



Statistical Modelling of Complex Urban Soundscapes

Conceptual developments, tools, and applications

Andrew James Mitchell

A thesis presented for the degree of
Doctor of Philosophy

Institute for Environmental Design & Engineering
University College London (UCL)

February 7, 2022

Principal Supervisor: Prof. Jian Kang
Co-Supervisor: Dr. Phil Symonds

Declaration

I, Andrew Mitchell, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

draft *Need to update the abstract and impact statements* Urban noise pollution affects 80 million EU citizens with impacts on public health which are not well addressed by conventional noise control methods. These methods typically limit their focus to the reduction of unwanted noise, ignoring the benefits of positive sounds and remaining restricted by practical limitations of noise reduction. Modern approaches to achieve improved health outcomes and public satisfaction aim to incorporate the perception of an acoustic environment, an approach known as ‘Soundscape’.

When attempting to apply soundscape in practice, it is apparent that new methods of analysing soundscape perception in urban spaces are required; in particular, a predictive model of the users’ perceptual response to the acoustic environment is necessary. Whether to determine the impact of an environmental change, or to integrate large scale data at neighbourhood and city levels, a mathematical model of the interacting factors forms a vital component of the implementation of the soundscape approach.

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. The first and final studies present the conceptual developments, including a new protocol for conducting soundscape assessments, the creation of a large-scale international database, and a novel method of representing the soundscape of an urban space. The key chapters trace the development of a multilevel linear regression model of soundscape perception. The first iteration of this model revealed how soundscape pleasantness and eventfulness are mediated by psychological well-being and demographic factors. The second makes use of sound source information to inform the statistical relationships between psychoacoustic metrics and noise annoyance, demonstrating that loudness is a crucial factor, regardless of the sound type. Finally, a context-aware model was put into practical use to investigate how the soundscapes of urban spaces were impacted by the COVID-19 lockdowns in a situation where previous assessment methods were impractical.

Acknowledgements

This research was funded by the European Research Council.

People to thank:

- Prof Jian Kang
- Dr Francesco Aletta
- Dr Tin Oberman
- Dr Phil Symonds
- Ms Mercede Erfanian
- Ms Magdalena Kachlicka
- Mr Matteo Lionello
- Friends: Nicole Watson, Valentina

List of Studies

This doctoral thesis is based on the following studies:

Mitchell, A., Oberman, T., Aletta, F., Erfanian, M., Kachlicka, M., Lionello, M., & Kang, J. (2020) The Soundscape Indices (SSID) Protocol: A Method for Urban Soundscape Surveys – Questionnaires with Acoustical and Contextual Information. *Applied Sciences*, 10 (7), 2397. <https://doi.org/10.3390/app10072397>

Erfanian, M., **Mitchell, A. J.**, Kang, J., & Aletta, F. (2019). The Psychophysiological Implications of Soundscape: A Systematic Review of Empirical Literature and a Research Agenda. *International Journal of Environmental Research and Public Health*, 16(19), 3533. <https://doi.org/10.3390/ijerph16193533>

Erfanian, M., **Mitchell, A.**, Aletta, F., & Kang, J. (2020). Psychological Well-being and Demographic Factors can Mediate Soundscape Pleasantness and Eventfulness: A large sample study. *Environmental Psychology*. <https://doi.org/10.1016/j.jenvp.2021.101660>

Orga, F., **Mitchell, A.**¹, Freixes, M., Aletta, F., Alsina-Pagès, R. M., & Foraster, M. (2021). Multilevel Annoyance Modelling of Short Environmental Sound Recordings. *Sustainability*, 13(11), Article 11. <https://doi.org/10.3390/su13115779>

Mitchell, A., Oberman, T., Kachlicka, M., Aletta, F., Lionello, M., Erfanian, M., & Kang, J. (2021). Investigating Urban Soundsapes of the COVID-19 Lockdown: A predictive soundscape modelling approach. *JASA*.

Mitchell, A., Aletta, F., & Kang, J. (2021). How to analyse and represent soundscape perception. *JASA-EL*.

¹Joint first author

The following studies are related works which influenced this thesis and were completed as part of the same work but have not been included as key components:

Kang, J., Aletta, F., Oberman, T., **Mitchell, A.**, Erfanian, M., Tong, H., Torresin, S., Xu, C., Yang, T. (2021). Supportive Soundscapes are Crucial for Sustainable Cities and Communities. *Nature Sustainability*.

Lionello, M., Aletta, F., **Mitchell, A.**, & Kang, J. (2020). Introducing a Method for Intervals Correction on Multiple Likert Scales: A Case Study on an Urban Soundscape Data Collection Instrument. *frontiers in Psychology*.

Aletta, F., Oberman, T., **Mitchell, A.**, Tong, H., & Kang, J. (2020). Assessing the changing urban sound environment during the COVID-19 lockdown period using short-term acoustic measurements. *Noise Mapping*.

Vida Manzano, J., Almagro Pastor, J. A, García Quesada, R., Aletta, F., Oberman, T., Aletta, F., **Mitchell, A.**, & Kang, J. (2021). The "sound of silence" in Granada during the COVID-19 lockdown. *Noise Mapping*, 8(1), 16–31. <https://doi.org/10.1515/noise-2021-0002>

Impact Statement

The statement should describe, in no more than 500 words, how the expertise, knowledge, analysis, discovery or insight presented in your thesis could be put to a beneficial use. Consider benefits from **inside** and **outside** academia and the ways in which these benefits could be brought about.

Contents

1	Introduction	1
1.1	Background	1
1.2	The SSID Project	1
1.2.1	Project collaborators	2
1.2.2	Motivation for the SSID Project	2
1.3	Research Aims & Questions	2
1.4	Thesis Structure	2
2	Literature Review	4
2.1	Soundscape Studies	4
2.1.1	Soundscape Descriptors and Indices	4
2.1.2	Standardisation	4
2.1.3	Soundscape Descriptors	5
2.1.4	Soundscape Indicators	5
2.1.5	The need for predictive soundscape models	6
2.1.6	Demographic differences	8
2.2	Approaches to Soundscape in Engineering	9
2.2.1	The Quiet Areas approach	9
2.2.2	SSID approach	10
2.2.3	Qualitative / Community approach	10
2.2.4	Sound Art / Installations	10
2.3	Existing Predictive Models	10
2.3.1	Models based on non-acoustic data sources	II
2.3.2	Red and Green Soundscape Indices	II
2.3.3	Clustering/Classification Approach	II
3	Methodology Part 1 - Development of the SSID Protocol	13
3.1	Introduction	13
3.2	Purpose	15

Contents

3.3	Protocol Design and Equipment	16
3.3.1	Labelling and Data Organisation	17
3.3.2	Location and Measurement Point Selection	18
3.3.3	Equipment	19
3.4	Techniques for Field Data Collection	20
3.4.1	Questionnaire Surveys	20
3.4.2	Contextual and Environmental Factor Data Collection	22
3.5	Procedure	23
3.5.1	Assembling the Equipment	26
3.5.2	Recording Stage	27
3.5.3	Questionnaire Stage	28
3.6	Lessons from International Data Collection	30
3.6.1	Sampling	30
3.6.2	Data Collection	31
3.6.3	Equipment	31
3.6.4	Translation	32
3.7	The Database	32
4	Methodology Part 2 - Data Analysis and Modelling	33
4.1	Practical Applications for Predictive Modelling	33
4.1.1	Predicting soundscape of not-yet-existing spaces	34
4.1.2	Soundscape mapping	34
4.1.3	Soundwalks	36
4.1.4	Swedish Soundscape Quality Protocol	36
4.2	Questionnaires	36
4.2.1	Likert Responses	37
4.2.2	Circumplex Projection	37
4.3	Psychoacoustics and Auditory Perception	37
4.3.1	Psychoacoustic Parameters	37
4.3.2	Feature Selection	37
4.4	Machine Learning and Regression Techniques	37
4.4.1	Multi-level Linear Regression	38
4.4.2	Feature Selection	39
4.4.3	Clustering Analysis	39

Contents

5 Psychological Well-being and Demographic Factors can Mediate Soundscape Pleasantness and Eventfulness	40
5.1 Introduction	40
5.1.1 Psychophysiological studies	42
5.2 Discussion	42
5.2.1 Incorporating personal factors into a predictive soundscape model	42
6 Multilevel Annoyance Modelling of Short Environmental Sound Recordings Including Sound Source Information	52
6.1 Introduction	53
6.2 State of the Art of Annoyance Evaluation and Modelling	54
6.2.1 Evaluation of Annoyance	54
6.2.2 Annoyance Prediction	55
6.3 Methods	57
6.3.1 Dataset	57
6.3.2 Design of the Perceptual Tests	58
6.3.3 Psychoacoustic Data Analysis	60
6.3.4 Multi-Level Linear Regression Modelling	60
6.4 Results	62
6.4.1 Differences in Annoyance between Groups	62
6.4.2 Annoyance Model	62
6.5 Discussion	63
6.6 Conclusions	65
7 Investigating Urban Soundscapes of the COVID-19 Lockdown: A predictive soundscape modelling approach	67
7.1 Introduction	68
7.1.1 Review of the impacts of COVID-19	68
7.1.2 Review of predictive modelling	76
7.2 Materials and Methods	78
7.2.1 Onsite data: Questionnaires, binaural measurements, and recordings	79
7.2.2 Modelling	82
7.2.3 Online Survey	83
7.3 Results	85
7.3.1 Perceived sound source dominance	86

Contents

7.3.2	Model selection, performance, and application	87
7.4	Discussion	89
7.4.1	Interpretation of the results	89
7.4.2	Limitations of the study	96
7.5	Conclusion	96
8	How to Analyse and Represent Soundscape Perception	98
8.1	Abstract	98
8.2	Introduction	98
8.3	The current ISO 12913 framework	100
8.3.1	Coordinate transformation into the two primary dimensions .	100
8.3.2	Summarising the soundscape assessment of a location	102
8.3.3	Limitations of the ISO	102
8.4	The Way Forward: Probabilistic Soundscape Representation	104
8.5	Making Use of the Soundscape Circumplex	108
8.5.1	Limitations of the circumplex and quantitative analysis	109
8.6	Conclusions	110
8.7	Interpreting the Soundscape Circumplex	111
8.7.1	Application & Simulations	111
8.7.2	Circular space discussion	112
8.7.3	Normal distribution discussion	114
8.7.4	Non-continuous projected values	115
9	Conclusions	116
9.1	General Discussion	116
9.2	Contribution to Knowledge	116
9.3	Implications	116
9.4	Limitations	116
9.5	Recommendations for Future Research	116
9.6	Concluding Remarks	116
Appendix A: SSID Questionnaire		128
Appendix B: 1/f Analysis of complex sound environments		129
1	Abstract	129
2	Introduction	129
2.1	Perceptual Attributes	130

Contents

2.2	Importance of temporal structure on the sound environment	130
2.3	Justification	132
2.4	Frequency ranges, intervals, and regimes	132
3	Methods	133
3.1	Survey collection	133
4	Analysis	133
4.1	Recording Processing	133
4.2	$1/f^\alpha$ Analysis	133
4.3	Correlation Analysis	134
5	Results & Discussion	134
5.1	Identified time and frequency regimes	134
5.2	Slope values	134
6	Correlations	135
7	Concluding Remarks	135

Appendix C: Deep Learning Techniques for the detection of noise Annoyance (DeLTA)	136
--	------------

List of Figures

2.1	Draft figure of a clustering approach	12
3.1	Timeline of the on site soundscape protocol. RegentsParkJapan (RPJ) is used as an example. Abbreviations as defined in Table 3.3.3 – Questionnaires (QUE): Questionnaires; 360Video (VID): 360°video; Site pictures (PIC): Site pictures; Binaural (BIN): Binaural Recording; Ambisonic recording (AMB): Ambisonic recording; Sound Level Meter (SLM): Sound Level Meter (acoustical factors); Environmental factors (ENV): Environmental factors.	24
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 Figure 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 multi-channel recorder are stored in a bag under the tripod which can contain all of the pieces of equipment for easy transport.	25
6.1	Screenshot of the MUlti Stimulus test with Hidden Reference and Anchor (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 <i>not annoying at all</i> to <i>extremely annoying</i>	60
6.2	Estimated Marginal Means for Annoyance as a function of Areas of residence (left) and Gender (right).	62
6.3	Second-level effects figures representing the regression coefficients by types of sound (label) and for different psychoacoustic parameters.	63

List of Figures

7.1	Reproduced with permission from Aletta et al. (2020b). On the left: Sound levels distributions at the 11 London locations before and during the lockdown measures implementation; on the right: Sound levels distributions (aggregated across locations) and corresponding mean values before and during the lockdown measures implementation.	71
7.2	Reproduced with permission from Aletta et al. (2020b). Loudness distributions at the 11 London locations before and during the lockdown measures implementation.	71
7.3	Reproduced with permission from Aletta et al. (2020b). Sharpness distributions at the 11 London locations before and during the lockdown measures implementation.	72
7.4	Reproduced with permission from Aletta et al. (2020b). "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.	73
7.5	Reproduced with permission from Aletta et al. (2020b). Di-dimensional plot for the three-cluster solution. The clusters have been labelled as: Cluster 1 – Quiet Areas; Cluster 2 – Active Areas; Cluster 3 – Traffic-dominated Areas.	74
7.6	Reproduced with permission from Aletta et al. (2020b). Mean A-weighted equivalent sound level reductions between the pre- and during-lockdown conditions as a function of cluster membership (i.e. Area type).	76
7.7	The study flowchart indicating the data collection, analysis, modelling, and discussion throughout the study. The subsections in the text to which each box refers are indicated in italics.	85
7.8	Mean values per LocationID for the perceived dominance of the sound source types, for the 2019 on-site campaign.	86
7.9	Location-level scaled coefficients for the ISO Pleasant model.	89
7.10	Soundscape circumplex coordinates for (a) the mean ISO Pleasant and ISO Eventful 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.	90

List of Figures

7.II	The relative movement of soundscape perception in the circumplex due to the Coronavirus disease of 2019 (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.	92
8.1	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.	101
8.2	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 $< 65dB\text{L}_{Aeq}$ and $> 65dBA$. The introduction of other acoustic, environmental, and contextual data can reveal new and complex relationships with the soundscape perception.	105
8.3	Mockup placeholder simulation heatmap plot from Lionello 2020 and discrete vs binned transformation values. For new one, need to include distribution histograms along each axis.	112
1	Set of plots including the average line for each of the parameters, with a dotted line indicating the identified regime break.	134
2	Single plot of a sample power spectrum, with the regression lines of the regimes plotted.	134
3	Non-linear scatterplot for highly correlated feature(s)	135

List of Tables

3.1	Labelling system for on site data collection. Regents Park Japanese Garden is used as an example location. Abbreviations as defined in Table ?? - SLM: Sound Level Meter (acoustical factors); ENV: Environmental factors; QUE: Questionnaires; PIC: Site pictures.	18
3.2	Recommended equipment for implementing the Soundscape Indices (SSID) protocol. SLM: Sound Level Meter; AMB: Ambisonics; BIN: Binaural; QUE: Questionnaires	19
3.3	Table of recommended context and acoustic measurement factors.	20
6.1	Psychoacoustic parameters calculated for the 27 stimuli used in the listening experiment.	60
6.2	Random intercept-random slope multi-level model of psychoacoustic 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.	63
7.1	Reproduced with permission from Aletta et al. (2020b). Descriptive statistics of the psychoacoustic metrics for the three identified clusters. . .	75
7.2	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.	81
7.3	Pearson correlation coefficients between candidate acoustic features and ISOPleasant and ISOEventful across all 13 locations. Only statistically significant ($p < 0.01$) coefficients are shown.	82
7.4	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' multivariate linear regression with no location effects.	88

List of Tables

1	Summary statistics of slopes?	135
2	Correlation rankings between psychoacoustic and temporal features, and perceptual features	135

1 Introduction

1.1 Background

Urban noise pollution affects 80 million EU citizens with substantial impacts on public health which are not well addressed by conventional noise control methods. 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'.

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 and 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.

When attempting to apply soundscape in practical applications in the built environment, it is immediately 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 The SSID Project

The SSID Project is a five-year, multi-disciplinary project funded by a Horizon 2020 European Research Council grant (no. 740696).

1.2.1 Project collaborators

1.2.2 Motivation for the SSID Project

1.3 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. 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 formation and what are the driving interactions between acoustic features and soundscape assessment?
2. How does the sound source composition in a complex sound environment mediate this interaction and how can this effect be simplified and modelled?
3. How can the multiple levels of soundscape formation be simplified and integrated into a cohesive predictive model, and what interpretations about the cross-effects of these levels can be drawn from the model?
4. In what ways and to what extent can predictive soundscape modelling be applied to address future urban design challenges? How can these methods best be integrated into policy, design, noise mapping, and engineering practice?

1.4 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. The first study presents a protocol for conducting large-scale soundscape assessment describes the resulting publicly available database which includes 18 locations in 4 European cities. The second study reviews the current state of research on the relationships between soundscape features and psychophysiological health and presents an initial development of the multilevel modelling approach used throughout the thesis to investigate the influence of psychological wellbeing on soundscape perception.

Studies three and four developed two approaches to incorporating context into a predictive model. The first makes use of sound source information to inform the statistical

1 Introduction

relationships between psychoacoustic metrics and noise annoyance, demonstrating that loudness is a crucial factor, regardless of the sound type. The second model presents a multilevel model which incorporates contextual information about the location.

Towards answering these questions, the results of five peer-reviewed studies are presented. These studies represent a series of work to

1. Advance the conceptual development and practice of soundscape studies
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.

2 Literature Review

2.1 Soundscape Studies

2.1.1 Soundscape Descriptors and Indices

2.1.2 Standardisation

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.

2.1.3 Soundscape Descriptors

In order to consistently discuss soundscape and the factors which influence it, it is important to understand what terms have been used to describe soundscapes and to construct a consistent framework within which to work. Both the traditional focus on the epidemiological impacts of noise and the development of the soundscape concept have used many different terms in order to describe the perception of a sound environment.

Noise annoyance is perhaps the best researched aspect of environmental sound perception.

Pleasantness

Quietness / Tranquillity

Acoustic Comfort

Perceived sound level

Music-likeness

Restorativeness

Soundscape quality

Appropriateness

Perceived Affective Quality (PAQ)

2.1.4 Soundscape Indicators

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 (Berglund and Nilsson, 2006; Yang and Kang, 2005; Rychtáriková and Vermeir, 2013; Aumond et al., 2017; Alsina-Pagès et al., 2021). 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,

2 Literature Review

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.

Subsequent studies have shown that, even with large data sets and several possible acoustic indicators examined, models that are based on objective/measurable metrics underperform in predicting soundscape assessment when compared to models based on perceptual responses. Ricciardi et al. (2015), with a methodology based on smart phone recordings, achieved $R^2 = 0.21$ with acoustic input factors L_{50} and $L_{10} - L_{90}$, whereas the same dataset and model building method achieved $R^2 = 0.52$ with perceptual input factors overall loudness (OL), visual amenity (VA), traffic (T), voice (V), and birds (B). This indicates that merely examining the acoustic level is not sufficient for predicting the assessed soundscape quality, and that additional objective factors and a more holistic and involved method of characterizing the environment is required. 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.

2.1.5 The need for predictive soundscape models

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 an existing environment, where users of the space in question are surveyed regarding their experience of the acoustic environment (Engel et al., 2018; Zhang et al., 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 Literature Review

Developing soundscape indices is a process that requires consideration of how people perceive, experience, and understand the surrounding sound environment. For the purpose of modelling and comparisons,

Previous soundscape research has demonstrated that perception of the acoustic environment, 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 collaboration with The French Institute of Science and Technology for Transport, Development and Networks (IFSTTAR) to collect this database across a wide range of locations and soundscape types. This work has already been mostly completed and the database is now ready to be put to use in building the overall soundscape predictive model.

This approach is unique in that it:

1. fundamentally incorporates all identified factors of soundscape perception in a coherent manner;
2. is extensible and interpretable;
3. considers how soundscape change over both multi-hour and multi-day timescales and incorporates this dynamic behaviour for increased accuracy.

Where previous ground-breaking strategies toward practical urban soundscape design (Lacey et al., 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).

2.1.6 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 [cit Kou2020effects](#) 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.

Section on Erfanian et al. 2020, Psychological Well-being

Low-income and minority evidence A consistent limitation of soundscape studies investigating the influence of personal factors is a sampling bias towards majority ethnicities (typically White British for UK studies and ethnic Chinese for Chinese studies) and middle-class and highly educated groups. This results in not only incomplete information about how demographics influence soundscape perception, but also represents a systemic under-representation of certain environments. While it may be unclear to what extent ethnicity and social class internally influence a person's perception, it is clear that these groups are exposed to different sound environments

and therefore studies which do not include under-represented groups are also by definition not including those sound environments which those groups inhabit.

A recent study by [cit Kou2020effects](#) was successful in making inroads in these under-represented environments by studying the Humboldt Park neighbourhood in Chicago, USA. Their study included

2.2 Approaches to Soundscape in Engineering

From this literature review, some conclusions about current approaches to incorporating the concept of "soundscape" into practical engineering and architectural design have been identified.

2.2.1 The Quiet Areas approach

This approach maintains a focus on "identifying and preserving quiet areas" (EEA, 2020) following the imperative given in the Environmental Noise Directive (END) (?). This approach is mostly rooted in a noise mindset, although the methods employed for identifying quiet areas vary across countries within the EEA. Background sound levels seem to play an important role in identifying quiet areas, in particular when attempting to produce maps of available quiet areas on a city- or agglomeration-scale such as that used in the EEA (2020), where quiet areas were defined as: "those with less than 55 dB L_{den} from road, rail, aircraft and industrial sources and were classified, depending on their land cover type, as quiet areas with green/blue land cover." However, several background noise thresholds are cited as being used by agglomerations for their definitions, along with non-acoustic criteria such as urban functionality, land cover type, location, size and accessibility of the area, visual qualities, and subjective judgement. Despite these attempts to incorporate multiple factors within the definition of quiet areas, this approach still tends toward a 1- or 2-dimensional focus, and struggles to take a holistic approach to people's

perception or response to the space.

Given that the Quiet Areas approach started with the 2002 END, predating the ISO 12913 series of technical specifications on soundscape, it has not yet moved in line with the conception of "soundscape" and the accompanying measurement methods and reporting requirements given in the ISO documents. There is therefore an open question of whether the directive to identify and preserve quiet areas would truly be considered soundscape, however it does represent the most successful foray into policy and is frequently cited as a success by soundscape researchers **cit Aletta, Guastavino, Kang, etc.**

2.2.2 SSID approach

2.2.3 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.*

2.2.4 Sound Art / Installations

These merit a necessary mention, however sound art and installations are typically considered distinct from 'engineering' and are not employed at every project. Therefore, these are not discussed further as their own approach, distinct from the other, more engineering-applied approaches.

2.3 Existing Predictive Models

Contrary to the hopes expressed by Aletta et al. (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)

(Lionello et al., 2020)

2.3.1 Models based on non-acoustic data sources

(Verma et al., 2020), (Gasco et al., 2020)

2.3.2 Red and Green Soundscape Indices

(Yang et al., 2022; Kogan et al., 2018)

2.3.3 Clustering/Classification Approach

One approach which may seem promising is to generate clusters of soundscapes. ***[draft Review ref on this -*i* Sun et al. (2019)]***. Soundscape perception of a space has many dimensions and can be difficult to adequately express. Clustering analysis provides a method to investigate the patterns present in a high-dimensional dataset and aid interpretation while imposing minimal starting assumptions. Within soundscape, this would enable the identification of 'soundscape types' based on some combination of perceptual, physical, and architectural data.

For the purposes of predictive soundscape modelling with the characteristics laid out above, clustering would allow us to identify perceptual types, into which a soundscape could be placed based on its physical features. This process is outlined in Fig. 2.1. Once the combined perceptual, contextual, and physical data are collected, unsupervised clustering analysis is performed on the perceptual data, including all Perceived Affective Quality (PAQ)s, the overall questions, and perhaps demographic data. Once the clusters are identified. prior knowledge and information about the architectural typologies and use cases are used to inform the selection of appropriate descriptive labels for the clusters. Then...

[draft Issue with this approach to expand on: Limitations of starting dataset - impossible to know if we've identified all possible classes. Highly dependent on the primary dataset, the identified clusters cannot extend beyond what was already present.]

2 Literature Review

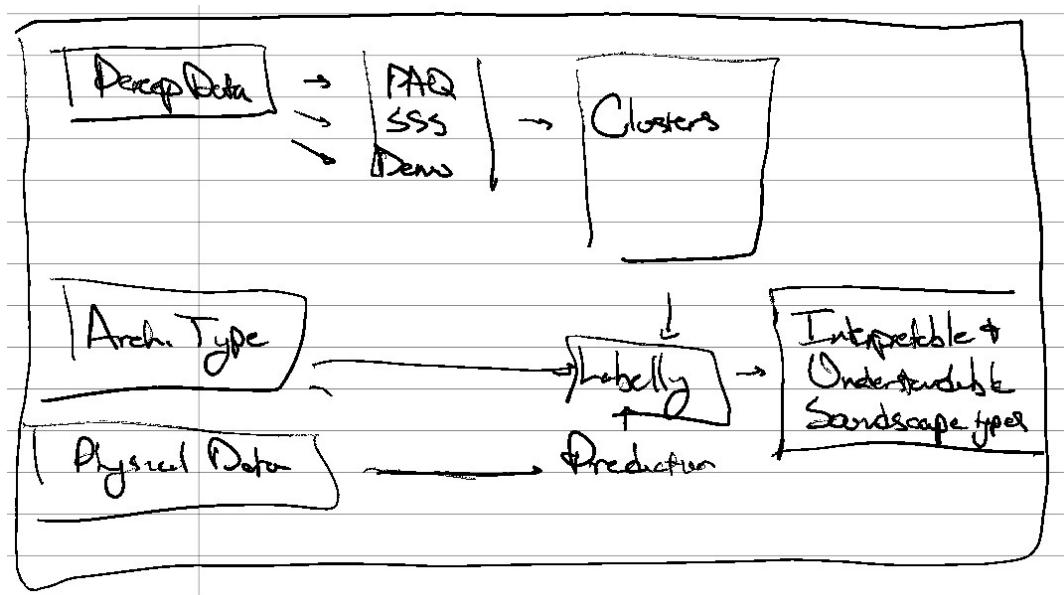


Figure 2.1: Draft figure of a clustering approach

3 Methodology Part 1 -

Development of The

Soundscape Indices (SSID)

Protocol and The International

Soundscape Database (ISD)

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.

Abstract

A protocol for characterizing urban soundscapes for use in the design of SSID and general urban research as implemented under the European Research Council (ERC)-funded SSID project is described in detail. 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 will form a large-scale, international soundscape database.

3.1 Introduction

Soundscape studies strive to understand the perception of a sound environment, in context, including acoustic, (non-acoustic) environmental, contextual, and persona factors.

3 Methodology Part 1 - Development of the SSID Protocol

These factors combine together to form a person's soundscape in complex interacting ways **cit 1**. 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.

The soundscape community is undergoing a period of increased methodological standardisation 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"; and soundscape indices can then be defined as "single value scales derived from either descriptors or indicators that allow for comparison across soundscapes" (Aletta and Kang, 2018).

This conception has recently been formalised and expanded upon with the adoption of the recent ISO 12913 standard series (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.

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; Sun et al., 2019; Oberman et al., 2018), pseudo-randomized experience sampling (Craig et al., 2017), and even non-participatory studies (Lavia et al., 2018). The protocol described in this paper was designed having in mind the need for a relatively large soundscape dataset that could be used for design and modelling

purposes, thus trying to expand the scope of soundwalks that typically deal with much smaller samples of participants (Engel et al., 2018). For the sake of comparability and standardization with these methods, we chose to refer to the soundscape attributes reported in the ISO Part 2 (Method A).

Several studies prior to the formalisation 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 [cit 1, 12-15](#). 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 dB(A)', 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.

3.2 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 various soundscapes, and so on [cit 31](#). 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.

3.3 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 SLM (as defined in IEC 61672-1:2013 [cit IEC61672](#)) 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 [cit 33](#), particularly when paired with simultaneous head-tracked virtual reality video [cit 22, 34, 35](#).

The on-site procedure to collect these data are separated into two stages, which will be outlined in detail in Section 3.5. 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.

3.3.1 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 ??, 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 difference 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 **GroupIDs**. One GroupID is assigned for *each group of participants*. This should correspond to a single binaural recording and a single 360°photo. This 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.

3 Methodology Part 1 - Development of the SSID Protocol

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

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 ?? - SLM: Sound Level Meter (acoustical factors); ENV: Environmental factors; QUE: Questionnaires; PIC: Site pictures.

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.

3.3.2 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 [cit 36-40](#), possibly contributing to soundscape assessment [cit 23, 41](#). This relies on the researcher's opinion-drive 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 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

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

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 3.5. It is important to avoid placing the recording equipment at a position where no users are expected (i.e. don't put 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.

3.3.3 Equipment

The equipment listed in Table 3.3.3 is designed to facilitate both the audio-visual recording of the location and the collection of objective environmental factors, as given in Table 3.3.3. 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.

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

3.4 Techniques for Field Data Collection

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

3.4.1 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 assessed 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 [draft Appendix X](#).

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 SSIo1 through SSIo4, respectively). These taxonomic categories of environmental sounds are based on the work done by [?](#) and [? cit 45](#).

Next are the 8 scales which make up the circumplex model of the Swedish Soundscape Quality Protocol (SSQP) [\(?\)](#), describing the 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 PAQo1 through PAQo8, 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 SSSo1 through SSSo5, respectively).

The fourth section comprises the WHO 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 [cit 47, 48](#). This information can provide additional insight into how exposure to pleasant or annoying soundscapes may impact psychological well-being as was investigated by

3 Methodology Part 1 - Development of the SSID Protocol

Aletta et al. (2019b) [cit 27](#) 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 [MISCO₃]) and a free response for the participant to provide any additional comments they would like to make on the sound environment [MISCO₁]. 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 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 [draft Appendix C](#). 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.

3.4.2 Contextual and Environmental Factor Data Collection

During each survey, the equipment listed in Section 3.3.3 is set up to capture the contextual and environmental data for the location. Table 3.3.3 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 recording 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 [cit. 11](#)), which, in the experience of the authors 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

3 Methodology Part 1 - Development of the SSID Protocol

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.

3.5 Procedure

Figure 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 mul-

3 Methodology Part 1 - Development of the SSID Protocol

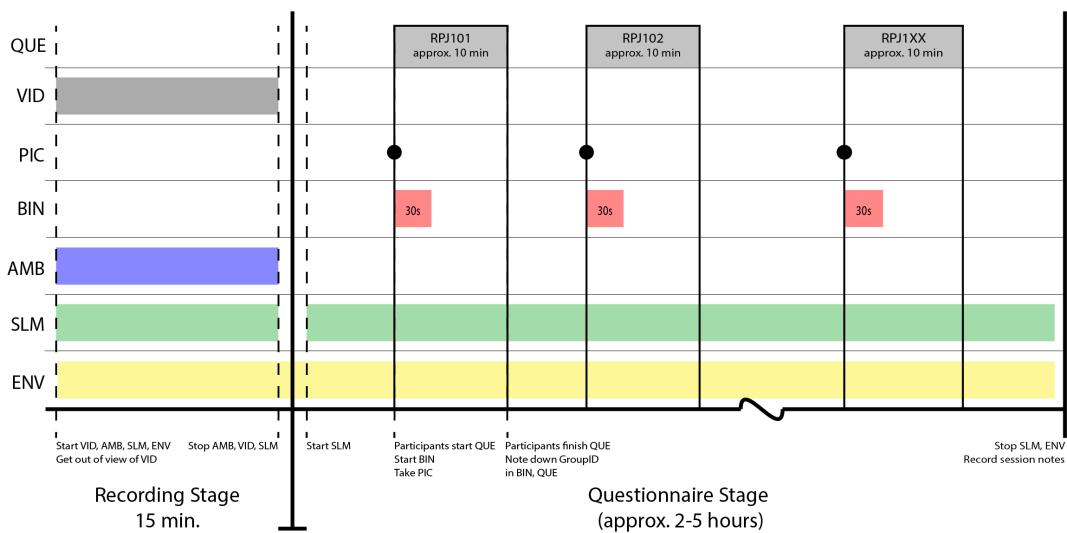


Figure 3.1: Timeline of the on site soundscape protocol. RegentsParkJapan (RPJ) is used as an example. Abbreviations as defined in Table 3.3.3 – QUE: Questionnaires; VID: 360°video; PIC: Site pictures; BIN: Binaural Recording; AMB: Ambisonic recording; SLM: Sound Level Meter (acoustical factors); ENV: Environmental factors.

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. Figure 3.2 shows an example of the recommended equipment setup.

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.

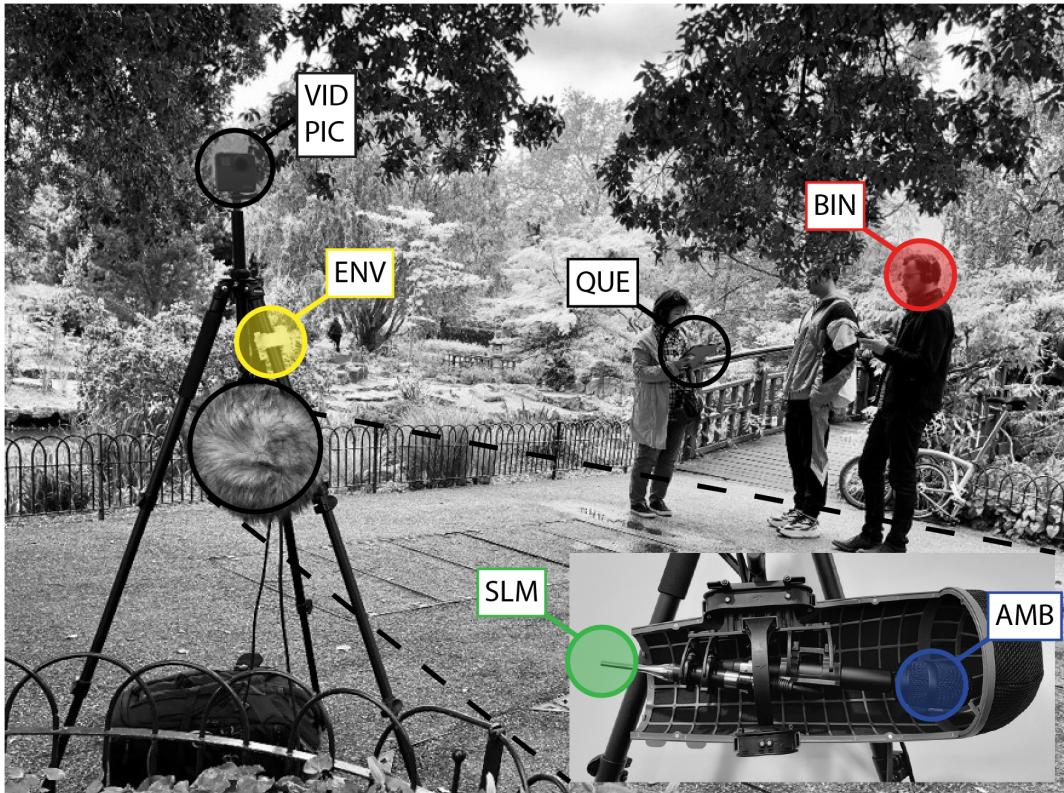


Figure 3.2: Photo of a full survey carried out in a park in London during the Questionnaire Stage. To the left is the equipment (colour-coded to match Figure 3.1), with the ambisonic microphone and SLM microphone in the windscreens, with the 360°camera on top of the tripod and to the right are one researcher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

3.5.1 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 Figure 3.2.

It is advisable to test the setup for video stitching issues and reconfigure if needed (e.g. the 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 3.3.3. The SLM should be mounted and positioned according to standard guidance for environmental noise measurements, like that given in Section 9 of ISO 1996-2:2017 [cit ISO1996](#) or Section 5 of ANSI/ASA S12.9-2013/Part 1 [cit ANSI](#). 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.

3.5.2 Recording Stage

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

3.5.3 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 3.3.2, 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 on 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

3 Methodology Part 1 - Development of the SSID Protocol

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 on 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.)

3.6 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, we have compiled the most common feedback and guidance to keep in mind when implementing this protocol.

3.6.1 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 com-

ments 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, the 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 excessive heat and cold to be negatively affecting the response rates. One research groups, which conducted the survey also in a residential area, distinguished privacy/ownership of the survey site as a major factor.

3.6.2 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.

3.6.3 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 asking members of the public not to touch or come near the measurement point as they experienced passers-by touching the wind-shield 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.

3.6.4 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. [cit 51-54](#)). For this reason, in the soundscape research community, there is a growing interest in testing and validating reliable translation of the ISO soundscape adjectives [cit Aletta SATP](#), 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.

3.7 The Database

[misc This section will form the start of the Scientific Data paper.](#)

The implementation of this protocol within the SSID project has resulted in the creation of a large dataset of urban soundscape assessments. The first version of the dataset has been released publicly as The International Soundscape Database (ISD). This database forms the basis for the majority of the work reported in this thesis.

[draft Technical description of the database, including organisation](#)

[draft Summary including figures of the data](#)

[draft The future of the database.](#)

4 Methodology Part 2 - Data Analysis and Modelling

4.1 Practical Applications for Predictive Modelling

draft *From surveys to modelling*

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. 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 Study III of this thesis **cit** *Lockdown paper* 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 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. 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 modeling'. In this context, predictive modeling involves predicting how physical acoustic environments would likely be perceived or assessed by the users of the space.

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 et al., 2018; Zhang et al., 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.

4.1.1 Predicting soundscape of not-yet-existing spaces

A challenge for the soundscape approach which frequently arises in practice is how it applies to the prospective design of new spaces. 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.

4.1.2 Soundscape mapping

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. **[draft]** *Should mention De Coensel's saliency summary as a solution here.*

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.

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) **[cit]** *other legislation and examples*, they can be difficult to translate into a soundscape approach. In general, noise maps are based on modelled traffic flows **[cit]** , from which decibel levels are extrapolated and mapped. 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 **[cit]** *need one to demonstrate failure*.

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.

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.

fig *Create a diagram illustrating how a sensor network could feed into a predictive model to produce a soundscape map.*

4.1.3 Soundwalks

Soundwalks, following World Soundscape Project (WSP) have focussed on the soundscape as 1) an individual's experience of a particular space or 2) as the sonic expression of a culture or community's relationship with the space (Droumova, 2021). Starting with Schafer's framing of the soundscape as a collective composition balancing background and foreground sounds, soundmarks and primary sound types, to the International Organization for Standardization (ISO) definition of a soundscape, the totality of the acoustic and contextual environment is processed and interpreted by an individual. Despite the ISO's expansion from the individual to a group highlighted by the phrasing "by a person, or people", the tools it presents – and, in particular, how they have been employed – fail when attempting to address the perception of many people.

This conceptual difficulty in dealing with the perception of many people has contributed to the problems associated with incorporating perception-focussed approaches in practice and regulation. **draft** *Read and integrate some section on the purpose of regulation. Peter John 2011 'Making Policy Work'*

4.1.4 Swedish Soundscape Quality Protocol

draft *Why questionnaires and modelling?*

4.2 Questionnaires

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.

4.2.1 Likert Responses

4.2.2 Circumplex Projection

4.3 Psychoacoustics and Auditory Perception

4.3.1 Psychoacoustic Parameters

Loudness

Zwicker and Fastl, Chap 8, see Mendeley notes and python-acoustics development notes.

4.3.2 Feature Selection

4.4 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.

As will be further expanded upon in chapter 8, using the trigonometric projection method provided in ISO/TS 12913-3:2019 (2019) 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 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.

4.4.1 Multi-level Linear Regression

Multi-level regression modelling is a technique commonly used in fields such as psychology [cit 42, 43 from Orga paper](#), and [draft other examples](#). Multi-Level Model (MLM)s are particularly useful when data is organised at one or more levels or groups. The concept behind MLMs can be built up starting from simple linear regression, as given by:

$$y_i = X_i\beta + \epsilon_i \quad (4.1)$$

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

$$y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \epsilon_i \text{ for } i = 1, \dots, n, \quad (4.2)$$

where the errors ϵ_i have independent normal distributions with mean 0 and standard deviation σ

Their primary feature 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

Repeated Measures

4.4.2 Feature Selection

4.4.3 Clustering Analysis

K-means

nbclust

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

[draft] This study will not be copied in wholesale. Instead, the goal is report the results in the context of working out the multi-level regression method and the influence of the non-acoustic factors. Would like to extend it by re-running the analysis with the full database, and controlling for binaural sound level.

Context

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.

5.1 Introduction

[draft] Wholesale from Lit review paper **[draft]** From a physiological point of view, a sound signal is a detectable change in our external environment that can elicit a series of unconscious responses, unbalancing our homeostasis (the dynamic state of equilibrium). These responses are similar among populations in a similar situation, are automatic, and regulated by the Sympathetic Nervous System (SNS). The Parasympathetic Nervous System (PSNS), on the contrary, is constantly active to maintain homeostasis. It has long been established that both the SNS and PSNS are part of the Auditory Nervous System (ANS), which is per se a division of the Peripheral Nervous System (PNS) **[cit]** 8 . Humans and soundscapes have a dynamic bidirectional relationship – while humans and their behaviour directly influence their soundscape, humans and their behaviour

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

are in turn influenced by their soundscape. Several scientific communities in the area of neuroscience and psychology, therefore, have begun to pay close attention to our day-to-day exposure to particular sounds and their impact on the mental and physical health of individuals. 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 **cit** 20-24. *In this regard, evidence indicates that positively perceived sounds (e.g. natural sounds) are tied with a high quality of life and enhanced physiological and physical health* **cit** 5, 25-27. *Subsequently, Attention Restoration Theory (ART) argues the impact of nature (e.g. being exposed to natural sounds such as waterfalls) on humans improved cognitive performance and stress recovery* **cit** 28-31. *Not only has spending time in nature been demonstrated to have positive effects on humans' nervous system but it has also been shown that humans innately tend to seek connections with nature, a hypothesis known as Biophilia* **cit** 32. *These theories and effects demonstrate the effect external environments have on the health and well-being of individuals, but this relationship is bidirectional within the mind as well. The prior mental and psychological state of the individual will also influence how that individual experiences the environment and can change their perception of it.*

[draft] *The physiological and psychological approaches are two sides of the same coin in the realm of soundscape research, strongly interconnected and equally important. The psychological approach, in soundscape research, strives to depict the acoustic environment through the human behavioural pattern by borrowing a more deductive approach. It translates the underlying mechanisms into explicit behavioural manifestations, resulting from the perception of the acoustic environment. The physiological approach investigates the impact of the acoustic environment through the investigation of fundamental mechanisms of Central Nervous System (CNS) and PNS by adopting a more inductive inferential approach. The physiological approach delineates the causation of the particular behaviour evoked by the environmental sounds.*

[draft] *The soundscape is composed of three main components – human interaction, acoustic environment, and perception – so it potentially draws attention across several life-science disciplines such as environmental psychology and public health, psychophysiology, and auditory neuroscience. To the best of the author's knowledge, no work previous to this review has highlighted the explicit psychophysiological underpinnings of the soundscape. The review in this chapter is the first work that reflects the fundamental mechanisms of the soundscape rather than its behavioural expressions.*

5.1.1 Psychophysiological studies

5.2 Discussion

5.2.1 Incorporating personal factors into a predictive soundscape model

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 in a multilevel acoustics-based predictive model, such as those presented in Chapters 6 to 7.

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 **[draft XX% or amount [need to check with results]]** less 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 **[draft -XX [0.65 x result]]** for a final demographics-corrected ISOPleasant prediction of **[draft XX]**.

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

Approach 2

Benefits and downsides of each approach



Psychological well-being and demographic factors can mediate soundscape pleasantness and eventfulness: A large sample study

Mercede Erfanian*, Andrew Mitchell, Francesco Aletta, Jian Kang

UCL Institute for Environmental Design and Engineering, The Bartlett, University College London (UCL), UK

ARTICLE INFO

Handling EditorLeila Scannell

Keywords:

Soundscape pleasantness
Soundscape eventfulness
Psychological well-being
Demographic factors
Acoustic environment

ABSTRACT

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. This study aims to assess the influence of psychological well-being and demographic factors including age, gender, occupation status, and education levels on the dimensions of the soundscape circumplex, i.e., Pleasantness and Eventfulness. Data was collected in eleven urban locations in London through a large-scale ($N = 1134$) soundscape survey according to the ISO 12913-2 technical specifications and incorporating the WHO-5 well-being index. Linear mixed-effects modelling applying backwards-step feature selection was used to model the interactions between internal factors including psychological well-being, age, gender, occupation status, education levels and the soundscape Pleasantness and Eventfulness, while accounting for the random effects of the survey location. The findings suggest that internal factors account for approximately 1.4% of the variance for Pleasantness and 3.9% for Eventfulness, while the influence of the locations accounted for approximately 34% and 14%, respectively. Psychological well-being is positively associated with perceived Pleasantness, while there is a negative association with Eventfulness only for males. Occupation status, in particular retirement as a proxy of age and gender, was identified as a significant factor for both dimensions. These findings offer empirical grounds for developing theories of the interaction between internal factors and soundscape formation whilst highlighting the importance of the location, namely: the context.

Introduction

Sound is a ubiquitous element in our daily lives. Despite a good deal of literature, it still strongly remains a centre of attention of many scientific communities. Looking deeper at the evolution of sound-related research in the field of engineering we see a considerable paradigm shift from noise mitigation to pleasant and restorative sound generation. This premise has been proposed with the hope to apply the existing environmental resources in order to provide a healthier and comforting acoustic environment and ultimately better quality of life (Kang et al., 2016; Kang et al., 2019). Hence, the soundscape concept, which places the emphasis on the human perception of the acoustic environment in context has emerged to support this premise.

Despite the strong evidence that research has brought for the soundscape, our understanding of the action of the Peripheral and Central Nervous System (PNS and CNS) associated with environmental sound interpretation and the factors influencing the perception of sound

is still evolving and a matter of dispute among scientific communities. Understanding of the soundscape is intimately tied to certain key factors known as primary factors of the soundscape comprising acoustic properties (physical features) of the sound such as frequency/pitch (Kumar, Forster, Bailey, & Griffiths, 2008; Patchett, 1979) and intensity/loudness (Kaya, Huang, & Elhilali, 2020) and secondary influences like emotions and personality traits (McDermott, 2012).

Pleasantness and eventfulness as key components of soundscape

Understanding the soundscape concept and its components largely depends on understanding the circumplex model of affect, proposed by James Russell (Russell, 1980). The circumplex model delineates the entanglement of the emotions and their neural substrates, opposing the classic model of discrete basic emotions (Panksepp, 2004; Tomkins, 1962).

This model suggests that all affective states, described with

* Corresponding author. Institute for Environmental Design & Engineering University College London, The Bartlett Central House, 14 Upper Woburn, London, WC1H 0NN, UK.

E-mail address: mercede.erfanianghasab.18@ucl.ac.uk (M. Erfanian).



descriptors such as alert, tense or serene, arise from cognitive interpretations of core physiological and neural sensations. These affective states are produced by two fundamental neurophysiological systems, including two orthogonal continuums: valence and arousal, which can be discerned as a linear combination or as fluctuating degrees of activation (Posner, Russell, & Peterson, 2005).

Valence refers to whether an emotion is experienced as pleasant/positive or unpleasant/negative and is distributed horizontally on the circumplex space (on the X-axis). Arousal refers to whether an emotion is physiologically activating (high arousal; e.g., excited) or deactivating (low arousal; e.g., calm) (on the Y-axis) (Russell, 1980). High arousal is associated with activation of the sympathetic components of the Autonomic Nervous System (ANS) (e.g., increased heart rate) whereas low arousal is attributed to parasympathetic activation (e.g., slower heart rate).

Similarly, the soundscape entails two main perceptual attributes: pleasantness and eventfulness that are different from the physical properties of the acoustic environment and by which the listeners appraise the quality of sounds (International Organization for Standardization Technical Specification, 12913–3:2019).¹ Soundscape pleasantness refers to the emotional magnitude of the sound perception, while soundscape eventfulness is attributed to the intensity of the sound perception (Erfanian, Mitchell, Kang, & Aletta, 2019). Like the Russell's model structure, the common model of representing soundscape is a bi-dimensional circumplex model with pleasantness on the X-axis and eventfulness on the Y-axis, proposed by Axelsson, Nilsson, and Berglund (2010).

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 component 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. In their final model, these attributes reduced to eight primary unidimensional scales of pleasant, vibrant, eventful, chaotic, annoying, monotonous, eventful and calm and the reduced attributes collapsed into two orthogonally positioned components of pleasantness and eventfulness (See 'Outcome variables').

Likewise in the current study, we employed the circumplex model reported in a two-dimensional scatter plot with coordinates for the two dimensions 'Pleasantness' plotted on X-axis and 'Eventfulness' plotted on Y-axis, taking into account the features of the locations. To differentiate these complex components from classic pleasantness and eventfulness in Axelsson's model, they will appear with the first letter capitalized throughout the text.

Psychological well-being and soundscape

There are understudied secondary factors that may be linked to the perception of the acoustic environment, such as psychological well-being (Aletta et al., 2019).

Individuals with an aberrant psychological state and poor mental health may experience environmental inputs differently to those people who do not experience such issues given that emotions, as one of the core components of psychological well-being, and sensory perceptions

are closely intertwined (Kelley & Schmeichel, 2014). As reported in the relevant literature, the impact of psychological well-being is consistent among all perceptual modalities such as vision (Zadra & Clore, 2011), tactile (Kelley & Schmeichel, 2014), olfactory (Krusmark, Novak, Gitelman, 2013), and auditory (Riskind, Kleiman, Seifritz, & Neuhoff, 2014). In parallel, studies in the field of psychopathology elucidated that individuals with poor psychological well-being, such as the clinically depressed, maintain bias and anomalous cognition, leading to inaccurate and distorted perception (Beck's cognitive theory) (Clark & Beck, 2010).

Demographic factors and soundscape

The perception of the acoustic environment or soundscape involves the sensation, identification, organization, and interpretation of ongoing omnipresent auditory information (Goldstein & Brockmole, 2016).

Soundscape does not always maintain consistency and shows a huge variation across individuals and on a general scale, among populations (Schneider & Wengenroth, 2009; Weinstein, 1978). There is evidence to suggest that the differences in the demographic characteristics like gender (Gulian & Thomas, 1986; Xiao & Hilton, 2019), age (Zhang & Kang, 2007), and educational background (Zhang & Kang, 2007) may determine the way we perceive environmental sounds. Additionally, these individual differences can potentially reflect in various perceptual properties, implying the difference between the encoding of certain auditory information between individuals such as pitch (Coffey, Colagrosso, Lehmann, Schönwiesner, & Zatorre, 2016) or loudness (Berthomieu, Koehl, & Paquier, 2021).

However, the results from past studies have, for a good part, remained inconclusive or inconsistent.

The current study

Whilst previous research has substantially advanced our knowledge of the soundscape determinants, past studies results are predominantly limited, often focussing on controlled laboratory-based experiments, individuals with psychopathology (i.e., depression) and investigating simple tones rather than complex sounds (Laufer, Israeli, & Paz, 2016; Riskind et al., 2014). In addition, the impact of psychological well-being in the context of the soundscape, by its current definition, has still largely been unexplored. So, our first aim is to understand if high levels of psychological well-being are associated with increased soundscape pleasantness and eventfulness.

The second aim of the study is to determine the associations between the soundscape and demographic factors, given there is insufficient consensus in the literature, studies are restricted to limited case studies (i.e., Peace Gardens in Sheffield – the UK) or a single ethnicity (i.e., Chinese) (Fang, Gao, Hedblom, Xu, Xiang, Hu, 2021; Ismail, 2014; Yang & Kang, 2005). We asked if age, gender, ethnicity, education level, and occupation are status associated with the soundscape Pleasantness and Eventfulness.

In this large-scale study, we explore the association of psychological well-being, demographic factors with soundscape among the members of the public with presumably no apparent psychopathology in an immersive environment with diverse demographic characteristics such as ethnicity (i.e., American, Italian) and occupation status (i.e., student, retired).

Methods

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

¹ International Organization for Standardization/Technical Specification (12913-3, 2019) deals with work still under technical progress/development, or where it is believed that there will be a future, but not immediate, possibility of agreement on an International Standard. A Technical Specification is published for immediate use, but it also provides a means to obtain feedback.

Participants

The present work is a large-scale study with data collected from the general members of the public in several locations in London with varying acoustic features. All passers-by at the data collection locations were approached in 11 locations/sites in London by the researchers and were asked if they were willing to participate in the study. Locations were selected which represented a variety of usage types, visual character, and acoustic characteristics. The minimum and maximum value of several acoustic metrics recorded at each location during the survey sessions are presented in Table B1 in Appendix B. Only individuals on the phone, with headphones on due to attention distraction, or individuals that were deemed to be younger than 18 years old (proxy consent required) were excluded from the data collection. The total number of surveys that were originally collected from the sites was 1467.

Measures and independent variables

The questionnaire, presented in full in Appendix A, comprising 38 items, is an adapted version of ISO/TS 12913–2:2018² Method ‘A’ (urban soundwalk method) (Axelsson, 2012; ISO, 2018) and WHO-5 well-being index (World Health Organization, 1998), as well as demographic information. In order to answer the questions raised in this study the authors only report some sections of the questionnaire which then undergo the statistical analyses.

Perceived affective quality/Perceptual attributes

The perceived affective quality (PAQ) of the sound environment as adopted in the method ‘A’, described in the ISO/TS 12913–2:2018, consists of category scales containing five response categories, based on the Swedish Soundscape Quality Protocol (SSQP; 41) (ISO, 2018). It includes a question ‘to what extent they agree/disagree that the present surrounding sound environment is ...’. The participants judged the quality of the acoustic environment by 8 adjectives: pleasant, chaotic, vibrant, uneventful, calm, annoying, eventful, or monotonous. The answers were presented in a 5-point Likert scale ranging from ‘strongly disagree = 1’ to ‘strongly agree = 5’. The perceptual attributes measure as a unidimensional measuring tool for the perception of the acoustic environment has not been validated to this date. The PAQs were utilized as aggregated values to construct the principal components of the soundscape (Pleasantness and Eventfulness) (See ‘Outcome variables’).

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.

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

WHO-5 well-being index asks how individuals have been feeling over the last two weeks such as ‘I have felt cheerful and in good spirits’. WHO-5 has been designed for multiple research and clinical purposes, covering a wide range of mental health domains namely perinatal mental health, the geriatrics mental health, endocrinology, clinical psychometrics, neurology, and psychiatric disorders screening.

The WHO-5 well-being index is known to be one of the most valid generic scales for quantification of general well-being. In terms of the construct validity of the scale, WHO-5 showed to have properties that are a coherent measure of well-being (Topp, Østergaard, Søndergaard, & Bech, 2015). With regards to relevant literature, WHO-5 confirmed that all items constitute an integrated scale in which items add up related information about the level of general psychological well-being among both youngsters and elderlies (Blom, Bech, Hogberg, Larsson, Serlachius, 2012; Lucas-Carrasco, Allerup, & Bech, 2012). 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 0 (the lowest level of well-being) to 100 (the highest level of well-being) (Topp et al., 2015).

Demographic characteristics

Demographic characteristics were presented such as age, gender (male, female), education level (some high school, high school, trade/technical/vocational training, university, and postgraduate), occupational status (employed, unemployed, retired, student, employed-student, other and rather not say), and ethnicity (Asian, black/Caribbean, middle eastern, white, and mixed). Some blank spaces were provided if they wanted to add further information. At the end of the survey, participants had the opportunity to write down any additional questions or remarks and were thanked for their participation.

Outcome variables (the soundscape pleasantness and eventfulness)

The soundscape data were analysed according to the procedure laid

Table 1
The sample demographic characteristics.

Demographic characteristics	N (%)
N = 1134	Age mean = 34.67 years ±15.11
Gender	
Female	610 (53.79)
Male	524 (46.2)
Age	
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)
Education 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 (44.2)
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)

² The ISO/TS 12913–2:2018 specifies requirements and provides supporting information on data collection and reporting for soundscape studies, investigations and applications.

out in Part 3 of the ISO 12913³ standard series. In order to ease data analysis and modelling the standard suggests a method to collapse the perceived affective quality responses for each of the 8 down to a 2-dimensional coordinate scatter plot with continuous values for 'Pleasantness' on the X-axis and 'Eventfulness' on the Y-axis. These coordinates are then normalized to between -1 and 1 (per the recommendation of ISO/TS 12913-3:2019). These dimensions were calculated as shown in Formulas (1 & 2) in [Appendix E](#).

Survey procedure

The participants were approached and asked if they were interested to participate in the study. All participants received information about the aim of the study, its procedures, confidentiality of research data, and how to contact the investigators, the supervisor of the project, or a member of the ethical committee. An informed consent document was given to participants, who declared to have read and understood the general information, take part voluntarily, and have understood the fact that they can stop their participation and withdraw their consent, anytime, and without any consequences. They could start filling in the questionnaire if the participant gave his/her consent. If they had no questions, they received either a paper version or an e-version of a questionnaire via a 10-inch tablet. The online questionnaires were collected and managed using REDCap electronic data capture tools hosted at UCL ([Harris et al., 2019](#)) and typically took between 5 and 10 min to complete. The goal of the researchers on-site was to collect a minimum of one-hundred questionnaires from each selected site/location, which was typically achieved over a period of 2–3 days each consisting of approximately a 4-h session. In some cases, either due to extenuating circumstances, time constraints, or excluded surveys, the full one hundred surveys were not achieved. The data was collected from 28th February 2019 to 18th October 2019 between 11 a.m. and 3 p.m.

During the survey period, acoustic and environmental metrics were simultaneously collected through binaural recordings, a calibrated sound level meter (SLM), and an environmental meter collected temperature, lighting level, and humidity data. The SLM was set up in the space in which the questionnaires were conducted and left running for the full duration of the survey in order to characterize the acoustic environment. The environmental metrics were not reported in this study since they were not in the scope of this paper but are included in the Appendices in order to provide context for the interested readers. The full protocol and data treatment as part of the SSID Database creation are described in detail by Mitchell and colleagues ([Mitchell et al., 2020](#)).

Data analytic analysis strategy

Missing data, checking for outliers and data scaling

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) (with no response) for gender, due to the very small sample size and to simplify the effects of gender (initial number of collected data = 1467, data included in the analysis = 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 removed them. For the most part, this was handled with our data quality method implemented in REDCap, to ensure the SSQP/perceptual

attributed 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 linearity between all pairs of variables including the predictors and outcome variables, Pearson correlation coefficient, Analysis of Variance (ANOVA) and Chi-square were performed between psychological well-being, age, gender, ethnicity, education level, occupation status and the soundscape Pleasantness and Eventfulness ([Table 2](#)).

Model specification (linear mixed-effects modelling)

Linear mixed-effects regression (LME) with random intercept and fixed slope, using backward stepwise feature selection was utilized to a) identify the association of our features of interest (FOIs) including psychological well-being, age, gender, education levels, ethnicity, occupation status, and their interaction terms with the soundscape Pleasantness and Eventfulness 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 Pleasantness and Eventfulness, and take the form (Formula 3 and 4) in [Appendix E](#).

In order to identify the significant FOIs within the multi-level structure, we 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-standardized 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, Brokhoff, & Christensen, 2017](#)) in R statistical software (version 4.0.3) ([R Core Team, 2013](#)). The summaries and plots were created using the 'sjPlot' package (version 2.8.6) ([Lüdecke, 2018](#)).

Results

The setup and procedures of this study allowed us to test a large group of participants with high diversity with rather various demographics including gender, age, education level, occupation status, and ethnicity (N = 1134) ([Table 1](#)).

Correlations

The correlation matrix for all study measures is demonstrated in [Table 2](#). Age was negatively correlated with Eventfulness, whereas it was positively correlated with Pleasantness. Gender appeared to be independent of Eventfulness but positively correlated with Pleasantness. Education was positively correlated with both Pleasantness and Eventfulness. Whilst psychological well-being exhibited positive and statistically significant correlations with Pleasantness, it was negatively correlated with Eventfulness. It is worth noting that occupation is significantly correlated with all other independent variables considered in the study and highly correlated with age, although it is not significantly correlated with either of dependant variables.

³ The ISO/TS 12913-3:2019 provides requirements and supporting information on analysis of data collected in-situ.

Table 2
Correlation coefficients for study variables.

Factors	Age	Education	Ethnicity	Eventful	Gender	Occupation	Pleasant
Age							
Education	0.32						
Ethnicity	0.23	0.04					
Eventful	-0.11***	0.1**	0.08				
Gender	0.1***	0.05	0.08*	0.05			
Occupation	0.71***	0.19***	0.13***	0.15	0.1**		
Pleasant	0.12***	0.11**	0.09	-0.91***	0.06*	0.16	
Psychological Well-being	0.12***	0.1	0.1*	-0.12***	0.02	0.16	0.14***

***p < 0.0005, **p < 0.005, *p < 0.05.

Linear mixed-effects modelling

The linear mixed-effects regression derived regularized models of the soundscape Pleasantness and Eventfulness. This model was then reduced via backward stepwise feature selection. Table 3 presents the soundscape Pleasantness and Eventfulness models, including non-standardized and standardized estimate values and 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 Pleasantness or Eventfulness 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 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 (See Appendix C).

The final models found that a higher level of psychological well-being and retirement are associated with higher Pleasantness. While individuals that do not rather report their occupation status showed 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 occupation status. Additionally, we detected that Eventfulness is positively associated with unemployment, whereas it is negatively associated with gender (male) and retirement (Table 3).

The marginal and conditional R² values are given in for each model in Table 3. In a mixed effects model, the marginal R² represents the

variance explained by the fixed effects (the individual-level independent variables) while the conditional R² represents the variance explained by both the fixed and random effects (Nakagawa & Schielzeth, 2013). From the conditional R², we can say that the full models explain 35.4% and 18.1% of the variance in Pleasantness and Eventfulness, respectively (Fig. 1& 2). While the majority of the variance is explained by

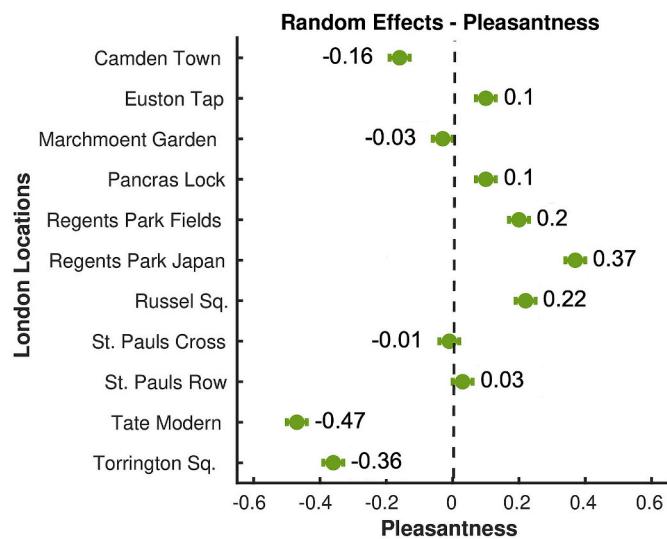


Fig. 1. The summary result demonstrated in the random-effects figures gives the average from the distribution of Pleasantness across locations.

Table 3

Fixed and random effects in a linear mixed model explaining variations in the soundscape Pleasantness and Eventfulness while controlling for psychological well-being and demographic factors. The standardized estimates are calculated by refitting the model on standardized data scaled by subtracting the mean and dividing by 1 SD, allowing a comparison of all features.

Predictor	Pleasantness			Eventfulness		
	Estimates	Std. Est	95% CI	Estimates	Std. Est	95% CI
Psychological Well-being	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
Psychological Well-being x Gender (male)	-	-	-	-0.001*	-0.04	-0.07, -0.00
Psychological Well-being x Occupation (Rather not say)	-	-	-	-0.01***	-0.21	-0.33, -0.09
Random Effects						
σ^2	0.11			0.08		
τ_{00}	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		

p < 0.05, **p < 0.01, ***p < 0.001.

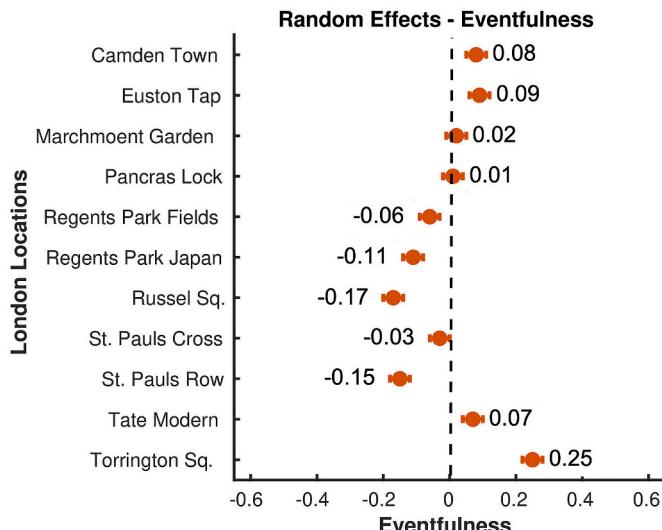


Fig. 2. The summary result demonstrated in the random-effects figures gives the average from the distribution of Eventfulness across locations.

location-level differences (as confirmed by the intraclass correlation coefficients (ICC)), 1.4% of variance in Pleasantness and 3.9% of variance in Eventfulness is explained by the FOIs (i.e., psychological well-being and age) included as fixed effects.

Discussion

For this study data of 1134 participants across 11 locations in London were included in the analysis. Our initial assumption was that an increased level of psychological well-being is associated with increased Pleasantness and Eventfulness assessments of the soundscape. Although the results showed that the psychological well-being was positively associated with Pleasantness, it was negatively associated with Eventfulness in males and individuals that did not report their occupations.

Then we hypothesized that differences in soundscape assessments are associated with demographic features. The results support this hypothesis to a certain degree. Occupation and gender appeared to be strong demographic factors influencing the Pleasantness and Eventfulness assessment. Retirement as occupation status showed to be positively attributed to the Pleasantness and negatively to the Eventfulness assessment. Further investigation revealed that the occupation (no occupation reported) was negatively associated with Pleasantness and gender (male) was negatively attributed to Eventfulness, whereas unemployment was positively associated with Eventfulness.

As expected, the majority of the total variance in the perceptual ratings is explained by the location-level differences (i.e., 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. Furthermore, an additional 3 percentage points of explained variance would represent a meaningful improvement in the performance of predictive soundscape models based on in-situ measurements of varying soundscape types (Lionello, Aletta, & Kang, 2020) and should therefore be considered when constructing these models.

Psychological well-being and its association with pleasantness and eventfulness

Our findings demonstrate a positive link between the perceived

Pleasantness and participants' psychological well-being, whereas the association between psychological well-being and Eventfulness is negative in males and individuals that did not report their occupations. Our results can be interpreted in light of previous research and it is consistent with the idea that psychological well-being underlies the perception of the external world (Kelley & Schmeichel, 2014) such as auditory input. While the enhanced global level of psychological state has a positive effect on auditory processing (Kumar, Sangamanatha, & Vikas, 2013), there is evidence that suggests an impairment of early auditory processing (analysing, blending, and acoustic input segmentation) in individuals with poor psychological well-being (Kähkönen et al., 2007). One of the potential trait biomarkers of poor psychological well-being such as depression (predominantly characterized by low mood and anhedonia (Erfanian, 2018) is the attenuation of neuronal activation in the auditory cortical area leading to alternations in auditory processing (Zwanzger et al., 2012).

Demographic factors and their associations with pleasantness and eventfulness

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 worthwhile to highlight that 'retirement' factor can be potentially a proxy for age (>65) and gender (male). To explore the effect of occupation/retirement deeper on Pleasantness and Eventfulness we removed the occupation factor from the model. Age ($\beta = 0.02, p = 0.05$) for Pleasantness ($\beta = -0.03, p = 0.01$) for Eventfulness and gender ($\beta = -0.04, p = 0.05$) for Eventfulness then came out significant (see Appendix C). This would indicate that occupation status, particularly 'retirement', represents a group of older male individuals. Even though incorporation of occupation into our model complicates the interpretation of our outcome, it results in a slightly better fitting model (R^2_c for Pleasantness (0.354) and Eventfulness (0.181) relative to (0.345) for Pleasantness and (0.165) for Eventfulness 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; Yang & Kang, 2005). Our 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 Çakir Aydin and Yilmaz (2016) in which they found that soundscape pleasantness reported by young individuals was significantly lower than the other age groups. The results withstood a control for the effect of age on the soundscape's pleasantness and eventfulness, suggesting that different neural and behavioural processes are responsible for the differences of soundscape appraisal in age.

One possibility is that age is associated with loss of function within the peripheral auditory system (hearing loss due to age or presbycusis) that may lead to the variation of the soundscape (Howarth & Shone, 2006). Higher tone frequencies have shown to be perceived less pleasant and more annoying relative to low tone frequencies (Landström, Kjellberg, SÖDerberg, & Nordström, 1994) and age-related hearing loss is most marked at higher frequencies, so missing higher frequencies (that can be potentially unpleasant) may lead to an increase in soundscape pleasantness. Second, since the human brain is highly plastic throughout the life span, by ageing, the auditory processing changes due to the temporal coding of the auditory cortex (Babkoff & Fostick, 2017; Bones & Plack, 2015). Temporal coding is the ability of the brain to encode sensory information to the action potentials that rely on precise timing.

Last, 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 (Yang & Kang, 2005).

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; Yang & Kang, 2005). These differences may be an indication of different auditory processing across genders. These differences are consistent with existing predictions of female top-down and male bottom-up strategies in spatial processing (ability to find where objects are in space) (Simon-Dack, Friesen, & Teder-Sälejärvi, 2009).

Soundscape pleasantness and eventfulness differences among locations

The Pleasantness and Eventfulness were significantly different among locations. The Pleasantness appeared to be highest in locations, dominating by nature sounds (i.e., Regents park Japan). In agreement with our results, Payne and colleagues (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. Also, Zhang (2014) reported a significant impact of natural soundscape on individuals' restorative experiences and boosting pleasantness. In the study by 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 mechanical 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 (i.e., Euston Tap) in our study, 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 (Axelsson et al., 2010; Erfanian et al., 2019).

These results insinuate the notion that there are multiple primary factors (McDermott, 2012) 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

Conclusion

We conducted a linear mixed-effects model to show the associations of psychological well-being, demographic factors with the 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. We further demonstrated that the occupation status, in particular retirement as a proxy of age and gender, was attributed to Pleasantness and Eventfulness. The findings of this study offer empirical grounds for developing and advancing theories on the influence of psychological well-being and demographic characteristics on the perception of the acoustic environment namely the soundscape.

Authorship contributions

Conception and design of study: M. Erfanian, A. Mitchell, F. Aletta, J. Kang; acquisition of data: M. Erfanian, A. Mitchell, F. Aletta; analysis and/or interpretation of data: M. Erfanian, A. Mitchell. Drafting the manuscript: M. Erfanian; revising the manuscript critically for important intellectual content: M. Erfanian, A. Mitchell, F. Aletta.

Supervision: J. Kang.

Approval of the version of the manuscript to be published (the names of all authors must be listed): M. Erfanian, A. Mitchell, F. Aletta, J. Kang

Acknowledgements

The authors are very grateful to Tin Oberman and other Soundscape Indices (SSID) project members that helped with data collection. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 740696).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jenvp.2021.101660>.

References

- Aletta, F., Oberman, T., Mitchell, A., Erfanian, M., Lionello, M., Kachlicka, M., & Kang, J. (2019). Associations between soundscape experience and self-reported wellbeing in open public urban spaces: a field study. *The Lancet*, 394, S17. [https://doi.org/10.1016/S0140-6736\(19\)32814-4](https://doi.org/10.1016/S0140-6736(19)32814-4)
- Axelsson, Ö. (2012). The ISO 12913 series on soundscape: An update. & ISO/TC 43/SC 1/WG 54. *The Journal of the Acoustical Society of America*, 131(4). <https://doi.org/10.1121/1.4708750>, 3381–3381 . (Accessed May 2012)
- Axelsson, Ö., Nilsson, M. E., & Berglund, B. (2010). A principal components model of soundscape perception. *The Journal of the Acoustical Society of America*, 128(5), 2836–2846. <https://doi.org/10.1121/1.3493436>
- Babkow, H., & Fostick, L. (2017). Age-related changes in auditory processing and speech perception: cross-sectional and longitudinal analyses. *European journal of ageing*, 14 (3), 269–281. <https://doi.org/10.1007/s10433-017-0410-y>
- Berthomieu, G., Koehl, V., & Paquier, M. (2021). Does loudness relate to the strength of the sound produced by the source or received by the ears? A review of how focus affects loudness. *Frontiers in Psychology*, 12, 13.
- Blom, E. H., Bech, P., Höglberg, G., Larsson, J. O., & Serlachius, E. (2012). Screening for depressed mood in an adolescent psychiatric context by brief self-assessment scales—testing psychometric validity of WHO-5 and BDI-6 indices by latent trait analyses. *Health and quality of life outcomes*, 10(1), 149. <https://doi.org/10.1186/1477-7525-10-149>
- Bones, O., & Plack, C. J. (2015). Losing the music: Aging affects the perception and subcortical neural representation of musical harmony. *Journal of Neuroscience*, 35(9), 4071–4080. <https://doi.org/10.1523/JNEUROSCI.3214-14.2015>
- Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology. *Cognitive neuroscience of emotion*, 25, 49–59. <https://doi.org/10.1017/CBO9780511546396.025>
- Çakir Aydin, D., & Yilmaz, S. (2016). Assessment of sound environment pleasantness by sound quality metrics in urban spaces. *AIZ ITU Journal of the Faculty of Architecture*, 13(2), 87–99.
- Clark, D. A., & Beck, A. T. (2010). Cognitive theory and therapy of anxiety and depression: Convergence with neurobiological findings. *Trends in Cognitive Sciences*, 14(9), 418–424. <https://doi.org/10.1016/j.tics.2010.06.007>
- Coffey, E. B., Colagrosso, E. M., Lehmann, A., Schönwiesner, M., & Zatorre, R. J. (2016). Individual differences in the frequency-following response: Relation to pitch perception. *PloS One*, 11(3), Article e0152374.
- Croome, D. J. (1977). *Noise, buildings and people*. United Kingdom.
- Erfanian, M. (2018). Childhood trauma: A risk for major depression in patients with psoriasis. *Psychiatry and Clinical Psychopharmacology*, 28(4), 378–385. <https://doi.org/10.1080/24750573.2018.1452521>
- Erfanian, M., Mitchell, A. J., Kang, J., & Aletta, F. (2019). The psychophysiological implications of soundscape: A systematic review of empirical literature and a research agenda. *International Journal of Environmental Research and Public Health*, 16 (19), 3533. <https://doi.org/10.3390/ijerph16193533>
- Fang, X., Gao, T., Hedblom, M., Xu, N., Xiang, Y., Hu, M., & Qiu, L. (2021). Soundscape perceptions and preferences for different groups of users in urban recreational forest parks. *Forests*, 12(4), 468.
- Goldstein, E. B., & Brockmole, J. (2016). *Sensation and perception*. Cengage Learning.
- Gulian, E., & Thomas, J. R. (1986). The effects of noise, cognitive set and gender on mental arithmetic performance. *British Journal of Psychology*, 77(4), 503. <https://doi.org/10.1111/j.2044-8295.1986.tb02214.x>
- Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., & Duda, S. N. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, 95, 103208. <https://doi.org/10.1016/j.jbi.2019.103208>
- Howarth, A., & Shone, G. R. (2006). Ageing and the auditory system. *Postgraduate Medical Journal*, 82(965), 166–171. <https://doi.org/10.1136/pgmj.2005.039388>
- Hume, K., & Ahtamad, M. (2013). Physiological responses to and subjective estimates of soundscape elements. *Applied Acoustics*, 74(2), 275–281. <https://doi.org/10.1016/j.apacoust.2011.10.009>
- International Organization for Standardization ISO/TS 12913-2:2018 acoustics — soundscape — part 2: Data collection and reporting requirements.
- International Organization for Standardization ISO/TS 12913-3. (2019). *Acoustics — soundscape — part 3: Analysis of data related to Method A*.
- Ismail, M. R. (2014). Sound preferences of the dense urban environment: Soundscape of cairo. *Frontiers of architectural research*. 3(1), 55–68. <https://doi.org/10.1016/j foar.2013.10.002>

- Kähkönen, S., Yamashita, H., Rytsälä, H., Suominen, K., Ahveninen, J., & Isometsä, E. (2007). Dysfunction in early auditory processing in major depressive disorder revealed by combined MEG and EEG. *Journal of Psychiatry & Neuroscience: Journal of Psychiatry & Neuroscience*, 32(5), 316.
- Kang, J., Aletta, F., Gjestland, T. T., Brown, L. A., Botteldooren, D., Schulte-Fortkamp, B., & Coelho, J. L. B. (2016). Ten questions on the soundscapes of the built environment. *Building and Environment*, 108, 284–294. <https://doi.org/10.1016/j.buildenv.2016.08.011>
- Kang, J., Aletta, F., Oberman, T., Erfanian, M., Kachlicka, M., Lionello, M., et al. (2019). towards soundscape indices. In *Proceedings of the international congress on acoustics—ICA, Aachen, Germany* (pp. 9–13), 9.
- Kaya, E. M., Huang, N., & Elhilali, M. (2020). Pitch, timbre and intensity interdependently modulate neural responses to salient sounds. *Neuroscience*. <https://doi.org/10.1016/j.neuroscience.2020.05.018>
- Kelley, N. J., & Schmeichel, B. J. (2014). The effects of negative emotions on sensory perception: Fear but not anger decreases tactile sensitivity. *Frontiers in Psychology*, 5, 942. <https://doi.org/10.3389/fpsyg.2014.00942>
- Krusemark, E. A., Novak, L. R., Gitelman, D. R., & Li, W. (2013). When the sense of smell meets emotion: Anxiety-state-dependent olfactory processing and neural circuitry adaptation. *Journal of Neuroscience*, 33(39), 15324–15332. <https://doi.org/10.1523/JNEUROSCI.1835-13.2013>
- Kumar, S., Forster, H. M., Bailey, P., & Griffiths, T. D. (2008). Mapping unpleasantness of sounds to their auditory representation. *Journal of the Acoustical Society of America*, 124(6), 3810–3817. <https://doi.org/10.1121/1.3006380>
- Kumar, U. A., Sangamanatha, A. V., & Vikas, J. (2013). Effects of meditation on temporal processing and speech perceptual skills in younger and older adults. *Asian Journal of Neuroscience*. <https://doi.org/10.1155/2013/304057>. 2013.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/JSS.V082.I13>
- Landström, U., Kjellberg, A., SÖDerberg, L., & Nordström, B. (1994). Measures against ventilation noise—which tone frequencies are least and most annoying? *Journal of Low Frequency Noise, Vibration and Active Control*, 13(3), 81–88. <https://doi.org/10.1177/026309239401300301>
- Laufer, O., Israeli, D., & Paz, R. (2016). Behavioral and neural mechanisms of overgeneralization in anxiety. *Current Biology*, 26(6), 713–722. <https://doi.org/10.1016/j.cub.2016.01.023>
- Lionello, M., Aletta, F., & Kang, J. (2020). A systematic review of prediction models for the experience of urban soundscapes. *Applied Acoustics*, 170, 107479. <https://doi.org/10.1016/j.apacoust.2020.107479>
- Lucas-Carrasco, R., Allerup, P., & Bech, P. (2012). The validity of the WHO-5 as an early screening for apathy in an elderly population. *Current gerontology and geriatrics research*. <https://doi.org/10.1155/2012/171857>. 2012.
- Lüdecke, D. (2018). sjPlot: Data visualization for statistics in social science. *R package version*, 2(1).
- McDermott, J. H. (2012). Auditory preferences and aesthetics: Music, voices, and everyday sounds. In *Neuroscience of preference and choice* (pp. 227–256). Academic Press. <https://doi.org/10.1016/B978-0-12-381431-9.00020-6>.
- Mitchell, A., Oberman, T., Aletta, F., Erfanian, M., Kachlicka, M., Lionello, M., et al. (2020). The soundscape indices (SSID) protocol: A method for urban soundscape surveys—questionnaires with acoustical and contextual information. *Applied Sciences*, 10(7), 2397. <https://doi.org/10.3390/app10072397>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in ecology and evolution*, 4(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press.
- Patchett, R. F. (1979). Human sound frequency preferences. *Perceptual & Motor Skills*, 49 (1), 324–326.
- Payne, S. R. (2013). The production of a perceived restorativeness soundscape scale. *Applied Acoustics*, 74(2), 255–263. <https://doi.org/10.1016/j.apacoust.2011.11.005>
- Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3), 715.
- Ren, X., Kang, J., & Liu, X. (2016). Soundscape perception of urban recreational green space. *Landscape Architecture Frontiers*, 4(4), 42–56. <https://doi.org/10.1121/1.3693644>
- Riskind, J. H., Kleiman, E. M., Seifritz, E., & Neuhoff, J. (2014). Influence of anxiety, depression and looming cognitive style on auditory looming perception. *Journal of Anxiety Disorders*, 28(1), 45–50. <https://doi.org/10.1016/j.janxdis.2013.11.005>
- Russell, J. (1980). A circumplex model of emotions. *Journal of Personality and Social Psychology*, 39, 1161–1178.
- Schneider, P., & Wengenroth, M. (2009). The neural basis of individual holistic and spectral sound perception. *Contemporary Music Review*, 28(3), 315–328.
- Simon-Dack, S. L., Friesen, C. K., & Tedér-Sälejärvi, W. A. (2009). Sex differences in auditory processing in peripersonal space: An event-related potential study. *NeuroReport*, 20(2), 105. <https://doi.org/10.1097/WNR.0b013e32831befc1>
- Team, R. C. (2013). *R: A language and environment for statistical computing*.
- Tomkins, S. S. (1962). Affect, imagery. In *Consciousness: Cognition: Duplication and transformation of information* (Vol. 4). Springer Publishing Company.
- Topp, C. W., Østergaard, S. D., Sondergaard, S., & Bech, P. (2015). The WHO-5 well-being index: A systematic review of the literature. *Psychotherapy and Psychosomatics*, 84(3), 167–176. <https://doi.org/10.1159/000376585>
- Weinstein, N. D. (1978). Individual differences in reactions to noise: A longitudinal study in a college dormitory. *Journal of Applied Psychology*, 63(4), 458. <https://doi.org/10.1037/0021-9010.63.4.458>
- World Health Organization. (1998). *Quality control methods for medicinal plant materials*. World Health Organization.
- Xiao, J., & Hilton, A. (2019). An investigation of soundscape factors influencing perceptions of square dancing in urban streets: A case study in a county level city in China. *International Journal of Environmental Research and Public Health*, 16(5), 840. <https://doi.org/10.3390/ijerph16050840>
- Yang, W., & Kang, J. (2005). Soundscape and sound preferences in urban squares: A case study in Sheffield. *Journal of Urban Design*, 10(1), 61–80. <https://doi.org/10.1080/13574800500062395>
- Zadra, J. R., & Clore, G. L. (2011). Emotion and perception: The role of affective information. *Wiley interdisciplinary reviews: Cognitive Science*, 2(6), 676–685. <https://doi.org/10.1002/wcs.147>
- Zhang, Y. (2014). Research on soundscape restorative benefits of urban open space and promotion strategy of the acoustic environment quality. *New Archit*, 165, 18–22. https://doi.org/10.4103/nah.NAH_73_16
- Zhang, M., & Kang, J. (2007). Towards the evaluation, description, and creation of soundscapes in urban open spaces. *Environment and Planning B: Planning and Design*, 34(1), 68–86. <https://doi.org/10.1068/b31162>
- Zwanzer, P., Zavorotny, M., Diemer, J., Ruland, T., Domschke, K., Christ, M., & Pfleiderer, B. (2012). Auditory processing in remitted major depression: A long-term follow-up investigation using 3T-fMRI. *Journal of Neural Transmission*, 119(12), 1565–1573. <https://doi.org/10.1007/s00702-012-0871-2>

6 Multilevel Annoyance Modelling of Short Environmental Sound Recordings Including Sound Source Information

Published as: Orga, F., **Mitchell, A.**¹, Freixes, M., Aletta, F., Alsina-Pagès, R. M., & Foraster, M. (2021). Multilevel Annoyance Modelling of Short Environmental Sound Recordings. *Sustainability*, 13(11), Article 11. <https://doi.org/10.3390/su13115779>

draft *This chapter will be re-worded and somewhat reorganised to better fit my thinking and to highlight the analysis work. Its focus within the thesis is on the use of sound source information to improve the prediction of annoyance. Right now this is wholesale copied from the published paper.*

Abstract

draft *The recent development and deployment of Wireless Acoustic Sensor Networks (WASN) present new ways to address urban acoustic challenges in a smart city context. A focus on improving quality of life forms the core of smart-city design paradigms and cannot be limited to simply measuring objective environmental factors, but should also consider the perceptual, psychological, and health impacts on citizens. This study therefore makes use of short (1 - 2.7s) recordings sourced from a WASN in Milan which were grouped into various environmental sound source types and given an annoyance rating via an online survey with $N = 100$ participants. A multilevel psychoacoustic model was found to achieve an overall $R^2 = 0.64$ which incorporates Sharpness as a fixed effect regardless of the sound source type and Roughness, Impulsiveness, and Tonality as random effects whose coefficients vary depending on the sound source. These results present*

¹Joint first author

a promising step toward implementing an on-sensor annoyance model which incorporates psychoacoustic features and sound source type, and is ultimately not dependent on sound level.

6.1 Introduction

[draft] Noise has been proven to have a wide impact on the social and economic aspects of citizens' lives (?) and is regarded as one of the primary environmental health issues referenced in the new environmental noise guidelines (?). Over the past few years, several research teams have analysed the causes and the impact of this noise, revealing that it causes more than 48,000 new cases of ischemic heart disease and around 12,000 deaths in Europe each year (?). Furthermore, it leads to chronic high annoyance for more than 22 million people, and sleep disturbance for more than 6.5 million people (?). One of the main noise sources according to research is road traffic noise (?), causing psychological reactions in citizens ? and even cardiovascular diseases (?).

[draft] Other studies analyse the effects of aircraft noise on sleep [cit] and learning impairments in children [cit] . Also, railway noise has proven to cause annoyance due to its huge variety of sounds, e.g. rail breaks, whistles, squeals, and vibrations [cit] 8, 9 . Most of the literature focuses on sound level measurements and the corresponding annoyance [cit] 10 , but other acoustical and psychoacoustical characteristics could be taken into account, e.g. loudness or sharpness [cit] 11 , in order to understand the degree of noise annoyance and identify the characteristics of sounds that may be more detrimental to psychological well-being and consequently for health. Such knowledge is relevant for policy makers and urban planners in order to create healthy environments.

[draft] Several tests used in studies to evaluate the effects of environmental noise for citizens [cit] 12 can be used to design this model. This study uses real-life data and its sound characterisation, thus focusing on noise sensitivity was not the closest approach to the problem. The tests used as a basis in this work have been defined with the purpose of finding new ways of analysing the impact of sound – usually traffic – on citizens in urban environments [cit] 13, 14 , in order to model the annoyance perception [cit] 15, 16 .

[draft] The perceptual tests were designed to measure the annoyance in people relating to different urban sounds and their characteristics [cit] 17, 18 , by means of short excerpts of raw acoustic audio obtained from the DYNAMAP project [cit] 19 . The most representative audio excerpts were selected, using a wide range of sound types (sirens, airplanes, people talking, dogs barking, etc.) [cit] 20, 21 . However, sound annoyance

depends on the acoustic characterisation of each sample, and it is possible to classify the acoustic excerpts depending on their characterisation, which can be the basis to ask participants about their perceptions. The characterisation is based on the psychoacoustic measurements of loudness, sharpness, and others defined by Zwicker [cit 11](#).

draft *The researchers asked more than 100 people to conduct the perceptual tests* [cit 18](#). *Some preliminary results of the three tests conducted were published in* [cit 17](#) *in which the relationship between sharpness and annoyance was analysed by means of an A/B test* [cit 22](#), *and later on in* [cit 18](#), *where some of the research questions were formulated. In this study, I aim to determine the parameters that have an effect in the individual annoyance scores. For this reason, a multilevel psychoacoustic model is trained using the results of the MUSHRA* [cit 23](#) *test, essentially focused on annoyance evaluation by the participants over several different types of sound, while loudness and sharpness were kept constant. The results show that the differences in annoyance perception between the different demographic groups is not statistically significant and that sharpness is the main predictor for annoyance.*

The chapter is structured as follows: Section 6.2 details the state-of-the-art of annoyance modelling by means of subjective data collection. Section 3.5 describes the procedure followed in this work, including the dataset and the design of the perceptual test. In section ?? the results obtained from the perception tests are presented and discussed, and the annoyance model is proposed. Section ?? contains the discussion and, finally, Section ?? presents the conclusions of the study.

6.2 State of the Art of Annoyance Evaluation and Modelling

In this section I gather a short synthesis of the most relevant contributions of the state-of-the-art on which the design of the tests and the modelling of perceptual annoyance have been based.

6.2.1 Evaluation of Annoyance

The evaluation of annoyance can be found in literature by means of the objective parameters related to sound and noise [cit 10](#). Nevertheless, when the goal is to measure the perception – the real annoyance experienced by people – one of the most frequently used methods is to conduct a survey to measure the degree of annoyance produced by different sounds [cit 24, 25,26](#). Following the recommendation of the International Committee

for the Biological Effects of Noise (ICBEN), this evaluation should be done in a qualitative way, using a verbal scale; this can be translated into *not at all, slightly, moderately, very* and *extremely*, just to give a few examples. Also an 11-point scale – also from an ICBEN recommendation – can be used, where in this case, zero corresponds to *not at all* and 10 corresponds to *extremely disturbing*.

Furthermore, taking advantage of the experience in soundscape evaluation [cit 27](#) citizens can be asked about other aspects besides annoyance. To this end a perceptual assessment based on a Likert scale [cit 28](#) could be used. This scale defines five levels of agreement with a given statement: *Strongly disagree, Disagree, Neither agree nor disagree, Agree* and *Strongly agree*. This scale was used in [cit 17,18](#) to evaluate several types of noise sources according to a small group of attributes such as *loud, shrill, noisy, disturbing, sharp, exciting, calming* and *pleasant* (see the complete list of adjectives in [cit 27](#)).

Borrowing from the subjective assessment of audio quality, the MUSHRA method has been also used for the evaluation of annoyance in [cit 17, 18](#). MUSHRA was described and designed by ITU-R under the recommendation ITU-R BS.1534-3 [cit 23](#). This recommendation gives guidelines on listening tests and subjective assessment, as well as audio quality (among other applications), assuming that the best way to evaluate audio quality is by means of subjective listening.

Listening tests can be conducted in a controlled scenario (e.g. in an anechoic chamber) thus allowing the organiser to have control over the setup and experimental design. Nevertheless, this approach is expensive and time consuming. Alternatively, online listening tests have been widely used in the perceptual evaluation of audio quality or speech synthesis systems, even resorting to crowdsourcing strategies [cit 29](#). These tests can be run in parallel and anywhere, thereby reducing costs and allowing researchers to reach a wider audience [cit 30](#).

6.2.2 Annoyance Prediction

After the design and execution of the perceptual tests, the resulting evaluation coming from participants are used to generate a model that can predict the annoyance value depending on the type and the parameters of the noise excerpt under study. One of the most representative examples of annoyance modelling is found in [cit 15](#), where a model based on the hypothesis that annoyance is primarily determined by the detection of intruding sounds is presented. The model takes into account several measurable elements:

1. signal-to-noise ration (SNR);
2. indoor background level;

3. the activity conducted by the listener, assuming that in the conducted tests, their main activity is not listening to events.

The model is obtained from the results of a test evaluating annoyance and acoustic data from a field experiment in a natural setting.

Another reference model for annoyance prediction is found in [cit 16](#), where the authors model and predict road traffic noise annoyance based on:

1. noise perception;
2. noise exposure levels;
3. demographics.

The authors apply machine-learning algorithms in order to conduct the prediction and measure error rates, which give them a good trade-off in the prediction of the traffic noise annoyance, with a strong dependence on subjective noise perception and predicted noise exposure levels, assuming that the classical statistical approaches fail in their predictions in terms of accuracy.

A model of annoyance based on a combination of psychoacoustic metrics was proposed by Zwicker and Fastl (2007). Generated from laboratory-collected data, this model attempts to provide a method to directly calculate the relative annoyance values of single-source sounds from the psychoacoustic Loudness, Roughness, Sharpness, and Fluctuation Strength. This model has also been further expanded upon to include a term for the Tonality of the sound [cit 31](#). However, this model was developed based on laboratory studies of generated, simple sounds (i.e. not real recorded sounds) and does not take into account the semantic information associated with the real environmental sounds present in an urban environment.

In [cit 32](#), the authors led us to a better understanding of the transportation noise-annoyance response, in three different and relevant approximations:

1. to unravel the factors that affect the annoyance response of people in reference to the mixed transportation noise;
2. to contrast the noise-annoyance dependence in situations where road traffic and railway noise dominate;
3. to detail the differences between those two using structural equation modelling.

As expected, the results show that annoyance is largely determined by noise disturbance and the noisiness perceived by citizens. Finally, in [cit 33](#) an approach to develop a road traffic noise prediction model is presented, and it takes into account:

1. social aspects
2. characteristics of traffic, and
3. urban development

It is based on the creation of a local model, with a pilot in Istanbul (Turkey), which uses all the information gathered for the creation of the noise maps as an input, and provides annoyance levels prediction as an output, complementing the noise maps which provide no subjective indicator.

6.3 Methods

In this section, I detail the several methods applied in this experiment from the perceptual test design based on an urban sound dataset [cit 21](#) to the multilevel linear regression modelling applied to obtain the annoyance prediction.

6.3.1 Dataset

In order to obtain a proper representation of the acoustic environment in the design of the perceptual tests, a large quantity of recorded data is needed. The data gathered in this project belongs to different recording times and urban locations, using the Wireless Acoustic Sensor Network (WASN) deployed in Milan (Italy) in the framework of the LIFE DYNAMAP project [cit 19, 21](#).

Gathering the data through a WASN facilitates the collection of a wide and accurate representation of the acoustic events, because it keeps the same recording conditions in every node and allows the retrieval of data 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 [cit 34](#). After that, experts from the DYNAMAP development team labelled the acoustic events of the recordings manually to obtain a 151-h dataset [cit 21](#). Due to the nature of the project, that 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 [cit 20](#). During the labelling process, the DYNAMAP developers found up to 26 types of anomalous events, which 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) [cit 35](#).

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 [cit 36](#). 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 [cit 17](#). The reason for that choice was double:

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 are the most probable to generate annoyance due to their repetitiveness.

The comparison between events was only carried out with sounds collected using the same sensor, in order to ensure the same recording conditions. For this reason, if the chosen events for the perceptive tests belong to a sensor or another, depends on the availability of the classes to be compared in each sensor. In all the cases, measures were taken to ensure that the sensor containing the events has enough variety of samples with various psychoacoustic parameters, to ensure a proper representation of each category. To satisfy these requirements, only data from four sensors have been used to make the comparisons, as they provide enough information to carry out the perceptual test, i.e. hbi15, hbi24, hbi27, and hbi33 [cit 20](#). More details about the event selection process and the availability of the study sensors are detailed in [cit 17](#), and the time of each event in the sensors is depicted in [cit 18](#).

6.3.2 Design of the Perceptual Tests

In order to assess the degree of annoyance produced by the aforementioned classes of sounds, an online test has been conducted using the Web Audio Evaluation Tool [cit 30](#). Specifically, the MUSHRA test method [cit 23](#) – which was originally designed for the

evaluation of audio codecs – has been adapted for that purpose. Participants were given a clear explanation of what they were asked, including detailed instructions on the operation of the test. No training phase was therefore considered. A demographics survey was included at the beginning of the test for all 100 participants, asking for them to identify 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 audios, 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 annoyance that they perceive. Each audio is represented with a green bar with a "play" icon on it and the audios are sorted randomly along the MUSHRA scale (see Figure 6.3.2). An audio is reproduced when the corresponding bar is clicked. The system ensures the participant listens to all the audios moves all the bars before they jump to the next set of audios. 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. Nonetheless, since stimuli pertaining to different classes are compared, no audio reference was included, thus avoiding biases towards a certain audio class. Moreover, 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 6.3.2.

Figure 6.1: 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*.

Table 6.1: Psychoacoustic parameters calculated for the 27 stimuli used in the listening experiment.

6.3.3 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* [cit 11](#); values for these parameters are reported in Table 6.3.2. 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 [cit 37, 38](#). 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 [cit 39](#). 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, in a free-field [cit 39](#). 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 [cit 40](#), 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), which is based on the hearing model of Sottek, with a frequency range of 20 Hz to 20 kHz [cit 41](#).

6.3.4 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 [cit 42, 43](#) and for applied prediction models [cit 44](#),

[45](#). MLM allows 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 [cit 46](#), is thus:
 annoyance $N_5 + R + S + T + I + (\text{I} - \text{user}) + (\text{I} + N_5 + R + S + T + I - \text{label})$

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. [draft Equation 1](#)). The feature with the highest p -value (least significant) 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 Variance Inflation Factor (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 [cit 47](#).

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 [cit 47](#). The model fitting and feature selection was performed using the 'step' function from

Figure 6.2: Estimated Marginal Means for Annoyance as a function of Areas of residence (**left**) and Gender (**right**).

‘lmerTest’ (v. 3.1.3) [cit 48](#) in the R statistical software (v. 4.0.5) [cit 49](#). The summaries and plots were created using the ‘sjPlot’ package (v. 2.8.7) [cit 50](#) and the multi-level R^2 values were calculated using ‘MuMIn’ (v. 1.43.17) [cit 51](#).

6.4 Results

6.4.1 Differences in Annoyance between Groups

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 6.4.1.

6.4.2 Annoyance Model

The modelling process returned some interesting results about the parameters that have an effect in predicting the individual annoyance scores. 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

Table 6.2: Random intercept-random slope multi-level model of psychoacoustic 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.

Figure 6.3: Second-level effects figures representing the regression coefficients by types of sound (label) and for different psychoacoustic parameters.

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 6.4.2. This implies that, regardless of any other factors, the sharper the sounds, the more annoying these are perceived to be.

The second-level effects presented in Figure 6.4.2 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. 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, accordingly. 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.

6.5 Discussion

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 [cit 52](#). A relevant issue to consider from the WASN perspective, is that previous studies conducted in both urban [cit 21](#) and [cit 20](#) environments, there is a clear influence of the type of environment around the sensor location on the types of noise detected. Not all the urban or suburban locations for sensors have frequent sirens or horns, it depends on the more common activities (leisure, hospitals, etc.), the type of road (wide, narrow) and even the type of building or house existing in the surroundings, the types of noise detected in the street and their frequency of occurrence varies widely. In the design of a generalist model for quality of life, the number of occurrences, together with the duration and the annoyance caused by all and each noise source should be taken into account, so the former variables in cities and suburban environments is considered.

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 [cit 53](#). 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.

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, begin 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 im-

possible for fixed sensors to take into account, so once again it is preferable not to rely on loudness as a predictor.

Positivity (or the absence of negativity?)

[draft NEW!!] From the experience of the previous studies which are highly focused on the existing environmental acoustic and psychoacoustic metrics, one (of many) potential limitations has been revealed. For the most part, these metrics were designed to characterise various negative qualities of the sound. Certainly, they therefore have a negative correlation with positive assessments of the sound, but the simple fact is that they were conceived of and implemented in an attempt to quantify some sonic characteristic that was assumed by the researchers to contribute to a negative perception. Hence why in Zwicker's empirical formula for Psychoacoustic Annoyance (Zwicker and Fastl, 2007) $PA = N_5(1 + \sqrt{\omega_S^2 + \omega_{FR}^2})$, all of the constituent parts have positive coefficients.

While this would not theoretically hinder a formula for describing positive aspects of the sound, it creates a sort of conceptual barrier. If all of these metrics are designed to capture negative aspects of the sound, then it is insufficient to use them create a formula to describe a positive sound, since that formula would only represent the 'absence of negativity', not necessarily positivity.

6.6 Conclusions

In this study, 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 us 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 impulsive-

ness, roughness, and tonality as random effects allowed to vary according to the type of sound (the second level) as predictors for perceived noise annoyance.

Taken together, the results of this study encourage us to continue our research work at all the stages described in this paper. The improvement of real-time algorithms to automatically detect the predefined sound sources under study is the first stage to gathering the most relevant samples in all and each of the sensors of a WASN. The application of the annoyance modelling can give the WASN a dimension without precedent: the availability of the objective acoustic measurements conducted by the sensors, and the estimation of annoyance in a real-time evaluation by means of the model. We can start to think about a dynamic annoyance map, which could be more far-reaching than a dynamic noise map.

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

This paper is part of a special issue on COVID-19 Pandemic Acoustic Effects.

draft *Will update to match published version, but essentially won't need to change.*

Abstract

The unprecedented lockdowns due to COVID-19 in spring 2020 triggered changes in human activities in public spaces. A predictive modelling approach was developed to characterise the resulting change in the perception of the sound environment when people could not be surveyed. Building on a database of soundscape questionnaires ($N = 1,136$) and binaural recordings ($N = 687$) collected in 13 locations across London and Venice during 2019, new recordings ($N = 571$) were made in the same locations during the 2020 lockdowns. Using these 30-second-long recordings, linear multi-level models were developed to predict the soundscape pleasantness ($R^2 = 0.85$) and eventfulness ($R^2 = 0.715$) during the lockdown and compare the changes for each location. The performance was above average for comparable models. An online listening study also investigated the change in sound sources within the spaces. Results indicate:

1. human sounds were less dominant and natural sounds more dominant across all locations;
2. contextual information is important for predicting pleasantness but not for eventfulness;

3. in general, perception shifted towards less eventful soundscapes and to more pleasant soundscapes for previously traffic-dominated locations, but not for human- and natural-dominated locations.

This study demonstrates the usefulness of predictive modelling and the importance of considering contextual information when discussing the impact of sound level reductions on the soundscape.

7.1 Introduction

7.1.1 Review of the impacts of COVID-19

The global emergency cause by the 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 et al., 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 (?). 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 (?) and conurbations in the south of France (?). Impacts have also been reported at a scale of a multimillion city such as Madrid (Asensio et al., 2020b) or Barcelona (?) as well as at a more local, city-centre or even public space-scale in cities such as Stockholm (?), London (Aletta et al., 2020b), Girona (?), 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.

Those studies were mostly focussed around the L_{Aeq} , as well as a standardization approach to reporting subsequent changes in soundscape proposed by Asensio et al. (2020a). They were 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 ability 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 (Lee and Jeong, 2021). They 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 (?).

? 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 ? 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.

The impacts of the lockdown in London

Aletta et al. (2020b) explored the impacts of the COVID-19 lockdowns on the acoustic environment in London in particular, through many short-term (30s) binaural recordings and presented an important early analysis of the changes in acoustic features during the lockdown. The data from this study form the basis for the modelling presented in this chapter; the results and analysis of (Aletta et al., 2020b) were performed by the author of this thesis and are reviewed in depth.

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 "staying 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 [cit 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 officers 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 et al., 2021).

Taking advantage of short-term acoustic measurements carried out in 2019 which formed the ISD, before the implementation of the lockdown measures related to COVID-19, a number of urban locations were selected in London where data were available and measurements were performed again according to the same protocols to assess the extent of sound level variation achieved at each site.

Data analysis From the binaural recording datasets (both the 2019 and 2020 series), the following acoustic parameters were computed for the left and right channels and the arithmetic average was presented: L_{Aeq} , L_{A10} , L_{A90} . There is still no clear consensus on how to merge binaural psychoacoustic readings into single values; in the case presented in Aletta et al. (2020b) an arithmetic average level was deemed to be acceptable as the interaural level difference was typically very small (less than 1 dB) [cit 27](#).

The same procedure was followed for the psychoacoustic metrics of Loudness (N_5 , sone) and Sharpness (S , acum). All parameters were computed using the ArtemiS Suite software (v11.5, HEAD acoustics GmbH). Loudness was calculated according to the ISO 532-1 standard for time-varying sounds, in a free-field, with the remaining analysis options left to their default (ISO 532-1:2017, 2017). 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, in a free-field, with the remaining analysis options left to their default [cit DIN45692](#). It was decided to include psychoacoustic parameters as they can often provide more nuances about the perception of the acoustic environment by people, as well as offer more insights into spectral features that could reflect changes in sound sources [cit 30,31](#). Considering the necessity of keeping the computational time limited in this initial study, Loudness and Sharpness were selected because they have been reported to provide a sensible representation of perceptual aspects. [cit 32](#)

Effect of the urban setting on sound levels reduction In addition to strictly documenting the changes in the sound environment in London, the study also aimed to

Figure 7.1: Reproduced with permission from Aletta et al. (2020b). On the left: Sound levels distributions at the 11 London locations before and during the lockdown measures implementation; on the right: Sound levels distributions (aggregated across locations) and corresponding mean values before and during the lockdown measures implementation.

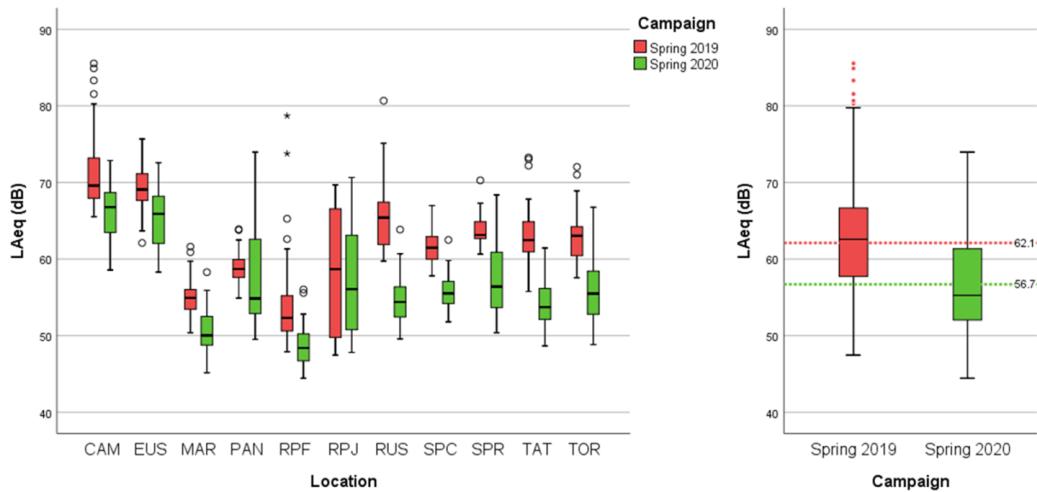


Figure 7.2: Reproduced with permission from Aletta et al. (2020b). Loudness distributions at the 11 London locations before and during the lockdown measures implementation.

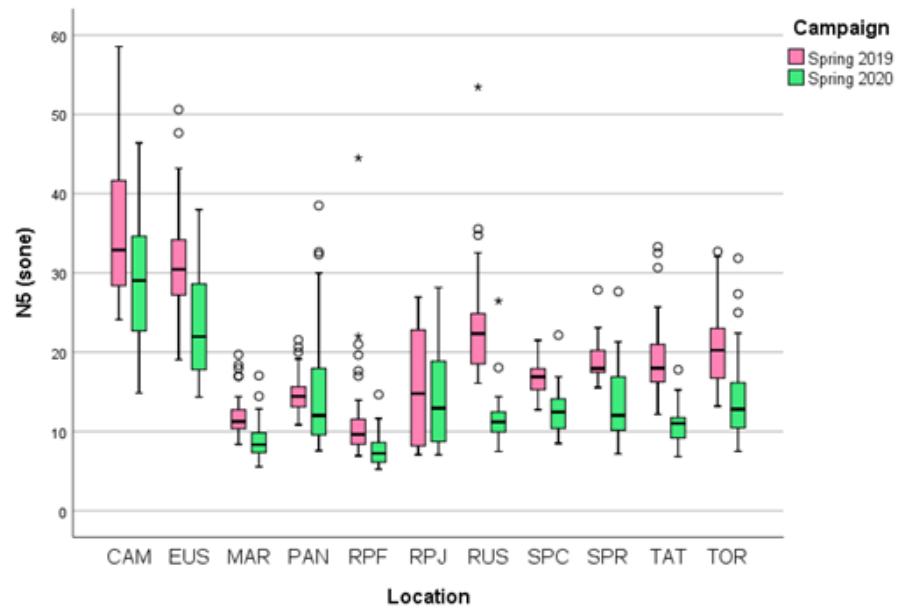
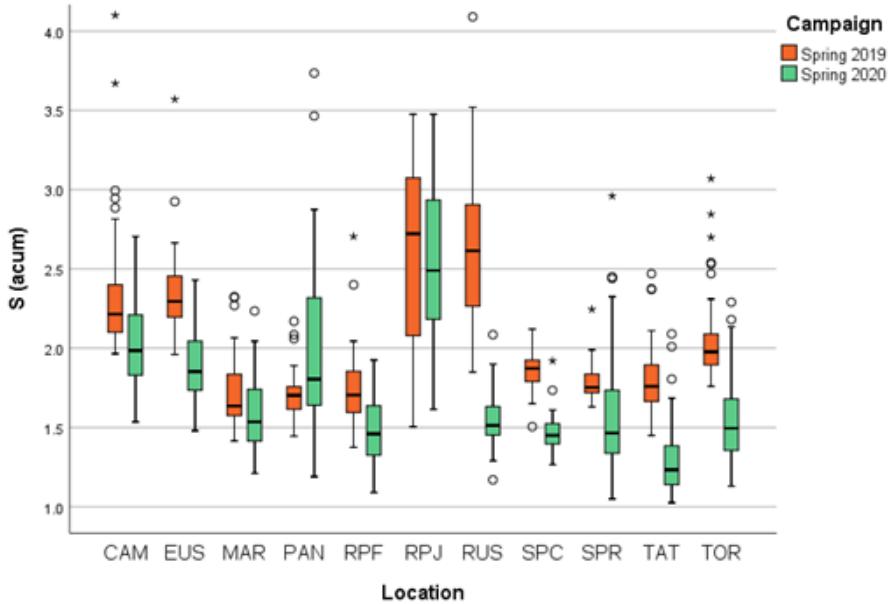


Figure 7.3: Reproduced with permission from Aletta et al. (2020b). Sharpness distributions at the 11 London locations before and during the lockdown measures implementation.



investigate whether the lockdown measures would result in different sound level reductions depending on the urban 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 , 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 report in Figure 7.1.1 ("elbow method") [cit 33](#). The analysis was conducted in R [cit RStats](#) and figures were produced using the package factoextra [cit factoextra](#).

Figure 7.1.1 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 [cit 36, 37](#).

Figure 7.4: Reproduced with permission from Aletta et al. (2020b). "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.

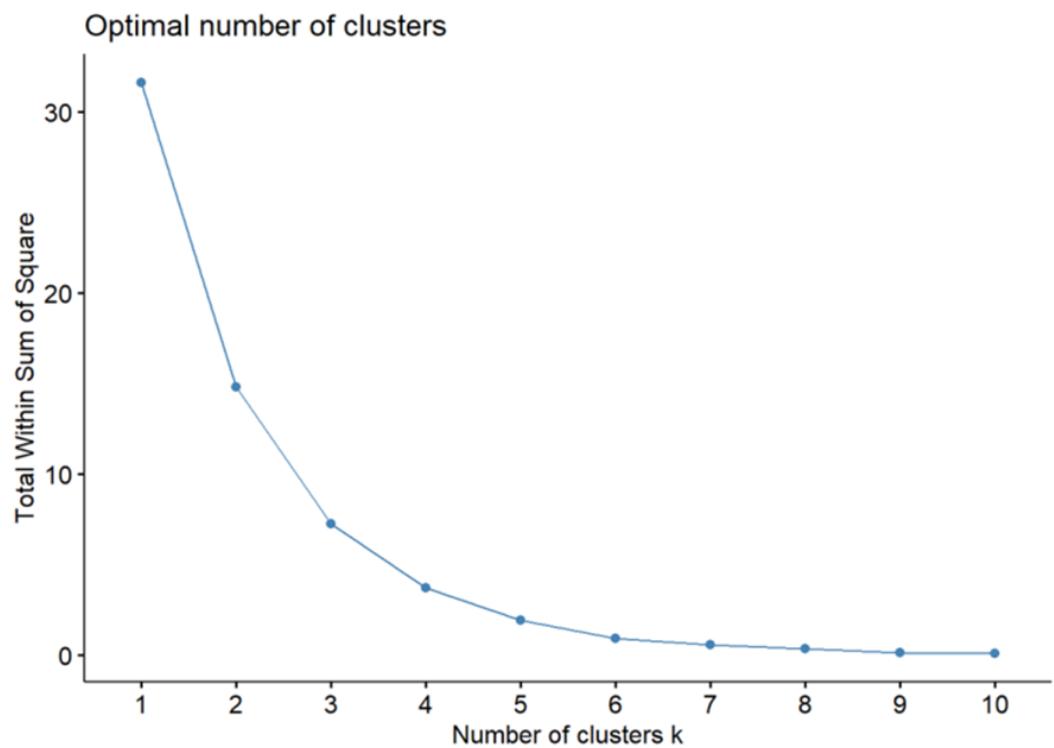


Figure 7.5: Reproduced with permission from Aletta et al. (2020b). Di-dimensional plot for the three-cluster solution. The clusters have been labelled as: Cluster 1 – Quiet Areas; Cluster 2 – Active Areas; Cluster 3 – Traffic-dominated Areas.

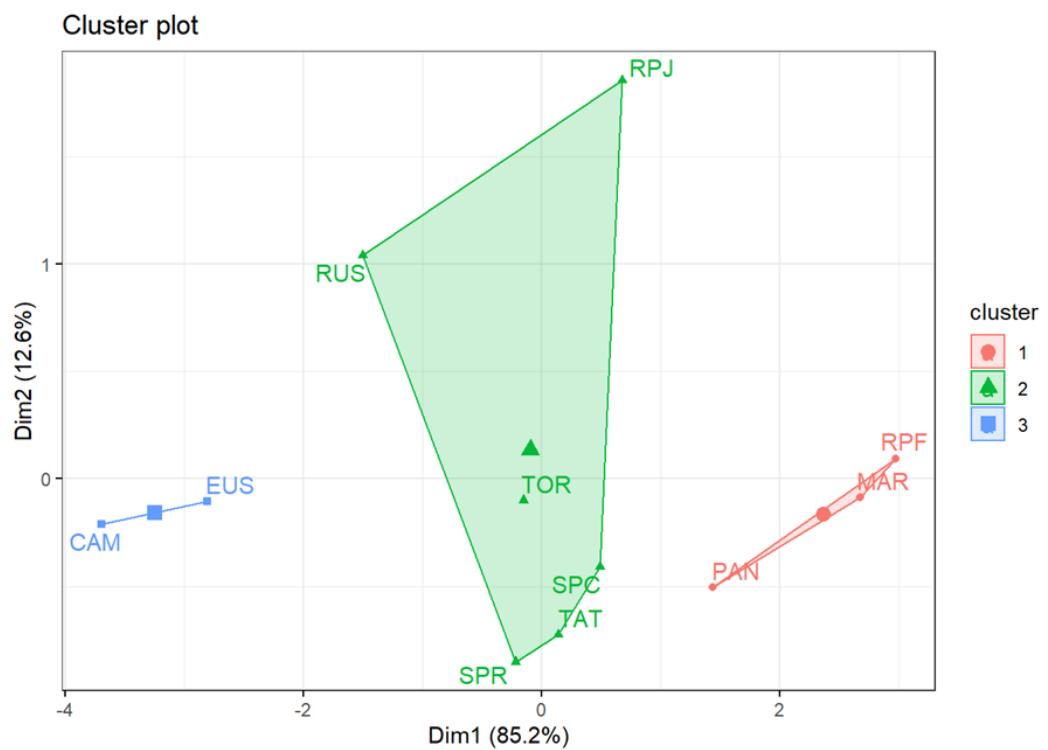


Table 7.1: Reproduced with permission from Aletta et al. (2020b). Descriptive statistics of the psychoacoustic metrics for the three identified clusters.

Table 7.1.1 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 [cit 26](#), the three clusters could be labelled as: *Traffic-dominated areas* (locations: Camden Town, Euston Tap), *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 Fields). Traffic-dominated areas are on major roads, where traffic noise is the dominant sound source. 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).

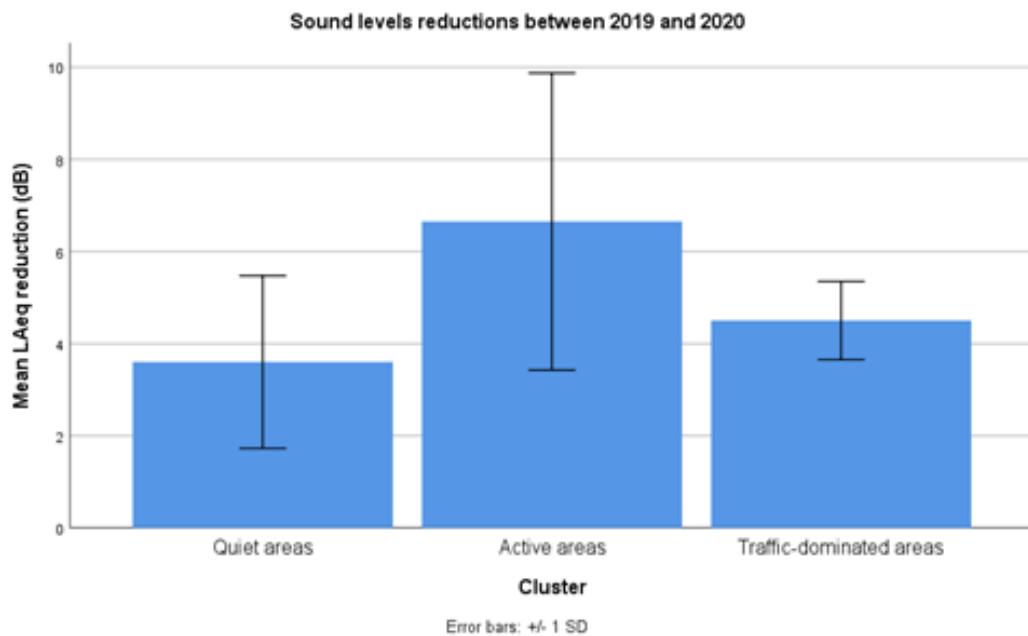
When considering the mean L_{Aeq} reductions between 2019 and 2020 as a function of Area type, it can be observed that they vary across the three clusters, as shown in Figure 7.1.1. The biggest reductions are for Active areas ($M = 6.6$ dB; $SD = 3.2$ dB), followed by Traffic-dominated areas ($M = 4.5$ dB; $SD = 0.8$ dB), and Quiet areas ($M = 3.6$ dB; $SD = 1.9$ dB). A possible explanation for this is that road 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.

This study revealed that average reductions in the various locations considered ranged from 10.7 dB (L_{Aeq}) to 1.2 dB, with an overall average reduction of 5.4 dB. This metric-reporting focussed approach left the following research questions unanswered:

1. How would people have perceived these 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)

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:

Figure 7.6: Reproduced with permission from Aletta et al. (2020b). Mean A-weighted equivalent sound level reductions between the pre- and during-lockdown conditions as a function of cluster membership (i.e. Area type).



- 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?

7.1.2 Review of predictive modelling

These questions arise out of the soundscape approach, which is characterised by prioritising the perceptual effect of an acoustic environment by taking into account the interaction of sound sources, context, and the person perceiving it (ISO 12913-1:2014, 2014; Truax, 1999), bringing together objective and subjective factors. The soundscape approach to noise mitigation and management is being recognised as a response to arising environmental requirements on noise pollution and sustainability, such as the regulation of quiet areas in Europe (?Aletta and Kang, 2018; Radicchi et al., 2021). This has been further formalised in ISO/TS 12913-2:2018 (2018) via the adoption of the circumplex model of soundscape (Axelsson et al., 2010), in which the perception of a soundscape can be described in terms of its pleasantness and eventfulness, as one of the standard methods of

soundscape assessment.

Soundscape research is therefore traditionally rooted in environmental acoustics and environmental psychology, typically dealing with outdoor spaces (Torresin et al., 2020) and urban open spaces, where parks and squares are often used as case study sites (?). A soundscape assessment typically requires people to be surveyed but the presence of people at a location influences assessment (Aletta and Kang, 2018) and 'quiet places' usually require low numbers of users to remain quiet, which limits the possibility of an assessment. Even in a crowded public space, soundscape surveys are demanding as they require significant resources to carry out at scale, limiting their widespread application (Mitchell et al., 2020). Therefore, a need for a predictive model arises to overcome this limitation and improve the implementation of the soundscape approach into everyday planning and management practices.

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 et al., 2021; Tarlao et al., 2020) or social media (Aiello et al., 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 et al. (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.

Therefore, a third research question arises: 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)?

7.2 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 et al., 2020b), including objective acoustic data (2019 and 2020) and subjective responses (2019 only). The full in situ dataset, as described in this section, has been made publicly available as "The International Soundscape Database (Vo.2.1)" on Zenodo (<https://zenodo.org/record/5654747>) (Mitchell et al., 2021a).

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 supplementary files. 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 (?).

7.2.1 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. Given the nature of this study as addressing a scenario when people could not be surveyed, no additional 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 (Aletta et al., 2020b; Mitchell et al., 2020), 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 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 ISO Pleasant and ISO Eventful and can be considered to form an x-y coordinate point ($x = \text{ISO Pleasant}$, $y = \text{ISO Eventful}$) as explained in detail in Lionello et al. (2021).

Data cleaning

The cleaning of the samples was conducted using the ArtemiS SUITE II. The researcher 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 7.2.1. 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 note 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 the supplementary files, 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 II to calculate the suite of 11 acoustic and psychoacoustic features given in 7.2 to be used as initial predictors.

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 7.1. In all, they are expected to reflect the sound level (L_{Aeq}), perceived sound level (glsn ζ), spectral content (S , $L_{Ceq} - L_{Aeq}$, T), temporal character or

Table 7.2: 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 (5th percentile)	N_5	sones	ISO 532-1:2017 (2017)
Sharpness	S	acum	ISO 532-1:2017 (2017)
Roughness	R	asper	Sottek and Genuit (2005)
Impulsiveness	I	iu	Sottek and Genuit (2005)
Fluctuation Strength	FS	vacil	Sottek and Genuit (2005)
Tonality	T	tuHMS	Sottek and Genuit (2005)
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)

predictability (I , FS , RA), and overall annoyance (PA). These metrics have been proposed as indicators to predict perceptual constructs of the soundscape (Aletta et al., 2017, 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 7.3 shows the Pearson correlation coefficient between each of the candidate acoustic features and the outcome pleasantness and eventfulness. For ISO Pleasant (*ISOPl*), we can perhaps 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 ISO Eventful (*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

Table 7.3: Pearson correlation coefficients between candidate acoustic features and ISO-Pleasant 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}	$L_{A10} - L_{A90}$
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

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.

7.2.2 Modelling

Two linear multi-level models (MLM) were computed to predict: 1) ISOPleasant, and 2) ISOEventful. The inherent grouped structure of the SSID database 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 improve-

ment 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) (?). To check for multicollinearity among the selected features, the 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 [draft Appendix B](#) 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, we use the Mean Absolute Error (MAE), which is in the scale of the response feature – for ISO Pleasant and ISO Eventful 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, we 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) (?) in R statistical software (v.4.0.3) (?). The summaries and plots were created using the 'sjPlot' package (v.2.8.6) (?) and 'seaborn' (v.0.11.1) (Waskom, 2021).

7.2.3 Online Survey

A online listening test was conducted using the Gorilla Experiment Builder (?). 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 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 report in [draft Appendix A](#).

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 (?) and a headphone reproduction level adjustment test (?). 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 ? 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

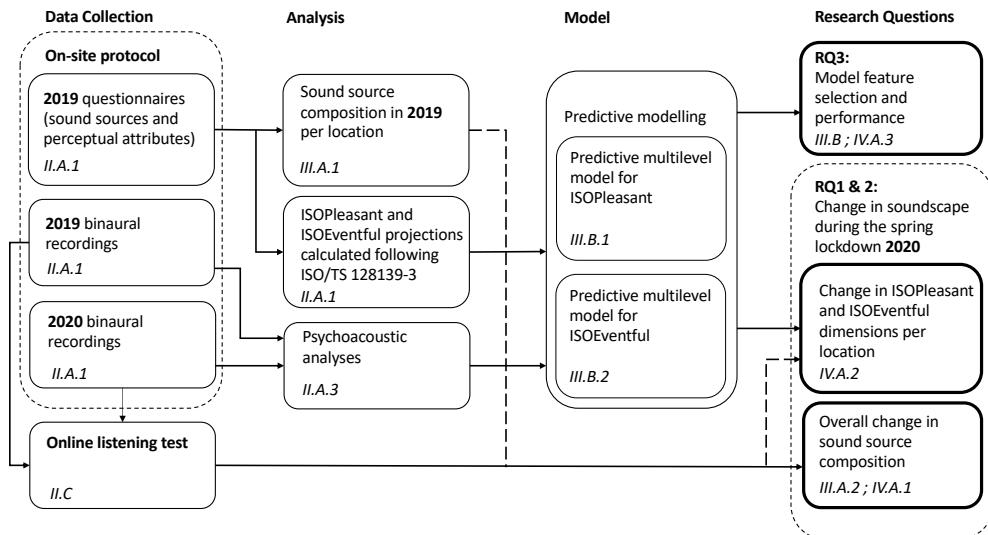


Figure 7.7: The study flowchart indicating the data collection, analysis, modelling, and discussion throughout the study. The subsections in the text to which each box refers are indicated in italics.

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.

Figure 7.2,3 illustrates and summarises the framework and sections described above.

7.3 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 (ISO Pleasant and ISO Eventful) and the lockdown-related changes.

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

7.3.1 Perceived sound source dominance

2019 sound source composition per location

Questionnaire data was collected English, Italian, and Spanish in both cities. The respective questionnaires can be found in the supplementary files and Mitchell et al. (2020). Data presented here was aggregated per LocationID.

According to the highest scored mean value of the dominant sound source type, as shown in Figure 7.3.1, 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 [draft](#) ([see Appendix A](#)). The frequency of occurrence, generated using the Word-Clouds web app, is shown in Figure 7.3.1, for the 2019 and the 2020 sets respectively. The most frequent words from both 2019 and 2020 groups are: noise, car/traffic, bird/birds, talk/voice, and (foot)steps.

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 (?). 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.

7.3.2 Model selection, performance, and application

ISOPleasant model selected

Following the feature selection, the ISOPleasant model (given in Table 7.3.2) 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 Figure 7.3.2. 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, including the LocationID factor itself. Therefore the final ISOEventful model is a 'flat' multi-variate linear 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

Table 7.4: Scaled linear regression models of ISO Pleasant and ISO Eventful for 13 locations in London and Venice. The ISO Pleasant model is a multi-level regression model with one level for individual effects and a second level for LocationID effects, while the ISO Eventful model is a 'flat' multi-variate linear regression with no location effects.

Predictors	ISO Pleasant			ISO Eventful		
	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						
Observations	13	$LocationID$				
MAE Test, Train	914			914		
	0.258	0.259		0.233		
				0.231		

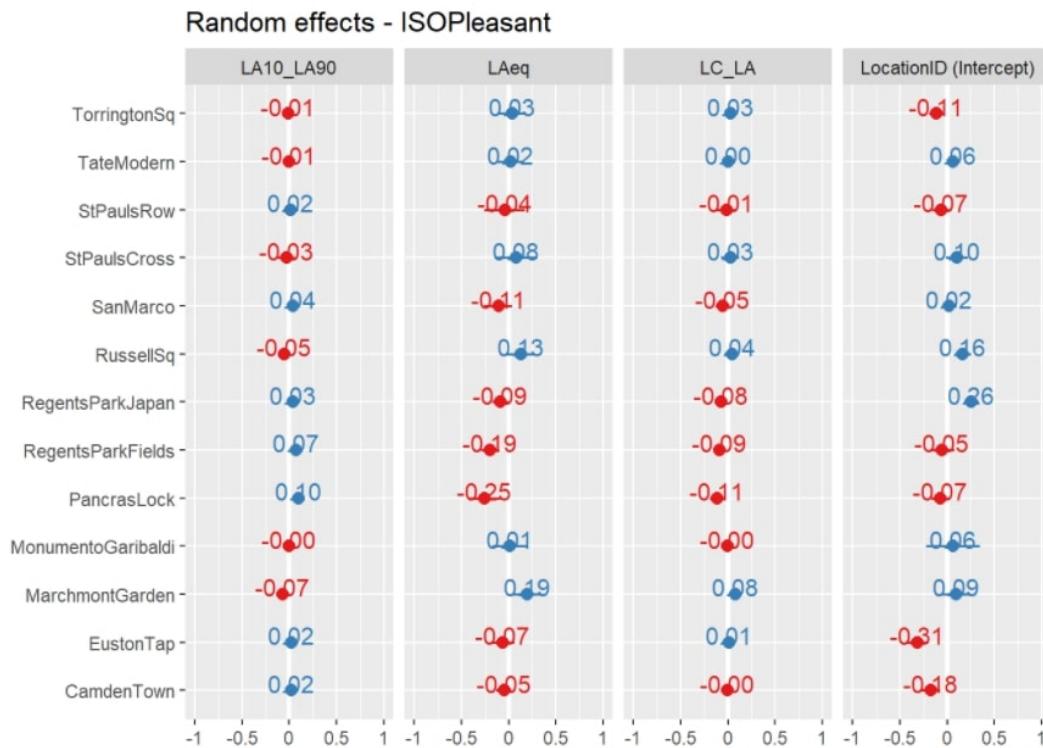
predicting the mean location responses, but still performs well ($R^2_{train} = 0.873$; $R^2_{test} = 0.715$).

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. Figure 7.3.2(a) shows the pre-lockdown ISO coordinates for each location and Figure 7.3.2(b) 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 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

Figure 7.9: Location-level scaled coefficients for the ISO Pleasant model.



are investigated further in the Discussion section below.

7.4 Discussion

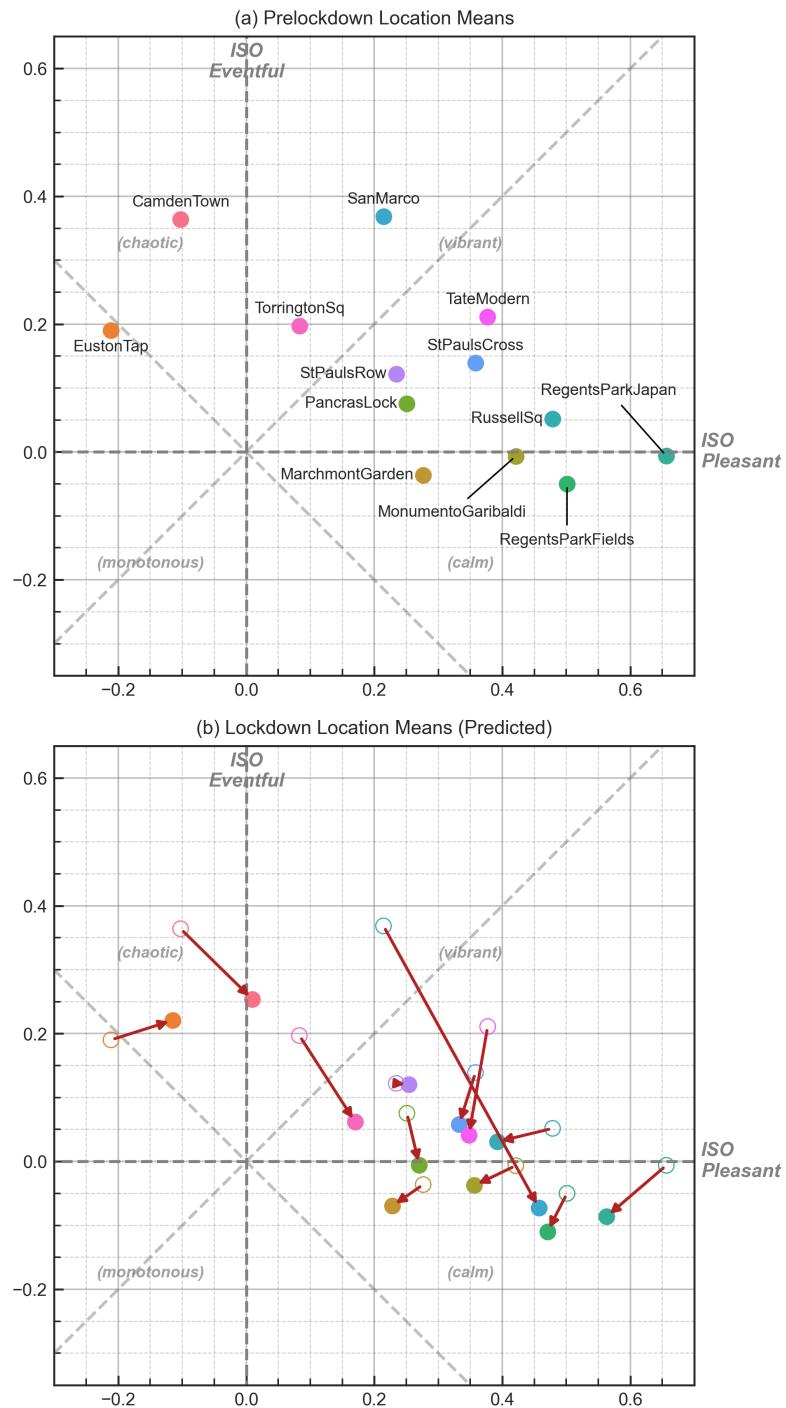
7.4.1 Interpretation of the results

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 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.

Figure 7.10: Soundscape circumplex coordinates for (a) the mean ISO Pleasant and ISO Eventful 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.



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 et al. (2020b), 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 (?). 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.

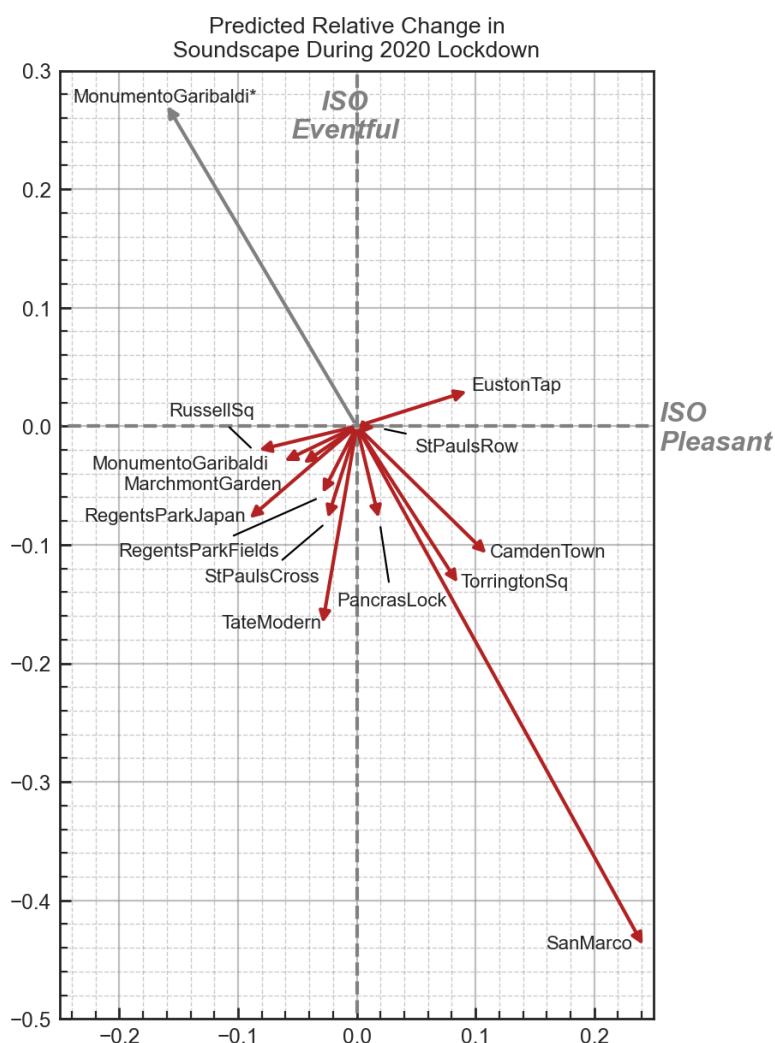
Movement of soundscapes

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 Figure 7.4.1. 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 Figure 7.3.2,
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

Figure 7.11: 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.



measures imposed in Italy, during the lockdown measurement period the square was almost entirely devoid of people. What is promising is that, without any of this contextual information about the presence or absence of people, our 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: EustonTap, 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 and 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 sug-

gest it is driven by human presence and sounds (Aletta and 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 fitted to the data from the two sessions independently and the session which was impacted by the landscaping works is shown in Figure 7.4.1 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 can 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 predicted 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 pre-

dicted 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, Yang and Kang (2005) 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 [cit](#) [add references](#). 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, our 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. The authors believe this modelling approach better addresses the practical goal of predictive soundscape modelling and reflects the structure of the data collection.

7.4.2 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.

7.5 Conclusion

This study 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 applica-

tions such as smart city sensing, urban planning, and cost-effective, sustainable design. To demonstrate the usefulness and feasibility of such an approach, we apply our 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.

8 How to Analyse and Represent Soundscape Perception

8.1 Abstract

This study first examines the methods presented in ISO 12913 for analysing and representing soundscape perception by applying them to a large existing database of soundscape assessments. The key issue identified is the inability of the standard methods to summarise the soundscape of locations and groups. The presented solution inherently considers the variety of responses within a group and provides an open-source visualisation tool to facilitate a nuanced approach to soundscape assessment and design. Several demonstrations of the soundscape distribution of urban spaces are presented, along with proposals for how this approach can be used and developed.

8.2 Introduction

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 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

8 How to Analyse and Represent Soundscape Perception

series on soundscape. Part 1 (ISO 12913-1:2014) 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.

In a recent editorial paper on Soundscape Assessment, Axelsson and colleagues observe that it is important to critically discuss current theories and models in soundscape studies and to examine their effectiveness, while also looking at how to integrate different methods and perspectives for the discipline to make further advancements (Axelsson et al., 2019). This work was mainly aimed at addressing the issue of meaningful comparability and representation of soundscape assessments. Part 2 of the ISO 12913 standard itself does not provide ultimate answers: the technical specifications recommend multiple methods, as consensus around a single protocol could not be reached. This diversity of methodological approaches should be interpreted as a fact that soundscape theory is still under development and, for this reason, the standardisation work should probably take a step back and focus on developing a reference method for comparability among soundscape studies, rather than a single protocol for soundscape data collection. Some attempts have indeed already been made in literature for the different methods proposed in the ISO/TS 12913-2:2018 (Aletta et al., 2019a; Jo et al., 2020).

This study thus aims to review the consequences of these methods for larger datasets and provide concrete examples for how soundscapes should be represented. In particular, we aim to strengthen the practices for characterising the soundscape of a location, as a collective perception by the users of the location. We also demonstrate how the progress of these tools from their initial scope (measuring and discussing the individual perception of a soundwalk participant) have not kept up with recent advances and requirements for larger-scale soundscape datasets. We question whether there are some issues related to the data collection instruments and data analysis methods as recommended, and examine the results of the model framework and mathematical transformations laid out in the ISO technical specifications in order to provide guidance on the interpretation of the soundscape circumplex.

To examine these tools and the questions raised, we apply them to an existing large

scale, real-world dataset of soundscape assessments collected according to the ISO methods. Finally, we propose a more holistic and advanced method of representing soundscapes as a probabilistic distribution of perceptions within the circumplex and provide a toolbox for others to use.

8.3 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 study we 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 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 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.

8.3.1 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

8 How to Analyse and Represent Soundscape Perception

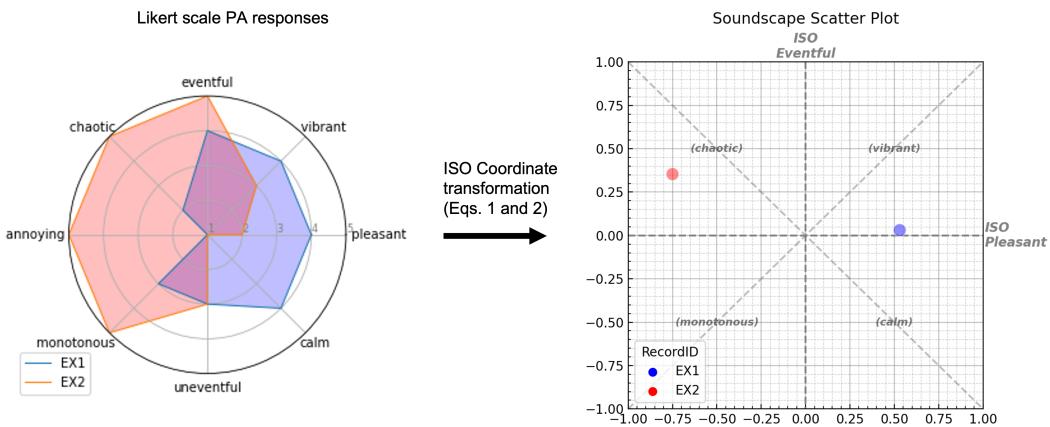


Figure 8.1: Example of representations of two soundscape assessments. Left: Radar plot of two example perceptual attribute (PA) ratings on the Likert scales (1 to 5). Right: Scatter plot of the same assessments on the soundscape circumplex, transformed according to ISO 12913 Part 3.

coordinates are thus calculated by:

$$\begin{aligned} ISO\text{Pleasant} = & [(pleasant - annoying) + \cos 45^\circ * (calm - chaotic) \\ & + \cos 45^\circ * (vibrant - monotonous)] * 1/(4 + \sqrt{32}) \end{aligned} \quad (8.1)$$

$$\begin{aligned} ISO\text{Eventful} = & [(eventful - uneventful) + \cos 45^\circ * (chaotic - calm) \\ & + \cos 45^\circ * (vibrant - monotonous)] * 1/(4 + \sqrt{32}) \end{aligned} \quad (8.2)$$

where the PAs are arranged around the circumplex as shown in Figure 8.1. The $\cos 45^\circ$ 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 Figure 8.1. 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.

8.3.2 Summarising the soundscape assessment of a location

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 Equations 8.1 and 8.2 presented above to calculate the ISOPleasant and ISOEventful values, which can then be plotted in a two-dimensional scatter plot.

8.3.3 Limitations of the ISO

How these 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. 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

8 How to Analyse and Represent Soundscape Perception

the range should be used, we therefore apply this analysis to an existing real-world soundscape dataset to determine whether it provides a useful measure of dispersion. Here we use the data contained in the International Soundscape Database (ISD) (Mitchell et al., 2021a), which includes 1,300+ individual responses collected across 13 locations in London and Venice, according to the SSID Protocol, which is based on the ISO methods explored in this paper (Mitchell et al., 2020).

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. 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).

8.4 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.

We therefore present a visualisation in Figure 8.2 of the soundscape assessments of several urban spaces included in the ISD (Mitchell et al., 2021a) 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 (Mitchell et al., 2021b)), 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 ISO Pleasant, ISO Eventful 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, 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 (i).

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:

- i. The distribution across both pleasantness and eventfulness, including the central tendency, the dispersion, and any skewness in the response;

8 How to Analyse and Represent Soundscape Perception

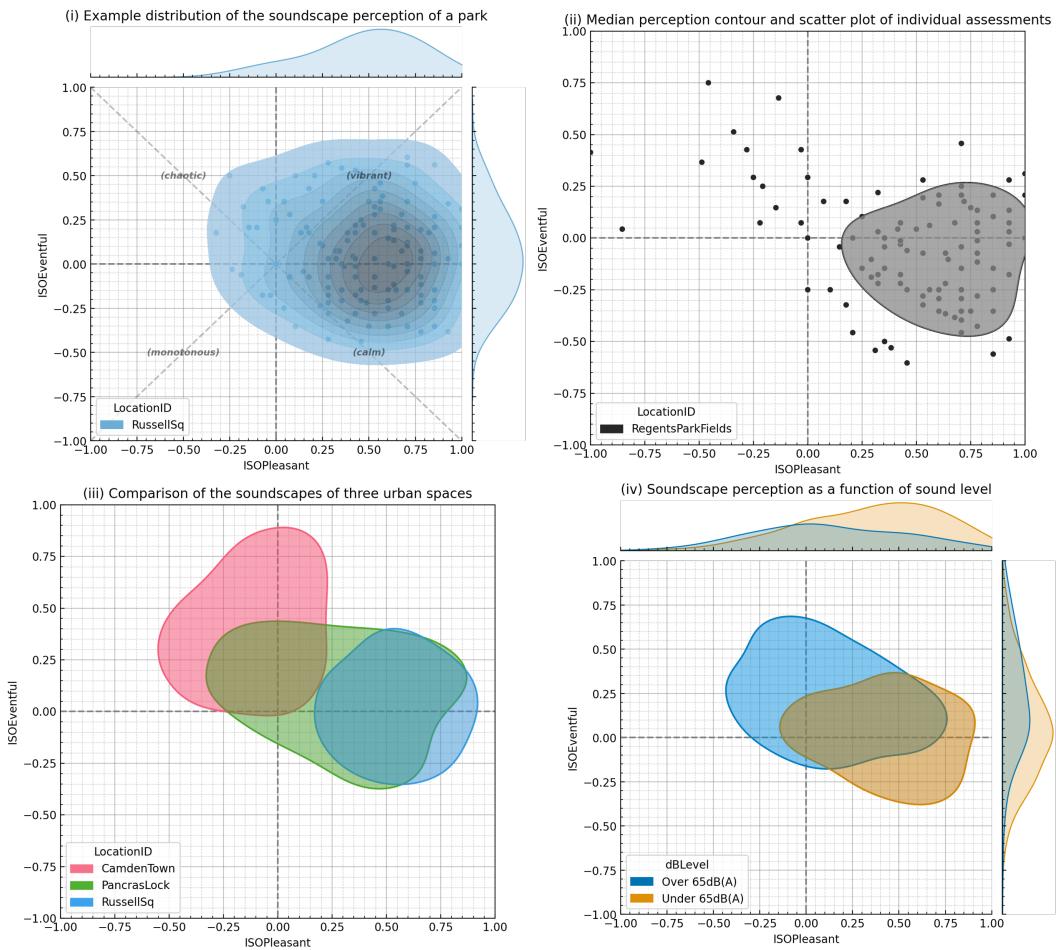


Figure 8.2: A demonstration of some use cases of representing soundscape perception as probabilistic distributions. Data is drawn from the International Soundscape Database (ISD) and is used for demonstration only. (i) Demonstrates a high-level of detail for presenting the bivariate distribution of soundscape 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 $< 65dB_{Aeq}$ and $> 65dB_A$. The introduction of other acoustic, environmental, and contextual data can reveal new and complex relationships with the soundscape perception.

2. The general shape of the soundscape within the space - in this case Russell Sq 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 Sq, as demonstrated by the compactness of the distribution, but this is not the case for every location.

Fig (i) 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 our 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 (ii) where the shaded portion contains 50% of the responses. This simplified view of the distribution presents several advantages, as is demonstrated in Figs. (iii and iv) and takes inspiration from the recommendation in the ISO standard to use the median as a summary statistic. In our 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 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 (ii) 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 nearly 20 minutes.

Fig (iii) 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

8 How to Analyse and Represent Soundscape Perception

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 (iv) 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 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 (iv) 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 65 dB L_{Aeq} and a set of more than 65 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 et al., 2021b). 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 library of plotting functions for producing the type of plots presented here has been made available as part of the ISD at <https://zenodo.org/record/5705908>. In addition, an interactive code notebook for this paper is included which provides a tutorial for using

this code, working with the ISD data, and recreating the figures of this paper. The figures were created using the seaborn (Waskom, 2021) and Matplotlib (Hunter, 2007) plotting libraries in Python v3.9.2 (Van Rossum and Drake, 2009).

8.5 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.

Although the visualisations shown in Figure 8.2 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), we 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 to summarise locations down to a single point to compare many different loca-

tions and to easily investigate how the soundscape assessment has generally changed over time. For this purpose, the mean of the ISO Pleasant and ISO Eventful 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 Figure 8.2. 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 Figure 5 of Mitchell et al. (2021b). The key to all of these analysis methods, whether they be the distributional plots shown in Figure 8.2, 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.

8.5.1 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 socio-logical 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 et al., 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 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 ISO Pleasant and ISO Eventful outcomes.

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 et al., 2020a) as long as the dimensional relationships of the circumplex are maintained.

8.6 Conclusions

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, we 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

8.7 Interpreting the Soundscape Circumplex

The circumplex model of soundscape, as originally defined by 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 (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 et al., 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. Studies in the field show that data collection campaigns rarely return extreme values for soundscape dimensions (Mancini et al., 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 (Sun et al., 2019).

8.7.1 Application & Simulations

In order to investigate the shape of the circumplex coordinate space generated by this transformation, a dataset of 3 million randomly simulated PA responses was generated. For each of the 8 PAs, an integer value from 1 to 5 is randomly generated from a uniform distribution, meaning each of the five responses is equally likely. These simulated data are specifically not intended to include any information about correlations between the various PAs when actually answered by respondents (see (Lionello et al., 2021) for more on this discussion), instead the PA responses are completely uncorrelated as they each have their own random distribution. Therefore, the simulated dataset represents a theoretical uniform coverage of the 8 dimensional PA space.

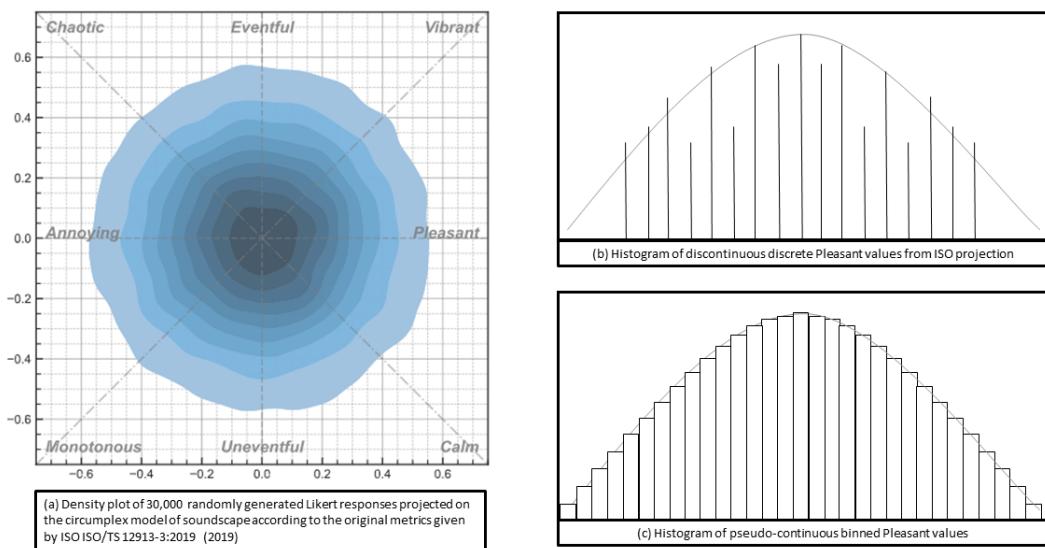


Figure 8.3: Mockup placeholder simulation heatmap plot from Lionello 2020 and discrete vs binned transformation values. For new one, need to include distribution histograms along each axis.

We then apply the ISO transformations given in Equations 8.1 and 8.2, resulting in 3 million coordinate pairs with a range of (-1, 1) in the x and y axes. A heatmap of the resulting two-dimensional circumplex space is shown in Figure 8.3, along with histograms of the individual dimension distributions. These distributions then represent the theoretical available circumplex space generated by the ISO transformation on uniform survey responses.

Two important observations can be made about the shape of the resulting two-dimensional distribution. The first is that the shape of the available space is a circle. It should be noted that, despite what the term 'circumplex' may indicate, the perceptual dimensions are not necessarily intended to circumscribe a circle. The second is that, in each dimension, the responses are normally distributed, centered around zero. These points will be discussed in detail below.

8.7.2 Circular space discussion

Visualisations of the circumplex model in soundscape literature tend to present it as circumscribing a circle (see Figure 1 in (Axelsson et al., 2010) and Figure 3 in (Torresin et al., 2020)), and this shape is further emphasised by the initial figure in Russell (1980)'s original formulation of the concept. However, it should be emphatically noted that all of these presentations are in fact artefacts of the analysis methods which generated them, not some sort of revealed pattern in the component attributes which make up the cir-

8 How to Analyse and Represent Soundscape Perception

complex. In Russell (1980), this first figure is generated by asking respondents to place each of the 27 attributes around a circle, according to their perceived spatial relationships - the circle shape was pre-imposed on the study. In both Axelsson et al. (2010) and Törresin et al. (2020), the figures are generated via Principle Components Analysis (PCA) which, again, presents these results superimposed on a circle. It is perhaps a weakness of these two, otherwise strong and impactful, papers, that they did not recognise this consequence and challenge the circular arrangement.

If we turn back to Russell's original work on the circumplex model of affect, we can see some indications that a circle does not, in fact, describe the spatial relationship of the perceptual attributes. Fig. 4 of (Russell, 1980), which did not pre-impose the circular arrangement in its analysis, instead most closely resembles a square with rounded corners. Continuing from this conception, when Russell presents a graphical method of assessing the two dimensions of affect (pleasure and arousal) (Russell et al., 1989), they use a square grid. This is all to say that, although the term 'circumplex' and the foundational analyses which lead to a soundscape circumplex may lead us to assume it must take the form of a circle, both the framework laid down by Russell and the common treatment of the spatial relationships of the attributes actually describe a square, instead.

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.

From this, we would infer that a maximally vibrant soundscape is both maximally pleasant and maximally eventful. However, when the projection transformation is applied it imposes certain limitations on the relationships between the dimensions which do not conform with this assumption. As shown in Figure 8.1, when a soundscape is maximally vibrant (i.e. a diagonal vector distance of 1), the maximum pleasantness value it can have is determined by the $\cos 45^\circ$ term, giving a max pleasantness value of ~ 0.7071 . The implication of this is that no soundscape can be both maximally pleasant and maximally eventful at the same time, meaning that these dimensions are not in fact considered as orthogonal, and that a highly vibrant soundscape cannot be considered highly pleasant or highly eventful. Similarly, if a soundscape were to begin at a maximum Eventfulness, with neutral Pleasantness, in order for the soundscape to become more pleasant, it must

by definition become less eventful. This is not conceptually correct or borne out in the treatments of previous literature. These same relationships and violations hold true for the other diagonal dimensions, chaotic, calm, and monotonous.

This implication violates both the assumptions made within the formulation of the circumplex model and the way that soundscape practitioners have understood and presented the interpretations of soundscapes within the circumplex space. In cases where the PA dimensions are referred to directly (?Steele et al., 2019) and those which have made use of the Part 3 transformation to 2-dimensional coordinates (Mancini et al., 2021; Lionello et al., 2021; Manzano et al., 2021), *Check Manzano2021importance* the conflation of maximal values on the diagonal axes with maximal values on the primary axes is made, as in the assumptions made by the standard. This is the first of the common understandings of the circumplex which are violated by the trigonometric transformation.

8.7.3 Normal distribution discussion

We can also see from the histograms included along the axes of Figure 8.3 that the projection creates a normal distribution in both dimensions. It is important here to remember that the input to the projection formulas were uniform distributions for each of the 8 PAs, and it is the projection into the two primary dimensions which results in this normal distribution. From the simulated distributions, we can derive a normal probability density distribution (PDF) for each of the dimensions.

$$f_X(x) = \frac{x^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$$

with a mean $\mu = 0$ and standard deviation $\sigma = 0.3$.

Realistic max values When looking at the distribution heatmap in Figure 8.3, it is useful to picture the gradients as representing the available space in the circumplex model. The probability of reaching a given result decreases as we move farther away from the origin. This means, for example, that the probability of getting a pleasantness value between 0.2 and 0.3 is nearly 4 times the probability of getting a value between 0.5 and 0.6. This may not seem important, but the consequence is that, as a result of strictly the projection calculation, neutral values within the soundscape circumplex are much more likely and the space available to compare soundscapes within is truncated.

When we start to think about real-world urban soundscape data collection, where the discussion of the soundscape of a space is not limited to a single person's perception, we need to start thinking in statistical terms. Theoretically, the limits of the projected Pleas-

antness are (-1, +1), however according to the PDF calculated above less than 10% of values fall outside the range (-0.5, +0.5). It may be argued that as long as +1 can theoretically be reached, this should be what is considered the maximum value for that dimension. However, in any situation which involves using multiple individual soundscape assessments in order to characterize the overall soundscape of a location, this max will effectively never be reached. According to the large-scale, multi-location data set reported in our previous study, it appears that the effective maximum values for Pleasantness and Eventfulness for the combined assessment of multiple people for a space is in reality approximately (-0.6, +0.6) (Lionello et al., 2021).

As such, extreme values on each of the perceptual dimensions are less likely to occur than are coordinate values which place the soundscape in the neutral areas of the circumplex space. This means an extremely calm (or chaotic, or vibrant, or pleasant) coordinate is significantly less likely to occur than a neutral coordinate.

8.7.4 Non-continuous projected values

An implicit assumption of the transformation is that the resulting coordinates are now continuous values, which allows linear regression and correlation methods to be used. Indeed, the transformation of the 8-dimensional ordinal Likert scale data to the two-dimensional coordinates creates a higher resolution of intervals, which would appear to be pseudo-continuous. Upon further investigation, the transformation actually results in XX discrete possible values. Figure 8.3(b) shows a histogram of this raw output from the transformation, demonstrating that these discrete values, while following the general normal distribution discussed above, are not evenly filled - some adjacent values may be much more or less likely than their neighbors. This poses potential issues for further analysis which assumes either continuous or equally-spaced discrete values.

9 Conclusions

9.1 General Discussion

9.2 Contribution to Knowledge

9.3 Implications

9.4 Limitations

9.5 Recommendations for Future Research

9.6 Concluding Remarks

Glossary

AIC Akaike Information Criterion. 83

AMB Ambisonic recording. xiii, xvi, 19, 22, 24, 26

ANE Anomalous Noise Event. 57

ANOVA ANalysis Of VAriance. 62

ANS Auditory Nervous System. 40

ART Attention Restoration Theory. 41

BIN Binaural. xiii, xvi, 19, 24, 29, 30

CNS Central Nervous System. 41

COMPLX complex. 58

COVID-19 Coronavirus disease of 2019. xv, 68–70, 78, 92

END (Environmental Noise Directive) *DIRECTIVE 2002/49/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 25 June 2002 relating to the assessment and management of environmental noise.* Policy directive within the EU setting out priorities and requirements of member nations for ensuring health and environmental protection as it relates to noise. Incorporates requirements for agglomerations to produce noise maps and identify and preserve quiet areas.. 9, 10

ENV Environmental factors. xiii, 24, 28

environmental unit An area within a public space in which environmental factors are consistent and which is typically perceived to constitute a single distinct area.. 17, 19, 22

FS Fluctuation Strength. 81, 87

Glossary

I Impulsiveness. 61, 81

ISD International Soundscape Database. 32, 70

ISO International Organization for Standardization. 36, 38, 98

ISOEventful The value along the primary eventfulness dimension of the soundscape circumplex, calculated via a trigonometric projection of the other PAQs, as defined in ?. xiv, xvi, 80–83, 85, 87, 88, 90, 94, 95

ISOPleasant The value along the primary pleasantness dimension of the soundscape circumplex, calculated via a trigonometric projection of the other PAQs, as defined in ?. xiv, xvi, 42, 80–83, 85, 87–90, 95

L_{A10} . 70, 72

$L_{A10} - L_{A90}$. 81, 87

L_{A90} . 70, 72, 84

L_{Aeq} . 68, 70, 72, 75, 80–82, 87, 94

$L_{Ceq} - L_{Aeq}$. 80, 81, 87

MAE Mean Absolute Error. 83, 87

MLM Multi-Level Model. 38, 60, 61

MUSHRA MUlti Stimulus test with Hidden Reference and Anchor. xiii, 55, 58–60, 65

N_5 Psychoacoustic Loudness. 60, 61, 70, 72, 81, 82, 87

PA Zwicker Psychoacoustic Annoyance. 81, 82

PAQ Perceived Affective Quality. 11, 20, 79, 80, 118

PIC Site pictures. xiii, 24, 29

PNS Peripheral Nervous System. 40, 41

PSNS Parasympathetic Nervous System. 40

QUE Questionnaires. xiii, xvi, 19, 24

Glossary

R Psychoacoustic Roughness. 61, 81

RA Relative Approach. 81

RTN Road Traffic Noise. 57

S Psychoacoustic Sharpness. 61, 70, 72, 80, 81, 87

SLM Sound Level Meter. xiii, xvi, 16, 19, 23–28, 35

SNS Sympathetic Nervous System. 40

SSID Soundscape Indices. xvi, 1, 13, 15, 19, 32, 78, 82

SSQP Swedish Soundscape Quality Protocol. 20

T Tonality. 61, 80, 81, 87

VID 360°Video. xiii, 24

VIF Variance Inflation Factor. 61, 83

VR Virtual Reality. 22

WASN Wireless Acoustic Sensor Network. 57, 63–66

WHO-5 WHO Well-being Index. 20, 21

WSP World Soundscape Project. 36

Bibliography

- Aiello, L. M., Schifanella, R., Quercia, D., and Aletta, F. (2016). Chatty maps: constructing sound maps of urban areas from social media data. *Royal Society Open Science*, 3(3):150690.
- Aletta, F., Astolfi, A., Di Blasio, S., Shtrepi, L., Kang, J., and Molinero, L. (2019a). Exploring associations between soundscape assessment, perceived safety and well-being: a pilot field study in Granary Square, London. In *23rd International Congress on Acoustics*, Aachen.
- Aletta, F., Axelsson, Ö., and Kang, J. (2014). Towards acoustic indicators for soundscape design. In *Forum acusticum*, number c, page 6.
- Aletta, F., Axelsson, Ö., and Kang, J. (2017). Dimensions underlying the perceived similarity of acoustic environments. *Frontiers in Psychology*, 8:1–11.
- Aletta, F. and Kang, J. (2018). Towards an urban vibrancy model: A soundscape approach. *International Journal of Environmental Research and Public Health*, 15(8):1712.
- Aletta, F., Kang, J., and Axelsson, Ö. (2016). Soundscape descriptors and a conceptual framework for developing predictive soundscape models. *Landscape and Urban Planning*, 149:65–74.
- Aletta, F., Oberman, T., Axelsson, Ö., Xie, H., Zhang, Y., Lau, S.-K., Tang, S.-K., Jambrošić, K., De Coensel, B., van den Bosch, K., et al. (2020a). Soundscape assessment: Towards a validated translation of perceptual attributes in different languages. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 261, pages 3137–3146. Institute of Noise Control Engineering.
- Aletta, F., Oberman, T., Mitchell, A., Erfanian, M., Lionello, M., Kachlicka, M., and Kang, J. (2019b). Associations between soundscape experience and self-reported well-being in open public urban spaces: A field study. *The Lancet*, 394:S17.

Bibliography

- Aletta, F., Oberman, T., Mitchell, A., Tong, H., and Kang, J. (2020b). Assessing the changing urban sound environment during the COVID-19 lockdown period using short-term acoustic measurements. *Noise Mapping*, 7(1):123–134.
- Alsina-Pagès, R. M., Freixes, M., Orga, F., Foraster, M., and Labairu-Trenchs, A. (2021). Perceptual evaluation of the citizen's acoustic environment from classic noise monitoring. *Cities & Health*, 5(1–2):145–149.
- Asensio, C., Aumond, P., Can, A., Gascó, L., Lercher, P., Wunderli, J.-M., Lavandier, C., de Arcas, G., Ribeiro, C., Muñoz, P., and Licitra, G. (2020a). A Taxonomy Proposal for the Assessment of the Changes in Soundscape Resulting from the COVID-19 Lockdown. *International Journal of Environmental Research and Public Health*, 17(12):4205.
- Asensio, C., Pavón, I., and de Arcas, G. (2020b). Changes in noise levels in the city of Madrid during COVID-19 lockdown in 2020. *The Journal of the Acoustical Society of America*, 148(3):1748–1755.
- Aumond, P., Can, A., Coensel, B. D., Botteldooren, D., Ribeiro, C., and Lavandier, C. (2017). Modeling soundscape pleasantness using perceptual assessments and acoustic measurements along paths in urban context. *Acta Acustica united with Acustica*, 103(3):430–443.
- Axelsson, Ö., Guastavino, C., and Payne, S. R. (2019). Editorial: Soundscape assessment. *Frontiers in Psychology*, 10:2514.
- Axelsson, Ö., Nilsson, M. E., and Berglund, B. (2010). A principal components model of soundscape perception. *The Journal of the Acoustical Society of America*, 128(5):2836–2846.
- Berglund, B. and Nilsson, M. E. (2006). On a Tool for Measuring Soundscape Quality in Urban Residential Areas. *Acta Acustica united with Acustica*, 92(6):938–944.
- Bild, E., Pfeffer, K., Coler, M., Rubin, O., and Bertolini, L. (2018). Public Space Users' Soundscape Evaluations in Relation to Their Activities. An Amsterdam-Based Study. *Frontiers in Psychology*, 9:1593.
- Botteldooren, D., De Coensel, B., and De Muer, T. (2006). The temporal structure of urban soundscapes. *Journal of Sound and Vibration*, 292(1–2):105–123.

Bibliography

- Craig, A., Moore, D., and Knox, D. (2017). Experience sampling: Assessing urban soundscapes using in-situ participatory methods. *Applied Acoustics*, 117:227–235.
- de Coensel, B., Botteldooren, D., and de Muer, T. (2003). 1/f Noise in Rural and Urban Soundscapes. *Acta Acustica united with Acustica*, 89:287–295.
- Droumeva, M. (2021). The sound of the future: listening as data and the politics of soundscape assessment. *Sound Studies*, 0(0):1–17.
- EEA (2020). Environmental noise in Europe, 2020. (European Environment Agency, Copenhagen, Denmark, 2020).
- Engel, M. S., Fiebig, A., Pfaffenbach, C., and Fels, J. (2018). A Review of Socio-acoustic Surveys for Soundscape Studies. *Current Pollution Reports*, 4(3):220–239.
- Erfanian, M., Mitchell, A., Aletta, F., and Kang, J. (2021). Psychological well-being and demographic factors can mediate soundscape pleasantness and eventfulness: A large sample study. *Journal of Environmental Psychology*, 77:101660.
- Gasco, L., Schifanella, R., Aiello, L. M., Quercia, D., Asensio, C., and de Arcas, G. (2020). Social Media and Open Data to Quantify the Effects of Noise on Health. *Frontiers in Sustainable Cities*, 2:41.
- Gelman, A. (2006). Multilevel (hierarchical) modeling: What It can and cannot do. *Technometrics*, 48(3):432–435.
- Hadjidemetriou, G. M., Sasidharan, M., Kouyialis, G., and Parlidak, A. K. (2020). The impact of government measures and human mobility trend on COVID-19 related deaths in the UK. *Transportation Research Interdisciplinary Perspectives*, 6:100167.
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95.
- IEC 61672-1:2013 (2013). Electroacoustics – Sound level meters – Part 1: Specifications. (International Electrotechnical Commission, Geneva, Switzerland, 2013).
- ISO 12913-1:2014 (2014). Acoustics – Soundscape – Part 1: Definition and conceptual framework. International Organization for Standardization, Geneva, Switzerland, 2014.

Bibliography

- ISO 1996-1:2016 (2016). Acoustics – Description, measurement and assessment of environmental noise – Part 1: Basic quantities and assessment procedures. (International Organization for Standardization, Geneva, Switzerland, 2016).
- ISO 532-1:2017 (2017). Acoustics – Methods for calculating loudness – Part 1: Zwicker method. (International Organization for Standardization, Geneva, Switzerland, 2017).
- ISO/TS 12913-2:2018 (2018). Acoustics – Soundscape – Part 2: Data collection and reporting requirements. International Organization for Standardization, Geneva, Switzerland, 2018.
- ISO/TS 12913-3:2019 (2019). Acoustics – Soundscape – Part 3: Data analysis. International Organization for Standardization, Geneva, Switzerland, 2019.
- ISO/TS 15666:2021 (2021). Acoustics – assessment of noise annoyance by means of social and socio-acoustic surveys. International Organization for Standardization, Geneva, Switzerland, 2021.
- Jeon, J. Y., Hong, J. Y., Lavandier, C., Lafon, J., Axelsson, Ö., and Hurtig, M. (2018). A cross-national comparison in assessment of urban park soundscapes in France, Korea, and Sweden through laboratory experiments. *Applied Acoustics*.
- Jeon, J. Y., Lee, P. J., Hong, J. Y., and Cabrera, D. (2011). Non-auditory factors affecting urban soundscape evaluation. *The Journal of the Acoustical Society of America*, 130(6):3761–3770.
- Jo, H. I., Seo, R., and Jeon, J. Y. (2020). Soundscape assessment methods: Compatibility of questionnaires and narrative interview based on ISO 12913-2. In *Inter-Noise 2020*, Seoul.
- Kang, J., Aletta, F., Gjestland, T. T., Brown, L. A., Botteldooren, D., Schulte-Fortkamp, B., Lercher, P., van Kamp, I., Genuit, K., Fiebig, A., Bento Coelho, J. L., Maffei, L., and Lavia, L. (2016). Ten questions on the soundscapes of the built environment. *Building and Environment*, 108:284–294.
- Kang, J., Aletta, F., Oberman, T., Erfanian, M., Kachlicka, M., Lionello, M., and Mitchell, A. (2019). Towards soundscape indices. In *Proceedings of the 23rd International Congress on Acoustics*, volume integrating 4th EAA Euroregio 2019 : 9-13 September 2019, pages 2488–2495, Aachen. RWTH Aachen University.

Bibliography

- Kogan, P., Arenas, J. P., Bermejo, F., Hinalaf, M., and Turra, B. (2018). A Green Soundscape Index (GSI): The potential of assessing the perceived balance between natural sound and traffic noise. *Science of The Total Environment*, 642:463–472.
- Kruize, H., Van Kamp, I., Van Den Berg, M., Van Kempen, E., Wendel-vos, W., Ruijsbroek, A., Swart, W., Maas, J., Gidlow, C., Smith, G., Ellis, N., Hurst, G., Masterson, D., Triguero-mas, M., Cirach, M., Gražulevičienė, R., Van Den Hazel, P., and Nieuwenhuijsen, M. (2019). Exploring mechanisms underlying the relationship between the natural outdoor environment and health and well-being – Results from the PHENOTYPE project. *Environment International*, (October 2018):105173.
- Lacey, J., Pink, S., Harvey, L., and Moore, S. (2019). Noise transformation: A critical listening-based methodology for the design of motorway soundscapes. *Qualitative Research Journal*, 19(1):49–64.
- Lavia, L., Witchel, H. J., Aletta, F., Steffens, J., Fiebig, A., Kang, J., Howes, C., and Healey, P. G. T. (2018). Non-participant observation methods for soundscape design and urban panning. In Aletta, F. and Xiao, J., editors, *Handbook of Research on Perception-Driven Approaches to Urban Assessment and Design*, number April.
- Lee, P. J. and Jeong, J. H. (2021). Attitudes towards outdoor and neighbour noise during the COVID-19 lockdown: A case study in London. *Sustainable Cities and Society*, 67:102768.
- Lionello, M., Aletta, F., and Kang, J. (2020). A systematic review of prediction models for the experience of urban soundscapes. *Applied Acoustics*, 170.
- Lionello, M., Aletta, F., Mitchell, A., and Kang, J. (2021). Introducing a Method for Intervals Correction on Multiple Likert Scales: A Case Study on an Urban Soundscape Data Collection Instrument. *Frontiers in Psychology*, 11:3943.
- Mancini, S., Mascolo, A., Graziuso, G., and Guarnaccia, C. (2021). Soundwalk, Questionnaires and Noise Measurements in a University Campus: A Soundscape Study. *Sustainability*, 13(2).
- Manzano, J. V., Pastor, J. A. A., and Quesada, R. G. (2021). The importance of changing urban scenery in the assessment of citizens' soundscape perception. on the need for different time-related points of view. *Noise Mapping*, 8(1):138–161.

Bibliography

- Mitchell, A., Oberman, T., Aletta, F., Erfanian, M., Kachlicka, M., Lionello, M., and Kang, J. (2020). The Soundscape Indices (SSID) Protocol: A Method for Urban Soundscape Surveys—Questionnaires with Acoustical and Contextual Information. *Applied Sciences*, 10(7):2397.
- Mitchell, A., Oberman, T., Aletta, F., Erfanian, M., Kachlicka, M., Lionello, M., and Kang, J. (2021a). The International Soundscape Database: An integrated multimedia database of urban soundscape surveys – questionnaires with acoustical and contextual information.
- Mitchell, A., Oberman, T., Aletta, F., Kachlicka, M., Lionello, M., Erfanian, M., and Kang, J. (2021b). Investigating urban soundscapes of the COVID-19 lockdown: A predictive soundscape modeling approach. *The Journal of the Acoustical Society of America*, 150(6):4474–4488.
- Oberman, T., Šćitaroci, B. B. O., and Jambrošić, K. (2018). Towards a virtual soundwalk. In *Handbook of Research on Perception-Driven Approaches to Urban Assessment and Design*, pages 317–343. IGI Global.
- Orga, F., Mitchell, A., Freixes, M., Aletta, F., Alsina-Pagès, R. M., and Foraster, M. (2021). Multilevel Annoyance Modelling of Short Environmental Sound Recordings. *Sustainability*, 13(11):5779.
- Radicchi, A., Cevikayak Yelmi, P., Chung, A., Jordan, P., Stewart, S., Tsaligopoulos, A., McCunn, L., and Grant, M. (2021). Sound and the healthy city. *Cities & Health*, 5(1-2):1–13.
- Ricciardi, P., Delaitre, P., Lavandier, C., Torchia, F., and Aumond, P. (2015). Sound quality indicators for urban places in Paris cross-validated by Milan data. *The Journal of the Acoustical Society of America*, 138:2337–2348.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.
- Russell, J. A., Weiss, A., and Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, 57(3).
- Rychtáriková, M. and Vermeir, G. (2013). Soundscape categorization on the basis of objective acoustical parameters. *Applied Acoustics*, 74(2):240–247.

Bibliography

- Sakagami, K. (2020). How did the 'state of emergency' declaration in Japan due to the COVID-19 pandemic affect the acoustic environment in a rather quiet residential area? *UCL Open Environment*, (1):1–9.
- Schafer, R. M. (1977). *The Soundscape: Our sonic environment and the tuning of the world*.
- Sottek, R. and Genuit, K. (2005). Models of signal processing in human hearing. 59(3):157–165.
- Southworth, M. (1969). The sonic environment of cities. *Environment and Behavior*.
- Steele, D., Bild, E., Tarlao, C., and Guastavino, C. (2019). Soundtracking the Public Space: Outcomes of the Musikiosk Soundscape Intervention. *International Journal of Environmental Research and Public Health*, 16(10):1865.
- Stuart Russell, P. N. (2021). *Artificial Intelligence: A Modern Approach, eBook, Global Edition*. Pearson ITP.
- Sun, K., De Coensel, B., Filipan, K., Aletta, F., Van Renterghem, T., De Pessemier, T., Joseph, W., and Botteldooren, D. (2019). Classification of soundscapes of urban public open spaces. *Landscape and Urban Planning*.
- Tarlao, C., Steffens, J., and Guastavino, C. (2020). Investigating contextual influences on urban soundscape evaluations with structural equation modeling. *Building and Environment*, 188.
- Tong, H., Aletta, F., Mitchell, A., Oberman, T., and Kang, J. (2021). Increases in noise complaints during the COVID-19 lockdown in Spring 2020: A case study in Greater London, UK. *Science of The Total Environment*, 785:147213.
- Torresin, S., Albatici, R., Aletta, F., Babich, F., Oberman, T., Siboni, S., and Kang, J. (2020). Indoor soundscape assessment: A principal components model of acoustic perception in residential buildings. *Building and Environment*, 182:107152.
- Truax, B. (1999). *Handbook for Acoustic Ecology*. Cambridge Street Publishing, Cambridge, MA.
- Van Rossum, G. and Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.

Bibliography

- Verma, D., Jana, A., and Ramamritham, K. (2020). Predicting human perception of the urban environment in a spatiotemporal urban setting using locally acquired street view images and audio clips. *Building and Environment*, 186.
- Vida Manzano, J., Almagro Pastor, J. A., García Quesada, R., Aletta, F., Oberman, T., Mitchell, A., and Kang, J. (2021). The "sound of silence" in Granada during the COVID-19 lockdown. *Noise Mapping*, 8(1):16–31.
- Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021.
- Westerkamp, H. (2002). Linking soundscape composition and acoustic ecology. *Organised Sound*, 7(1):51–56.
- Yang, D., Cao, X., and Meng, Q. (2022). Effects of a human sound-based index on the soundscapes of urban open spaces. *Science of The Total Environment*, 802:149869.
- Yang, M., De Coensel, B., and Kang, J. (2015). Presence of 1/f noise in the temporal structure of psychoacoustic parameters of natural and urban sounds. *The Journal of the Acoustical Society of America*, 138(2):916–927.
- Yang, W. and Kang, J. (2005). Acoustic comfort evaluation in urban open public spaces. *Applied Acoustics*, 66(2):211–229.
- Zhang, X., Ba, M., Kang, J., and Meng, Q. (2018). Effect of soundscape dimensions on acoustic comfort in urban open public spaces. *Applied Acoustics*, 133:73–81.
- Zwicker, E. and Fastl, H. (2007). *Psychoacoustics: facts and models*. Springer, Berlin ; New York, third ed. edition.

Appendix A: SSID Questionnaire

Appendix B: $1/f$ Analysis of complex sound environments

draft *Massive edits to make*

1 Abstract

It has been proposed that the temporal structure of acoustic environments exhibit a $1/f^\alpha$ behaviour, like many other time series in nature. To investigate the relative randomness within urban soundscapes, the slope coefficient (α) of the temporal structure of a series of psychoacoustic metrics is calculated for recordings made in 215 locations across Europe. The first stage of this analysis involves identifying at which frequency regimes this α_f metric exhibits different behaviours for the various psychoacoustic metrics. Finally, the relationships between this temporal structure and the assessed soundscape pleasantness and eventfulness are examined and the most significant features are identified.

2 Introduction

General justification for soundscape (no more than 3 sentences?). The impact of noise pollution on general health and well-being has been recognised for decades and various strategies have been developed to incorporate contributing effects on perception when considering the impact of the noise environment and have developed and improved tools for assessing the perceived quality of a sound environment, known as the soundscape.

The soundscape approach to assessing and addressing urban noise pollution aims to take a more holistic view of the sound environment than more traditional methods. This is typically done through in situ questionnaires, soundwalks, and interviews, as laid out recently in Section 5 of ISO/TS 12913-2:2018 (2018).

2.1 Perceptual Attributes

The model of soundscape characterization described by Axelsson et al. Axelsson et al. (2010) and recently standardised by ISO 12913-2 (ISO/TS 12913-2:2018, 2018) comprises 8 perceptual attributes. These perceptual attributes form a 2-dimensional circumplex axis in which Pleasant/Unpleasant is orthogonal to Eventful/Uneventful, which is subsequently 45° offset from the orthogonal axis containing Exciting/Monotonous and Chaotic/Calm. Any measured soundscape can thus be plotted on the circumplex axis according to its rated perceptual attributes.

2.2 Importance of temporal structure on the sound environment

Previous work on identifying the key acoustic and psychoacoustic parameters has so far yielded conclusive results only in the realm of soundscape identification Rychtáriková and Vermeir (2013) and has indicated that traditional acoustic parameters and analysis are insufficient metrics for subjective assessment of a soundscape's overall pleasantness Aletta et al. (2014). However, progress has been made in determining the relationships between (psycho)acoustic parameters and more targeted perceptual attributes such as vibrancy and eventfulness Aletta and Kang (2018); Jeon et al. (2011).

Over the years, attempts to improve or reform traditional methods of noise assessment have produced a multitude of available noise indicators. A vital subset of these have provided information regarding the time series of the acoustic environment, typically serving one or both of two primary purposes: (1) to generalise sound levels or the response to sound levels over an extended period of time, and (2) to indicate the amount of variability of the sound level over the time period.

Those which fall into only the first category includes 24 hour metrics such as Community Noise Equivalent Level (CNEL) and the Day-Night or Day-Evening-Night Levels (L_{dn} and L_{den} , respectively); also included are levels meant for work-safety assessments, such as Sound Exposure Level (SEL). ***[draft] Note: Psychoacoustic metrics were developed for industrial assessments of single source, product noise. This limits their applicability to multi-source complex sound environments.*** ***[draft] Discuss current methods of assessing urban noise, i.e. Leq, Ldn, SEL, Ln, and how they handle temporal structure.*** ***Discussion of statistical methods e.g. Venaklasen method for CALGreen?*** ***[draft] Discussion of perceptual attributes and their collection and uses. Relate perceptual attributes to acoustical metrics above. Any discussion of temporality in perceptual attributes?***

Introduce and review items of previous research in the area (obligatory).

[draft] *Music and Self-organised criticality* Jeon et al. (2011) investigated the idea of an overall temporal structure in music,

Key points:

- Found a $1/f$ structure in the spectral density of fluctuations of the audio power of many musical selections and English speech.
- $1/f$ structure of audio power holds down to $5 * 10^{-4}$ Hz (2000s period).
- $1/f$ structure of frequency fluctuations of music holds down to the inverse to the length of the sample.
- $1/f$ structure of frequency fluctuations of English speech show different characteristics, with a break of around 0.1s, the length of a single syllable.

The spectral density $S_v(f)$ of a quantity $V(t)$ fluctuating with time t is a **measure of the mean squared variation** $\langle V^2 \rangle$ in a unit bandwidth centered on the frequency f .

de Coensel et al. (2003) were the first to investigate the presence of this $1/f$ structure in environmental sound sources. Using a selection of 6 recordings from quiet rural areas in Flanders, Belgium and 12 from urban spaces in Ghent, Belgium, the authors examine the log-log spectra of A-weighted sound pressure, psychoacoustic loudness, and pitch fluctuations (i.e. zero-crossing rate) of both rural and urban soundscapes.

Botteldooren et al. (2006) presented a temporal indicator based on this $1/f$ structure. However, ...

Yang et al. (2015) investigated the presence of a $1/f^\alpha$ behaviour in 102 recordings made in the countryside, natural parks, and urban areas in England between 1994 and 2010. They extended the previous investigations beyond simply considering the α value, to also consider whether different frequency regimes (the authors use the term 'range' instead) demonstrated different behaviours and to include the quadratic deviation from the best-fitted straight line. By visual inspection of the $1/f$ plots of the various psychoacoustic metrics, the authors noted two breaks in the frequency regimes, at 10^{-1} Hz and at 10^0 Hz

Through these three primary studies various approaches to understanding this behaviour in soundscapes have been advanced. However, as de Coensel, Botteldooren, and de Muer explicitly stated, none of them have attempted to "link our observations to perception of the soundscape by the human observer". This is where this study picks up.

2.3 Justification

Characterising the dynamic nature of the sound environment, and the influence this will have on perception, is key to extending our assessments beyond what can be assessed in person by researchers.

To bring soundscape back to its roots as a consideration of sound environments as akin to a musical composition, when considering and discussing a piece of music, the three primary components are Melody, Harmony, and Rhythm. In this way, we could consider melody as the foreground, sound mark, or keynote sounds; harmony as the background or ambient sound; and rhythm as the temporal structure of the sound environment, both of the foreground and background sounds. Although the first two domains are fairly well researched in the soundscape literature, there is a lack of understanding of the influence of temporal structure on perception in soundscape.

Rhythm and structure in music can be distinguished into several levels or domains: the short-term is the rhythm of individual notes; the medium-term is the rhythm and structure of musical phrases (which could include the repetition of short-term rhythms); the long-term is the overall structure of a piece and organisation of movements. In the same way, the temporal structure of a sound environment can be distinguished into several domains: short-term (*iос?*) being the time structure of individual sound events (e.g. repeated beeps or knocking); the medium-term ...

In order to move towards an understanding of the temporal structure of complex, real-world soundscapes and towards the eventual development of a metric to describe this temporal structure, an analysis of the $1/f^\alpha$ structures of urban soundscapes has been conducted. By drawing from the large Soundscape Indices (SSID) database (Mitchell et al., 2020), which includes 1,600 socio-acoustic surveys and accompanying 30s binaural recordings characteristics of the $1/f^\alpha$ structure in the time series of calculated psychoacoustic metrics are identified. Correlations between this structure and the soundscape perception as assessed by the in-situ surveys are then calculated to identify the influence the temporal structure may have on perception in real-world urban soundscapes.

2.4 Frequency ranges, intervals, and regimes

zBZoth De Coensel and Yang identified frequency ranges or intervals over which the behaviour of the spectrum differs. de Coensel et al. (2003) divided the frequency interval into two parts: $I_1 = [0.002\text{Hz}, 0.2\text{Hz}]$ and $I_2 = [0.2\text{Hz}, 5\text{Hz}]$. I_1 which ranges from 200ms to 5s is therefore taken to represents the behaviour of the source itself, while I_2 , which ranges from 5s to 10min is influenced by behaviours *between* sources.

If this hypothesis is correct, wherein the high frequency interval is *within* sources and the low frequency interval is *between* sources, we would then expect to see similar interval separation across several different metrics.

3 Methods

3.1 Survey collection

SSID Protocol This study makes use of the database collected as part of the (SSID) project. This SSID database consists of soundscape assessments and corresponding recordings carried out in 17 locations across the United Kingdom, the Netherlands, Spain, and Italy. The ISO 12913-1:2014 (2014) series was consulted for reporting on soundscape data. ISO/TS 12913-2:2018 (2018) and ISO/TS 12913-3:2019 (2019) were used as the basis of the data collection protocol and data analysis used. For a full description of the SSID Protocol and of the publicly available database, please see Mitchell et al. (2020) and ?.

Binaural Recorder

Sessions

Sound sources and PAQ questions

4 Analysis

The temporal variation of several psychoacoustic parameters has been analyzed for each of the included recordings. Based on previous soundscape literature (?)

4.1 Recording Processing

Psychoacoustic calc

Channel max

4.2 $1/f^\alpha$ Analysis

Power Spectrum analysis Welch's method for noisy signals. nperseg determination.

Figure 1: Set of plots including the average line for each of the parameters, with a dotted line indicating the identified regime break.

Figure 2: Single plot of a sample power spectrum, with the regression lines of the regimes plotted.

Slope (α) calculation Regression calculation for the given regimes.

Regime Breaks The regime break is defined as the point in the spectrum at which there is a noticeable change in the slope. This change indicates that the regime at frequencies below the break has a different α from the regime at frequencies above the break and therefore exhibits a different characteristic randomness behaviour.

Figure 1 The spectra for all 1,000+ recordings is plotted, then the geometric mean line of these spectra is plotted in order to visualise the general pattern. The a mean line is used as it is otherwise impractical to investigate the behaviour across such a large number of samples. The geometric mean is used to account for the log scale relationship for the y-axis.

This is done for each parameter independently.

4.3 Correlation Analysis

Partial Distance Correlation using dcor package from Python. Or Mutual information. With control for SessionID.

draft Only explain the method of correlation analysis, leave justification (i.e. non-linear) for the results

5 Results & Discussion

5.1 Identified time and frequency regimes

Discuss What does the regime above and below the break represent? How do we interpret this? How is it useful?

5.2 Slope values

Not -1 slope, not pink noise

Table 1: Summary statistics of slopes?

Rank	ISOPleasant	ISOEventful	Traffic	Natural	Human
1	SIL (.41)	SIL (.43)
2

Table 2: Correlation rankings between psychoacoustic and temporal features, and perceptual features

Discussion Problems with using fft versus periodogram resulted in incorrect interpretation of previous results.

6 Correlations

Explain why MI is better for this. This is where the nrmse / deviation discussion / limitation can go.

7 Concluding Remarks

This theme appeared in literature 10 years ago, then just went quiet.

Figure 3: Non-linear scatterplot for highly correlated feature(s)

Appendix C: Deep Learning Techniques for the detection of noise Annoyance (DeLTA)