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)
June 21, 2021

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

Abstract

List of Studies

This doctoral thesis is based on the following studies:

Commentary: 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*.

Protocol: **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`

Study I: 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*.

Study II: 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`

Study III: **Mitchell, A.**, Oberman, T., Kachlicka, M., Aletta, F., Lionello, M., Erfanian, M., & Kang, J. (2021). Applied Predictive Soundscape Modelling: A Case Study Investigating Changes from the COVID-19 Lockdown. *JASA*.

Study IV: **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*.

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

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 Impacts of Urban Noise on Health and Wellbeing

1.2 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.

1.3 Soundscape - Theory and Application

1.3.1 Soundscape Descriptors and Indices

(Aletta et al., 2016)

1.3.2 The ISO 12913 Standard Series

1.4 Environmental Acoustics and Psychoacoustics Analyses

1.5 Practical Applications for Predictive Modelling

1.6 General Aim

2 General Method

2.1 In-situ Surveys

2.2 An Open International Urban Soundscape Database

2.3 Predictive Modelling

2.3.1 Multi-level Regression

2.3.2 Deep Learning

3 Studies & Summaries

The following studies form the core of this thesis, with each contributing a key component of the soundscape modelling process. They are presented in a logical, rather than chronological order and are each preceded by a brief summary.

3.1 Commentary: Supportive Soundscapes are Crucial for Sustainable Cities and Communities

Background and aims

Presented as a *Comment* paper in Nature Sustainability, this paper provides an opininated view towards the placement of soundscape in future sustainability research and development. Heavily shaped and primarily drafted by me, it puts forth the argument that cities and communities cannot be sustainably designed without a consideration of 1) how noise impacts the community, and 2) how the community impacts the existing soundscape. **misc** *add a bit about the agenda presented*

Although published towards the end of my PhD, it is placed at the beginning as it offers our strongest argument for the necessity of soundscape in future sustainable design. It situates noise as a key environmental concern and provides a justification for a soundscape approach, which requires and incorporates predictive tools to be applied in engineering, research, and design contexts.

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

Background and aims

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.

Result and conclusions

applied sciences

MDPI

Protocol

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

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

UCL Institute for Environmental Design and Engineering, The Bartlett, University College London, Central House, 14 Upper Woburn Place, London WC1H 0NN, UK; t.oberman@ucl.ac.uk (T.O.); mercede.erfanianghasab.18@ucl.ac.uk (M.E.); magdalena.kachlicka.18@ucl.ac.uk (M.K.); matteo.lionello.18@ucl.ac.uk (M.L.)

* Correspondence: andrew.mitchell.18@ucl.ac.uk (A.M.); f.aletta@ucl.ac.uk (F.A.); j.kang@ucl.ac.uk (J.K.)

Received: 26 February 2020; Accepted: 24 March 2020; Published: 1 April 2020

check for updates

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 to previous methodologies for soundscape characterization 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.

Keywords: soundscape; SSID; binaural recordings; ambisonic recording; questionnaire; visual factors; ISO12913

1. Introduction

Soundscape studies strive to understand the perception of a sound environment, in context, including acoustic, (non-acoustic) environmental, and contextual, and personal factors. These factors combine together to form a person's soundscape in complex interacting ways [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 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. [2] 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" [3]. (Please find the link to the Soundscape Indices (SSID) ERC-funded website in the Supplementary Materials.)

This conception has recently been formalized and expanded upon with the adoption of the recent ISO 12913 standard series [4–6]. 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) [5], 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 analyzing data gathered using only those methods [6]. However, the range of possible methodological approaches to soundscape data collection is much broader and it includes, for instance, laboratory experiments [2,7,8], pseudo-randomized experience sampling [9], and even non-participatory studies [10]. 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 modeling purposes, thus trying to expand the scope of soundwalks that typically deal with much smaller samples of participants [11]. 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 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 [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 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. [16], 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. This protocol is trying to extend the scope of objective measurements that are being collected in conjunction with perceptual responses by including other environmental and visual data. 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, and no non-acoustic factors, generally limiting the generalizability of their results beyond the investigated locations.

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 [11,17]. 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 modeling and comparisons, it is important that such indices are numerical entities and that these quantities are collected consistently across all investigated spaces and soundscapes. Although the soundscape approach taken in this protocol represents a step-change away from existing methods of noise exposure measurements, strong cues particularly in the realm of acoustic measurement methods should be taken from existing standards both to make use of the significant knowledge and experience that has gone into the creation of these standards and to facilitate compatibility between soundscape and traditional measurements. In general, the measurement methods and best practice given in environmental noise standards such as ISO 1996-1:2016 and ISO 1996-2:2017 [18,19] should be followed wherever possible, including the use of standardized acoustic equipment such as standard sound level meters.

An European Research Council (ERC) Advanced Grant project is ongoing to develop the proposed "Soundscape Indices" (SSID), which adequately reflect levels of human comfort and preference while integrating measurable and observable quantities. The framework proposed for the SSID project is laid out in detail by Kang et al. [3], the first step of which is generating a large-scale and coherent database of the required soundscape characterization data. Given the already recognized differences in soundscape assessment across various countries and cultures [20,21] and the success of existing international soundscape efforts such as the Soundscapes of the World project [22], the collection of soundscapes from many different countries and in many different contexts is an important component of the SSID project.

Therefore, the following protocol has been conceived and implemented within the SSID framework to collect data about urban soundscapes for use in general soundscape research and toward the design of Soundscape Indices. Thus far, the collected database includes nearly 4000 participants' responses from 59 locations in 10 cities and provinces across the UK, China, Spain, and Italy. This protocol has been refined and adjusted as needed during this extensive data collection process to arrive at this final version. This work was conducted by nine associated research groups and coordinated by the SSID group based at University College London and has already produced several pieces of published work towards the creation of Soundscape Indices [23–30]. Additional collaborations and data collection efforts are currently underway in France, the Netherlands, and Croatia.

Purpose

This protocol was designed to achieve two primary goals: (1) gather in situ soundscape assessments from the public, which can be further analyzed and utilized 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 analyzed on a large scale, or used for training in machine learning modeling. 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 [31]. Laboratory experiments with controlled environments are required to address these aspects. Toward the development of a coherent SSID, however, 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.

2. Protocol Design and Equipment

The first goal is achieved by conducting in situ questionnaires using a slightly altered version of Method A (questionnaire) from Annex C of the ISO/TS 12913-2:2018 technical specification [5] 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-h 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 [32]) 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 [33], particularly when paired with simultaneous head-tracked virtual reality video [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. The stage during which the spatial 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 perceptual responses can be matched with the appropriate corresponding environmental and acoustic data even when some information is lost or forgotten.

2.1. Labeling and Data Organization

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 labeling system is suggested. This system is focused 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. [14] demonstrated the importance of addressing multiple levels of factors which influence perception, form individual-, to session-, to location-level. The successful pleasantness models building 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 1.

At the top level is the **Location** information. This includes information about the location which does not change day-to-day, and generally characterizes the architectural character of the space, or typical climate conditions for the area. As described in Section 2.2, each 'environmental unit' should be considered a new location. Therefore, if researchers want to investigate the differences in soundscape assessment in the middle of a small urban park and along the road next to the same park, these would be considered different locations since they would (typically) have different environmental factors, and should be given different names. The name chosen should be concise, but it should be obvious what location is referred to.

The next level is information which is specific to each session, labeled with a **SessionID**. This SessionID should contain the name of the location and a numerical index which will increase with each repeated session at that location. The SessionID is associated with the data collected during the Recording Stage, and with the data which are continuous throughout the Questionnaire Stage, SLM, and ENV data. For easy automatic processing, correct spelling and consistency with the format is crucial so that data can be filtered according to the SessionID or the location, as is often necessary.

In addition, for ease of automatic processing, it is recommended not to include spaces in the SessionID to avoid string splitting issues in analysis code.

Underneath each SessionID will be a set of **GroupIDs**. One GroupID is assigned for *each group of participants*. This should correspond to a single binaural recording and a single 360° photo. This will be used to (1) relate multiple surveys taken simultaneously and (2) link the recording and photo with the surveys. The GroupID is particularly crucial as it allows commonly missing data to be shared across multiple collection methods. 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 recording 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 labeled 'RPJ210'. This format ensures that, if the location or SessionID are not recorded for a questionnaire, it is still obvious which session it belongs to.

Table 1. Labeling system for on site data collection. Regent's Park Japanese Garden is used as an example location. SLM: Sound Level Meter (acoustical factors); ENV: Environmental factors; BIN: Binaural; QUE: Questionnaires; PIC: Site pictures.

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

2.2. Location and Measurement Point Selection

To select the appropriate measurement point, it should be ensured that the following contextual factors representative of the site are present in the spatial recording: openness, greenness, presence of landmarks, dominant use (walking, staying), and social presence (related to the dominant use). These are identified as objective metrics often used in urban and landscape research [36–40], possibly contributing to soundscape assessment [23,41]. This relies on researcher's opinion-driven assessment—it is advised to observe the location for a moment and then choose the point representative of the context and the first-person user experience. For instance, in a park, it would probably be near a bench in the central area near the fountain; in a busy square, it would be a place where most people gather and have the best view on the landmark. While doing so, the placement too near the prominent vertical objects such as a statue, a wall, or a mast should be avoided as it might cause issues in later handling the visual data (3 m 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 in acoustic shadows. Further guidance on this is given in Point 4 of Section 4.1. 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 judgment of the researchers on site to select an appropriate measurement point to best capture the character of the environmental unit.

2.3. Equipment

The equipment listed in Table 2 is designed to facilitate both the audio-visual recording of the location and the collection of objective environmental factors, as given in Table 3. 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 colors as opposed to high-visibility colors and selecting compact equipment.

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

Table 2. Recommended equipment for implementing the SSID protocol. SLM: Sound Level Meter; AMB: Ambisonics; BIN: Binaural; QUE: Questionnaires.

Equipment	**Requirements**
Tripod stand	With add-on hooks/holders for AMB microphone, SLM, environmental meter(s) and 360° camera with suitable suspension for microphones
360° camera	4 K, 5.1 K or better resolution video, with suitable battery life and optional remote control
Spatial audio/Ambisonics (AMB) microphone system	Min. quality should be First-order Ambisonics (FOA) capability, however systems which achieve higher-order ambisonics would be preferred where available.
Multi-channel field recorder	Min. inputs to accommodate output from AMB microphone
Windshield(s) for AMB and SLM microphones	This can be a single large windshield which can accomodate both microphones or separate windscreens for each microphone
Sound Level Meter (SLM)	class 1 (preferred) or class 2 with omnidirectional pattern measurement microphone
Binaural recording system	Portable, worn by the researcher or with a mounted binaural head
Sound calibrator for SLM, AMB microphones and binaural system	According to IEC 60942: 2017 Electroacoustics—Sound calibrators [42]
Environmental meter(s)	See Table 3 for the recommended metrics
Tablets and/or printed questionnaires	Internet connectivity or offline app to submit the questionnaires on site

Table 3. Table of recommended context and acoustic measurement factors.

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

* The recommended acoustic data settings are given here in order of importance. In cases where researchers do not have access to a meter capable spectral logging, L_{Aeq} logging should be prioritized over spectral analysis. During both stages, spectral data can typically be extracted from the audio recordings, but accurately tracking the sound level is crucial. ** The recommended environmental factors are given here in order of importance. More flexibility is allowed in selecting which factors to record and investigate (compared to the acoustic data) as it is still unclear how and to what extent environmental factors influence soundscape assessment. However, previous studies have indicated visual (i.e., lighting levels) and temperature are significant factors [43].

3. Techniques for Field Data Collection

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

3.1. Questionnaire Surveys

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

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 (labeled SSI01 through SSI04, respectively). These taxonomic categories of environmental sounds are based on the work done by Guastavino [44] and Brown, Kang, and Gjestland [45].

Next are the 8 scales which make up the circumplex model of the Swedish Soundscape Protocol [46], 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 (labeled 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 (labeled SSS01 through SSS05, respectively).

The fourth section comprises the WHO-5 well-being index, 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 [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. [27] or, alternatively, how a person's current psychological status may influence their perception of the sound environment as recently investigated by Erfanian, Mitchell, Aletta, and Kang [49]. Each of the five WHO questions (labeled 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 behavior of the participants, indexing and labeling metadata, and space for any additional notes. More information and guidance on this information is included below.

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

General notes for conducting the questionnaires:

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

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

3.2. Contextual and Environmental Factor Data Collection

During each survey, the equipment listed in Section 2.3 is set up to capture the contextual and environmental data for the location. Table 3 lists the factors to be collected and at what stage they should be collected.

3.2.1. Spatial Audio-Visual Recordings

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

Typically, questionnaire surveys are carried out over a period of several days at the same location. The goal of these multiple sessions is to capture as many questionnaire responses as needed (100 for a particular soundscape is typically recommended [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 is useful to use the spatial recording as a chance to gain a different perspective on the space under investigation. For instance, if the questionnaires are conducted in the middle of a large urban park, the first session could collect a spatial recording within the environmental unit of the questionnaire site, but the subsequent returns to the site could collect spatial recordings in a different environmental unit, say, along a road bounding the park, or in a space in the park which does not typically have many people. This enables the simultaneous expansion of the questionnaire database and the gathering of additional environments to investigate in a laboratory setting.

General notes for spatial recordings:

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

3.2.2. Reference Recordings

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

4. Procedure

Figure 1 shows the whole process of the on site soundscape protocol. The relevant equipment in each row should be operating when the row is colored 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 2 shows an example of the recommended equipment setup.

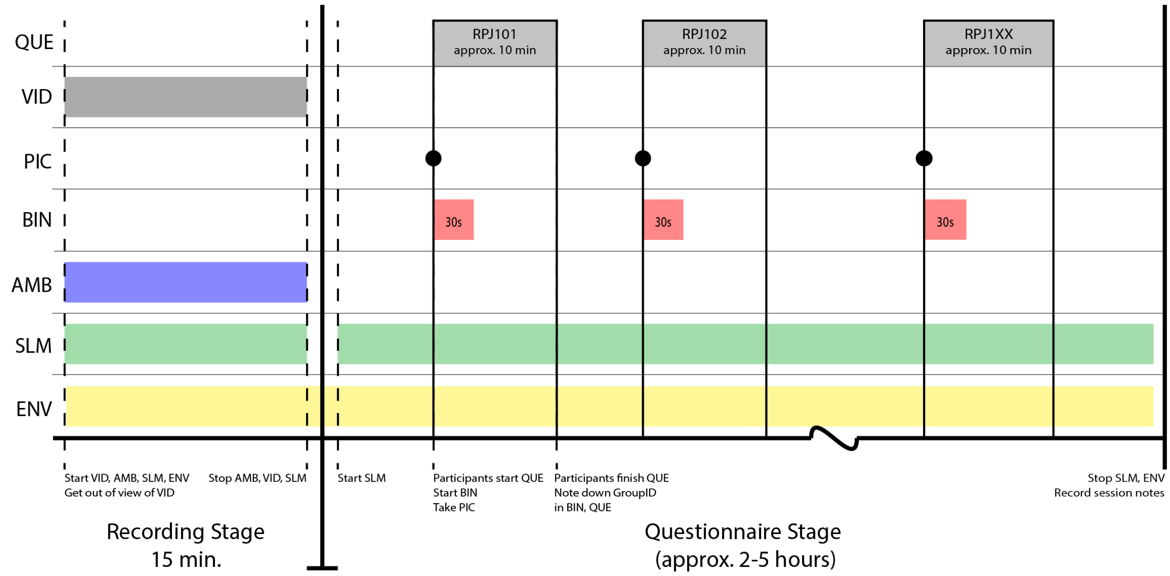

Figure 1. Timeline of the on site soundscape protocol. RegentsParkJapan (RPJ) is used as an example. Abbreviations as defined in Table 3—QUE: Questionnaires; VID: 360° video; PIC: Site pictures; BIN: Binaural Recording; AMB: Ambisonic recording; SLM: Sound Level Meter (acoustical factors); ENV: Environmental factors.

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 sound level meters 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 finalized 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.

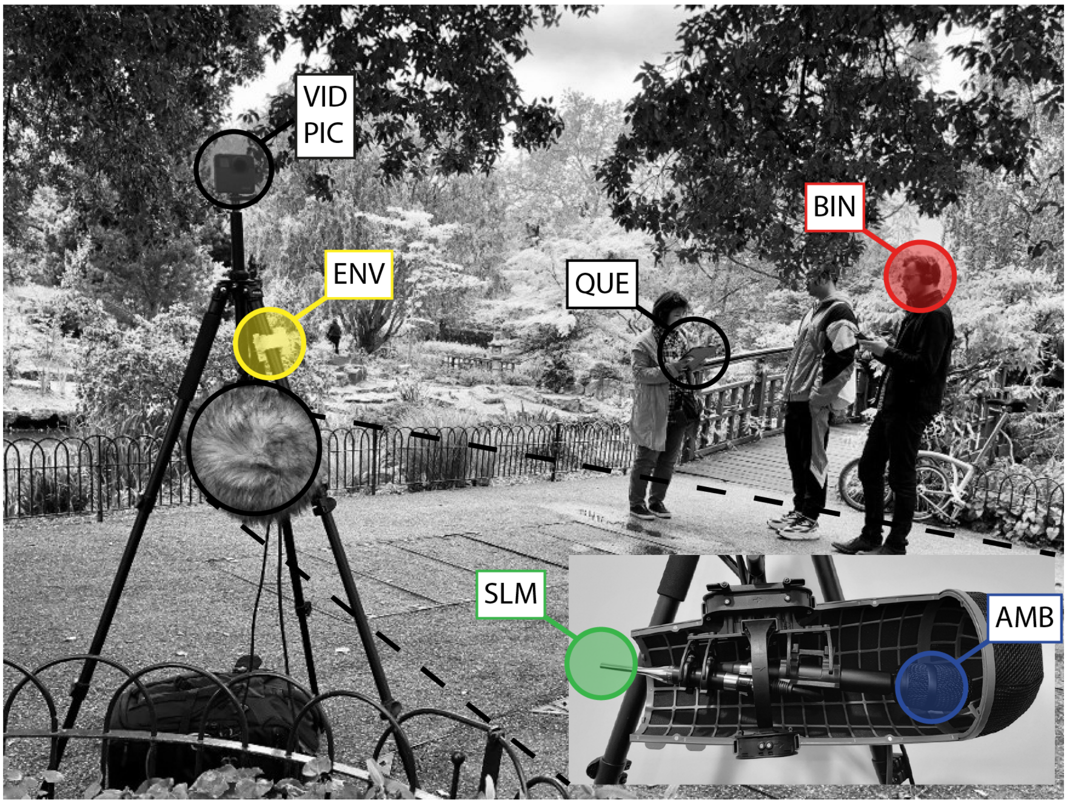

Figure 2. Photo of a full survey carried out in a park in London during the Questionnaire Stage. To the left is the equipment (color-coded to match Figure 1), with the ambisonic microphone and SLM microphone in the windscreen, with the 360° camera on top of the tripod and to the right are one researcher 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.

4.1. Assembling the Equipment

1. Set up the equipment by prioritizing the position of the 360° camera and position the lens at the average eye level 160–180 cm, as shown in Figure 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 MixPre 10), this can be set by locking the channel gain settings to a

single channel. Many devices also offer ambisonic plugins which simplify these settings and automatically link the gain settings—these should be used where available.

4. Set the SLM to log sound levels and simultaneously record .wav audio. The recommended logging settings are given in Table 3. 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 [19] or Section 5 of ANSI/ASA S12.9-2013/Part 1 [50]. Generally, the microphone should be a minimum of 1.2 m above the ground and a minimum of 1 m 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.

4.2. Recording Stage

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

4.3. 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 2.2, while making sure the location selected is representative of where the respondents will be stopped. Wherever possible, the equipment should be assembled and located so as not to draw the attention of the respondents and particularly to avoid influencing their perception of the space.

1. Double check all settings and file save locations on the recording equipment. If starting this stage immediately after the Recording Stage, make sure to rename or advance the index on the filenames for the SLM and environmental meters.
2. Start recording on the SLM and environmental 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. Emphasize 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 focused 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 binaural recordings and 360° photo.
6. Once the participant has started answering the questionnaire, start recording the binaural 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 30 s 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 30 s of good audio. The goal is to capture the sound environment which the participant was 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 binaural 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 binaural recording, or make a manual note of the binaural recording file name and the GroupID separately.
8. Take one 360° photo 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 researcher 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 versions) 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 binaural equipment can then shift between the groups once they have finished the 30 s 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.)

5. Lessons from International Data Collection

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

5.1. Sampling

The research groups were instructed to try keeping the structure of respondents well-balanced. This often led to longer times and larger sample sizes required as most comments from five research groups addressed age and type of location as the most influential factors for participant sampling. However, while some reported higher response rates from younger (students) members of public, the others reported higher response rates in case of older high educated people. A common observation was that public parks are the locations with the highest response rates, most probably 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 group, which conducted the survey also in a residential area, distinguished privacy/ownership of the survey site as a major factor.

5.2. Data Collection

A group of three researchers seems to be the minimum number needed to conduct the survey, as observed by the partner research groups. The 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 emphasizes the focus on the actual acoustic environment perceived at the moment, additional care should be made to ensure the proper understanding of that concept while approaching the participants. Researcher's 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/misleading data.

5.3. Equipment

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

5.4. Translation

Regarding the on-site 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 standardized 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, [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 [24], 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.

Supplementary Materials: Please find the link to the Soundscape Indices (SSID) ERC-funded website at: https://www.ucl.ac.uk/bartlett/environmental-design/soundscape-indices-ssid.

Author Contributions: Conceptualization, A.M., T.O., F.A., and J.K.; methodology, A.M., T.O., F.A., M.E., M.L., and J.K.; validation, A.M., T.O., M.E., M.K., and M.L.; investigation, A.M., T.O., F.A., M.E., M.K., and M.L.; resources, T.O., F.A., and J.K.; data curation, A.M. and T.O.; writing—original draft preparation, A.M.; writing—review and editing, A.M., T.O., F.A., and J.K.; visualization, A.M.; supervision, J.K.; project administration, F.A. and J.K.; funding acquisition, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: 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, 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.

Acknowledgments: The authors would like to thank our partner research groups, particularly those in Universidad de Granada, Harbin Institute of Technology, Harbin Institute of Technology (Shenzhen), Shenyang Jianzhu University, Tianjin University, Northwest A&F University, Huazhong University of Science and Technology, and Chongqing University for their valuable feedback and for their data collection efforts. We would also like to thank the Masters students, visiting researchers, and other PhD students at UCL who have helped contribute to the refinement of this protocol and collection of the current database. Study data were collected and managed using REDCap electronic data capture tools hosted at University College London (UCL) [55,56]. REDCap (Research Electronic Data Capture) is a web-based software platform designed to support data capture for research studies.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SSID	Soundscape Indices
FOA	First Order Ambisonics

PAQ	Perceived Affective Quality
ERC	European Research Council
VR	Virtual Reality
QUE	Questionnaires
VID	360° Video
PIC	Site Pictures
BIN	Binaural Recording
AMB	Ambisonic Video
SLM	Sound Level Meter
ENV	Environmental Data

Appendix A. Sample Information Sheet

PARTICIPANT INFORMATION SHEET: PERCEIVED SOUNDSCAPE

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

I am XXXXX YYYYY and I am a student at ZZZZZZ University. I am working on a project about "soundscapes" and the aim of this questionnaire is collecting data on how people perceive urban acoustic environments, and what are the relationship between acoustic environment and well-being. Results from this survey will help us to gather further insights into these relationships and the knowledge we'll gain from this research should inform urban sound planners. Other participants will be randomly approached on site and will be invited to take part, like you. You can participate to this survey only if you are over eighteen years old.

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep. Please observe that participating in a scientific experiment is voluntary and you are free to withdraw at any time without giving any reason. Should you choose to withdraw before completing the questionnaire, your partial responses will be deleted. However, once your response is submitted, all data are completely anonymised, and it will not be possible for us to identify or remove your response. The questionnaire is designed to take approximately 10 minutes to complete, and it will allow you to evaluate the surrounding acoustic environment in the public area where you are now, related with your well-being. Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will raise environmental awareness for the soundscapes of our cities.

This study has been approved via the ethics review procedure of the Bartlett School of Environment, Energy and Resources. If you wish to raise a complaint you should contact the principal researcher of this project, Prof Jian Kang at the e-mail kang@ucl.ac.uk or telephone number 020 3108 7338. If you feel your compliant has not been handled to your satisfaction, then you can contact the Chair of the UCL Research Ethics Committee (ethics@ucl.ac.uk).

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications. All the data collected will be stored in on UCL secure servers with password protection and at the end of the project all data collected will be securely archived.

This project is funded through a European Research Council (ERC) Advanced Grant (no. 740696) on "Soundscape Indices" (SSID) (Principal Investigator: Prof Jian Kang).

If you experience any problems or need further information, contact me on email: uclssid@gmail.com

Thank you in advance for your time and participation!

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

Background and aims

Some summary text here.

Result and conclusion

Abstract

Sound forms a key component of our everyday environment but is not a purely physical phenomenon. Soundscape studies are an attempt to consider the holistic perception of a sound environment, including both the physical environment and how this is mediated by internal factors. The importance of the internal factors and how they interact with the sound environment to form soundscape is not well understood. 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 standard and incorporating the WHO-5 well-being index. Linear mixed-effects modelling applying backwards-step feature selection was used to model the interactions between the internal factors 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.

Keywords: Soundscape pleasantness, soundscape eventfulness, psychological well-being, demographic factors, acoustic environment

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

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, Aletta, Gjestland, Brown, Botteldooren, Schulte-Fortkamp et al., 2016; Kang, Aletta, Oberman, Erfanian, Kachlicka, Lionello 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, 1998; Tomkins, 1962).

This model suggests that all affective states, described with 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 (e.g., increased heart rate) whereas low arousal is associated with 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 of Standardization Technical Specification, 2019) [1]. Soundscape pleasantness refers to the emotional magnitude of

[1] International Organization for Standardization/Technical Specification (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.

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, 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 pleasantness and eventfulness (See 'Outcome variables').

or the current study, we employed the circumplex model reported in a two-dimensional scatter plot with coordinates for the two dimensions 'leasantness' plotted on -axis and 'ventfulness' plotted on -axis, taking into account the features of the locations. o differentiate these complex components from classic pleasantness and eventfulness in xelsson'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, Oberman, Mitchell, Erfanian, Lionello, Kachlicka 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 show a huge variation among populations (Weinstein,1978). There is evidence to suggest that the differences in the demographic characteristics like gender (Xiao & Hilton, 2019; Gulian & Thomas 1986), age (Zhang & Kang, 2007), and educational background (Zhang & Kang, 2007) may determine the way we perceive sounds. 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 (Riskind et al., 2014; Laufer, Israeli, Paz, 2016). 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, Chinese) and occupation status (i.e., student, retired).

Methods

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

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 B.1 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.

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

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

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 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

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

recommendation of ISO/TS 12913-3:2019). These dimensions were calculated as shown in Formulas (1 & 2):

$$Pleasantness\ (P) = \sum_{i=1}^{8} PAQ_i * \cos\theta_i$$

(1)

$$Eventfulness\ (E) = \sum_{i=1}^{8} PAQ_i * \sin\theta_i$$

(2)

where, PAQ_1 = pleasant, $\theta_1 = 0°$; PAQ_2 = vibrant, $\theta_2 = 45°$; PAQ_3 = eventful, $\theta_3 = 90°$; PAQ_4 = chaotic, $\theta_4 = 135°$; PAQ_5 = annoying, $\theta_5 = 180°$; PAQ_6 = monotonous, $\theta_6 = 225°$; PAQ_7 = uneventful, $\theta_7 = 270°$; PAQ_8 = calm, $\theta_8 = 315°$.

Survey procedure

The participants were approached and asked if they were interested to participate in the study. All participants received information abo.ut 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, Taylor, Minor, Elliott, Fernandez, O'Neal et al., 2019) and typically took between 5 and 10 minutes 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-hour 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 am to 3 pm.

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, Oberman, Aletta, Erfanian, Kachlicka, Lionello 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):

$$Pleasantness_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \cdots + \beta_n x_{nij} + \varepsilon_{ij}$$

(3)

$$Eventfulness_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \cdots + \beta_n x_{nij} + \varepsilon_{ij}$$

(4)

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

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

(5)

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

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, 2020). 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).

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)

Demographic characteristics	N (%)
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)

Table 1. The sample demographic characteristics.

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.

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*

Table 2. Correlation coefficients for study variables.

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 R^2 values are given in for each model in Table 3. In a mixed effects model, the marginal R^2 represents the variance explained by the fixed effects (the individual-level independent variables) while the conditional R^2 represents the variance explained by both the fixed and random effects (Nakagawa & Schielzeth, 2012). From the conditional R^2, we can say that the full models explain 35.4% and 18.1% of the variance in Pleasantness and Eventfulness, respectively (Figure 1& 2). While the majority of the variance is explained by 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.

	Pleasantness			**Eventfulness**		
Predictor	**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 $_{\text{Location}}$			0.01$_{\text{Location}}$		
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*

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.

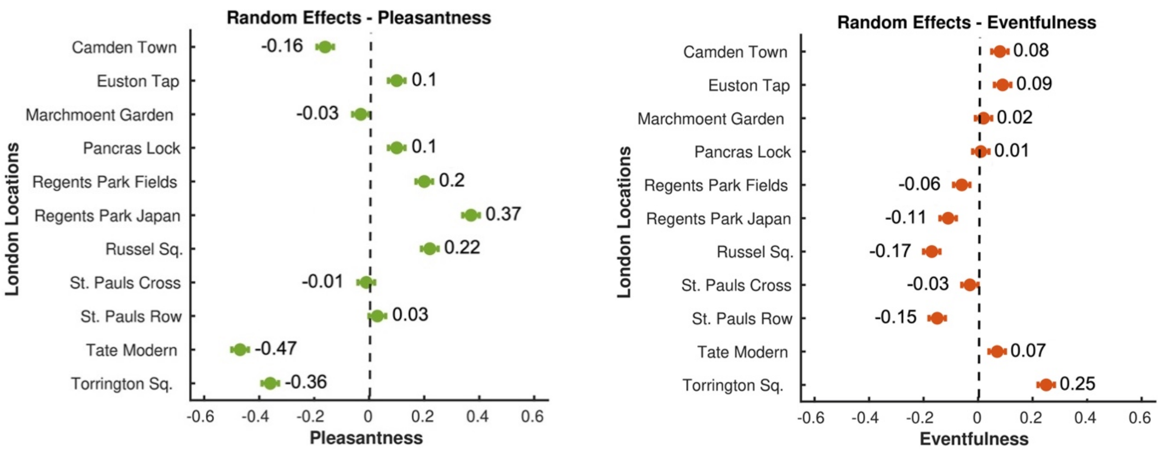

Figure 1 and 2. The summary result demonstrated in the random-effects figures gives the average from the distribution of Pleasantness (left) and Eventfulness (right) across locations.

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 men 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 men 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, Yamashita, Rytsälä, Suominen, Ahveninen, Isometsä, 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, Zavorotnyy, Diemer, Ruland, Domschke, Christ et al., 2012).

Demographic factors and their associations with Pleasantness and Eventfulness

Occupation status

According to our findings, occupation status, in particular 'retirement' and to a lesser degree, gender (male) were important factors in the pattern of soundscape assessments. It is worthwhile to highlight that 'retirement' factor can be potentially a proxy for age (>65) and gender (male). To explore the effect of occupation/retirement deeper on Pleasantness and Eventfulness we removed the occupation factor from the model. Age ($\beta = 0.02, p = 0.05$) for Pleasantness ($\beta = -0.03, p = 0.01$) for Eventfulness and gender ($\beta = -0.04, p = 0.05$) for Eventfulness then came out significant (see Appendix C). This would indicate that occupation status, particularly 'retirement', represents a group of older male individuals. Even though incorporation of occupation into our model complicates the interpretation of our outcome, it results in a slightly better fitting model (R^2_c for Pleasantness (0.354) and Eventfulness (0.181) relative to (0.345) for Pleasantness and (0.165) for Eventfulness in the model without occupation status which is why it is selected by the feature selection process. These findings are in line with previous research, suggesting significant differences among age groups in the soundscape of different acoustic environments (Ren, Kang, Liu, 2016; Yang & Kang, 2005). Our findings imply that an increase in age leads to an increase in the positive appraisal of the soundscape Pleasantness. This is supported by a study by Çakir Aydin & Yilmaz (2016) in which they found that soundscape pleasantness reported by young individuals was significantly lower than the other age groups. The results withstood a control for the effect of age on the soundscape's pleasantness and eventfulness, suggesting that different neural and behavioural processes are responsible for the differences of soundscape appraisal in age.

One possibility is that age is associated with loss of function within the peripheral auditory system (hearing loss due to age or *presbycusis*) that may lead to the variation of the soundscape (Howarth and 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 (Bones & Plack, 2015; Babkoff & Fostick, 2017). 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).

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 & Lang (2000) and Hume & 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 and colleagues (2019), throughout the soundscape literature, arousal has been applied as the equivalent of Eventfulness and indicated on the Y-axis of the circumplex model (Erfanian et al, 2019; Axelsson et al., 2010).

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

In sum, 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 men and individuals that did not report their occupations. We further demonstrated that the occupation status 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.

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., & ISO/TC 43/SC 1/WG 54. (2012). The ISO 12913 series on soundscape: An update, May 2012. *The Journal of the Acoustical Society of America*, *131*(4), 3381-3381. https://doi.org/10.1121/1.4708750

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. doi: 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. doi: 10.1007/s10433-017-0410-y

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Ime4: Linear Mixed-EffectsMoels Using Eigen and S4.

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. doi: 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. doi: 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. doi: 10.1016/j.tics.2010.06.007.

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. doi: 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.

Gelman, A., & Hill, J. (2006). Data analysis using regression and multilevel/hierarchical models. Cambridge university press.

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. doi:10.1016/J.JBI.2019.103208.

Howarth, A., & Shone, G. R. (2006). Ageing and the auditory system. Postgraduate medical journal, 82(965), 166-171. http://dx.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. 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: JPN*, *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., & Mitchell, A. (2019, September). Towards soundscape indices. In *Proceedings of the International Congress on Acoustics—ICA, Aachen,* Germany (pp. 9-13).

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. doi: 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. doi: 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. The 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*, *2013*. https://doi.org/10.1155/2013/304057

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. DOI: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. DOI: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*, *2012*. DOI: 10.1155/2012/171857

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., & 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

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 and 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.

Team, R. C. (2013). R: A language and environment for statistical computing.

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. doi: 10.1016/j.janxdis.2013.11.005.

Russel, J. (1980). A circumplex model of emotions. Journal of Personality and Social Psychology, 39, 1161-1178.

Tomkins, S. S. (1962). Affect, Imagery, 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. doi: 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 influencingperceptions 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. doi: 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. doi: 10.1002/wcs.147.

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

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. doi: 10.4103/nah.NAH_73_16

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. doi: 10.1007/s00702-012-0871-2.

3.4 Study II: Multilevel Annoyance Modelling of Short Environmental Sound Recordings

Background and aims

The soundscape approach cannot be limited to only the cases where detailed acoustical or social surveys can be conducted. The rest of the papers in this thesis rely on an extensive survey methodology (as outlined in Section 3.2 and (Mitchell et al., 2020)), however a practical application of a more holistic approach is key to more sustainable future development. One of the goals, therefore, is to be able to implement such a model on the nodes of a Wireless Acoustic Sensor Network (WASN). **draft** *Add more here*

This study therefore aimed to make use of data which is more limited than the SSID database, but to still demonstrate the potential for an annoyance model based on psychoacoustic metrics which is 1) independent of sound level, and 2) is developed in the context of real-world sounds, rather than simplified artificial sounds.

Result and conclusion

sustainability

MDPI

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

check for updates

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.

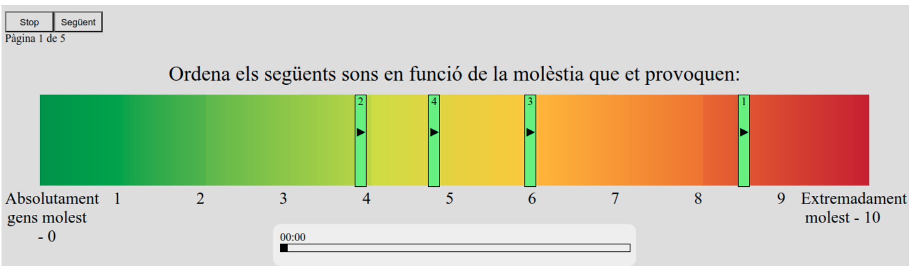

Figure 1. Screenshot of the MUSHRA test conducted to assess the annoyance provoked by different sounds. Title: sort the following sounds according to the caused annoyance. The scale ranges from *not annoying at all* to *extremely annoying*.

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:

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

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.

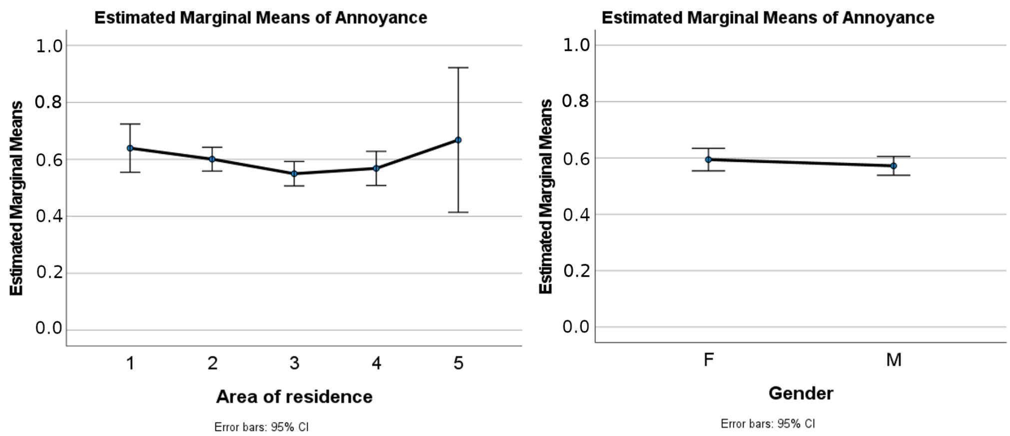

Figure 2. Estimated Marginal Means for Annoyance as a function of Areas of residence (**left**) and Gender (**right**).

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		

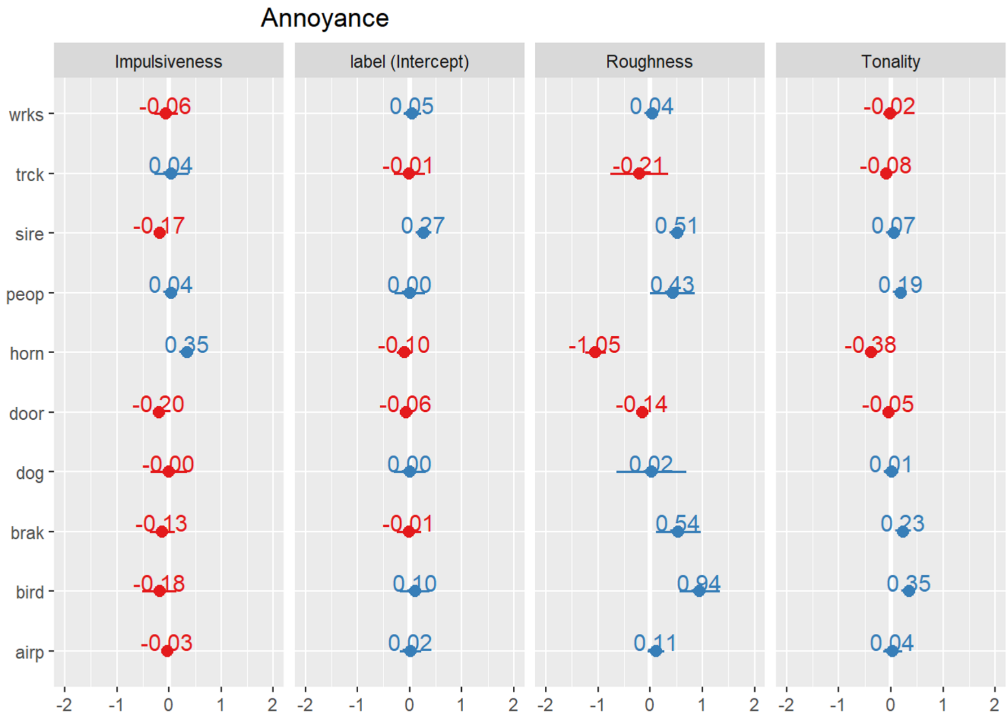

Figure 3. Second-level effects figures representing the regression coefficients by types of sound (label) and for different psychoacoustic parameters.

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]

3.5 Study III: Applied Predictive Soundscape Modelling: A Case Study Investigating Changes from the COVID-19 Lockdown

Background and aims

Some summary text here.

Result and conclusion

3.6 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/.

3.7 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:

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

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

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

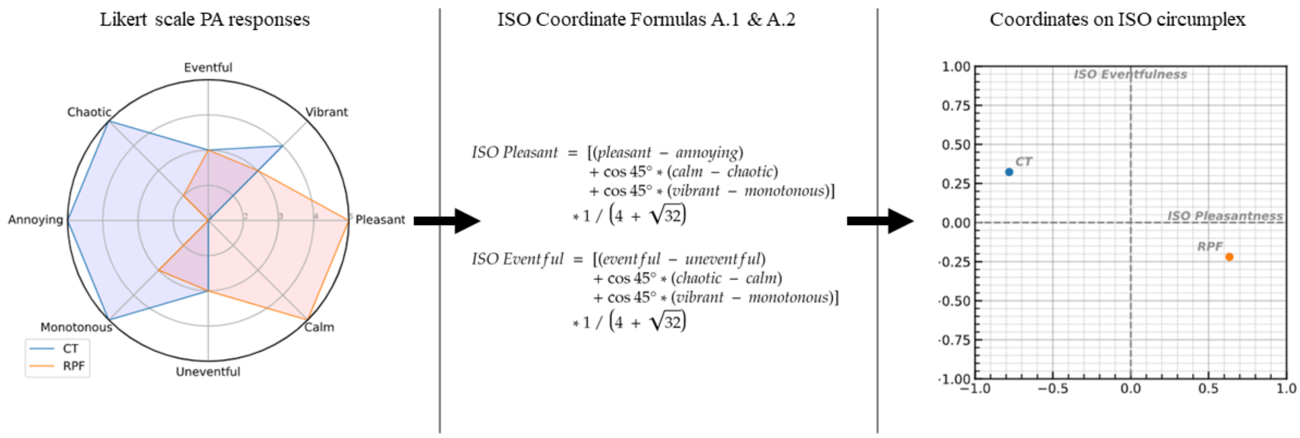

FIG. 1. Placeholder radar plot and projection. Need to remove the middle set of equations, since they are being included in the text

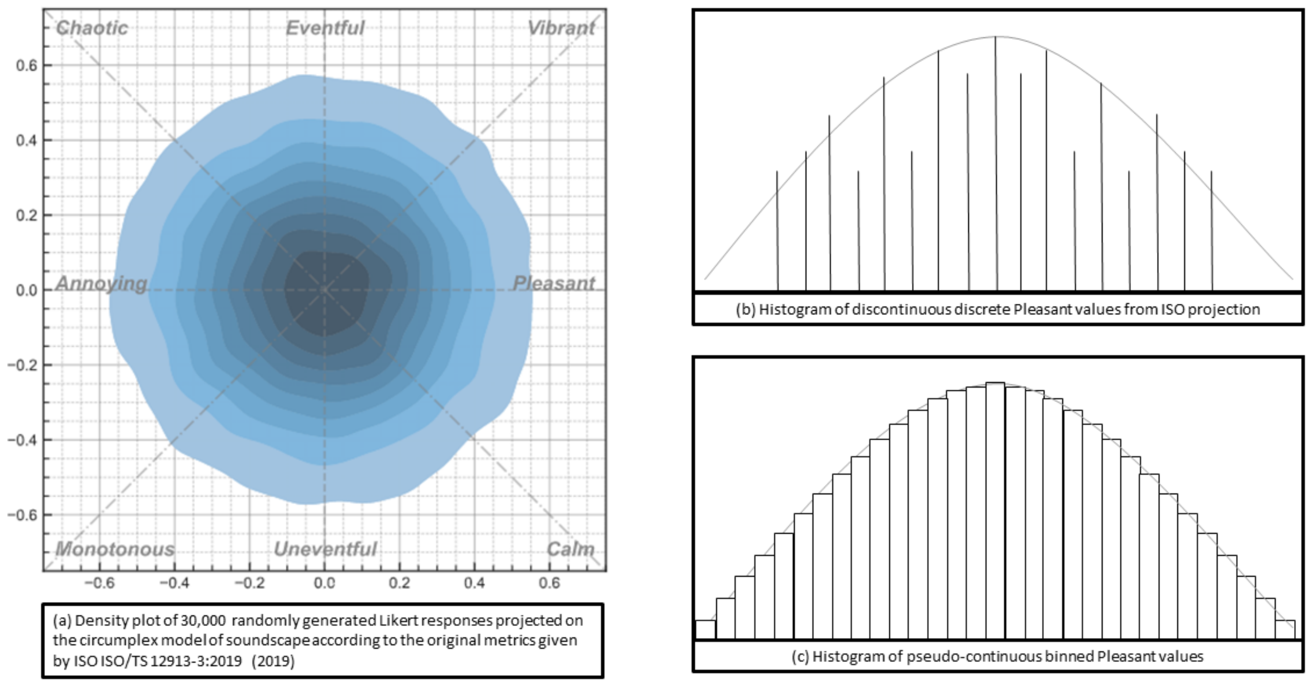

FIG. 2. Mockup placeholder simulation heatmap plot from Lionello 2020 and discrete vs binned transformation values. For new one, need to include distribution histograms along each axis.

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 $\sim$ 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.

4 Conclusions

4.1 General Discussion

4.2 Implications

4.3 Limitations and Recommendations for Future Research

4.4 Concluding Remarks

Bibliography

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.

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

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.

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