Predictive Modelling of Complex Urban Soundscapes

Multi-level Regression and Deep Learning Approaches

Andrew James Mitchell

A thesis presented for the degree of
Doctor of Philosophy

Institute for Environmental Design & Engineering
University College London (UCL)
September 6, 2021

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

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 typically limit their focus to the reduction of unwanted noise, ignoring the 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 the perception of an acoustic environment, an approach known as 'Soundscape'.

When attempting to apply soundscape in practical applications, 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 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.

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 and type of 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 non-acoustic factors is required. Through the combination of in-person field questionnaires and multi-factor characterisation of the environment, my research will focus on building up the levels of analysis and, finally, on integrating these levels and factors into an overall soundscape predictive model which focusses on interpretability and potential for practical use.

Acknowledgements

This research was funded by the European Research Council.

People to thank:

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

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*.

Orga, F., **Mitchell, A.**, 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 Soundscapees of the COVID-19 Lockdown: A predictive soundscape modelling approach. *JASA*.

Mitchell, A., Soelitsyo, C., Erfanian, M., Xue, J-H., Oberman, T., Kang, J., & Aletta, F. (2021). A Temporal Convolutional Neural Network for Multi-label Sound Recognition and Annoyance Detection of Complex Soundscapes. *IEEE*.

Mitchell, A., Aletta, F., Chalabi, Z., & Kang, J. (2021). From Deterministic to Probabilistic Soundscapes: A critical tour around the soundscape circumplex. *JASA-EL*.

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*.

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:

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`

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*.

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.

COVID-19 Statement

In March of 2020, 18 months into the development of this thesis, the COVID-19 pandemic hit the UK, forcing it into lockdowns which would continue for over a year. Solely by good fortune and a tendency to speed ahead with too-little thought, the primary data collection had fortunately been completed prior to the first lockdown. However, this work was impacted in three ways:

1. Further in-situ data collection could not be completed, reducing the range of soundscape types we could include;

2. The unprecedented and stressful world of the pandemic had a significant mental health and social impact, the effects of which cannot be quantified, nor overstated;

3. In response to the unique scientific opportunity of a world-wide transportation and social lockdown, new, unplanned studies were carried out.

In particular, this final point has had an impact on the structure and content of this thesis. Certain aspects of the research, in particular the model development and building, were accelerated and put into practice to investigate the impacts of the COVID lockdowns, before being returned to and further developed. The initial research plan would have followed a more logical path of nailing down the model development first, then moving on to a first implementation. In addition, new work was added to this thesis which may appear incongruous or unrelated, but represents a great deal of necessary work which further informed the key strains of the thesis.

Contents

1 Introduction

1.1 Research Summary

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. 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. 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. How does the soundscape of a place vary over time, is this variation driven by environmental features or by context, and can this variation be predicted?

1.2 The SSID Project

1.2.1 Project collaborators

1.2.2 Motivation for the SSID Project

1.3 Research Aims

1.4 Soundscape Indices and Metrics

1.4.1 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 (International Organization for Standardization, 2014, 2018, 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.

1.4.2 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)

1.4.3 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, 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 under-perform 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.

1.4.4 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.

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.

draft *(In contrast to Lacey (2016))* Where previous ground-breaking strategies toward practical urban soundscape design (Lacey, 2016), 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 incredible 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, 2016) 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).

1.5 Practical Applications for Predictive 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 draft *cite 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.

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.

draft *Predicting soundscape of not-yet-existing spaces*

draft *Soundscape mapping*

draft *Where to put this in this section?* 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 International Organization for Standardization (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.

1.6 General Aim

2 Literature Review

2.1 Impact of Urban Noise on Health and Wellbeing

cit *Environmental noise in Europe 2020*

Give a full formal background to why noise control is important for public health.

2.2 Current Methods of Assessing and Addressing Urban Noise

The approach to a practical predictive soundscape model arrived at within this thesis is heavily based on past environmental acoustics approaches. I will therefore begin with a brief summary of these past approaches.

2.2.1 Acoustical Parameters

2.2.2 ISO Environmental Acoustics Standards

ISO 1996-1, esp sections on annoyance, e.g. Annex F, G, H

2.2.3 EU Noise Mapping

cit *Environmental noise in Europe 2020*

2.2.4 Shortcomings

2.3 Soundscape Studies

2.3.1 Soundscape Descriptors and Indices

2.3.2 World Soundscape Project

2.3.3 Swedish Soundscape Quality Protocol

2.3.4 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

2.4 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.

2.4.1 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) (EU, 2002). 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.* .

2.4.2 SSID approach

2.4.3 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.*

2.4.4 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.

2.5 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)

2.5.1 Models based on non-acoustic data sources

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

2.5.2 Red and Green Soundscape Indices

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

3 Methods

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.

3.1 Questionnaires

The full protocol developed for this thesis is outlined in Chapter 4. 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.

3.1.1 Likert Responses

3.1.2 Circumplex Projection

3.2 Psychoacoustics and Auditory Perception

3.2.1 Psychoacoustic Parameters

Loudness

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

3.2.2 Feature Selection

3.3 Machine Learning and Regression Techniques

3.3.1 Feature Selection

Mutual Information

draft *It appears that mutual information is related to the Bayes formula. I still need to read more into this, but it appears based on relative and overlapping probability distributions between the variables in question.*

From scholarpedia: draft *Based on entropy, where the uncertainty about a variable can be expressed as "the number of yes/no questions it takes to guess a random variable, given knowledge of the underlying distribution and taking the optimal question-asking strategy". "The mutual information is therefore the* reduction *in uncertainty about variable X, or the expected reduction in the number of yes/no questions needed to guess X after observing Y.".*

draft *"Mutual Information is just one way among many of measuring how related two variables are. However, it is a measure ideally suited for analyzing communication channels. Abstractly, a communication channel can be visualized as a transmission medium which receives an input x and produces an output y. If the channel is* noiseless, *the output will be equal to the input. However, in general, the transmission medium is noisy and an input x is converted to an output y with probability $P_{Y|X}(y|x)$.* misc *This seems very useful for my conception of sound perception / auditory processing, where the perception system is a noisy communication channel.*

Conditional Mutual Information

The Mutual Information between two variables, given another variable as a control.

3.3.2 Clustering Analysis

K-means

nbclust

3.3.3 Modelling Likert-type Data

Multiple Linear Regression

Ordinal Logistic Regression

Multi-output Regression

3.3.4 Multi-level Models

3.3.5 Bayesian Regression

4 The Soundscape Indices (SSID) Protocol: A Method for Urban Soundscape Surveys – Questionnaires with Acoustical and Contextual Information

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 (International Organization for Standardization, 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 Soundscape Indices (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.

4.1 Introduction

Soundscape studies strive to understand the perception of a sound environment, in context, including acoustic, (non-acoustic) environmental, contextual, and persona factors. 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 (International Organization for Standardization, 2014, 2018, 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) (International Organization for Standardization, 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 (International Organization for Standardization, 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.

4.2 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.

4.3 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 (International Organization for Standardization, 2018) collected either via handheld tablets or paper copies of the questionnaire. Typically, a minimum of 100 responses are collected at each location during multiple 2-5 hr sessions over several days. During the survey sessions, acoustic data are collected via a stationary class 1 or class 2 Sound Level Meter (SLM) (as defined in IEC 61672-1:2013 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 4.5. 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.

4.3.1 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 4.1.

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

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 4.1: 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.

consistency with the format is crucial so that data can be filtered according to the SessionID or the location, as is often necessary. In addition, for ease of automatic processing, it is recommended not to include spaces in the SessionID to avoid string splitting issues in analysis code.

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

4.3.2 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

Table 4.2: Recommended equipment for implementing the SSID protocol. SLM: Sound Level Meter; AMB: Ambisonics; BIN: Binaural; QUE: Questionnaires

Table 4.3: Table of recommended context and acoustic measurement factors.

wall, or a mast should be avoided as it might cause issues in later handling the visual data (3m is considered a safe distance from these features). Similar concerns are also true for the audio data and careful attention should be paid to avoid placing the recording equipment near extraneous noisy equipment or acoustic shadows. Further guidance on this is given in Point 4 of Section 4.5. 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.

4.3.3 Equipment

The equipment listed in Table 4.3.3 is designed to facilitate both the audio-visual recording of the location and the collection of objective environmental factors, as given in Table 4.3.3. 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.

4.4 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.

4.4.1 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 assess on a 5-point Likert scale, coded from 1 to 5. A sample codebook to demonstrate the recommended variable naming and response coding is included in 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 SSI01 through SSI04, respectively). These taxonomic categories of environmental sounds are based on the work done by Guastavino et al. (2005) and Brown and De Coensel (2018) cit *45*.

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

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

The fourth section comprises the WHO 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 Aletta et al. (2019) 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 [MISC03]) and a free response for the participant to provide any additional comments they would like to make on the sound environment [MISC01]. It is important to note that the section on ethnicity, and to a lesser extent education level, will need to be adjusted to ensure the available responses are appropriate for the location where the survey is being conducted.

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

This questionnaire is intended to collect a consistent core set of perceptual responses and information about the participant, with space to add additional questions as required by specific research goals. Some examples of this 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.

4.4.2 Contextual and Environmental Factor Data Collection

During each survey, the equipment listed in Section 4.3.3 is set up to capture the contextual and environmental data for the location. Table 4.3.3 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 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 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,

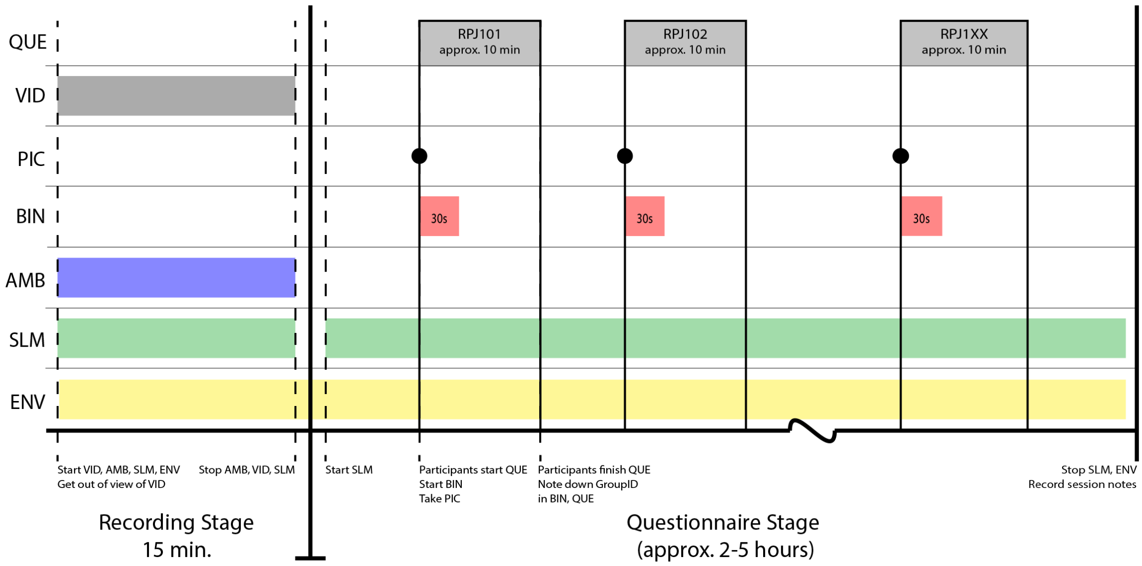


Figure 4.1: Timeline of the on site soundscape protocol. RegentsParkJapan (RPJ) is used as an example. Abbreviations as defined in Table 4.3.3 – QUE: Questionnaires; VID: 360°video; PIC: Site pictures; BIN: Binaural Recording; AMB: Ambisonic recording; SLM: Sound Level Meter (acoustical factors); ENV: Environmental factors.

this has been achieved through the use of separate calibrated binaural recordings and measurements made with a calibrated SLM.

4.5 Procedure

Figure 4.1 shows the whole process of the on site soundscape protocol. The relevant equipment in each row should be operating when the row is coloured in, such that when multiple rows are shaded this means that multiple pieces of equipment should be running during that time period. The following section prepares step-by-step instructions for conducting the in-situ surveys, including the Recording Stage and Questionnaire Stage. Figure 4.2 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.

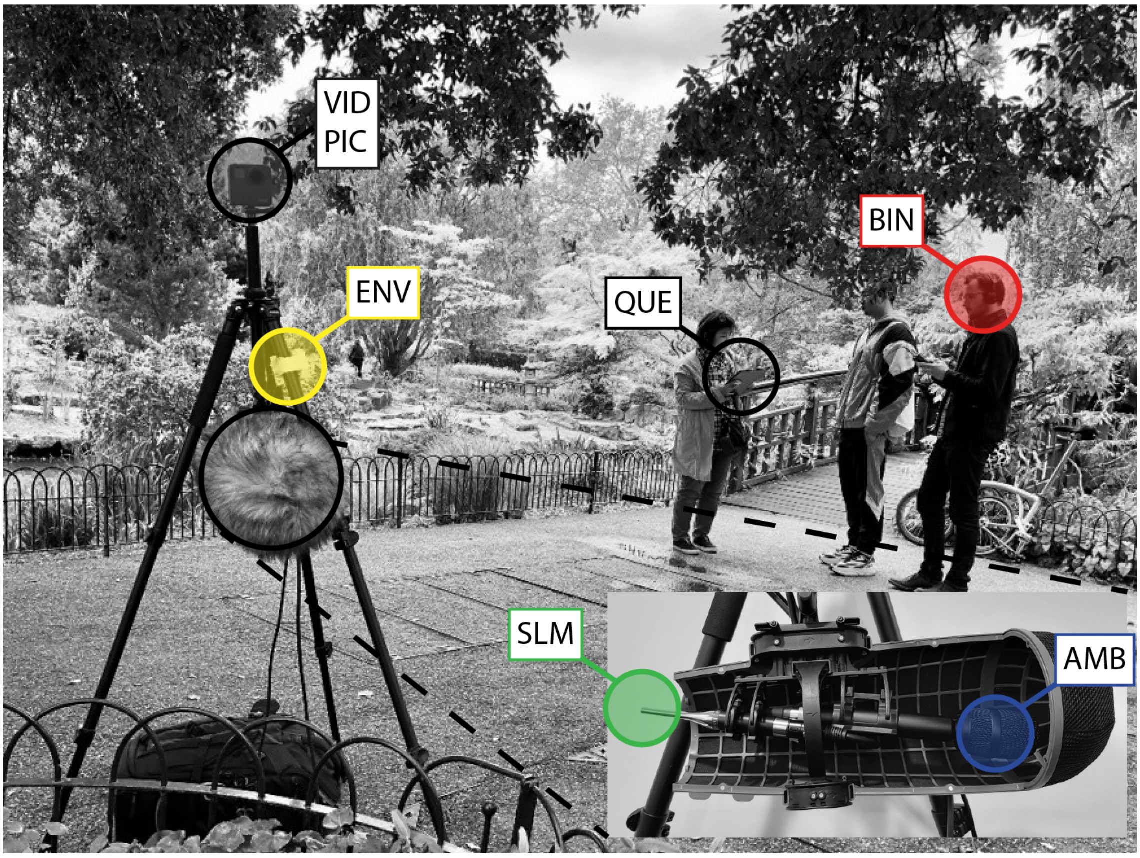


Figure 4.2: Photo of a full survey carried out in a park in London during the Questionnaire Stage. To the left is the equipment (colour-coded to match Figure 4.1), with the ambisonic microphone and SLM microphone in the windscreen, with the 360°camera on top of the tripod and to the right are one researcher 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.

4.5.1 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 4.2.

   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 4.3.3. 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.

4.5.2 Recording Stage

The following section prepares step-by-step instructions for conducting the Recording Stage of the on site protocol, as shown in Figure 4.1.

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.

4.5.3 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 4.3.2, 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 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.)

4.6 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.

4.6.1 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 comments from five research groups addressed age and type of location as the most influential factors for participant sampling. However, while some reported higher response rates from younger (student) members of the public, 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.

4.6.2 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.

4.6.3 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 windshield out of curiosity.

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

4.6.4 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.

5 Study I: Psychological Well-being and Demographic Factors can Mediate Soundscape Pleasantness and Eventfulness: A large sample study

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.*

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.

Journal of Environmental Psychology 77 (2021) 101660

Contents lists available at ScienceDirect

Journal of Environmental Psychology

journal homepage: www.elsevier.com/locate/jep

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

A R T I C L E I N F O

Handling Editor Leila Scannell

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

A B S T R A C T

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).

https://doi.org/10.1016/j.jenvp.2021.101660
Received 24 February 2020; Received in revised form 22 July 2021; Accepted 27 July 2021
Available online 31 July 2021
0272-4944/

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 (Krusemark, 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 12913–2:2018 [2] 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.

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)

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

[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 12913 [3] 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 (LMER) 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

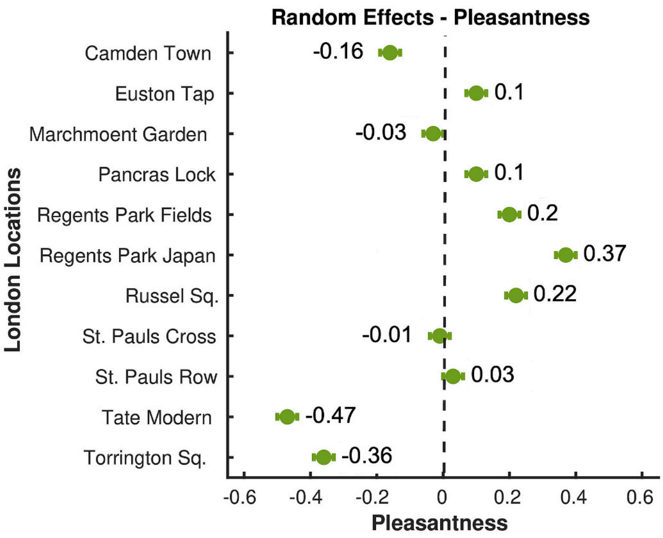


Fig. 1. The summary result demonstrated in the random-effects figures gives the average from the distribution of Pleasantness across locations.

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		

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

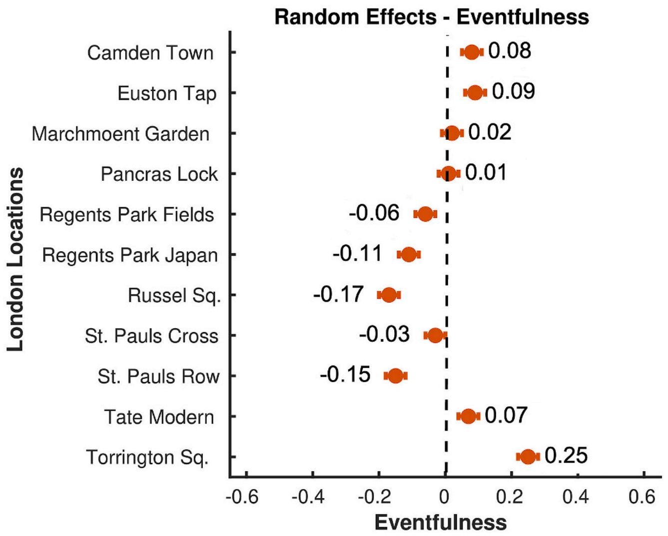


Fig. 2. The summary result demonstrated in the random-effects figures gives the average from the distribution of Eventfulness across locations.

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 Aydın and Yılmaz (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 the 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

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

Babkoff, 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., Högberg, 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 Aydın, D., & Yılmaz, S. (2016). Assessment of sound environment pleasantness by sound quality metrics in urban spaces. *A|Z 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

Russel, 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., & Teder-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., Søndergaard, 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

Zwanzger, P., Zavorotnyy, 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 Study II: Multilevel Annoyance Modelling of Short Environmental Sound Recordings

Published as: Orga, F., **Mitchell, A.**, 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`

Abstract

The recent development and deployment of Wireless Acoustic Sensor Networks (WASN) present new ways to address urvan 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 a promising step torward 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

Noise has been proven to have a wide impact on the social and economic aspects of citizens' lives (Goines and Hagler, 2007) and is regarded as one of the primary environmental health issues referenced in the new environmental noise guidelines (Nations, 2018). Over the past few years, several research teams have analysed the causes and the impact of this noise, revealling that it causes more than 48,000 new cases of ischemic heart disease and around 12,000 deaths in Europe each year (Blanes et al., 2017). Furthermore, it leads to chronic high annoyance for more than 22 million people, and sleep disturbance for more than 6.5 million people (Ndrepepa et al., 2011). One of the main noise sources according to research is road traffic noise (Ouis, 2001), causing psychological reactions in citizens Basner et al. (2006) and even cardiovascular diseases (Ndrepepa et al., 2011).

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, squeels, 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 charactersitics 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.

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 closes 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 .

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 .

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 quesstions 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 sesveral 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 detailes the state-of-the-art of annoyance modelling by means of subjective data collection. Section 4.5 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 annoyanced 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 inn 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 corresponse 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 accordig to a small group of attributes such as *loud, shrill, noisy, disturbing, sharp, excting, 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, which stands for *MUlti Stimulus test with Hidden Reference and Anchor*, 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 annoyanced 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 responnse, 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

sustainability

Article

Multilevel Annoyance Modelling of Short Environmental Sound Recordings

Ferran Orga [1,†], Andrew Mitchell [2,†], Marc Freixes [1], Francesco Aletta [2], Rosa Ma Alsina-Pagès [1,*] and Maria Foraster [3,4,5,6]

1 Grup de recerca en Tecnologies Mèdia, La Salle—URL, Quatre Camins, 30, 08022 Barcelona, Spain; ferran.orga@salle.url.edu (F.O.); marc.freixes@salle.url.edu (M.F.)
2 Institute for Environmental Design and Engineering, The Bartlett, University College London (UCL), Central House, 14 Upper Woburn Place, London WC1H 0NN, UK; andrew.mitchell.18@ucl.ac.uk (A.M.); f.aletta@ucl.ac.uk (F.A.)
3 ISGlobal, Parc de Recerca Biomèdica de Barcelona (PRBB), Doctor Aiguader, 88, 08003 Barcelona, Spain; maria.foraster@isglobal.org
4 Universitat Pompeu Fabra (UPF), 08018 Barcelona, Spain
5 CIBER Epidemiología y Salud Pública (CIBERESP), 28029 Madrid, Spain
6 PHAGEX Research Group, Blanquerna School of Health Science, Universitat Ramon Llull, 08025 Barcelona, Spain
* Correspondence: rosamaria.alsina@salle.url.edu; Tel.: +34-932-902-425
† These authors contributed equally to this work and share the first author position.

Abstract: 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 (1–2.7 s) 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 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.

Keywords: noise; annoyance evaluation; citizen; perceptive test; smart-city; annoyance modelling; wireless acoustic sensor network

Citation: Orga, F.; Mitchell, A.; Freixes, M.; Aletta, F.; Alsina-Pagès, R.M.; Foraster, M. Multilevel Annoyance Modelling of Short Environmental Sound Recordings. *Sustainability* **2021**, *13*, 5779. https://doi.org/10.3390/su13115779

Academic Editors: Cinzia Buratti, Juan Miguel Navarro and Jaume Segura-Garcia

Received: 22 April 2021
Accepted: 17 May 2021
Published: 21 May 2021

1. Introduction

Noise has been proven to have a wide impact on the social and economic aspects of citizens' lifes [1] and is regarded as one of the primary environmental health issues referenced in the new environmental noise guidelines [2]. Over the past few years, several research teams have analyzed 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 [2]. Furthermore, it leads to chronic high annoyance for more than 22 million people, and sleep disturbance for more than 6.5 million people [3]. One of the main noise sources according to research is road traffic noise [4], causing psychological reactions in citizens [5] and even cardiovascular diseases [4]. Other studies analyze the effects of aircraft noise on sleep [6] and learning impairments on children [7]. Also railway noise has proven to cause annoyance due to its huge variety of sounds, e.g., rail breaks, whistles, squeels and vibrations [8,9]. Most of the literature focuses on sound level measurements and the corresponding annoyance [10], but other acoustical and psychoacoustical characteristics could be taken into account, e.g., loudness or sharpness [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.

Several tests used in studies to evaluate the effects of environmental noise for citizens [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 analyzing the impact of sound -usually traffic- on citizens in urban environments [13,14], in order to model the annoyance perception [15,16].

The perceptual tests were designed to measure the annoyance in people relating to different urban sounds and their characteristics [17,18], by means of short excerpts of raw acoustic audio obtained from the DYNAMAP project [19]. The most representative audio excerpts were selected, using a wide range of sound types (sirens, airplanes, people talking, dogs barking, etc.) [20,21], keeping the constants of location and sensor calibration. However, sound annoyance depends on the acoustic characterization of each sample, and it is possible to classify the acoustic excerpts depending on their characterization, 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 [11].

The authors asked more than 100 people to conduct the perceptual tests [18]. Some preliminary results of the three tests conducted were published in [17] in which the relationship between sharpness and annoyance was analyzed by means of an A/B test [22], and later on in [18], where some of the research questions were formulated. In this paper, we 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 [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 paper is structured as follows: Section 2 details the state-of-the-art of annoyance modelling by means of subjective data collection. Section 3 describes the procedure followed in this work, including the dataset and the design of the perceptual test. In Section 4, the results obtained from the perceptive tests are presented and discussed, and the annoyance model is proposed. Section 5 contains for the discussion and finally, Section 6 presents the conclusions of the paper.

2. State of the Art of Annoyance Evaluation and Modelling

In this section we 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 the perceptual annoyance have been based.

2.1. Evaluation of Annoyance

The evaluation of annoyance can be found in literature by means of the use of objective parameters related to sound and noise [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 [24–26]. 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 numeric 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 soundscapes evaluation [27] citizens can be asked about other aspects besides annoyance. To this end, a perceptual assessment based on a Likert scale [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 [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 [27]).

Borrowing from the subjective assessment of audio quality, the MUSHRA method has been also used for the evaluation of annoyance in [17,18]. MUSHRA, which stands for *MUlti Stimulus test with Hidden Reference and Anchor*, was described and designed by ITU-R under the recommendation ITU-R BS.1534-3 [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 organizer to have control over all the setup. 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 [29]. These tests can be run in parallel and anywhere, thereby reducing costs and allowing to reach a wider audience [30].

2.2. Annoyance Prediction

After the design and the execution of the perceptual tests, the resulting evaluations coming from participants are used to generate an 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 [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: *(i)* signal-to-noise ratio (SNR), *(ii)* indoor background level, *(iii)* the activity conducted by the listener—assuming that in the conducted tests, their main activity is not listening to events—among others. 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 [16], where the authors model and predict road traffic-noise annoyance based on: *(i)* noise perception, *(ii)* noise exposure levels and *(iii)* demographics. The authors apply machine-learning algorithms in order to conduct the prediction and measure the 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 [11]. 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 [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 [32], the authors led us to a better understanding of the transportation noise-annoyance response, in three different and relevant approximations: *(i)* to unravel the factors that affect the annoyance response of people in reference to the mixed transportation noise, *(ii)* to contrast the noise-annoyance dependence in situations where road traffic and railway noise dominate and *(iii)* 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 [33] an approach to develop a road traffic noise annoyance prediction model is presented, and it takes into account: *(i)* social aspects, *(ii)* characteristics of traffic and *(iii)* 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 that provide no subjective indicator.

3. Methods

In this section we detail the several methods applied our experiment from the perceptual test design based on an urban sound dataset [21] to the multilevel linear regression modelling applied to obtain the annoyance prediction described as contribution in this paper.

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 [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 [34]. After that, experts from the DYNAMAP developing team labelled the acoustic events of the recordings manually to obtain a 151-h dataset [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 RTN (Road Traffic Noise) that belongs to the 83.7% of the total time of the dataset, and ANE (Anomalous Noise Event) with the 8.7% of the total time. Another class was used to include overlapping and unidentified events: COMPLX (complex) with 7.6% of the total time [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) [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 [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 [17]. The reason for that choice was double: *(i)* 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, and *(ii)* 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 the events is only be carried with sounds collected using the same sensor, in order to respect 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 variate 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 the perceptual test, i.e., hb115, hb124, hb127 and hb133 [20]. More details about the event selection process and availability study of the sensors are detailed in [17], and the time of each event in the sensors is depicted in [18].

3.2. Design of the Perceptual Tests

In order to assess the degree of annoyance produced by the aforementioned classes of sounds, an on-line test has been conducted using the Web Audio Evaluation Tool [30]. Specifically, the MUSHRA test method [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 the 100 participants, asking for 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 customized including a color 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 1). An audio is reproduced when the corresponding bar is clicked. The system ensures the participant listens to all the audios and 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 undertok this test, 59 men and 41 women, with an average 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 *(i)* obtain an individual score of annoyance for each audio and *(ii)* carry 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 1.

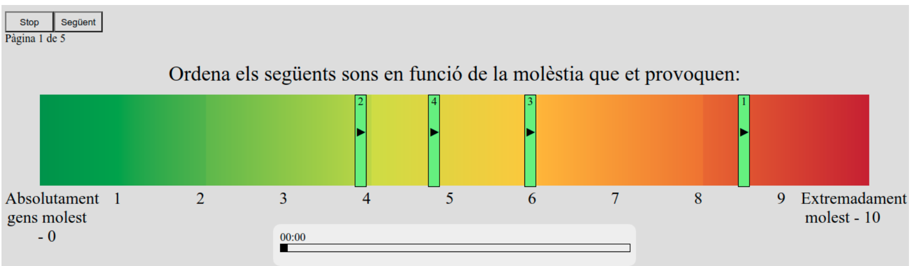


Figure 1. Screenshot of the MUSHRA test conducted to assess the annoyance provoked by different sounds. Title: sort the following sounds according to the caused annoyance. The scale ranges from *not annoying at all* to *extremely annoying*.

Table 1. Psychoacoustic parameters calculated for the 27 stimuli used in the listening experiment.

Sensor	Label	Psychoacoustic Parameters				
		Loudness (N_5 sone)	Sharpness (acum)	Roughness (asper)	Tonality (tuHMS)	Impulsiveness (iu)
hb133	peop	15.1	1.46	0.032	0.204	0.270
hb133	door	16.8	1.43	0.029	0.113	0.354
hb133	dog	13.1	1.22	0.033	0.373	0.266
hb133	brak	16.0	1.76	0.030	0.326	0.241
hb133	bird	12.6	1.73	0.024	0.283	0.214
hb133	airp	13.0	1.27	0.060	0.438	0.231
hb127	sire	17.7	1.56	0.045	1.540	0.178
hb127	peop	16.1	1.62	0.035	0.410	0.417
hb127	horn	18.1	1.56	0.028	0.666	0.260
hb127	door	19.8	1.72	0.037	0.037	0.479
hb127	brak	19.0	1.95	0.034	0.251	0.281
hb127	sire	20.1	1.73	0.046	1.670	0.288
hb127	peop	22.0	1.96	0.036	0.322	0.452
hb127	horn	19.9	2.16	0.034	1.290	0.336
hb127	brak	21.0	1.81	0.030	1.170	0.275
hb127	airp	24.4	1.65	0.056	0.172	0.446
hb115	wrks	20.3	1.97	0.054	0.227	0.267
hb115	trck	24.4	1.60	0.033	0.040	0.276
hb115	sire	19.5	1.46	0.054	0.861	0.333
hb115	peop	25.1	1.79	0.032	0.411	0.331
hb115	horn	22.3	2.00	0.032	0.806	0.155
hb115	door	26.3	1.62	0.038	0.045	0.397
hb115	brak	20.6	1.93	0.034	0.216	0.313
hb115	wrks	24.6	1.92	0.064	0.447	0.317
hb115	sire	26.6	1.77	0.044	0.626	0.290
hb115	horn	29.5	2.35	0.039	0.486	0.262
hb115	door	31.3	1.88	0.048	0.223	0.402

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 in the ArtemiS Suite software (v. 11.5, HEAD acoustics GmbH) and the following psychoacoustic parameters were computed: *loudness*, *sharpness*, *roughness*, *tonality*, and *impulsiveness* [11]; values for these parameters are reported in Table 1. The rationale for selecting a relatively large set of psychoacoustic metrics is that they are often used as indicators to predict perceptual constructs (such as annoyance) in perceptual studies, as shown in recent soundscape literature [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 [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 [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 [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 the ECMA-74 (17th edition), which is based on the hearing model of Sottek, with a frequency range of 20 Hz to 20 kHz [41].

3.4. Multi-Level Linear Regression Modelling

The analysis for this study utilizes multi-level linear regression modelling (MLM), with a random intercept and a random slope, using backward step feature selection. MLM~~'~~s are commonly used in psychological research for repeated measures studies [42,43] and for applied prediction models [44,45]. Multi-level modelling allows for the incorporation of nested and non-nested group 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 ~~is~~ 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 [46], is thus:

$$\begin{aligned} \text{annoyance} \sim \text{Loudness} + \text{Roughness} + \text{Sharpness} + \text{Tonality} + \text{Impulsiveness} \\ +(1 \mid \text{user}) + (1 + \text{Loudness} + \text{Roughness} + \text{Sharpness} + \text{Tonality} + \text{Impulsiveness} \mid \text{label}) \end{aligned} \tag{1}$$

Feature Selection

The MLM is initially fitted with all of the potential features included within both levels. In order to reduce the complexity of the model, a backwards step feature 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., 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 standardized estimates terms. Finally, quantile plots of the residuals and random effects are examined to confirm they are normally distributed [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 [47]. The model fitting and feature selection was performed using the 'step' function from 'lmerTest' (v. 3.1.3) [48] in the R statistical software (v. 4.0.5) [49]. The summaries and plots were created using the 'sjPlot' package (v. 2.8.7) [50] and the multi-level R^2 values were calculated using 'MuMIn' (v. 1.43.17) [51].

4. Results

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 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 measure 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 2.

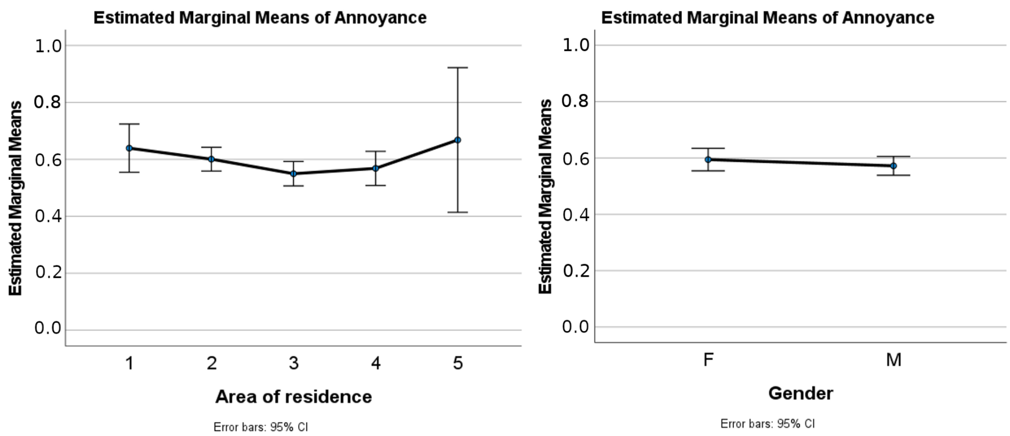


Figure 2. Estimated Marginal Means for Annoyance as a function of Areas of residence (**left**) and Gender (**right**).

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 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 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 3 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. Among those, the effects of Impulsiveness, 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 they tonal component is 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.

Table 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.

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

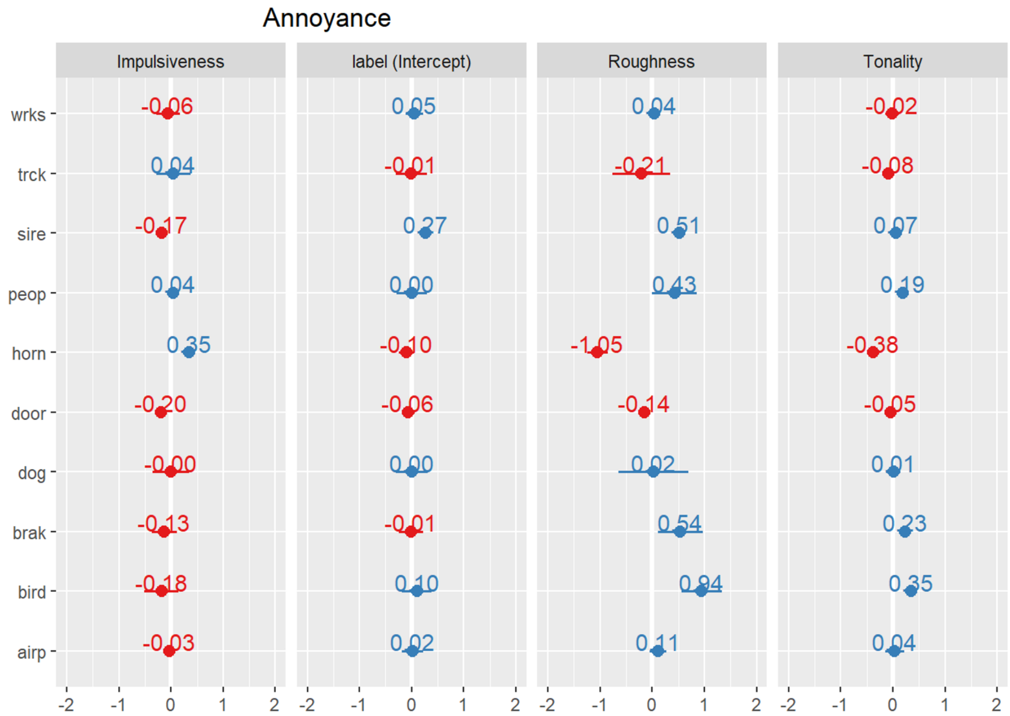


Figure 3. Second-level effects figures representing the regression coefficients by types of sound (label) and for different psychoacoustic parameters.

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 wireless acoustic sensor network (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 [52]. A relevant issue to consider from the WASN perspective, is that previous studies conducted in both urban [21] and suburban [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 [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., 1–3 s) 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 any loudness-related parameter was selected by the feature selection algorithm. To some extent, this is possibly due to fact that, being an online experiment, it was not possible for the research team to actually calibrate the loudness playback level accurately for the remote participants. On the other hand, considering this aspect from the WASN implementation perspective, this could be seen as an encouraging finding, since calibrating a diffuse acoustic monitoring network may not be practical in real-world scenarios, so it is good to have models that can achieve up to 64% of variance explained regardless of actual levels. Furthermore, in complex acoustic environments, loudness would likely vary over time depending on the relative positions between sound sources and (human) listeners in ways in which the other psychoacoustic parameters such as sharpness and tonality are less likely to. This is something that is impossible for fixed sensors to take into account, so once again it is preferable not to rely on loudness as a predictor.

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 off-line perceptual test;
- when considering short recordings of single-source environmental sounds, no significant differences in noise annoyance were observed as a function of demographic factors, such as gender and self-reported area of residence (i.e., from very quiet to very noisy);
- the multi-level linear regression model derived from this case study achieved an overall $R^2 = 0.64$, using sharpness as a fixed effect (the first level), and impulsiveness, roughness, 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 the all the stages described in this paper. The improvement of the 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 estimated 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.

Author Contributions: Conceptualization, F.A., R.M.A.-P., M.F. (Maria Forster); methodology, F.A., A.M., R.M.A.-P., M.F. (Maria Forster); software, F.O., M.F. (Marc Freixes), A.M.; investigation, M.F. (Marc Freixes), F.O., A.M., R.M.A.-P., F.A., M.F. (Maria Forster); data curation, F.O., A.M., F.A.; writing–review and editing, M.F. (Marc Freixes), F.O., A.M., R.M.A.-P., F.A., M.F. (Maria Forster); visualization, A.M., M.F. (Marc Freixes); supervision, M.F. (Maria Forster), F.A., R.M.A.-P.; project administration, F.A., R.M.A.-P.; funding acquisition, F.A., R.M.A.-P. All authors have read and agreed to the published version of the manuscript.

Funding: The UCL authors are funded by the European Research Council (ERC) under the European Union's Horizon 2020 research programme (grant agreement No. 740696). La Salle authors would like to thank Secretaria d'Universitats i Recerca from the Departament d'Empresa i Coneixement (Generalitat de Catalunya) and Universitat Ramon Llull, under the grant 2020-URL-Proj-054 (Rosa Ma Alsina-Pagès).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data is available here: Ferran Orga, Marc Freixes, Rosa Ma. Alsina-Pagès, Alexandra Labairu-Trenchs. (2021). Audio dataset for perceptive studies in DYNAMAP project [Data set]. Zenodo. https://doi.org/10.5281/zenodo.4775328, accessed on 22 April 2021.

Acknowledgments: The authors would like to thank to all the participants who volunteered to conduct the listening experiment.

Conflicts of Interest: The authors declare no conflict of interest. The Funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ANE	Anomalous Noise Event
ANOVA	Analysis of variance
ICBEN	International Committee for the Biological Effects of Noise
L_{eq}	Equivalent Level
MLM	Multi-level Linear regression Modelling
MUSHRA	MUlti Stimulus test with Hidden Reference and Anchor
RTN	Road Traffic Noise
VIF	Variance Inflation Factor
WASN	Wireless Acoustic Sensor Network

References

1. Goines, L.; Hagler, L. Noise Pollution: A Modern Plague. *South. Med. J. Birm. Ala.* **2007**, *100*, 287–294. [CrossRef]
2. World Health Organization. *Environmental Noise Guidelines for the European Region*; Technical Report; World Health Organization: Geneva, Swizterland, 2018.
3. Blanes, N.; Fons, J.; Houthuijs, D.; Swart, W.; de la Maza, M.; Ramos, M.; Castell, N.; van Kempen, E. Noise in Europe 2017: Updated Assessment. In *European Topic Centre on Air Pollution and Climate Change Mitigation (ETC/ACM)*; European Environment Agency: Bilthoven, The Netherlands, 2017.
4. Ndrepepa, A.; Twardella, D. Relationship between noise annoyance from road traffic noise and cardiovascular diseases: A meta-analysis. *Noise Health* **2011**, *13*, 251. [PubMed]
5. Ouis, D. Annoyance from road traffic noise: A review. *J. Environ. Psychol.* **2001**, *21*, 101–120. [CrossRef]

6. Basner, M.; Samel, A.; Isermann, U. Aircraft noise effects on sleep: Application of the results of a large polysomnographic field study. *J. Acoust. Soc. Am.* **2006**, *52*, 109–123. [CrossRef] [PubMed]
7. Hygge, S.; Evans, G.W.; Bullinger, M. A prospective study of some effects of aircraft noise on cognitive performance in schoolchildren. *Psychol. Sci.* **2002**, *13*, 469–474. [CrossRef] [PubMed]
8. Licitra, G.; Fredianelli, L.; Petri, D.; Vigotti, M.A. Annoyance evaluation due to overall railway noise and vibration in Pisa urban areas. *Sci. Total Environ.* **2016**, *568*, 1315–1325. [CrossRef] [PubMed]
9. Gidlöf-Gunnarsson, A.; Ögren, M.; Jerson, T.; Öhrström, E. Railway noise annoyance and the importance of number of trains, ground vibration, and building situational factors. *Noise Health* **2012**, *14*, 190. [CrossRef] [PubMed]
10. Berglund, B.; Lindvall, T.; Schwela, D.H. *Guidelines for Community Noise*; World Health Organization: Geneva, Switzerland, 1995.
11. Zwicker, E.; Fastl, H. *Psychoacoustics: Facts and Models*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; Volume 22.
12. Guski, R.; Schreckenberg, D.; Schuemer, R. WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Annoyance. *Int. J. Environ. Res. Public Health* **2017**, *14*, 1539. [CrossRef]
13. Botteldooren, D.; Dekoninck, L.; Gillis, D. The influence of traffic noise on appreciation of the living quality of a neighborhood. *Int. J. Environ. Res. Public Health* **2011**, *8*, 777–798. [CrossRef]
14. Jakovljevic, B.; Paunovic, K.; Belojevic, G. Road-traffic noise and factors influencing noise annoyance in an urban population. *Environ. Int.* **2009**, *35*, 552–556. [CrossRef]
15. De Muer, T.; Botteldooren, D.; De Coensel, B.; Berglund, B.; Nilsson, M.; Lercher, P. A model for noise annoyance based on notice-events. In Proceedings of the 2005 International Congress and Exposition on Noise Control Engineering (Internoise 2005)-Paper 2033, Rio de Janeiro, Brazil, 7–10 August 2005.
16. Bravo-Moncayo, L.; Lucio-Naranjo, J.; Chávez, M.; Pavón-García, I.; Garzón, C. A machine learning approach for traffic-noise annoyance assessment. *Appl. Acoust.* **2019**, *156*, 262–270. [CrossRef]
17. Labairu-Trenchs, A.; Alsina-Pagès, R.M.; Orga, F.; Foraster, M. Noise Annoyance in Urban Life: The Citizen as a Key Point of the Directives. *Proceedings* **2019**, *6*, 1. [CrossRef]
18. Alsina-Pagès, R.M.; Freixes, M.; Orga, F.; Foraster, M.; Labairu-Trenchs, A. Perceptual evaluation of the citizen's acoustic environment from classic noise monitoring. *Cities Health* **2021**, *5*, 145–149. [CrossRef]
19. Sevillano, X.; Socoró, J.C.; Alías, F.; Bellucci, P.; Peruzzi, L.; Radaelli, S.; Coppi, P.; Nencini, L.; Cerniglia, A.; Bisceglie, A.; et al. DYNAMAP—Development of low cost sensors networks for real time noise mapping. *Noise Mapp.* **2016**, *3*, 172–189. [CrossRef]
20. Alías, F.; Orga, F.; Alsina-Pagès, R.M.; Socoró, J.C. Aggregate Impact of Anomalous Noise Events on the WASN-Based Computation of Road Traffic Noise Levels in Urban and Suburban Environments. *Sensors* **2020**, *20*, 609. [CrossRef] [PubMed]
21. Alías, F.; Socoró, J.C.; Alsina-Pagès, R.M. WASN-Based Day–Night Characterization of Urban Anomalous Noise Events in Narrow and Wide Streets. *Sensors* **2020**, *20*, 4760. [CrossRef] [PubMed]
22. Bech, S.; Zacharov, N. *Perceptual Audio Evaluation-Theory, Method and Application*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
23. Recommendation, I. *1534-1: Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems*; International Telecommunication Union: Geneva, Switzerland, 2003.
24. Fields, J.; De Jong, R.; Gjestland, T.; Flindell, I.; Job, R.; Kurra, S.; Lercher, P.; Vallet, M.; Yano, T.; Research Team at Ruhr University; et al. Standardized general-purpose noise reaction questions for community noise surveys: Research and a recommendation. *J. Sound Vib.* **2001**, *242*, 641–679. [CrossRef]
25. Björk, J.; Ardö, J.; Stroh, E.; Lövkvist, H.; Östergren, P.O.; Albin, M. Road traffic noise in southern Sweden and its relation to annoyance, disturbance of daily activities and health. *Scand. J. Work. Environ. Health* **2006**, *32*, 392–401. [CrossRef]
26. Lim, C.; Kim, J.; Hong, J.; Lee, S. The relationship between railway noise and community annoyance in Korea. *J. Acoust. Soc. Am.* **2006**, *120*, 2037–2042. [CrossRef]
27. Lercher, P.; van Kamp, I.; von Lindern, E.; Botteldooren, D. Perceived Soundscapes and Health-Related Quality of Life, Context, Restoration, and Personal Characteristics: Case Studies. In *Soundscape and the Built Environment*; CRC Press: Boca Raton, FL, USA, 2016; [CrossRef]
28. Likert, R. A technique for the measurement of attitudes. *Arch. Psychol.* **1932**, *22*, 55.
29. Buchholz, S.; Latorre, J.; Yanagisawa, K. Crowdsourced Assessment of Speech Synthesis. In *Crowdsourcing for Speech Processing*; John Wiley & Sons, Ltd.: Oxford, UK, 2013; pp. 173–216. [CrossRef]
30. Jillings, N.; Moffat, D.; De Man, B.; Reiss, J.D. Web Audio Evaluation Tool: A browser-based listening test environment. In Proceedings of the 12th Sound and Music Computing Conference, Maynooth, Ireland, 26 July–1 August 2015.
31. Di, G.Q.; Chen, X.W.; Song, K.; Zhou, B.; Pei, C.M. Improvement of Zwicker's psychoacoustic annoyance model aiming at tonal noises. *Appl. Acoust.* **2016**, *105*, 164–170. [CrossRef]
32. Lam, K.C.; Chan, P.K.; Chan, T.C.; Au, W.H.; Hui, W.C. Annoyance response to mixed transportation noise in Hong Kong. *Appl. Acoust.* **2009**, *70*, 1–10. [CrossRef]
33. Ascigil-Dincer, M.; Demirkale, S.Y. Model development for traffic noise annoyance prediction. *Appl. Acoust.* **2021**, *177*, 107909. [CrossRef]
34. Alsina-Pagès, R.M.; Alías, F.; Socoró, J.C.; Orga, F. Detection of Anomalous Noise Events on Low-Capacity Acoustic Nodes for Dynamic Road Traffic Noise Mapping within an Hybrid WASN. *Sensors* **2018**, *18*, 1272. [CrossRef] [PubMed]
35. Alías, F.; Socoró, J.C. Description of anomalous noise events for reliable dynamic traffic noise mapping in real-life urban and suburban soundscapes. *Appl. Sci.* **2017**, *7*, 146. [CrossRef]

36. Orga, F.; Alías, F.; Alsina-Pagès, R.M. On the Impact of Anomalous Noise Events on Road Traffic Noise Mapping in Urban and Suburban Environments. *Int. J. Environ. Res. Public Health* **2017**, *15*, 13. [CrossRef] [PubMed]
37. Aletta, F.; Kang, J.; Axelsson, Ö. Soundscape descriptors and a conceptual framework for developing predictive soundscape models. *Landsc. Urban Plan.* **2016**, *149*, 65–74. [CrossRef]
38. Aletta, F.; Axelsson, Ö.; Kang, J. Dimensions underlying the perceived similarity of acoustic environments. *Front. Psychol.* **2017**, *8*, 1162. [CrossRef] [PubMed]
39. Calculation of Loudness Level and Loudness from the Sound Spectrum—Zwicker Method—Amendment 1: Calculation of the Loudness of Time-Variant Sound. 17.140.01—Acoustic Measurements and Noise Abatement in General. Available online:https://www.iso.org/obp/ui/#!iso:std:63078:en (accessed on 4 April 2021).
40. Sottek, R. Sound quality evaluation of noises with spectro-temporal patterns. In Proceedings of the INTER-NOISE and NOISE-CON Congress and Conference Proceedings, Hong Kong, China, 27–30 August 2017; Volume 255, pp. 3927–3936.
41. *ECMA-74 Measurement of Airborne Noise Emitted by Information Technology and Telecommunications Equipment*, 17th ed.; December 2019. Available online: https://www.ecma-international.org/publications-and-standards/standards/ecma-74/ (accessed on 4 April 2021).
42. Quené, H.; van den Bergh, H. On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Commun.* **2004**, *43*, 103–121. [CrossRef]
43. Volpert-Esmond, H.I.; Page-Gould, E.; Bartholow, B.D. Using multilevel models for the analysis of event-related potentials. *Int. J. Psychophysiol.* **2021**, *162*, 145–156. [CrossRef]
44. Gelman, A. Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. *Technometrics* **2006**, *48*, 432–435. [CrossRef]
45. Frees, E.W.; Kim, J.S. Multilevel Model Prediction. *Psychometrika* **2006**, *71*, 79–104. [CrossRef]
46. Wilkinson, G.N.; Rogers, C.E. Symbolic Description of Factorial Models for Analysis of Variance. *J. R. Stat. Soc. Ser. Appl. Stat.* **1973**, *22*, 392–399. [CrossRef]
47. Harrison, X.A.; Donaldson, L.; Correa-Cano, M.E.; Evans, J.; Fisher, D.N.; Goodwin, C.E.; Robinson, B.S.; Hodgson, D.J.; Inger, R. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* **2018**, *6*, e4794. [CrossRef]
48. Kuznetsova, A.; Brockhoff, P.B.; Christensen, R.H.B. lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw.* **2017**, *82*, 1–26. [CrossRef]
49. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
50. Lüdecke, D. sjPlot: Data Visualization for Statistics in Social Science. R Package Version 2.8.7. 2021. Available online: https://strengejacke.github.io/sjPlot/ (accessed on 20 April 2021).
51. Barton, K. MuMIn: Multi-Model Inference. R Package Version 1.43.17. 2020. Available online: https://r-forge.r-project.org/R/?group_id=346 (accessed on 20 April 2021).
52. Kang, J.; Aletta, F.; Margaritis, E.; Yang, M. A model for implementing soundscape maps in smart cities. *Noise Mapp.* **2018**, *5*, 46–59. [CrossRef]
53. Kang, J.; Aletta, F.; Gjestland, T.T.; Brown, L.A.; Botteldooren, D.; Schulte-Fortkamp, B.; Lercher, P.; van Kamp, I.; Genuit, K.; Fiebig, A.; et al. Ten questions on the soundscapes of the built environment. *Build. Environ.* **2016**, *108*, 284–294. [CrossRef]

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.

Abstract

The unprecedented lockdowns enforced around the world to fight 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 ($N = 1,318$) and binaural recordings ($N = 693$) collected in 13 locations across London and Venice during 2019, new recordings ($N = 608$) were made in the same locations during the lockdowns in 2020. Using these 30-second-long recordings, linear multi-level models were developed to predict the pleasantness and eventfulness of the soundscapes during the lockdown and compare changes for each location. 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 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

The global emergency cause by the Coronavirus disease of 2019 (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 (Ren, 2020). The explorations in environmental acoustics of lockdown conditions across the world have revealed various degrees of impact on the acoustic environment, both at a city-scale (Asensio et al., 2020b; Bonet-Solà et al., 2021; Hornberg et al., 2021; Munoz et al., 2020; Rumpler et al., 2021) and at a more local, public space-scale (Aletta et al., 2020; Alsina-Pagès et al., 2021; Bonet-Solà et al., 2021; 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.

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 (Bartalucci et al., 2021; Lee and Jeong, 2021). Garrido-Cumbrera et al. (2021) relied on an online survey deployed in England, Ireland, and Spain to explore the perceived change in natural environments in particular. They observed a consistent increase in the perceived presence of natural sounds across all major cities and rural areas respectively in these three countries. A very similar trend was observed in Argentina, also based on an online questionnaire without a listening task (Maggi et al., 2021). By combining field recordings and focus groups, Sakagami (2020) and Lenzi et al. (2021) observed changes in the sound source composition and the affective quality of soundscape in a residential area in Kobe, Japan and a public space in Getxa, Spain, respectively, during the different stages of the lockdown period. Following the easing of lockdown measures, a decrease in animal and traffic sounds was observed in Kobe, while an increase in eventfulness, loudness, and presence of human sound sources, followed by a decrease in pleasantness, was shown in Getxa.

Aletta et al. (2020) explored the impacts of the COVID-19 lockdowns on the acoustic environment in London in particular, through many short-term (30s) binaural recordings. 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:

- 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?

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 (International Organization for Standardization, 2014) cit *truax1999* , 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 (EEA, 2020; Aletta and Kang, 2018) cit *Radicchi2021* . This has been further formalised in International Organization for Standardization (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 cit *Kang 2007*. 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 (Kajihara et al., 2017), personal/demographic (Erfanian et al., 2021) cit *tarlao2021* 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), however this information must also be obtained from people via a survey and therefore are unsuitable for predictive modelling where surveys are not possible. 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).

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

The sites were selected to provide a mixture of sizes and uses, varying in typology ranging from paved squares to small and large parks to waterside spaces across both cities. Throughout the text they are indexed via a LocationID based on the location's name (e.g. CamdenTown, SanMarco), while a more in-depth overview of each is given in 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.

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

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. 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. Each of the 13 locations was characterised by between 14 to 80 recordings and between 32 to 155 questionnaire responses. Mean age of the participants was 33.9 (45% male, 53.8% female, 0.4% non-conforming, 0.9% prefer-not-to-say).

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 608 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 (International Organization for Standardization, 2018), as described in (Aletta et al., 2020; 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 (International Organization for Standardization, 2019). In order to distinguish the projected values from the Likert-scale PAQ responses, the projected values will be referred to as ISOPleasant and ISOEventful and can be considered to form an x-y coordinate point (x = ISOPleasant, y = ISOEventful) as explained in detail in Lionello et al. (2021).

Data cleaning

The cleaning of the samples was conducted using the ArtemiS SUITE 11. 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 supplementary files.

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 International Organization for Standardization (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 11 to calculate the following suite of 11 acoustic and psychoacoustic features to be used as initial predictors:

1. Loudness (N_5, sones, per ISO 532-1:2017)
2. Sharpness (S, acum, per ISO 532-1:2017)
3. Roughness (R, asper)
4. Impulsiveness (I)
5. Fluctuation Strength (FS, vacil)
6. Tonality (T, tuHMS)
7. Zwicker Psychoacoustic Annoyance (PA, per Zwicker and Fastl (1999))
8. L_{Aeq}, 30s (dB)
9. $L_{A10} - L_{A90}$ (dB)
10. $L_{Ceq} - L_{Aeq}$ (dB)
11. Relative Approach (RA, per Genuit (1996))

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 Section 7.1. In all, they are expected to reflect the sound level (L_{Aeq}), perceived sound level (N_5), spectral content (S, $L_{Ceq} - L_{Aeq}$, T), temporal character or predictability (I, FS, RA), and overall annoyance (PA). These metrics have been proposed as indicators to predict perceptual constructs of the soundscape (Aletta et al., 2017, 2016) and have shown promise when combined together to form amore 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 suggest in ISO/TS 12913-3:2019 (International Organization for Standardization, 2019).

Table **??** shows the Pearson correlation coefficient between each of the candidate acoustic features and the outcome pleasantness and eventfulness. For ISOPleasant ($ISOPl$), 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 ISOEventful ($ISOEv$), these tiers are:

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

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

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 improvement is seen by removing more features. This process is carried out first on the location-level features (including the potential to remove all features including LocationID, resulting in a 'flat' or standard multivariate linear regression model), then on the individual-level features. The performance criterion used for this process was the Akaike Information Criterion (AIC) (Akaike, Hirotugu, 1974). To check for multicollinearity among the selected features, the Variance Inflation Factor (VIF) was calculated and a threshold of $VIF < 5$ was set. Any features which remained after the backwards step-wise selection and which exceeded this threshold were investigated and removed if they were highly collinear with the other features.

All of the input features are numeric values, in the units described above. Before conducting feature selection, the input features are z-scaled to enable proper comparison of their effect sizes. After the feature selection, the scaled coefficients are used in the text when reporting the final fitted models to facilitate discussion and comparison between the features. The unscaled model coefficients are reporting 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 ISOPleasant and ISOEventful this means our response can range from -1 to $+1$. However, since the end-goal of the model is to predict the soundscape assessment of the location as a whole, rather than the individual responses, 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) (Kuznetsova et al., 2017) in R statistical software (v.4.0.3) (R Core Team, 2021). The summaries and plots were created using the 'sjPlot' package (v.2.8.6) (Lüdecke, 2021) and 'seaborn' (v.0.11.1) (Waskom, 2021).

7.2.3 Online Survey

A online listening test was conducted using the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). 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 (Woods et al., 2017) and a headphone reproduction level adjustment test (Gontier et al., 2019). The level adjustment was performed using an 11-second-long pink noise sample matched to the lowest and the highest L_{A90} values from the experimental set. Participants were asked to adjust their listening level to clearly hear the quieter sample while keeping the level low enough, so they don't find the louder sample disturbing. The headphone screening test followed, featuring a stereo signal of 1-second-long 100 Hz sin tone, generated with Izotope RX6 application, played at a 3 dB difference where one of the equally loud pairs had its phase inverted. A 100 Hz sin was used because the pilot tests revealed that the 200 Hz sin tone proposed by Woods et al. (2017) created a higher uncertainty varying across different laptop models and would likely contribute to the chances of a participant fooling the test. It was expected that participants using speakers would not be able to either hear the sin wave or would be fooled by the inverted phase effect and therefore not able to pass the trials, unless they were indeed using headphones. The participant needed to recognise the quietest of the 3 samples in a trial of 6 attempts. Only participants correctly answering 5 or more out of 6 trials were allowed to proceed with the experiment. Participants were asked not to change their audio output settings during the rest of the experiment. This was introduced to ensure that a participant is using a headphone playback system which allows a listener to clearly recognise a 3 dB difference at 100 Hz as a proxy for sufficient audio quality playback.

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

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 (ISOPleasant and ISOEventful) and the lockdown-related changes.

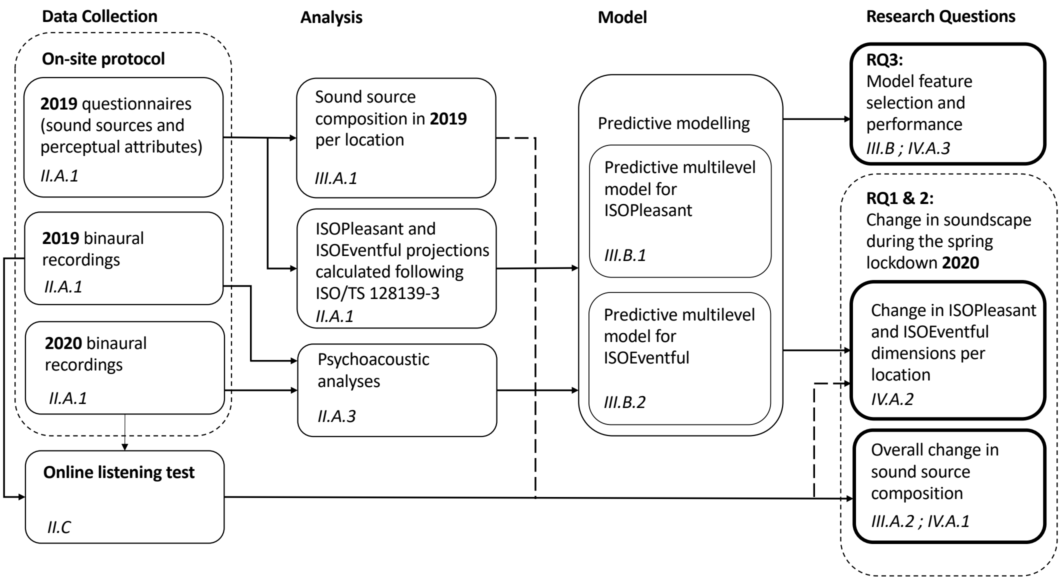


Figure 7.1: The study flowchart indicating the data collection, analysis, modelling, and discussion throughout the study. The subsections in the text to which each box refers are indicated in italics.

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 show 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 frequency 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.

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

Levene's test for equality of variances resulted in highly statistically significant values for all 4 sound sources investigated (< 0.001). Therefore, a Mann-Whitney U-test was used as a non-parametric equivalent to the t-test to investigate the change in the perceived dominance of the four sound source types (McKnight and Najab, 2010). The results for human sounds indicated that the perceived dominance was greater for the 2019 sample ($M = 3.82$) than for the 2020 sample ($M = 2.62, U = 41,656, p < 0.001$). The results for natural sounds indicated the perceived dominance increased from 2019 ($M = 2.00$) to 2020 ($M = 2.54, U = 63,797, p < 0.001$). However, the differences for the noise sources (traffic and other) were not statistically significant.

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

Table 7.1: Scaled linear regression models of ISOPleasant and ISOEventful for 13 locations in London and Venice. The ISOPleasant model is a multi-level regression model with one level for individual effects and a second level for LocationID effects, while the ISOEventful model is a 'flat' multi-variate linear regression with no location effects.

Figure 7.3: Location-level scaled coefficients for the ISOPleasant model.

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 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 are investigated further in the Discussion section below.

Figure 7.4: Soundscape circumplex coordinates for (a) the mean ISOPleasant and ISOEventful responses for each location; and (b) the mean predicted responses based on recordings made during the lockdown and the location's movement in the circumplex.

7.4 Discussion

7.4.1 Interpretation of the results

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

Andrew Mitchell,[1] Tin Oberman,[1] Francesco Aletta,[1] Magdalena Kachlicka,[1] Matteo Lionello,[1] Mercede Erfanian,[1] and Jian Kang[1]

Institute for Environmental Design & Engineering, The Bartlett, University College London, London, United Kingdom[a)]

(Dated: 5 August 2021)

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

The unprecedented lockdowns enforced around the world to fight COVID-19 in spring 2020 triggered changes in human activities in public spaces. A predictive modeling approach was developed to characterize the resulting change in the perception of the sound environment when people could not be surveyed. Building on a database of soundscape questionnaires ($N = 1,318$) and binaural recordings ($N = 693$) collected in 13 locations across London and Venice during 2019, new recordings ($N = 608$) were made in the same locations during the lockdowns in 2020. Using these 30-second-long recordings, linear multi-level models were developed to predict the pleasantness and eventfulness of the soundscapes during the lockdown and compare changes for each location. 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 traffic-dominated locations, but not for human- and natural-dominated locations. This study demonstrates the usefulness of predictive modeling and the importance of considering contextual information when discussing the impact of sound level reductions on the soundscape.

[a] j.kang@ucl.ac.uk;

I. INTRODUCTION

The global emergency caused by the COVID-19 pandemic 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 (Ren, 2020). The explorations in environmental acoustics of lockdown conditions across the world have revealed various degrees of impact on the acoustic environment, both at a city-scale (Asensio *et al.*, 2020b; Bonet-Solà *et al.*, 2021; Hornberg *et al.*, 2021; Munoz *et al.*, 2020; Rumpler *et al.*, 2021) and at a more local, public space-scale (Aletta *et al.*, 2020; Alsina-Pagès *et al.*, 2021; Bonet-Solà *et al.*, 2021; Vida Manzano *et al.*, 2021). In general, these studies have demonstrated a decrease in urban noise levels and indicated a difference in the amount the level decrease depending on the type of space investigated (e.g. parks, urban squares, etc.) and the type of human activity characteristic for the space.

Those studies were mostly focused 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 also because of: 1) the lack of subjective data for the exact or comparable locations in previous years; and 2) the lack of participants present in public spaces during the lockdown, hence the inability to collect soundscape data in situ. Attempts have been made to bridge this gap by using social networks to source subjective data but resulted in a focus on indoor conditions following the shift in the citizens' behavior, i.e. spending more time indoors (Bartalucci *et al.*, 2021; Lee and Jeong, 2021). Garrido-Cumbrera *et al.* (2021) relied on an online survey deployed in England, Ireland and Spain to explore the perceived change in natural environments in particular. They observed a consistent increase in the perceived presence of natural sounds across all major cities and rural areas respectively in these three countries. A very similar trend was observed in Argentina, also based on an online questionnaire without a listening task (Maggi *et al.*, 2021). By combining field recordings and focus groups, Sakagami (2020) and Lenzi *et al.* (2021) observed changes in the sound source composition and the affective quality of soundscape in a residential area in Kobe, Japan and a public space in Getxa, Spain, respectively, during the different stages of the lockdown period. Following the easing of lockdown measures, a decrease in animal and traffic sounds was observed in Kobe, while an increase in eventfulness, loudness, and presence of human sound sources, followed by a decrease in pleasantness, was shown in Getxa.

Aletta *et al.* (2020) explored the impacts of the COVID-19 lockdowns on the acoustic environment in London in particular, through many short-term (30s) binaural recordings. 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 focused approach left the following research questions unanswered: how would people have perceived these spaces as a result of this change in acoustic environment (RQ1), and 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: how was the sound source composition influenced by the change; how would the affective response to the acoustic environment in lockdowns change; and could this demonstrate the effect of human activities on the perception of an acoustic environment in general?

These questions arise out of the soundscape approach, which is characterized by prioritizing 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 recognized as a response to arising environmental requirements on noise pollution and sustainability, such as the regulation of quiet areas in Europe (European Environment Agency, 2020; Kang and Aletta, 2018; Radicchi *et al.*, 2021). This has been further formalized 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 (Kang, 2007). 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 (Kajihara *et al.*, 2017), personal/demographic (Erfanian *et al.*, 2020; Tarlao *et al.*, 2021) 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), however this information must also be obtained from people via a survey and therefore are unsuitable for predictive modeling where surveys are not possible. 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).

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)?

II. 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 Soundscape Indices

(SSID) project (Mitchell *et al.*, 2020) and in 2020 during the strictest part of the lockdowns (Aletta *et al.*, 2020), including objective acoustic data (2019 and 2020) and subjective responses (2019 only). Using both 2019 and 2020 binaural recordings, an online listening experiment was conducted to provide an understanding about the change in sound source composition. The 2019 onsite questionnaire data were used to define the dominant sound source at each location as a starting point for interpreting 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.

The ISO/TS 12913 (ISO/TS 12913-2:2018, 2018) series were consulted for reporting on soundscape data. A detailed description of the 2019 survey campaigns is featured throughout the paper and in the supplementary files. This study was approved by departmental UCL IEDE Ethics Committee on 17$^{\text{th}}$ July 2018 for onsite data collection and on the 2$^{\text{nd}}$ of June 2020 for the on-line listening experiment and is conducted in adherence to the ethical requirements of the Declaration of Helsinki (World Medical Association, 2013).

A. 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 28$^{\text{th}}$ of February and the 21$^{\text{st}}$ of June 2019, with additional sessions in July and October 2019. 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. Each of the 13 locations was characterized by between 14 to 80 recordings and between 32 to 155 questionnaire responses. Mean age of the participants was 33.9 (45% male, 53.8% female, 0.4% non-conforming, 0.9% prefer-not-to-say).

The subsequent measurement campaign in 2020 mimicked the binaural recording strategy applied in the initial campaign and was performed between the 6$^{\text{th}}$ and the 25$^{\text{th}}$ of April 2020 in both cities, this time excluding the questionnaire. An additional 608 binaural recordings were collected on-site in 2020.

1. 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 (2018), as described in (Aletta *et al.*, 2020; 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 perceived affective quality (PAQ).

Participants are first asked to rate the perceived dominance of several sound sources, as assessed via a 5-point Likert scale, coded from 1 (Not at all) to 5 (Dominates completely). The sound sources are split into four categories: Traffic noise, Other noise, Human sounds, and Natural sounds and each is rated separately. Next are the 8 PAQs which make up the circumplex model of soundscape (Axelsson *et al.*, 2010): pleasant, chaotic, vibrant, uneventful, calm, annoying, eventful, and monotonous. These are assessed on a 5-point Likert scale from 1 (Strongly disagree) to 5 (Strongly agree). In order to simplify the results and allow for modeling the responses as continuous values, the 8 PAQs undergo a trigonometric projection to reduce them onto the two primary dimensions of pleasant and eventful, according to the procedure outlined in Part 3 of the ISO 12913 series (ISO/TS 12913-3:2019, 2019). In order to distinguish the projected values from the Likert-scale PAQ responses, the projected values will be referred to as ISOPleasant and ISOEventful and can be considered to form an x-y coordinate point (x = ISOPleasant, y = ISOEventful) as explained in detail in Lionello *et al.* (2021).

The calibrated binaural device SQobold with BHS II by Head Acoustics was used in both campaigns at all the locations by various operators to capture acoustic data, as mentioned in the acknowledgments. Following the established onsite protocol (Mitchell *et al.*, 2020), when participants were stopped in a group and filled in their responses simultaneously, a single binaural recording was used to capture their experience as a group. The purpose behind this sampling strategy was to obtain data from the perspective of a typical user, corresponding to a range of individual experiences available within an urban open space. These recordings are indexed by a GroupID such that the recording for each group is matched up to each of the corresponding respondents and their individual survey responses.

2. Data cleaning

The cleaning of the samples was conducted using the ArtemiS SUITE 11. 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 then processed further, as described in the section B.2. Psychoacoustic analyses and shown in supplementary files.

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

The site characterization 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.

3. Psychoacoustic analyses

The binaural recordings were analyzed in ArtemiS SUITE 11 to calculate the following suite of 11 acoustic and psychoacoustic features to be used as initial predictors:

1. Loudness (N_5, sones, per ISO 532-1:2017)
2. Sharpness (acum, per ISO 532-1:2017)
3. Roughness (asper)
4. Impulsiveness (iu)
5. Fluctuation Strength (vacil)
6. Tonality (tuHMS)
7. Zwicker Psychoacoustic Annoyance (per Zwicker and Fastl (2007))
8. L_{Aeq}, 30s (dB)
9. $L_{A10} - L_{A90}$ (dB)

10. $L_{Ceq} - L_{Aeq}$ (dB)

11. Relative Approach (per Genuit (1996))

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 Section I. In all, they are expected to reflect the sound level (L_{Aeq}), perceived sound level (N_5), spectral content (Sharpness, $L_{Ceq} - L_{Aeq}$, Tonality), temporal character or predictability (Impulsiveness, Fluctuation Strength, Relative Approach), and overall annoyance (Psychoacoustic Annoyance). 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 (2019).

Table 1 shows the Pearson correlation coefficient between each of the candidate acoustic features and the outcome pleasantness and eventfulness. For ISOPleasant (*ISOPl*), we can perhaps see three tiers of correlations: the more highly correlated tier ($|r| > 0.28$) consists of RA, L_{Aeq}, R, N_5, and PA; the low correlation tier consists of $L_{A10} - L_{A90}$, T, and I; while $L_{Ceq} - L_{Aeq}$, I, and S show no correlation. For ISOEventful (*ISOEv*), these tiers are: RA, L_{Aeq}, T, R, and N_5 comprise the most correlated tier ($|r| > 0.30$); $L_{Ceq} - L_{Aeq}$, $L_{A10} - L_{A90}$, FS, and PA show low correlations; I and S show no correlation.

Among the correlations for the psychoacoustic metrics considered for inclusion as input features, we can see several highly correlated features. As expected, PA, L_{Aeq}, and N_5 are

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

Parameter	**ISOPl**	**ISOEv**	**PA**	N_5	**S**	**R**	**I**	**FS**	**T**	L_{Aeq}	$L_{A10} - L_{A90}$	$L_{Ceq} - L_{Aea}$
ISOPleasant												
ISOEventful	-0.24											
PA	-0.28	0.24										
N5	-0.37	0.33	0.94									
S			0.71	0.56								
R	-0.36	0.32	0.63	0.74	0.11							
I			-0.10		-0.37	0.24						
FS	-0.11	0.14	0.37	0.43		0.46	0.55					
T	-0.21	0.30	0.58	0.63	0.12	0.54	0.16	0.52				
L_{Aeq}	-0.34	0.37	0.84	0.93	0.56	0.72	-0.09	0.37	0.57			
$L_{A10} - L_{A90}$	-0.18	0.15	0.21	0.33	-0.20	0.31	0.36	0.44	0.40	0.23		
$L_{Ceq} - L_{Aeq}$		-0.20	-0.49	-0.49	-0.54	-0.31		-0.27	-0.28	-0.61	-0.22	
RA	-0.34	0.31	0.60	0.74	0.18	0.71	0.31	0.63	0.58	0.73	0.23	-0.14

highly correlated, meaning that careful consideration is paid to these features to ensure they do not contribute to multicollinearity in the final model.

B. 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 modeling and analysis approach which considers the differing relationships between the objective acoustic features and the soundscape's perceived affective quality ratings across the various locations and contexts. The individual-level of the models is made up of the acoustic features calculated from the binaural recordings made during each respondent's survey period, while the group-level includes the categorical ‘LocationID‘ variable indicating the location in which the survey was taken, acting as a non-auditory contextual factor.

A separate backwards-step feature selection was performed for each of the outcome models in order to identify the minimal feature set to be used for predicting each outcome. In this feature selection process, an initial model containing all of the candidate features was fit. Each feature was then removed from the model one at a time, then the best-performing model is selected and the procedure continues step-wise until no improvement is seen by removing more features. This process is carried out first on the location-level features (including the potential to remove all features including LocationID, resulting in a ‘flat’ or standard multivariate linear regression model), then on the individual-level features. The performance criterion used for this process was the Akaike Information Criterion (AIC) (Akaike, Hirotugu, 1974). To check for multicollinearity among the selected features, the variance inflation factor (VIF) was calculated and a threshold of VIF < 5 was set. Any features which remained after the backwards stepwise selection and which exceeded this threshold were investigated and removed if they were highly collinear with the other features.

All of the input features are numeric values, in the units described above. Before conducting feature selection, the input features are z-scaled to enable proper comparison of their effect sizes. After the feature selection, the scaled coefficients are used in the text when reporting the final fitted models to facilitate discussion and comparison between the features. The unscaled model coefficients are reported in Appendix 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 was 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 ISOPleasant 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) (Kuznetsova *et al.*, 2017) in R statistical software (v4.0.3) (R Core Team, 2020). The summaries and plots were created using the ‘sjPlot‘ package (v2.8.6) (Lüdecke, 2021) and ‘seaborn‘ (v0.11.1) (Waskom, 2021).

C. Online survey

An online listening test was conducted using the Gorilla Experiment Builder (`www.gorilla.sc`) (Anwyl-Irvine *et al.*, 2020). The participants were exposed to a random selection of 78 binaural recordings (39 from 2019 and 39 from 2020, 6 recordings per each 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 chances of participants complying to 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 characterization 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 the Part 3 of the online experiment is reported in 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 (Woods *et al.*, 2017) and a headphone reproduction level adjustment test (Gontier *et al.*, 2019). The level adjustment was performed using an eleven-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 headphones screening test followed, featuring a stereo signal of one-second-long 100 Hz sine tone, generated with Izotope RX 6 application, played at a 3 dB difference where one of the equally loud pairs had its phase inverted. A 100 Hz sine was used because the pilot tests revealed the 200 Hz sine tone proposed by Woods *et al.* (2017) created a higher uncertainty varying across different laptop models and would likely contribute to the chances of a participant fooling the test. It was expected that participants using speakers would not be able to either hear the sine 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 recognize 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 headphones playback system which allows a listener to clearly recognize a 3 dB difference at 100 Hz as a proxy for sufficient audio quality playback.)

Online questionnaire data was collected between the 9$^{\text{th}}$ of June and the 9$^{\text{th}}$ of August 2020. Within the Gorilla Experiment Builder, a total of 250 attempts to complete the experiment were recorded, where 165 participants were excluded either on the basis of not passing the headphones 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 1 illustrates and summarizes the framework and sections described above.

III. RESULTS

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

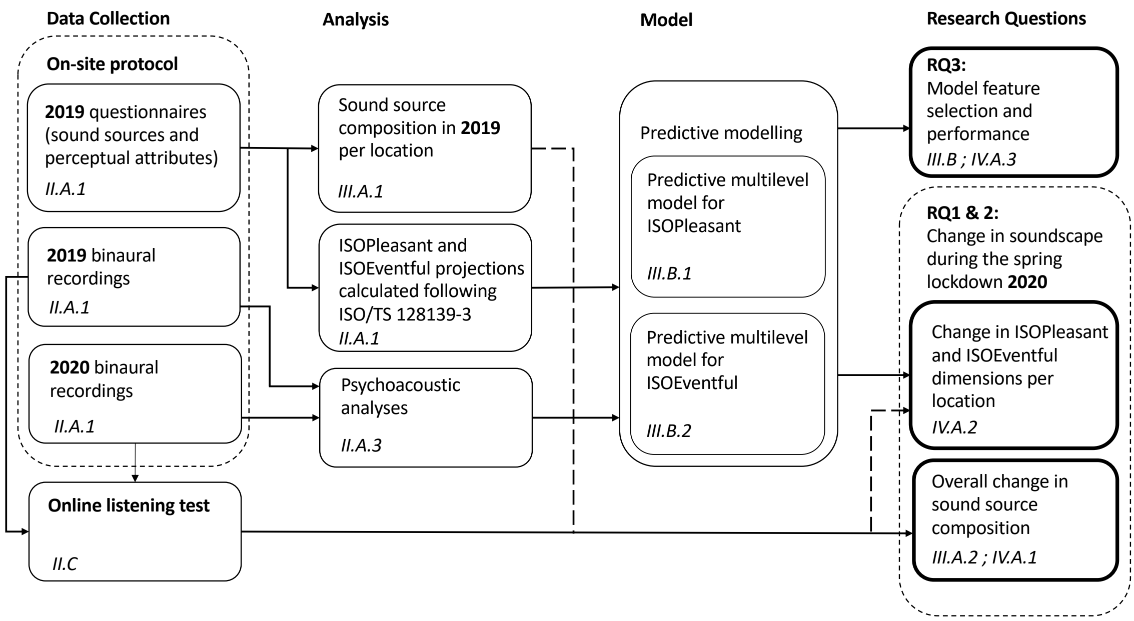


FIG. 1. The study flowchart indicating the data collection, analysis, modeling, and discussion throughout the study. The subsections in the text to which each box refers are indicated in italics.

A. Perceived sound source dominance

1. 2019 sound source composition per location

Questionnaire data was collected in 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 2, the locations can be grouped into: natural sounds dominated (RegentsPark-

Japan, RegentsParkFields, RussellSq), human sounds dominated (SanMarco, TateModern, StPaulsRow, StPaulsCross, MonumentoGaribaldi), noise (traffic and other noise) sounds dominated (CamdenTown, EustonTap, TorringtonSq, PancrasLock).

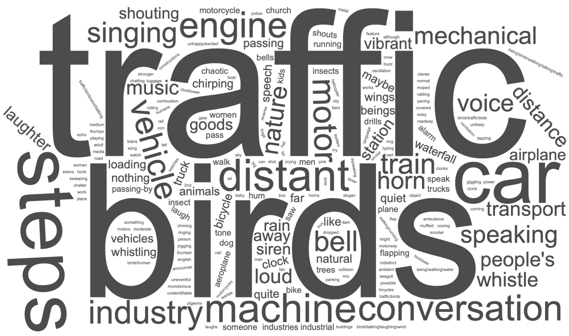
Locations	Traffic	Human	Natural	Other
CamdenTown	3.8	3.3	1.3	2.7
EustonTap	3.7	2.6	1.7	3.0
MarchmontGarden	2.7	2.7	2.6	2.5
MonumentoGaribaldi	1.9	3.4	3.0	2.0
PancrasLock	2.4	2.5	2.4	3.3
RegentsParkFields	2.4	2.9	3.1	1.9
RegentsParkJapan	1.9	2.5	4.0	1.5
RussellSq	2.8	3.0	3.3	2.1
SanMarco	1.4	4.0	2.2	1.9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

FIG. 2. (Color online) Mean values per Location ID for the perceived dominance of the sound source types, for the 2019 on-site campaign.

2. Overall change in the perceived sound source dominance during lockdown

1803 words describing the sound sources present in the 2019 recordings and 1395 words related to the 2020 recordings were input by participants in response to the open-ended question Q1 (see Appendix A). The frequency of occurrence, generated using the Word-Clouds web app, is shown in the Figure 3, 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.

(2019)



(2020)

FIG. 3. A graphic illustrating the frequency of occurrence of the sound sources reported by the participants of the online study across all locations, shown for recordings from the 2019 (above) and 2020 (below).

The results from the listening tests deployed online were analyzed using the SPSS Statistics v. 25. Levene's test for equality of variances resulted in highly statistically significant values for all 4 sound sources investigated (less than 0.001). Therefore, a Mann-Whitney U-test test was used as a non-parametric equivalent to the T-test to investigate the change in the perceived dominance of the four sound source types (McKnight and Najab, 2010). The results for human sounds indicated that the perceived dominance was greater for the 2019 sample (M=3.82), than for the 2020 sample (M=2.62), U=41,656, $p<0.001$. The results for

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

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

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.

B. Model selection, performance, and application

1. *ISOPleasant model selected*

Following the feature selection, the ISOPleasant model (given in Table III) 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 ($R^2_{train} = 0.998$; $R^2_{test} = 0.85$).

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

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

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 4. 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.

FIG. 4. (Color online) Location-level scaled coefficients for the ISOPleasant model.

2. ISOEventful model selected

Through the group-level feature selection, all of the group-level coefficients were removed, including the LocationID factor itself. Therefore the final ISOEventful model is a 'flat' multivariate linear regression model, rather than a multi-level model. The ISOEventful model is a linear combination of S, FS, T, L_{Aeq}, and $L_{Ceq} - L_{Aeq}$. The training and testing MAE are very similar, indicating that the model is not over-fit to the training data ($MAE_{train} = 0.233$; $MAE_{test} = 0.231$). The model performs slightly worse than the ISOPleasant at predicting the mean location responses, but still performs well ($R^2_{train} = 0.873$; $R^2_{test} = 0.715$).

3. 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 5(a) shows the pre–lockdown ISO coordinates for each location and Figure 5(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. San Marco, Tate Modern). Camden Town and Euston Tap, 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 general positive move along the 'pleasant' dimension and general negative move along the 'eventful' dimension, but several different patterns of movement can be noted. These are investigated further in the Discussion section below.

IV. DISCUSSION

A. Interpretation of the results

To interpret the results addressing the 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.

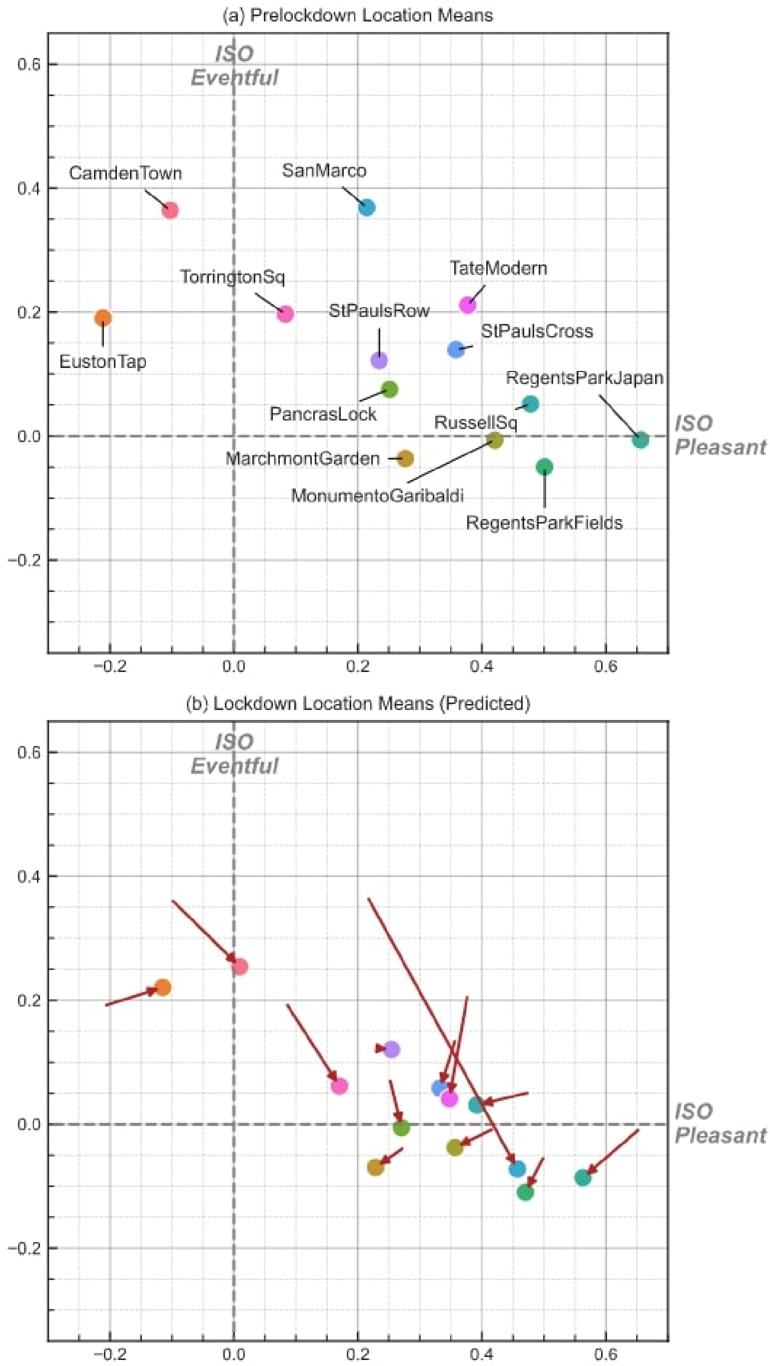


FIG. 5. (Color online) Soundscape circumplex coordinates for (a) the mean ISOPleasant and ISOEventful responses for each location; and (b) the mean predicted responses based on recordings made during the lockdown and the location's movement in the circumplex.

1. 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 (2018) revealed a statistically significant reduction in human sound sources and a significant increase in the perceived dominance of natural sound sources.

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

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

2. *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 6. This clearly shows a few different patterns of movement due to the effects of the 2020 lockdown. These can be further looked into depending on 1) the magnitude of change; 2) the direction of change; 3) shift between the quadrants shown in Figure 5; 4) sound source composition.

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 centers around the square. By contrast, due to the particularly strict measures imposed in Italy, during the lockdown measurement period, the square was almost entirely devoid of people. What is promising is that, 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.

The next locations of interest are those which, in the 2019 survey data, were rated as being dominated by traffic noise: Euston Tap, Camden Town, Torrington Square, and Pancras Lock. These are the only locations (besides San Marco) which show a predicted

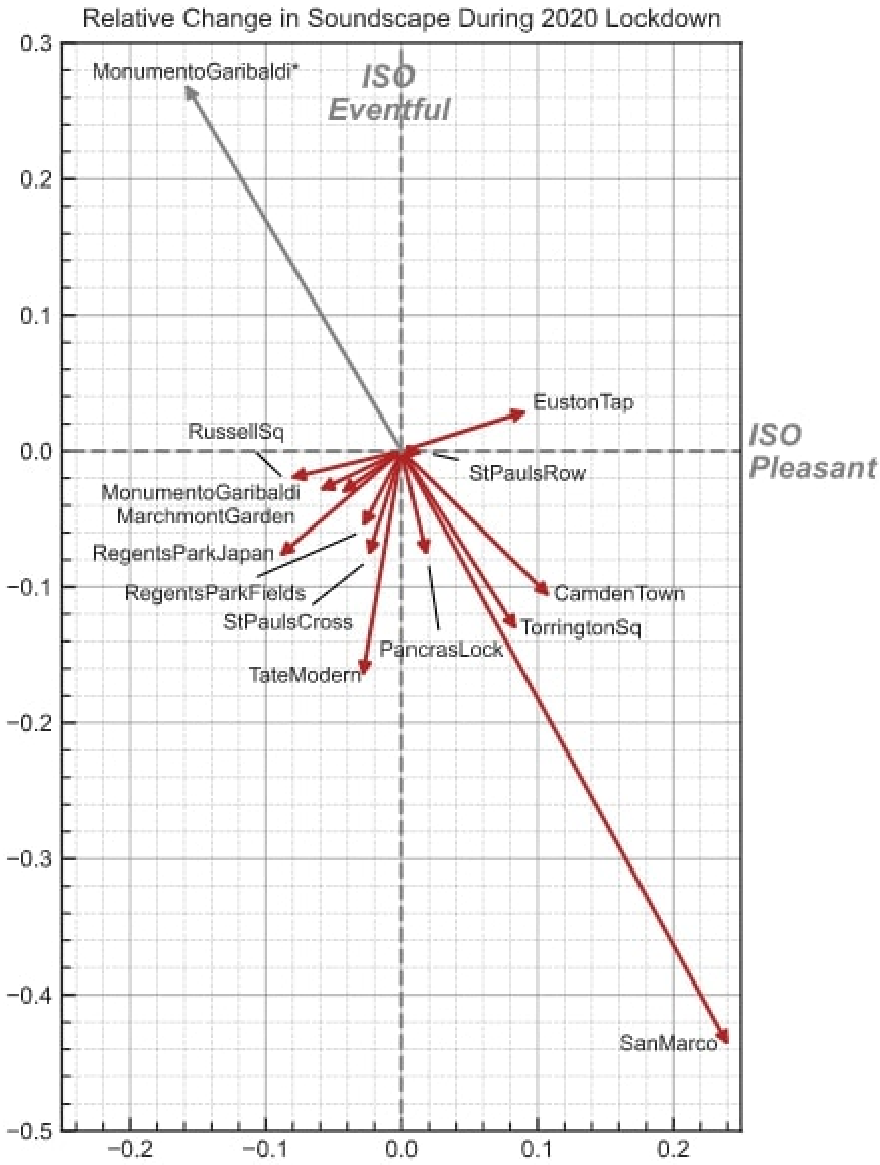


FIG. 6. (Color online) The relative movement of soundscape perception in the circumplex due to the COVID-19 lockdowns, represented as vectors centered on the origin. *The lawn-works dominated session is shown separately as MonumentoGaribaldi* with a gray arrow to indicate that this is distinct from the effects of the lockdown changes.

increase in pleasantness. Of these traffic-dominated spaces, the two which were most heavily dominated by traffic noise (Camden Town and Euston Tap) showed the most increase in pleasantness, with Torrington Square having slightly less of an increase. Pancras Lock, which was also rated as having high levels of both Human and Natural sounds shows only a modest improvement in pleasantness.

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 (St Pauls Cross and Tate Modern) and Natural (Regents Park Japan, Regents Park Fields, Russell Square) 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.

An unexpected result is that Euston Tap 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, Euston Tap is a mostly-outdoor drinking venue located at the entrance to the Euston Train station and situated directly along a very busy central London road. During the 2020 survey, the researchers noted that the music and chatter of people from the pub was noticeably missing, but that the perceived reduction in road traffic was minimal. Based on the theory of vibrancy which would suggest it is driven by human presence and sounds (Aletta 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.

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 6 in gray and labeled MonumentoGaribaldi*, while the unaffected session is shown in red. In the latter case, the predicted change in soundscape as a result of the lockdown fits neatly into what would be expected and closely matches the predicted behavior of similar locations in London (i.e. Marchmont Garden and Russell Square). On the other hand, the session which was dominated by noise from the strimmers is predicted to have become much more chaotic, with a decrease in pleasantness of 0.16 and an increase in eventfulness of 0.27. This indicates that, although the model has no contextual information about the type of sound and in fact the training data never included sounds from similar equipment, just based on the psychoacoustic features of the sound it is able to reasonably predict the expected change in soundscape.

As a whole, the primary impact of the 2020 lockdowns on the soundscapes in London and Venice was an overall decrease in eventfulness. With the exception of Euston Tap, all of the sessions show some degree of reduction in eventfulness, reflecting the general decrease in sound levels and human sound sources across the locations. The impact of the lockdowns on pleasantness is more mixed and seems to be driven by the previous dominance of traffic noise in the space. However, it could also be noted that, while all locations experienced a reduction in sound level, those which are predicted to become more pleasant had an average L_{Aeq} above 60 dB in 2019. By contrast, the locations which were predicted to experience a decrease in pleasantness generally had sound levels below 60 dBA in 2019. This may indicate that reductions in sound level can improve pleasantness when the sound level exceeds some threshold of around 60 - 65 dBA 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 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.

3. *Model selection results*

The most immediately interesting result of the model-building and feature selection process, answering to the 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 (references). For this modeling, 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 modeling 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 focused 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 modeling 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 characterize the location itself, through the aggregated predicted responses of individuals. The authors believe this modeling approach better addresses the practical goal of predictive soundscape modeling and reflects the structure of the data collection.

B. Limitations of the study

The onsite sampling method was initially not intended as the ultimate characterization 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 characterize 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 longterm 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.

V. CONCLUSION

This study demonstrates an application of predictive modeling 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 modeling–focused method of this sort is a key component to the potential scalability of the soundscape approach to applications such as smart city sensing, urban planning, and cost-effective, sustainable design. To demonstrate the usefulness and feasibility of such an approach, 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 these results are limited in that they represent 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.

ACKNOWLEDGMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 740696, project title: Soundscape Indices – SSID). More information and related publications can be found at the CORDIS webpage of the project: https://cordis.europa.eu/project/rcn/211802/factsheet/en.

The authors would like to thank Zhongzhe Li for conducting binaural recordings at Euston Square Gardens and Torrington Square in 2019. The authors would like to thank Meihui Ba, Nicolas Assiotis, Veronica Rugeles Allan, Yu Wang and Hua Su for their help in conducting on-site surveys during the spring and the autumn of 2019.

On-site study data were collected and managed using REDCap electronic data capture tools hosted at University College London (UCL).

APPENDIX A: ONLINE QUESTIONNAIRE

APPENDIX B: MODEL RESULTS

Table V presents the unscaled coefficients for the ISOPleasant and ISOEventful predictive models. The scaled coefficients are presented in the body of the text to facilitate comparisons

TABLE IV. Questionnaire deployed via the Gorilla Experiment Builder

Q1	**While listening, please note any sound sources you can identify in this sound environment:**
Q2	**To what extent have you heard the following four types of sounds?**
	Traffic noise (e.g. cars, buses, trains, airplanes)
	Not at all / A little / Moderately / A lot / Dominates completely
	Other noise (e.g. sirens, construction, industry, loading of goods)
	Not at all / A little / Moderately / A lot / Dominates completely
	Sounds from human beings (e.g. conversation, laughter, children at play, footsteps)
	Not at all / A little / Moderately / A lot / Dominates completely
	Natural sounds (e.g. singing birds, flowing water, wind in vegetation)
	Not at all / A little / Moderately / A lot / Dominates completely

between the various factors. However, we feel it is important to present unscaled coefficients such that these models could be implemented and compared for future work.

TABLE V. Unscaled linear regression models of ISOPleasant and ISOEventful for 13 locations in London and Venice.

	ISOPleasant			**ISOEventful**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.39	0.28 - 0.50	**<0.001**	-0.77	-1.05 - -0.48	**<0.001**
N_5	-0.01	-0.01 - -0.00	**<0.001**			
S				-0.17	-0.23 - -0.12	**<0.001**
FS				-1.36	-2.61 - -0.11	**0.033**
T				0.24	0.08 - 0.39	**0.002**
L_{Aeq}				0.02	0.02 - 0.02	**<0.001**
$L_{Ceq} - L_{Aeq}$				-0.01	-0.02 - 0.00	0.052
Random Effects						
σ^2	0.11					
τ_{00}	$1.01_{LocationID}$					
τ_{11}	$0.00_{LocationID.L_{Aeq}}$					
	$0.00_{LocationID.L_{A10}-L_{A90}}$					
	$0.00_{LocationID.L_{Ceq}-L_{Aeq}}$					
ICC	0.90					
N	$13_{LocationID}$					
Observations	914			914		

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, https://royalsocietypublishing.org/doi/10.1098/rsos.150690, doi: 10.1098/rsos.150690 number: 3.

Akaike, Hirotugu (**1974**). "A New Look at the Statistical Model Identification," IEEE Transactions on Automatic Control **19**(6), 716–723 number: 6.

Aletta, F., Axelsson, Ö., and Kang, J. (**2017**). "Dimensions underlying the perceived similarity of acoustic environments," Frontiers in Psychology **8**(JUL), 1–11, doi: 10.3389/fpsyg.2017.01162.

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, http://www.mdpi.com/1660-4601/15/8/1712, doi: 10.3390/ijerph15081712 number: 8.

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, doi: 10.1016/j.landurbplan.2016.02.001.

Aletta, F., Oberman, T., Mitchell, A., Tong, H., and Kang, J. (**2020**). "Assessing the changing urban sound environment during the COVID-19 lockdown period using short-term acoustic measurements," Noise Mapping **7**(1), 123–134, https://www.degruyter.com/view/journals/noise/7/1/article-p123.xml, doi: 10.1515/noise-2020-0011 number: 1.

Alsina-Pagès, R. M., Bergadà, P., and Martínez-Suquía, C. (**2021**). “Changes in the soundscape of Girona during the COVID lockdown,” The Journal of the Acoustical Society of America **149**(5), 3416–3423, https://asa.scitation.org/doi/10.1121/10.0004986, doi: 10.1121/10.0004986 number: 5.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J. K. (**2020**). “Gorilla in our midst: An online behavioral experiment builder,” Behavior research methods **52**(1), 388–407.

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. (**2020**a). “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, https://www.mdpi.com/1660-4601/17/12/4205, doi: 10.3390/ijerph17124205 number: 12.

Asensio, C., Pavón, I., and de Arcas, G. (**2020**b). “Changes in noise levels in the city of madrid during covid-19 lockdown in 202,” The Journal of the Acoustical Society of America **148**, 1748–1755, https://asa.scitation.org/doi/10.1121/10.0002008, doi: 10.1121/10.0002008.

Axelsson, O., 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, http://asa.scitation.org/doi/10.1121/1.3493436, doi: 10.1121/1.3493436 number: 5.

Bartalucci, C., Bellomini, R., Luzzi, S., Pulella, P., and Torelli, G. (**2021**). “A survey on the soundscape perception before and during the COVID-19 pandemic in

Italy," Noise Mapping **8**(1), 65–88, https://www.degruyter.com/document/doi/10.1515/noise-2021-0005/html, doi: 10.1515/noise-2021-0005 number: 1.

Bonet-Solà, D., Martínez-Suquía, C., Alsina-Pagès, R. M., and Bergadà, P. (**2021**). "The Soundscape of the COVID-19 Lockdown: Barcelona Noise Monitoring Network Case Study," International Journal of Environmental Research and Public Health **18**(11), 5799, https://www.mdpi.com/1660-4601/18/11/5799, doi: 10.3390/ijerph18115799 number: 11.

Erfanian, M., Mitchell, A., Aletta, F., and Kang, J. (**2020**). "Psychological Well-being, Age and Gender can Mediate Soundscapes Pleasantness and Eventfulness: A large sample study," preprint , http://biorxiv.org/lookup/doi/10.1101/2020.10.16.341834, doi: 10.1101/2020.10.16.341834.

European Environment Agency (**2020**). *Environmental noise in Europe, 2020.* (EEA, Publications Office, LU), https://data.europa.eu/doi/10.2800/686249.

Garrido-Cumbrera, M., Foley, R., Braçe, O., Correa-Fernández, J., López-Lara, E., Guzman, V., Gonzalez Marín, A., and Hewlett, D. (**2021**). "Perceptions of change in the natural environment produced by the first wave of the covid-19 pandemic across three european countries. results from the greencovid study," Urban Forestry & Urban Greening https://linkinghub.elsevier.com/retrieve/pii/S1618866721002879, doi: 10.1016/j.ufug.2021.127260.

Genuit, K. (**1996**). "Objective evaluation of acoustic quality based on a relative approach," in *Proceedings-Institute of Acoustics*, Vol. 18, pp. 3233–3238, iSBN: 0309-8117.

Gontier, F., Lavandier, C., Aumond, P., Lagrange, M., and Petiot, J.-F. (**2019**). "Estimation of the Perceived Time of Presence of Sources in Urban Acoustic Environments Using Deep Learning Techniques," Acta Acustica united with Acustica **105**(6), 1053–1066, https://www.ingentaconnect.com/content/10.3813/AAA.919384, doi: 10.3813/AAA.919384 number: 6.

Gordo, O., Brotons, L., Herrando, S., and Gargallo, G. (**2021**). "Rapid behavioural response of urban birds to COVID-19 lockdown," Proceedings of the Royal Society B: Biological Sciences **288**(1946), 20202513, https://royalsocietypublishing.org/doi/10.1098/rspb.2020.2513, doi: 10.1098/rspb.2020.2513 number: 1946.

Hadjidemetriou, G. M., Sasidharan, M., Kouyialis, G., and Parlikad, 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, https://linkinghub.elsevier.com/retrieve/pii/S2590198220300786, doi: 10.1016/j.trip.2020.100167.

Hornberg, J., Haselhoff, T., Lawrence, B. T., Fischer, J. L., Ahmed, S., Gruehn, D., and Moebus, S. (**2021**). "Impact of the covid-19 lockdown measures on noise levels in urban areas-a pre/during comparison of long - term sound pressure measurements in the ruhr area, germany," International Journal of Environmental Research and Public Health **18**, 4653, https://www.mdpi.com/1660-4601/18/9/4653, doi: 10.3390/ijerph18094653.

ISO 12913-1:2014 (**2014**). "Acoustics – Soundscape – Part 1: Definition and conceptual framework" International Organization for Standardization, Geneva, Switzerland, 2014.

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.

Kajihara, Y., Dozono, S., and Tokui, N. (**2017**). "Imaginary Soundscape: Cross-Modal Approach to Generate Pseudo Sound Environments," in *31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA, https://nips2017creativity.github.io/doc/Imaginary%5FSoundscape.pdf.

Kang, J. (**2007**). *Urban sound environment* (Taylor & Francis, London ; New York), oCLC: ocm65301511.

Kang, J., and Aletta, F. (**2018**). "The Impact and Outreach of Soundscape Research," Environments **5**(5), 58, http://www.mdpi.com/2076-3298/5/5/58, doi: 10.3390/environments5050058.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (**2017**). "lmerTest package: Tests in linear mixed effects models," Journal of Statistical Software **82**(13), 1–26, doi: 10.18637/jss.v082.i13.

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, https://linkinghub.elsevier.com/retrieve/pii/S2210670721000603, doi: 10.1016/j.scs.2021.102768.

Lenzi, S., Sádaba, J., and Lindborg, P. (**2021**). "Soundscape in Times of Change: Case Study of a City Neighbourhood During the COVID-19 Lockdown," Frontiers in Psychology **12**, 570741, https://www.frontiersin.org/articles/10.3389/fpsyg.2021.570741/full, doi: 10.3389/fpsyg.2021.570741.

Lionello, M., Aletta, F., and Kang, J. (**2020**). "A systematic review of prediction models for the experience of urban soundscapes," Applied Acoustics **170**, 107479, https://linkinghub.elsevier.com/retrieve/pii/S0003682X20305831, doi: 10.1016/j.apacoust.2020.107479.

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**, 602831, https://www.frontiersin.org/articles/10.3389/fpsyg.2020.602831/full, doi: 10.3389/fpsyg.2020.602831.

Lüdecke, D. (**2021**). *sjPlot: Data Visualization for Statistics in Social Science*, https://CRAN.R-project.org/package=sjPlot, r package version 2.8.9.

Maggi, A. L., Muratore, J., Gaetán, S., Zalazar-Jaime, M. F., Evin, D., Pérez Villalobo, J., and Hinalaf, M. (**2021**). "Perception of the acoustic environment during COVID-19 lockdown in Argentina," The Journal of the Acoustical Society of America **149**(6), 3902–3909, https://asa.scitation.org/doi/10.1121/10.0005131, doi: 10.1121/10.0005131 number: 6.

McKnight, P. E., and Najab, J. (**2010**). "Mann-whitney u test," in *The Corsini Encyclopedia of Psychology*, edited by I. B. Weiner and W. E. Craighead (John Wiley & Sons, Inc.), p. corpsy0524, http://doi.wiley.com/10.1002/9780470479216.corpsy0524, doi: 10.

1002/9780470479216.corpsy0524.

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, https://www.mdpi.com/2076-3417/10/7/2397, doi: 10.3390/app10072397 number: 7.

Munoz, P., Vincent, B., Domergue, C., Gissinger, V., Guillot, S., Halbwachs, Y., and Janillon, V. (**2020**). "Lockdown during COVID-19 pandemic: impact on road traffic noise and on the perception of sound environment in France," Noise Mapping **7**(1), 287–302, https://www.degruyter.com/document/doi/10.1515/noise-2020-0024/html, doi: 10.1515/noise-2020-0024 number: 1.

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), https://www.mdpi.com/2071-1050/13/11/5779, doi: 10.3390/su13115779.

R Core Team (**2020**). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.

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, https://www.tandfonline.com/doi/full/10.1080/23748834.2020.1821980, doi: 10.1080/23748834.2020.1821980.

Ren, X. (**2020**). "Pandemic and lockdown: a territorial approach to COVID-19 in China, Italy and the United States," Eurasian Geography and Economics **61**(4-5), 423–

434, https://www.tandfonline.com/doi/full/10.1080/15387216.2020.1762103, doi: 10.1080/15387216.2020.1762103 number: 4-5.

Rumpler, R., Venkataraman, S., and Göransson, P. (**2021**). “Noise measurements as a proxy to evaluating the response to recommendations in times of crisis: An update analysis of the transition to the second wave of the CoViD-19 pandemic in Central Stockholm, Sweden,” The Journal of the Acoustical Society of America **149**(3), 1838–1842, https://asa.scitation.org/doi/10.1121/10.0003778, doi: 10.1121/10.0003778 number: 3.

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 **2**, https://ucl.scienceopen.com/document?vid=0dccbf42-eac3-4bda-b3c7-acf6b5dc62ed, doi: 10.14324/111.444/ucloe.000009.

Tarlao, C., Steffens, J., and Guastavino, C. (**2021**). “Investigating contextual influences on urban soundscape evaluations with structural equation modeling,” Building and Environment **188**, 107490, https://linkinghub.elsevier.com/retrieve/pii/S036013232030857X, doi: 10.1016/j.buildenv.2020.107490.

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, https://linkinghub.elsevier.com/retrieve/pii/S0360132320305266, doi: 10.1016/j.buildenv.2020.107152.

Truax, B. (**1999**). *Handbook for Acoustic Ecology* (Cambridge Street Publishing, Cambridge, MA), https://www.sfu.ca/sonic-studio-webdav/handbook/index.html.

Vida Manzano, J., Pastor, J. A. A., Quesada, R. G., 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, https://www.degruyter.com/document/doi/10.1515/noise-2021-0002/html, doi: 10.1515/noise-2021-0002 number: 1.

Waskom, M. L. (**2021**). "seaborn: statistical data visualization," Journal of Open Source Software **6**(60), 3021, https://doi.org/10.21105/joss.03021, doi: 10.21105/joss.03021.

Woods, K. J. P., Siegel, M. H., Traer, J., and McDermott, J. H. (**2017**). "Headphone screening to facilitate web-based auditory experiments," Attention, Perception, & Psychophysics **79**(7), 2064–2072, http://link.springer.com/10.3758/s13414-017-1361-2, doi: 10.3758/s13414-017-1361-2 number: 7.

World Medical Association (**2013**). "World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects," JAMA **310**(20), 2191–2194, doi: 10.1001/jama.2013.281053.

Yang, W., and Kang, J. (**2005**). "Acoustic comfort evaluation in urban open public spaces," Applied Acoustics **66**(2), 211–229, doi: 10.1016/j.apacoust.2004.07.011.

Zwicker, E., and Fastl, H. (**2007**). *Psychoacoustics: facts and models*, third ed. ed. (Springer, Berlin ; New York), https://link.springer.com/book/10.1007/978-3-540-68888-4.

8 Study IV: A Temporal Convolutional Neural Network for Multi-label Sound Recognition and Annoyance Detection of Complex Soundscapes

Background and aims

Positivity (or the absence of negativity?)

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.

Result and conclusion

A Temporal Convolutional Neural Network for Multi-label Sound Recognition and Annoyance Detection of Complex Soundscapes

Andrew Mitchell,[1,2,a)] Christopher Soelitsyo,[3] Mercede Erfanian,[1] Tin Oberman,[1] Jing-Hao Xue,[4,2] Jian Kang,[1] and Francesco Aletta[1]
[1] *Institute for Environmental Design & Engineering, University College London, UK*
[2] *The Alan Turing Institute, London, UK*
[3] *Charras Lab, Institute for the Physics of Living Systems, University College London, UK*
[4] *Department of Statistical Science, University College London, UK*

(Dated: 20 June 2021)

Increasing urban noise pollution and simultaneous improvements in smart city sensor technology and deployment have created a necessity for increasingly sophisticated approaches to automated noise recognition. Sensor networks which are focused primarily on sound level monitoring have proved to be insufficient to adequately identify harmful sound events or to reflect the human impact of noise in cities. Therefore, . . .

 [http://dx.doi.org(DOI number)]

[XYZ] Pages: 1–2

I. INTRODUCTION

Increasing urban noise pollution and simultaneous improvements in smart city sensor technology and deployment have created a necessity for increasingly sophisticated approaches to automated noise recognition. Sensor networks which are focused primarily on sound level monitoring have proved to be insufficient to adequately identify harmful sound events or to reflect the human impact of noise in cities. Therefore, the development of automated environmental sound recognition (ESR) systems has become a necessary component of next-generation approaches to noise pollution mitigation.

A. Importance of sound source and annoyance detection

B. AI for sound source recognition

1. Previous approaches

C. DCASE Challenge

D. SONYC

1. Datasets

(Cartwright *et al.*)

2. Component Parts

E. Empirical Models of Annoyance

1. Zwicker Psychoacoustic Annoyance

The field of psychoacoustics has had a particular focus on annoyance modelling, however this field presents some typical limitations. Firstly, from its inception it has made use of simple, simulated sounds for conducting laboratory tests. These are useful in that they enable much more control over the acoustic characteristics of the sound, allowing for isolated testing of the independent variables, in a conventional experimental approach. This is a limiting approach also taken in the field of auditory neuroscience ct *Need to add citations for this* . Second, the field is primarily developed towards and focussed on annoyance modelling of single sound sources, typically commercial products such as vacuum cleaners and high-end cars ct .

2. Soundscape Models

F. AI Models of Annoyance

II. METHODS

A. Temporal convolutional neural network

III. EXPERIMENTS

A. Datasets

1. DeLTA / SSID Binaural Dataset

Recording splitting In order to increase the available dataset and to make all of the recordings a consistent length, the original recordings were split into 15 seconds chunks. For all recordings, as many complete 15s chunks

a) andrew.mitchell.18@ucl.ac.uk

	dBFS	max_dBFS
count	2891	289
mean	-36.33	-18.92
std	7.45	7.68
min	-63.86	-46.34
25%	-40.04	-23.23
50%	-35.49	-18.40
75%	-31.42	-13.69
max	-15.35	-0.66

as possible were extracted and the remaining portion was excluded; for instance, for a 34s original recording, two sequential 15s chunks are extracted from the beginning, and the remaining 4s are not used. The original dataset of 1,453 recordings then results in 2,921 15s mp3s.

Gain Boost Due to the limitations of the means of delivery of the stimuli and to ensure the sounds did not exceed a safe level, we excluded the top 30 most loudest recordings as outliers. This was done by calculating the peak volume of the recording and excluding the top 1% (> -8.64 dBFS) loudest recordings. The peak value was used to ensure no recordings would clip. We then added a gain boost of 8 dB to all recordings, enabling us to include 250 very soft acoustic environments featuring little or no specific sound sources. This results in a total dataset of 2,891 recordings, with the relative volumes given in Table III A 1. The audio processing was done in Python, using pydub (Robert *et al.*, 2018).

2. DCASE 2018 / SONYC

B. Model architecture

C. Training and Evaluation

IV. DISCUSSION

"Cumulative annoyance due to compounding acute annoyance events."

V. CONCLUSION

And in conclusion. . .

ACKNOWLEDGMENTS

This research was supported by ...

Cartwright, M., Mendez, A. E. M., Dove, G., Cramer, J., Lostanlen, V., Wu, H.-H., Salamon, J., Nov, O., and Bello, J. P. "SONYC Urban Sound Tagging (SONYC-UST): a multi-label dataset from an urban acoustic sensor network" http://markcartwright.com/files/cartwright2019sonycust.pdf, doi: 10.5281/ZENODO.3338310.

Robert, J., Webbie, M. *et al.* (**2018**). "Pydub" http://pydub.com/.

9 Commentary: From Deterministic to Probabilistic Soundscapes: A critical tour around the soundscape circumplex

Some summary text here.

From Deterministic to Probabilistic Soundscapes: A critical tour around the soundscape circumplex

Andrew Mitchell,[1, a)] Francesco Aletta,[1, b] Zaid Chalabi,[1, c] and Jian Kang[1, d]
University College London, London, UK

(Dated: 20 June 2021)

The circumplex model of soundscape, as originally defined by Axelsson et al., is commonly understood to be ...
 [http://dx.doi.org(DOI number)]

[XYZ] Pages: 1–6

I. 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 and R.M. Schafer, 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 standardized tools to gather individual responses on the perception of urban acoustic environments are indeed desirable, to provide comparable datasets and soundscape characterizations across different locations and times and samples of people. This was actually one of the main drivers for the establishment of a Working Group at the International Organization for Standardization (ISO) back in 2008, which was named "Perceptual assessment of soundscape quality" (ISO/TC 43/SC 1/WG 54) that has so far published three documents within the ISO 12913 series on soundscape. Part 1 provides a general framework and definitions (ISO, 2014), while Part 2 and Part 3 offer guidance on how data should be collected and analyzed, accordingly (ISO, 2018, 2019). Different methods are proposed for data collection in Part 2 (ISO, 2018), but in the context of this study we focus on Method A, because it is the only one underpinned by a theoretical relationship among the items of the questionnaire that compose it, the circumplex model of soundscape (Axelsson *et al.*, 2010). This is in turn based on the Swedish Soundscape Quality Protocol (SSQP), originally developed at Stockholm University (Axelsson *et al.*, 2012).

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). However, in this work we question whether there are some issues related to the data collection instruments and data analysis methods per se.

A. Objectives

Several consequences of the current ISO standard implementation of the soundscape circumplex model are identified and discussed, in particular that of the coordinate transformation process given in Equations A.1 and A.2 of ISO 12913 Part 3 (ISO, 2019). These consequences arise, not out of any particular real-world implementation or data collection, but instead are strictly the result of the model framework and mathematical transformations laid out in the standard. We believe that the results presented here have not been fully discussed previously and may contradict much of the general understanding of the circumplex model within the field.

Once the existing consequences of the standard are identified and discussed, we then present two proposed treatments of the circumplex framework which may bring the model more closely in line with the current understanding within the field.

[a]Also at: The Alan Turing Institute, Headquartered at the British Library, London, UK;)andrew.mitchell.18@ucl.ac.uk
[b)]f.aletta@ucl.ac.uk
[c)]z.chalabi@ucl.ac.uk
[d)]j.kang@ucl.ac.uk

II. THE CURRENT ISO STANDARD

The core of the questionnaire-based soundscape assessment in ISO 12913 Part 2 (ISO, 2018) are 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. In Axelsson *et al.* (2010), a third primary dimension, *Familiarity* is also found, however this only accounted for 8% of the variance and is typically disregarded as part of the standard circumplex.

A. Coordinate transformation

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, 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 analyzable 2 dimensions. The ISO coordinates are thus calculated by:

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

$$\begin{aligned} ISOEventful = [&(eventful - uneventful) \\ &+ \cos 45° * (chaotic - calm) \\ &+ \cos 45° * (vibrant - monotonous)] \\ &* 1/(4 + \sqrt{32}) \end{aligned} \tag{2}$$

where the PAs are arranged around the circumplex as shown in Figure 1. The $\cos 45°$ term operates to project the diagonal terms down onto the x and y axes, and the $1/(4+\sqrt{32})$ scales the resulting coordinates to the range (-1, 1). The result of this transformation is demonstrated in Figure 1.

III. ASSUMPTIONS AND IMPLICATIONS OF THE ISO

A. 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.

We then apply the ISO transformations given in Equations 1 and 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 2, 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.

B. Circular space discussion

Visualisations of the circumplex model in soundscape tend to present it as circumscribing a circle (see Fig XX in (Axelsson *et al.*, 2010) and Fig XX in (**?**)), and this shape is further emphasised by the initial figure in **?**'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 circumplex. In **?**, 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 **?**, 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. XX of (**?**), 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

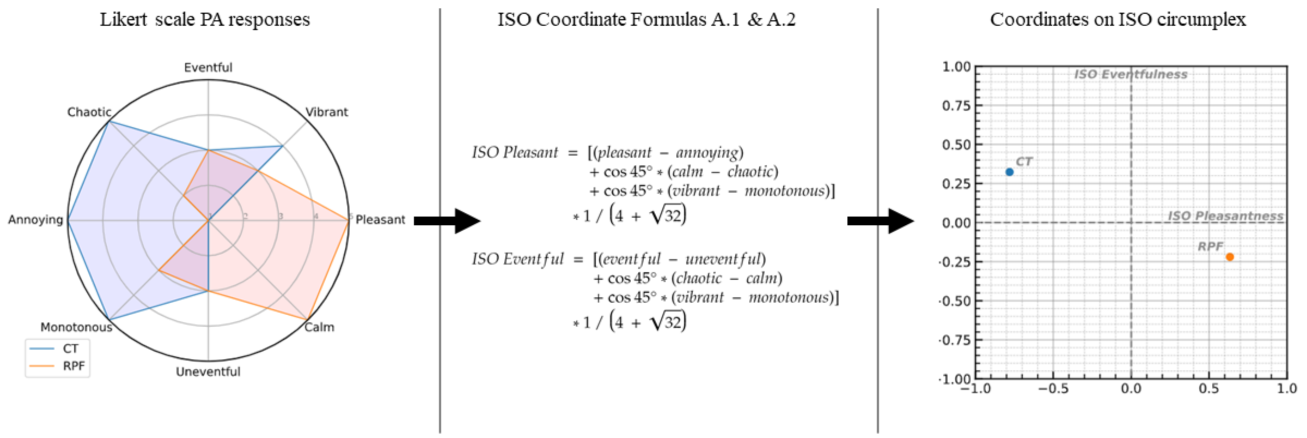


FIG. 1. Placeholder radar plot and projection. Need to remove the middle set of equations, since they are being included in the text

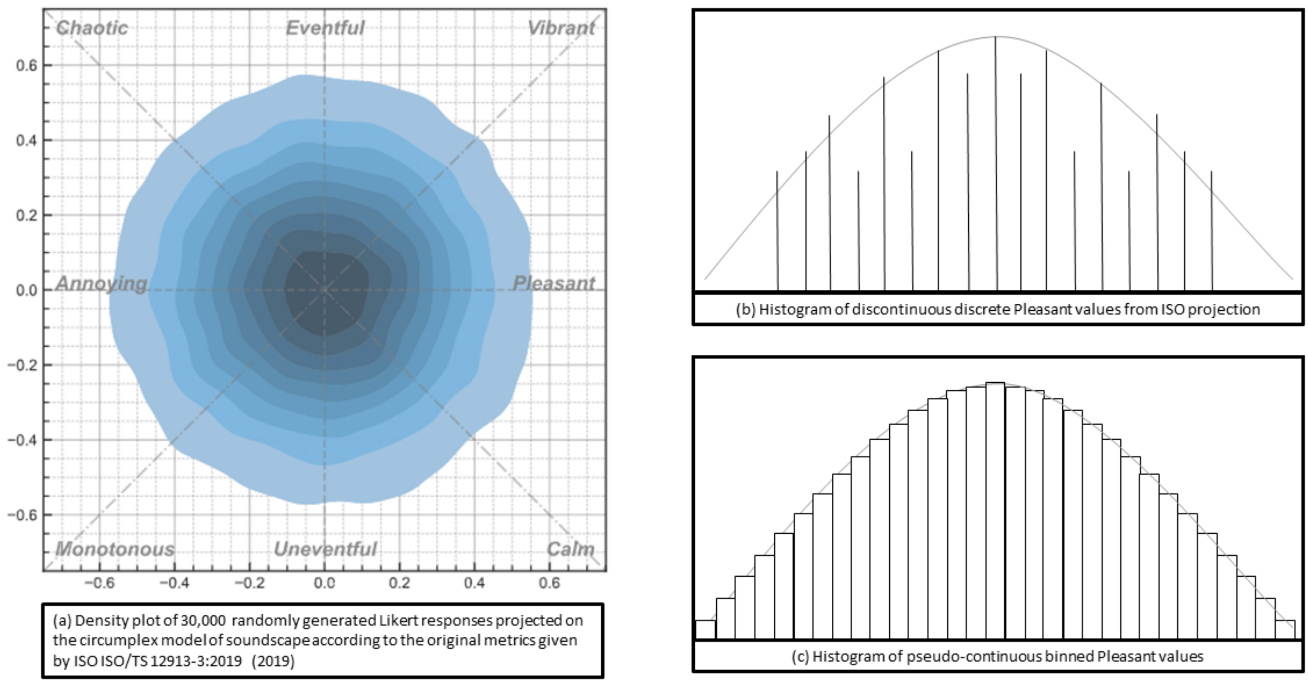


FIG. 2. Mockup placeholder simulation heatmap plot from Lionello 2020 and discrete vs binned transformation values. For new one, need to include distribution histograms along each axis.

dimensions of affect (pleasure and arousal) (?), they use a square grid. This is all to say, 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, 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 fi-

nally 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 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 $\cos 45°$ 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.*, 2016, 2019) and those which have made use of the Part 3 transformation to 2-dimensional coordinates (Lionello *et al.*, 2021; Mancini *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.

C. Normal distribution discussion

We can also see from the histograms included along the axes of Figure 2 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. When looking at the distribution heatmap in Figure 2, it is useful to picture the gradients as representing the available space in the circumplex model.

Need to add more here? Different transition?

Probability Density Function 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 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 Pleasantness are (-1, +1), however according to the PDF calculated above, the 3σ value is $\pm$0.XX. This means that only 0.3% of values fall outside the range (-0.XX, +0.XX). 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.

Non-linearity of movement around the space We can further use this as a demonstration of how we might conceive of a soundscape moving within the available space. * ease of getting to a certain area

* clustering near neutral

D. 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 2(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.

IV. PROPOSED SOLUTIONS

A. Probabilistic Distribution Thinking

The instruments described in the ISO 12913 Part 2 ISO (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 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 characterizing the overall soundscape of public spaces (Mitchell *et al.*, 2020). 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. Boiling 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 completely dismisses the reality of the space. Likewise, this overall soundscape of a public space cannot possibly be determined through a 10-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).

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.

B. Proposal for CDF projections

The CDF of the simulation of the ISO projections is thus:

$$\Phi(x) = \int_{-\infty}^{x} \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

V. DISCUSSION / CONCLUSIONS

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 standardization 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.*, 2019; Jo *et al.*, 2020).

ACKNOWLEDGMENTS

This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program (grant agreement No. 740696, project title: Soundscape Indices - SSID).

We would like to acknowledge Matteo Lionello for the helpful discussions, and (in alphabetical order) Mercede Erfanian, Magdalena Kachlicka, and Tin Oberman for the helpful discussions and the data collection support.

Aletta, F., Guattari, C., Evangelisti, L., Asdrubali, F., Oberman, T., and Kang, J. (**2019**). "Exploring the compatibility of method a and method b data collection protocols reported in the iso/ts 12913-2: 2018 for urban soundscape via a soundwalk," Applied Acoustics **155**, 190–203.

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, doi: 10.1016/j.landurbplan.2016.02.001.

Axelsson, Ã., Nilsson, M. E., and Berglund, B. (**2012**). "The swedish soundscape-quality protocol," The Journal of the Acoustical Society of America **131**(4), 3476–3476.

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, doi: 10.1121/1.3493436.

Axelsson, ., Guastavino, C., and Payne, S. R. (**2019**). "Editorial: Soundscape assessment," Frontiers in Psychology **10**, 2514, https://www.frontiersin.org/article/10.3389/fpsyg.2019.02514, doi: 10.3389/fpsyg.2019.02514.

ISO (**2014**). "ISO 12913-1:2014 Acoustics - Soundscape - Part 1: Definition and conceptual framework" .

ISO (**2018**). "ISO/TS 12913-2:2018 Acoustics - Soundscape - Part 2: Data collection and reporting requirements" .

ISO (**2019**). "ISO/TS 12913-3:2019 Acoustics - Soundscape - Part 3: Data analysis" https://bsol.bsigroup.com/Bibliographic/BibliographicInfoData/000000000030386393.

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 and NOISE-CON Congress and Conference Proceedings*, Institute of Noise Control Engineering, Vol. 261, pp. 3509–3518.

Kang, J., Aletta, F., Gjestland, T. T., Brown, L. A., Botteldooren, D., Schulte-Fortkamp, B., Lercher, P., van Kamp, I., Genuit, K., Fiebig, A. *et al.* (**2016**). "Ten questions on the soundscapes of the built environment," Building and environment **108**, 284–294.

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, https://www.frontiersin.org/article/10.3389/fpsyg.2020.602831, doi: 10.3389/fpsyg.2020.602831.

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), 841.

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, https://doi.org/10.1515/noise-2021-0011, doi: 10.1515/noise-2021-0011.

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 SurveysQuestionnaires with Acoustical and Contextual Information," Applied Sciences **10**(7), 2397, doi: 10.3390/app10072397.

Steele, D., Bild, E., and Guastavino, C. (**2016**). "Evaluation of an urban soundscape intervention with music: quantitative results from questionnaires," in *Inter-Noise and Noise-Con Congress and Conference Proceedings*, Institute of Noise Control Engineering, Vol. 253, pp. 4627–4637.

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.

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 **189**, 139–155.

10 Conclusions

10.1 General Discussion

10.2 Developing a framework for sustainable soundscapes

10.2.1 Defining Sustainability

10.3 Implications

10.4 Limitations and Recommendations for Future Research

10.5 Concluding Remarks

Glossary

AIC Akaike Information Criterion. 65

AMB Ambisonic recording. 20, 22, 24, 26

BIN Binaural. 20, 24, 28, 29

COVID-19 Coronavirus disease of 2019. 60, 62

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. 24, 27

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.. 18, 20, 23

FS Fluctuation Strength. 64, 65, 70

I Impulsiveness. 64, 65

ISOEventful The value along the primary eventfulness dimension of the soundscape circumplex, calculated via a trigonometric projection of the other PAQs, as defined in International Organization for Standardization (2019). 63, 65–67, 70

ISOPleasant The value along the primary pleasantness dimension of the soundscape circumplex, calculated via a trigonometric projection of the other PAQs, as defined in International Organization for Standardization (2019). 63, 65–67, 69, 70

$L_{A10} - L_{A90}$. 64, 65, 69

L_{A90} . 67

L_{Aeq} . 60, 64, 65, 69, 70

$L_{Ceq} - L_{Aeq}$. 64, 65, 69, 70

MAE Mean Absolute Error. 66, 69, 70

N_5 Psychoacoustic Loudness. 64, 65, 69

PA Zwicker Psychoacoustic Annoyance. 64, 65

PAQ Perceived Affective Quality. 21, 63, 64, 128

PIC Site pictures. 24, 28, 29

QUE Questionnaires. 20, 24

R Psychoacoustic Roughness. 64, 65

RA Relative Approach. 64, 65

S Psychoacoustic Sharpness. 64, 65, 70

SLM Sound Level Meter. 17, 20, 24–27

SSID Soundscape Indices. 15, 17, 20, 62, 65

SSQP Swedish Soundscape Quality Protocol. 21

T Tonality. 64, 65, 70

VID 360°Video. 24

VIF Variance Inflation Factor. 65, 66

VR Virtual Reality. 22

WHO-5 WHO Well-being Index. 21

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. Number: 3.

Akaike, Hirotugu (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6):716–723. Number: 6.

Aletta, F., Axelsson, Ö., and Kang, J. (2017). Dimensions underlying the perceived similarity of acoustic environments. *Frontiers in psychology*, 8:1162.

Aletta, F., Axelsson, s., and Kang, J. (2014). Towards acoustic indicators for soundscape design. In *Forum acusticum*, number c, page 6.

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., Mitchell, A., Erfanian, M., Lionello, M., Kachlicka, M., and Kang, J. (2019). Associations between soundscape experience and self-reported wellbeing in open public urban spaces: A field study. *The Lancet*, 394:S17.

Aletta, F., Oberman, T., Mitchell, A., Tong, H., and Kang, J. (2020). Assessing the changing urban sound environment during the COVID-19 lockdown period using short-term acoustic measurements. *Noise Mapping*, 7(1):123–134. Number: 1.

Alsina-Pagès, R. M., Bergadà, P., and Martínez-Suquía, C. (2021). Changes in the soundscape of Girona during the COVID lockdown. *The Journal of the Acoustical Society of America*, 149(5):3416–3423. Number: 5.

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.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods*, 52(1):388–407.

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. Number: 12.

Asensio, C., Pavón, I., and de Arcas, G. (2020b). Changes in noise levels in the city of madrid during covid-19 lockdown in 202. *The Journal of the Acoustical Society of America*, 148:1748–1755.

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, O., 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. Number: 5.

Axelsson, O., Nilsson, M. E., and Berglund, B. (2012). The Swedish soundscape-quality protocol. *The Journal of the Acoustical Society of America*, 131(4):3476–3476. Number: 4.

Bartalucci, C., Bellomini, R., Luzzi, S., Pulella, P., and Torelli, G. (2021). A survey on the soundscape perception before and during the COVID-19 pandemic in Italy. *Noise Mapping*, 8(1):65–88. Number: 1.

Basner, M., Samel, A., and Isermann, U. (2006). Aircraft noise effects on sleep: application of the results of a large polysomnographic field study. *The Journal of the Acoustical Society of America*, 52(5):109–123.

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.

Blanes, N., Fons, J., Houthuijs, D., Swart, W., de la Maza, M., Ramos, M., Castell, N., and van Kempen, E. (2017). Noise in europe 2017: Updated assessment. *European Topic Centre on Air Pollution and Climate Change Mitigation (ETC/ACM): Bilthoven, The Netherlands*.

Bonet-Solà, D., Martínez-Suquía, C., Alsina-Pagès, R. M., and Bergadà, P. (2021). The Soundscape of the COVID-19 Lockdown: Barcelona Noise Monitoring Network Case Study. *International Journal of Environmental Research and Public Health*, 18(11):5799. Number: 11.

Brown, A. and De Coensel, B. (2018). A study of the performance of a generalized exceedance algorithm for detecting noise events caused by road traffic. *Applied Acoustics*, 138:101–114.

Craig, A., Moore, D., and Knox, D. (2017). Experience sampling: Assessing urban soundscapes using in-situ participatory methods. *Applied Acoustics*, 117:227–235.

EEA (2020). Environmental noise in Europe – 2020. Technical report, European Environmental Agency.

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.

EU (2002). Directive 2002/49/EC of the European Parliament and the Council of 25 June 2002 relating to the assessment and management of environmental noise.

Garrido-Cumbrera, M., Foley, R., Braçe, O., Correa-Fernández, J., López-Lara, E., Guzman, V., Gonzalez Marín, A., and Hewlett, D. (2021). Perceptions of change in the natural environment produced by the first wave of the covid-19 pandemic across three european countries. results from the greencovid study. *Urban Forestry & Urban Greening*.

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.

Genuit, K. (1996). Objective evaluation of acoustic quality based on a relative approach. In *Proceedings-Institute of Acoustics*, volume 18, pages 3233–3238. ISBN: 0309-8117.

Goines, L. and Hagler, L. (2007). Noise Pollution: a Modern Plague. *Southern Medical Journal-Birmingham Alabama-*, 100(3):287–294.

Gontier, F., Lavandier, C., Aumond, P., Lagrange, M., and Petiot, J.-F. (2019). Estimation of the Perceived Time of Presence of Sources in Urban Acoustic Environments Using Deep Learning Techniques. *Acta Acustica united with Acustica*, 105(6):1053–1066. Number: 6.

Guastavino, C., Katz, B. F., Polack, J.-D., Levitin, D. J., and Dubois, D. (2005). Ecological Validity of Soundscape Reproduction. *Acta Acustica united with Acustica*, 91(2):333–341. Number: 2.

Hadjidemetriou, G. M., Sasidharan, M., Kouyialis, G., and Parlikad, 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.

Hornberg, J., Haselhoff, T., Lawrence, B. T., Fischer, J. L., Ahmed, S., Gruehn, D., and Moebus, S. (2021). Impact of the covid-19 lockdown measures on noise levels in urban areas-a pre/during comparison of long - term sound pressure measurements in the ruhr area, germany. *International Journal of Environmental Research and Public Health*, 18:4653.

International Organization for Standardization (2014). *ISO 12913-1:2014 Acoustics – Soundscape – Part 1: Definition and conceptual framework*. Geneva, CH.

International Organization for Standardization (2018). *ISO/TS 12913-2:2018 Acoustics – Soundscape – Part 2: Data collection and reporting requirements*. Geneva, CH.

International Organization for Standardization (2019). *ISO/TS 12913-3:2019 Acoustics - Soundscape - Part 3: Data analysis*. Geneva, CH.

Kajihara, Y., Dozono, S., and Tokui, N. (2017). Imaginary Soundscape: Cross-Modal Approach to Generate Pseudo Sound Environments. In *31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA.

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*, pages 2488–2495, Aachen.

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.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26.

Lacey, J. (2016). *Sonic Rupture: A Practice-led Approach to Urban Soundscape Design*. Bloomsbury Sound Studies. Bloomsbury Academic, London, UK.

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.

Lüdecke, D. (2021). *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.7.

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.

Lenzi, S., Sádaba, J., and Lindborg, P. (2021). Soundscape in Times of Change: Case Study of a City Neighbourhood During the COVID-19 Lockdown. *Frontiers in Psychology*, 12:570741.

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*, 11:602831.

Maggi, A. L., Muratore, J., Gaetán, S., Zalazar-Jaime, M. F., Evin, D., Pérez Villalobo, J., and Hinalaf, M. (2021). Perception of the acoustic environment during COVID-19 lockdown in Argentina. *The Journal of the Acoustical Society of America*, 149(6):3902–3909. Number: 6.

McKnight, P. E. and Najab, J. (2010). Mann-whitney u test. In Weiner, I. B. and Craighead, W. E., editors, *The Corsini Encyclopedia of Psychology*, page corpsy0524. John Wiley & Sons, Inc., Hoboken, NJ, USA.

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.

Munoz, P., Vincent, B., Domergue, C., Gissinger, V., Guillot, S., Halbwachs, Y., and Janillon, V. (2020). Lockdown during COVID-19 pandemic: impact on road traffic noise and on the perception of sound environment in France. *Noise Mapping*, 7(1):287–302. Number: 1.

Nations, U. (2018). World urbanization prospects: The 2018 revision. *Population Division of the UN Department of Economic and Social Affairs*.

Ndrepepa, A., Twardella, D., et al. (2011). Relationship between noise annoyance from road traffic noise and cardiovascular diseases: a meta-analysis. *Noise and health*, 13(52):251.

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).

Ouis, D. (2001). Annoyance from road traffic noise: a review. *Journal of environmental psychology*, 21(1):101–120.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ren, X. (2020). Pandemic and lockdown: a territorial approach to COVID-19 in China, Italy and the United States. *Eurasian Geography and Economics*, 61(4-5):423–434. Number: 4-5.

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.

Rumpler, R., Venkataraman, S., and Göransson, P. (2021). Noise measurements as a proxy to evaluating the response to recommendations in times of crisis: An update analysis of the transition to the second wave of the CoViD-19 pandemic in Central Stockholm, Sweden. *The Journal of the Acoustical Society of America*, 149(3):1838–1842. Number: 3.

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

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*, 2.

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*.

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.

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., Pastor, J. A. A., Quesada, R. G., 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. Number: 1.

Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021.

Woods, K. J. P., Siegel, M. H., Traer, J., and McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, 79(7):2064–2072. Number: 7.

World Medical Association (2013). World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*, 310(20):2191–2194.

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, 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. (1999). *Psychoacoustics: facts and models*. Number 22 in Springer series in information sciences. Springer, Berlin ; New York, 2nd updated ed edition.

Zwicker, E. and Fastl, H. (2007). *Psychoacoustics: facts and models*. Springer, Berlin ; New York, third ed. edition.