites truly 'pleasant', when considered from a holistic soundscape perspective. In addition, the transport reductions seen under COVID-19 resulted in negative impacts to other highly pleasant soundscapes, where the reduced traffic and human sounds resulted in less pleasant soundscapes.

If noise control engineering and urban design want to progress beyond a singular focus on reducing sound levels, it needs tools which can Predictive soundscape modelling can

10.2. Contribution to Knowledge

* Properly laying out a framework

10.3. Limitations and Future Work

Retrospective assessment methods also struggle to capture the dynamics of the soundscape in a space. Whether through the narrative interview method of draft *section of ISO12913-2*, through soundwalks, or through in-situ questionnaires (Mitchell et al., 2020), only the soundscape during the particular period which the researchers are actively investigating is captured. This makes it very difficult to determine diurnal, seasonal, or yearly patterns of the soundscape. These patterns may be driven by corresponding diurnal, seasonal, or yearly patterns in the acoustic or visual environment, or by variations in how people process and respond to the sound at different times of day/season/year. Currently the only way to investigate any of these patterns is through repeated surveys. Predictive modelling, on the other hand, could allow a trained soundscape model to be paired with longterm monitoring methods to track how a soundscape may change in response to changes in the acoustic environment.

Admittedly, this method would not be able to answer the second part of the question - how do people's responses to a given acoustic and visual environment change throughout the various daily/seasonal/yearly periods? **draft** *This part should maybe be moved to a discussion* One approach to answering this question which has not, as far as the author is aware, been employed is through an un-attended survey method. Such a method could involve creating and posting fliers asking users of a space to complete a soundscape survey (accessed through a QR code) and leaving these fliers installed for longer periods of time. It is unclear how successful such a general approach would be, in particular what response rate would be expected, but given the increasing familiarity with QR codes among the general public following their use for track-and-trace during COVID-19, it does appear promising. These un-attended surveys could also be paired with long-term acoustic and environmental monitoring via a WASN or powered SLM which could simultaneously track the acoustic environment. This would thus result in a time series of online soundscape questionnaires with a corresponding time series of acoustic and environmental information, allowing us to track the changes of each over long periods of time.

Qualitative / Community approach **draft** *An approach rooted in the qualitative and sociological relationships between people and their soundscapes. Focus on Sarah Payne and Edda Bild's work.*

Several criticisms of the sorts of questionnaire-based approaches highlighted in ISO/TS 12913-2:2018 (2018) and used throughout this thesis have been raised. Bild et al. (2018) notes

> [...] the questionnaires used as tools to gain insight on users' soundscape evaluations mostly employ categorical-based assessments and rarely include openended questions [...] thus representing a limited understanding of users' soundscape evaluations. Finally, these methods minimize or do not adequately account for the role of moderating factors, like activity, in influencing how people evaluate what they hear, despite increasing evidence on activity as a moderating activity for users' soundscapes.

In contrast, Bild et al. (2018) employs a mixed-methods approach which includes both 'reported' (i.e. questionnaire-based) and 'enacted' soundscape evaluations. Enacted evaluations are assessed by observing how people actually use the space under investigation.

10.4. Concluding Remarks

Where previous ground-breaking strategies toward practical urban soundscape design (Lacey, Pink, Harvey, & Moore, 2019), have been limited in their scope, providing methods of improving individual soundscapes or approaches which can be applied to bespoke projects, this work aims to move towards a generalised and widely applicable engineering-based approach. The goal is to promote a soundscape mindset as the 'standard', not just as an extra add-on for forward-thinking projects or as a localised sonic rupture which, while incredibly effective (and affective) within its radius, is not suited to being applied on a city- or national-policy scale. For this purpose, we require a standardised and implementable index and direction of best practice which can be implemented by trained technicians, engineers, designers, and planners across all aspects of urban design, from the billion dollar museum to the inner-city public elementary school. A desire for good and restorative soundscapes should be the baseline standard in a city's design, upon which art which highlights the 'mythic, imaginative and poetic relationships within the affective environments' (Lacey et al., 2019) can be implemented by the specialists. The goal of this work therefore, is not to critique or counter the creative approaches taken by those within sound art or acoustic ecology, but instead to move towards a new baseline, a new way of designing all environments of the city, from the lowest to the highest (but mostly at the lowest, where it is needed most).